*The Knowledge Engineering Review*, Vol. 29:1, 1–30. © Cambridge University Press, 2013
doi:10.1017/S026988891300009X
First published online 18 March 2013

# Repeated games for multiagent systems: a survey

ANDRIY BURKOV and BRAHIM CHAIB-DRAA

*Department of Computer Science and Software Engineering, Université Laval, Québec, QC G1V OA 6, Canada;*
*e-mail: burkov@damas.ift.ulaval.ca, chaib@ift.ulaval.ca*

## Abstract

Repeated games are an important mathematical formalism to model and study long-term economic interactions between multiple self-interested parties (individuals or groups of individuals). They open attractive perspectives in modeling long-term multiagent interactions. This overview paper discusses the most important results that actually exist for repeated games. These results arise from both economics and computer science. Contrary to a number of existing surveys of repeated games, most of which originated from the economic research community, we are first to pay a special attention to a number of important distinctive features proper to artificial agents. More precisely, artificial agents, as opposed to the human agents mainly aimed by the economic research, are usually bounded whether in terms of memory or performance. Therefore, their decisions have to be based on the strategies defined using finite representations. Furthermore, these strategies have to be efficiently computed or approximated using a limited computational resource usually available to artificial agents.

## 1 Introduction

Usually, repeated games (Fudenberg & Tirole, 1991; Osborne & Rubinstein, 1994; Mailath & Samuelson, 2006) are used as an important mathematical formalism to model and study long-term economic interactions between multiple self-interested parties (individuals or groups of individuals). They open attractive perspectives in modeling multiagent interactions (Littman & Stone, 2005; Conitzer & Sandholm, 2007). Repeated games have also been widely employed by the researchers for modeling fairness, reputation, and trust in the situations of repeated interactions between self-interested individuals (Ramchurn *et al.*, 2004; Mailath & Samuelson, 2006; Jong *et al.*, 2008).

Computer science (namely, the research on multiagent systems (MAS)) mainly considered repeated games as an environment in which multiagent algorithms evolved, often during only an initial phase of learning or mutual adaptation. As a matter of fact, the fundamental repetitive property of repeated games has been typically reduced to a way to give to an adaptive (or learning) algorithm a sufficient time to converge (often, jointly with another learning agent) to a fixed behavior (Bowling & Veloso, 2002; Banerjee & Peng, 2003; Conitzer & Sandholm, 2007; Burkov & Chaib-draa, 2009). Thus, the repetitive nature of the game has only been considered as a *permissive* property, that is, a property that permits implementing an algorithm. However, repeated games possess another important property. This is a *constraining* property, which, as we will demonstrate in this overview, generously expands and, at the same time, rigorously limits the set of strategies that rational[1] agents can adopt. A few exceptions are Papadimitriou (1992) and Papadimitriou and Yannakakis (1994), two important but relatively ancient papers on the complexity of computing

[1] At this point, let the word 'rational' simply stand for 'the one whose goal is to maximize its own payoff'.

| | | Player 2 | |
|---|---|---|---|
| | | $C$ | $D$ |
| Player 1 | $C$ | $2, 2$ | $-1, 3$ |
| | $D$ | $3, -1$ | $0, 0$ |

**Figure 1** The payoff matrix of Prisoner's Dilemma

best-response strategies for repeated games, Littman & Stone (2005) describing an efficient algorithm for solving two-player repeated games, and Borgs *et al.* (2008) presenting an important inefficiency result about the solvability of general multiplayer repeated games.

The behavior to which the majority of the existing multiagent algorithms for repeated games converge is usually a *stationary equilibrium*. The player's behavior is called *stationary* when it is independent of the history of the game: the player always draws its actions from the same probability distribution, regardless of the actions executed by the opponent players in the past. This means that a player that adopts a stationary behavior acts in a repeated game (consisting of the player sequentially playing a certain stage-game with a constant set of opponents) as if each repetition of the stage-game was played with an entirely new set of opponents.

By focusing on stationary strategies, we often omit solutions having a greater mutual utility for the players. Consider the example of Prisoner's Dilemma shown in Figure 1. In this specific game, there are two players, called Player 1 and Player 2. Each player has a choice between two actions: $C$ (for cooperation) and $D$ (for defection, i.e., non-cooperation). The players perform their actions simultaneously; a pair of actions is called an action profile. Each action profile induces a corresponding game outcome. For each outcome, a player-specific payoff function specifies a numerical payoff obtained by the corresponding player. For example, when Player 1 plays action $C$ and Player 2 plays action $D$, the action profile is ($C$, $D$) and the corresponding payoffs of players are, respectively, $-1$ and 3. To the so-called cooperative outcome ($C$, $C$) there corresponds the payoff profile (2, 2); and to the non-cooperative outcome ($D$, $D$) there corresponds the payoff profile (0, 0).

A combination of strategies, where each player always plays $D$ no matter what the other does is a stationary Nash equilibrium in the repeated Prisoner's Dilemma. As it was mentioned above, the majority of multiagent algorithms for repeated games, with the exception of the algorithm by Littman and Stone (2005), content themselves with this sort of solution.

Now let us suppose that both players playing Prisoner's Dilemma know that the game with the same opponent will continue forever (or that the probability of continuation is close enough to 1). Additionally, let us suppose that when choosing between actions $C$ or $D$, Player $i \in \{1, 2\}$, knows that its opponent will start by playing $C$, but whenever Player $i$ plays $D$, the opponent will start playing $D$ until Player $i$ reverts back to playing $C$. Such opponent's behavior is usually referred to in the literature as the 'Tit-For-Tat' strategy (TFT). If, before the game starts, both players were told that the opponent will behave according to TFT, no player would probably[2] try to play $D$. The reason for this would be that an infinite sequence of cooperative outcomes generates for each player an infinite sequence of that player's payoffs (2, 2, 2, 2,...). This corresponds to an average per stage payoff of 2. On the other hand, the player's behavior in which it plays $D$ once and then reverts to playing $C$ forever, yields the average payoff (AP) inferior to 2, because the sequence of the player's payoff will be (3, $-1$, 2, 2,...). Finally, playing $D$ forever will correspond to the sequence of payoffs (3, 0, 0, 0,...) whose AP tends to 0. It is easy to verify that no other behavior in which some player plays $D$ can bring to this player an AP superior or equal to 2 (assuming that the other player follows TFT). In such a situation, the rationality principle dictates playing $C$ whenever the other player does the same. Observe that in this repeated game, the pair of TFT strategies is also a Nash equilibrium: when one player behaves according to TFT, the other one cannot do better than

[2] This is true for certain choices of the players' long-term payoff criteria. Different criteria will be described further.

playing according to TFT as well. Furthermore, the per stage payoff of each player in this Nash equilibrium, 2, is higher than the payoff of the stationary Nash equilibrium, that is, 0.

Repeated game theory studies the cases similar to the above example. It provides theorems about the existence and the properties of equilibria similar to TFT in different settings. These settings vary upon the long-term payoff criteria adopted by the players, the observability by the player of the actions made by the opponents, and the patience of the players (i.e. how they value the future payoffs, 'promised' by the proposed behavior, compared with the present stage-game payoff). Furthermore, a number of existing approaches permit constructing, in different games, the strategies in the spirit of TFT (Abreu, 1988; Judd *et al.*, 2003; Littman & Stone, 2005; Burkov & Chaib-draa, 2010).

In this overview paper, we will talk about the solution notions of a repeated game; we see the latter as a general model for repeated multiagent interactions. We will present a number of important theoretical results originating from economics, the so-called folk theorems. The main property that distinguishes our overview of repeated game theory from a number of the existing references (Aumann, 1981; Pearce, 1992; Benoit & Krishna, 1999; Mailath & Samuelson, 2006; Gossner & Tomala, 2009), is its strict focus on the applicability of the existing theory to the *artificial agents*, that is, those that use a limited amount of memory and have limited computational capabilities.

## 2 Repeated game

The description of a repeated game starts with a *stage-game* (also referred to as a matrix game or a normal form game).

### 2.1 Stage-game

DEFINITION 1 **(Stage-game)** *A (finite) stage-game is a tuple $\left(N, \{A_i\}_{i\in N}, \{r_i\}_{i\in N}\right)$. In a stage-game, there is a finite set N, $|N| \equiv n$, of individual players that act (i.e. make a move in the game) simultaneously. Player $i \in N$, has a finite set $A_i$ of pure actions in its disposal. When each player i among N chooses a certain action $a_i \in A_i$, the resulting vector $a \equiv (a_1, \ldots, a_n)$ is called an action profile and corresponds to a specific stage-game outcome. Each action profile belongs to the set of action profiles $A \equiv \times_i A_i$. A player-specific payoff function $r_i$ specifies player i's numerical reward for different game outcomes, that is, $r_i : A \mapsto \mathbb{R}$.*

We denote the profile of payoff functions as $r \equiv \times_i r_i$. Given an action profile, $a$, $v = r(a)$ is called a payoff profile. A mixed action $\alpha_i$ of player $i$ is a probability distribution over player's actions, that is, $\alpha_i \in \Delta(A_i)$. The payoff function is extended to mixed actions by taking expectations.

The set of players' stage-game payoffs that can be generated by the pure action profiles is denoted as

$$F \equiv \{v \in \mathbb{R}^n : \exists a \in A \text{ s.t. } v = r(a)\}$$

The set $F^\dagger$ of feasible payoffs is the convex hull of the set $F$, that is, $F^\dagger = \text{co } F$. In other words, $F^\dagger$ is the smallest convex set containing $F$. Feasible payoffs is an important concept in the context of repeated games. Observe that any point belonging to $F^\dagger \backslash F$ is a convex combination of two or more points from the set $F$. Therefore, any expected per stage payoff profile that can be obtained by the players in the repeated game belongs to the set $F^\dagger$ of feasible payoffs.

A payoff profile $v \in F^\dagger$ is *inefficient* if $\exists v' \in F^\dagger$ s.t. $v'_i > v_i$, $\forall i \in N$. Otherwise, $v$ is called Pareto efficient.

An important concept in both MAS and economics is one of *individual rationality*. This concept is closely related to the notion of *minmax*. The *minmax payoff* $\underline{v}_i$ of player $i$ is defined as

$$\underline{v}_i \equiv \min_{\alpha_{-i} \in \times_{j\neq i}\Delta(A_j)} \max_{a_i \in A_i} r_i(a_i, \alpha_{-i})$$

The minmax payoff of player $i$ is the minimal payoff, which it can guarantee itself regardless of the strategies chosen by the opponents. A payoff profile $v$ is called *individually rational* if, for any $i$, $v_i \geq \underline{v}_i$. An important remark is that any player in any game can always obtain in expectation at least its minmax payoff given the knowledge of the opponents' action profile.

A mixed action minmax profile for player $i$, $\underline{\alpha}^i = (\underline{a}^i_i, \underline{\alpha}^i{}_{-i})$, is a mixed action profile with the property that $\underline{a}^i_i$ is player $i$'s stage-game best response to the mixed action profile $\underline{\alpha}^i{}_{-i}$ of the remaining players, and $r_i(\underline{\alpha}^i) = \underline{v}_i$.

We define the set of *feasible and individually rational* payoffs and the set *feasible and strictly individually rational* payoffs as, respectively, $F^{\dagger *} \equiv \{v \in F^\dagger : v_i \geq \underline{v}_i \ \forall i \in N\}$ and $F^{\dagger +} \equiv \{v \in F^\dagger : v_i > \underline{v}_i \ \forall i \in N\}$.

In certain situations, the players cannot be allowed to randomize over their actions. In this case, the *pure minmax payoff* of player $i$, is defined as

$$\underline{v}^p_i \equiv \min_{a_{-i} \in \times_{j \neq i} A_j} \max_{a_i \in A_i} r_i(a_i, a_{-i}).$$

A pure action minmax profile for player $i$, $\underline{a}^i = (\underline{a}^i_i, \underline{a}^i{}_{-i})$, is an action profile with the property that $\underline{a}^i_i$ is player $i$'s stage-game best response to the action profile $\underline{a}^i{}_{-i}$ of the remaining players, and $r_i(\underline{a}^i) = \underline{v}^p_i$.

Similarly, we define the set of *feasible and pure individually rational* payoffs as $F^\dagger \equiv \{v \in F^\dagger : v_i \geq \underline{v}^p_i \forall i \in N\}$ and the set of *feasible and strictly pure individually rational* payoffs as $F^{\dagger + p} \equiv \{v \in F^\dagger : v_i > \underline{v}^p_i \ \forall i \in N\}$.

Finally, we define the two most restrictive sets of individually rational payoffs, the set of *pure individually rational* payoffs and the set of *strictly pure individually rational* payoffs, respectively, as $F^{*p} \equiv \{v \in F : v_i \geq \underline{v}^p_i \ \forall i \in N\}$ and $F^{+p} \equiv \{v \in F : v_i > \underline{v}^p_i \ \forall i \in N\}$.

A number of theoretical results existing in the literature (Osborne & Rubinstein, 1994; Mailath & Samuelson, 2006) have only been proven when the players' payoffs were assumed to lie in the more restricted sets, such as $F^{\dagger * p}$ or $F^{+p}$. Further throughout the paper, we will always assume players to be individually rational in either sense. The specific context will be indicated where necessary.

For the example of Prisoner's Dilemma shown in Figure 1, the four sets $F$, $F^\dagger$, $F^{\dagger +}$ and $F^{\dagger *}$ are depicted in Figure 2. The set $F$ of pure action payoffs includes four bold dots denoted as $r(C, D)$, $r(C, C)$, $r(D, C)$ and $r(D, D)$ These dots represent the payoff profiles for the respective pure action profiles. The set $F^\dagger$ of feasible payoffs is the whole diamond-shaped area formed by the four dots and the bold lines that connects them. The set $F^{\dagger *}$ of feasible and individually rational

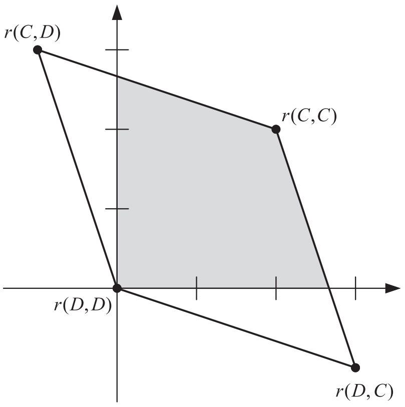


**Figure 2** The sets $F$, $F^\dagger$, $F^{\dagger *}$, and $F^{\dagger +}$ in the Prisoner's Dilemma from Figure 1. The set $F$ includes four bold dots denoted as $r(C, D)$, $r(C, C)$, $r(D, C)$, and $r(D, D)$. The set $F^\dagger$ is the whole diamond-shaped area formed by the four dots and the bold lines that connects them. The set $F^{\dagger *}$ is shown as the shaded sector inside this diamond-shaped area including the bounds. The set $F^{\dagger +}$ is a subset of $F^{\dagger *}$ that excludes the axes

payoffs is shown as the shaded sector inside this diamond-shaped area including the bounds. The set $F^{\dagger+}$ of feasible and strictly individually rational payoffs is a subset of $F^{\dagger*}$ that excludes the axes. As one can see, in Prisoner's Dilemma, the sets $F^{\dagger+p}$ and $F^{\dagger*p}$ coincide with, respectively, $F^{\dagger+}$ and $F^{\dagger*}$.

### *2.2 Repeated game*

In the (infinitely) repeated game, the same stage-game is played in periods $t = 0, 1, 2, \ldots$, also called repeated game iterations, repetitions or stages. At the beginning of each iteration, the players choose their actions, which consequently form an action profile. Then they simultaneously play this action profile, and collect the stage-game payoffs corresponding to the resulting stage-game outcome. Then repeated game passes to the next iteration.

The set of repeated game *histories* up to iteration $t$ is given by $H^t \equiv \times_t A$. The set of all possible histories is given by $H = \bigcup_{t=0}^{\infty} H^t$. For instance, a history $h^t \in H$ is a stream of action profiles arose starting from period 0 up to period $t-1$:

$$h^t = (a^0, a^1, a^2, \ldots, a^{t-1})$$

A pure strategy $\sigma_i$ of player $i$ in the repeated game is a mapping from the set of all possible histories to the set of player $i$'s actions, that is, $\sigma_i : H \mapsto A_i$. A mixed strategy of player $i$ is a mapping $\sigma_i : H \mapsto \Delta(A_i)$. Like in the stage-games, a pure strategy is a special case of a mixed strategy[3].

A subgame (or, continuation game) of an original repeated game is a repeated game based on the same stage-game as the original repeated game but started from a given history $h^t$. Imagine a subgame induced by a history $h^t$. The behavior of players in this subgame after a history $h^\tau \in H$ will be identical to the behavior of players in the original repeated game after the history $h^t h^\tau$, where $h^t h^\tau \equiv h^t \cdot h^\tau \equiv (h^t, h^\tau)$ is a concatenation of two histories. Given a strategy profile $\sigma \equiv (\sigma_i)_{i \in N}$ and a history $h^t$, we denote the subgame (or, continuation) strategy profile induced by $h^t$ as $\sigma|_{h^t}$.

An outcome path in a repeated game is an infinite stream of action profiles $\mathbf{a} \equiv (a^0, a^1, \ldots)$. A finite prefix of length $t$ of an outcome path corresponds to a history in $H^{t+1}$. A profile $\sigma$ of strategies of players induces an outcome path $\mathbf{a}(\sigma) \equiv \left(\mathbf{a}^0(\sigma), \mathbf{a}^1(\sigma), \mathbf{a}^2(\sigma), \ldots\right)$ as follows:

$$\begin{aligned} &a^0(\sigma) \sim \sigma(\varnothing), \\ &a^1(\sigma) \sim \sigma\big(a^0(\sigma)\big), \\ &a^2(\sigma) \sim \sigma\big(a^0(\sigma), a^1(\sigma)\big), \\ &\ldots, \end{aligned}$$

where we denote by $a^t(\sigma) \sim \sigma(h^t)$ the action profile played by the players at iteration $t$ after the history $h^t$ according to the strategy profile $\sigma$. Obviously, in any two independent runs of the same repeated game, a pure strategy profile deterministically induces the same outcome path. On the contrary, at each iteration $t$, the action profile $a^t(\sigma)$ belonging to the outcome path induced by a mixed strategy profile $\sigma$ is a realization of the random process $\sigma(h^t)$.

In order to compare two repeated game strategy profiles in terms of the payoff they induce to a player, we need a criterion permitting comparing infinite payoff streams. The literature (Myerson, 1991; Osborne & Rubinstein, 1994) usually suggests two criteria: (i) the average payoff (AP) criterion, called also 'the limit of the means' criterion, and (ii) the discounted AP (DAP) criterion.

Notice that to an infinite outcome path $\mathbf{a} = (a^0, a^1, \ldots)$, there uniquely corresponds an infinite sequence $\mathbf{v} = (v^0, v^1, \ldots)$ of stage-game payoff profiles. We can now introduce the notion of a long-term payoff criterion.

[3] Another definition of a mixed strategy is also possible. If we denote by $\Sigma_i$ the set of all pure strategies available to player $i$, then player $i$'s mixed strategy can be defined as a mapping $\rho_i : H \mapsto \Sigma_i$. It can be verified that these two definitions are equivalent (Mertens *et al.*, 1994).

DEFINITION 2 **(Average payoff criterion)** *Given an infinite sequence of payoff profiles* $\mathbf{v} = (v^0, v^1, \ldots)$, *the* average payoff (AP) $u_i(\mathbf{v})$ *of this sequence, for player i, is given by*

$$u_{\mathrm{i}}(\mathbf{v}) \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} v_i^t \tag{1}$$

DEFINITION 3 **(Discounted average payoff criterion)** *Given an infinite sequence of payoff profiles* $\mathbf{v} = (v^0, v^1, \ldots)$, *the* discounted average payoff (DAP) $u_i(\mathbf{v})$ *of this sequence, for player i, is given by*

$$u_i(\mathbf{v}) \equiv (1-\gamma)\sum_{t=0}^{\infty} \gamma^t v_i^t \tag{2}$$

where $\gamma \in [0, 1)$ is a so-called discount factor[4]. Observe that the DAP is normalized by the factor $(1 - \gamma)$. After the normalization, the player's payoffs computed according to the first and to the second criteria can be compared both between them and with the payoffs of the stage-game. Observe that regardless of the criterion, $u_i(\mathbf{v}) \in F^{\dagger}$, for any instance of $\mathbf{v}$. Notice that if a sequence of payoff profiles $\mathbf{a}$ corresponds to an outcome path $\mathbf{v}$, we can interchangeably and with no ambiguity write $u_i(\mathbf{v})$ and $u_i(\mathbf{a})$ referring to the same quantity.

There are two ways to interpret the discount factor $\gamma$. The first interpretation can for convenience be called 'economic'. The idea behinds it is the following. It has been observed by economists that individuals (or groups of individuals, such as private companies) value their current well-being, or the well-being in the near term, substantially more than in the long-term. Thus, for the economists, the power of the discount factor permits reflecting this phenomenon. Another interpretation, which is mathematically equivalent to the first one, can for convenience be called 'natural'. According to it, the discount factor $\gamma \in [0,1)$ is viewed as a probability that the repeated game will continue at the next iteration (similarly, $(1 - \gamma)$ can be viewed as the probability that the repeated game stops after the current iteration). This explanation is more convenient for artificial agents because it is generally questionable whether they have to value the future in a similar way as the humans do. The probability of continuation, in turn, seems to be more 'natural' because the machine has always a non-zero probability of fault at any moment of time.

If the economic interpretation of the discount factor is chosen, the discount factor is often called the player's *patience*. Each player is therefore supposed to have its own value of the discount factor. However, it appears that the main body of the research on the repeated games is done assuming that the players are equally patient (or equally impatient). The results we present in this paper are also based on this assumption. Different approaches can be found in Fudenberg *et al.* (1990), Lehrer and Pauzner (1999) and Lehrer and Yariv (1999).

To compare the strategy profiles, similar equations can be used. Let $\sigma$ be a pure strategy profile. Then the AP for player $i$ of the strategy profile $\sigma$ is given by

$$u_i(\sigma) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} r_i(a^t(\sigma))$$

The corresponding DAP can be defined as

$$u_i(\sigma) = (1-\gamma)\sum_{t=0}^{\infty} \gamma^t r_i(a^t(\sigma)) \tag{3}$$

As usual, when the players' strategies are mixed, one should take an expectation over the realized outcome paths.

[4] In the notation $\gamma^t$, $t$ is the power of $\gamma$ and not a superscript.

### 2.3 *Nash equilibrium*

In order to act effectively in a given environment, an agent needs to have a strategy. When we talk about rational agents, this strategy has to be optimal in the sense that it should maximize the agent's expected payoff with respect to the properties of the environment. In the single agent case, it can often be assumed that the properties of the environment do not change in response to the actions executed by the agent (Sutton & Barto, 1998). In this case, it is said that the environment is stationary. Obviously, in order to act optimally in a stationary environment, the agent has to solve the following optimization problem:

$$\sigma_i = \max_{a_i \in A_i} \mathrm{E}_{a_j \sim \alpha_j}\left[r_i\left(a_i, a_j\right)\right]$$

where $j$ denotes the environment as if it was a player always playing a mixed action $\alpha_j$.

When an agent plays a game with the other rational agents, it needs to optimize in the presence of the other optimizing players. This makes the problem non-trivial, since an optimal strategy for one player depends on the strategies chosen by the other players. When the opponents can change their strategies, the player's strategy cannot usually remain constant in order to remain optimal.

We have already seen that one important property for finding a solution in such decision problems is the individual rationality. In other words, given the strategies chosen by the opponent players, the strategy chosen by a rational agent should at least provide a payoff that is not lower than that player's minmax payoff. Recall that the minmax strategy for player $i$, that is, the strategy guaranteeing player $i$ at least its minmax payoff, is defined as follows:

$$\underline{\sigma}_i = \arg\max_{a_i \in A_i} r_i(a_i, \alpha_{-i})$$

$$\text{s.t.}\, \alpha_{-i} = \arg\min_{\alpha_{-i} \in \times_{j \neq i} \Delta(A_j)} \max_{a_i \in A_i} r_i(a_i, \alpha_{-i})$$

However, the fact that player $i$ plays its minmax strategy does not imply that the other players optimize with respect to player $i$. The concept of *equilibrium* describes strategies in which all players' strategic choices simultaneously optimize with respect to each other.

DEFINITION 4 **(Nash equilibrium)** *A strategy profile $\sigma$, such that $\sigma \equiv (\sigma_i, \sigma_{-i})$, is a Nash equilibrium if for all players $i \in N$ and strategies, $\sigma'_i$,*

$$u_i(\sigma) \geq u_i\left(\sigma'_i, \sigma_{-i}\right)$$

In other words, in the equilibrium no player can unilaterally change its strategy so as to augment its own utility.

One can wonder whether the Nash equilibrium strategy profile is a satisfactory concept with respect to the notion of individual rationality? Recall that any strategy profile is only satisfactory (in the sense of individual rationality) if the payoffs proposed to each agent by that strategy profile are not less than the agent's minmax payoff. The following lemma answers this question positively.

LEMMA 1 *If $\sigma$ is a Nash equilibrium, then for all i, $u_i(\sigma) \geq \underline{v}_i$.*

*Proof.* Let $\sigma$ be a strategy profile having a property of Nash equilibrium. After any history $h^t$, each player can simply play its best response to the action profile $a^t_{-i}(\sigma)$ of the other players. Such a strategy will bring to player $i$ a payoff of at least $\underline{v}_i$. But since, in equilibrium, no player can change its strategy so as to get its own utility increased, than $i$'s payoff in the equilibrium is at least $\underline{v}_i$. □

In conjunction with the notion of individual rationality, another notion is important when we talk about the strategies in repeated games. This is the notion of *sequential rationality* or *subgame-perfection*. First of all, let us formally define it.

| | | Player 2 | |
|---|---|---|---|
| | | $C$ | $D$ |
| Player 1 | $C$ | $2, 2$ | $-1, 3$ |
| | $D$ | $3, -1$ | $0, -2$ |

**Figure 3** A game in which a profile of two grim trigger strategies is not a subgame-perfect equilibrium

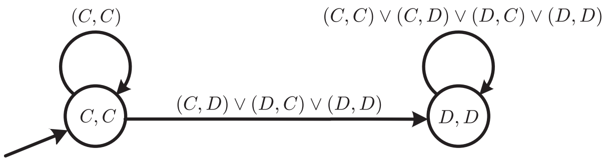
DEFINITION 5 **(Subgame-perfect equilibrium)** *A strategy profile $\sigma$ is a subgame-perfect equilibrium (SPE) in the repeated game if for all histories $h^t \in H$, the subgame strategy $\sigma|_{h^t}$ is a Nash equilibrium of the continuation game.*

Now let us informally explain why the notion of subgame-perfection is of such high importance in the repeated games. Consider a grim trigger strategy. This strategy is similar to TFT in the sense that two players start by playing $C$ in the first period and continue playing $C$ until the other player deviates. The difference with TFT is how the players act in the case of the opponent's deviation. In grim trigger, if the opponent deviates then starting from the next iteration the player always plays $D$ regardless the subsequent actions of the deviator. Let the game be as shown in Figure 3. Observe that in this game, the only reason why each player continues preferring to play the cooperative action $C$ while the opponent plays $C$ is that the profile of two grim trigger strategies is a Nash equilibrium. For example, let suppose that the players' payoff criterion in this game is AP 1. Let Player 1 'think' about a possibility of deviation to the action $D$ when Player 2 is supposed to play $C$. Player 1 knows that according to the strategy profile $\sigma$ (which is a profile of two grim trigger strategies) starting from the next iteration, Player 2 will play $D$ infinitely often. Thus, according to the AP criterion, after only one iteration at which the profile $(D, D)$ is played following the deviation, Player 1 looses all the additional gain it obtains owing to the deviation.

Now, let suppose that Player 1 still decides to deviate after a certain history $h^t$. It plays $D$ whenever Player 2 plays $C$ and gains the payoff of 3 instead of 2. The repeated game enters into the subgame induced by the history $h^{t+1} = (h^t, (D, C))$. Now, according to the strategy profile $\sigma|_{h^{t+1}}$ Player 2 is supposed to always play $D$ say, in order to 'punish' the deviator, Player 1. However, observe the rewards of Player 2. If it always plays $D$ as prescribed by the Nash equilibrium, it certainly obtains the AP of $-2$ in the subgame, since the rational opponent will optimize with respect to this strategy. But if it continues playing $C$, it obtains the average playoff of $-1$ in the subgame, while its opponent, the deviator, will continue enjoying the payoff of 3 at each subsequent period. As one can see, even if after the equilibrium histories the profile of two grim trigger strategies constitutes a Nash equilibrium, it cannot be a Nash equilibrium in an *out-of-equilibrium* subgame. Thus, due to this simple example it becomes clear why, in order to implement Nash equilibria in practice, one needs to have recourse to the subgame-perfect equilibria. While one rational player should have no incentive to deviate being informed about the strategy prescribed to the opponents (the property of Nash equilibrium), its rational opponents, in turn, need to have incentives to follow their prescribed strategies *after* the player's deviation (the property of subgame-perfection).

LEMMA 2 *A subgame-perfect equilibrium always exists.*

*Proof.* Consider a strategy profile $\sigma$ that prescribes playing the same Nash equilibrium of the stage-game after any history of the repeated game. According to Nash (1950), in any stage-game there exists such an equilibrium. By the definition of the latter, when player $i$ plays its part of a stage-game Nash equilibrium, it plays its immediate best response to the mixed action of the other players. At the same time, by the definition of $\sigma$, the future play is independent of the current actions. Therefore, playing a stage-game Nash equilibrium at every iteration will also be a best response in the repeated game after any history. The latter observation satisfies the property of subgame-perfection. □



$(C,C)$
$(C,C) \vee (C,D) \vee (D,C) \vee (D,D)$
$C,C$
$(C,D) \vee (D,C) \vee (D,D)$
$D,D$

**Figure 4** An example of an automaton implementing a profile of two grim trigger strategies. The circles are the states of the automaton; they are labeled with the action profiles prescribed by the profiles of decision functions. The arrows are the transitions between the corresponding states; they are labeled with outcomes

The two most important questions about subgame-perfect equilibria are (1) what is the set of subgame-perfect equilibrium payoffs of a repeated game; and (2) what can be the form of a subgame-perfect equilibrium strategy profile. We will tend to answer both questions in the following subsections. But first, we need to introduce a way to represent strategy profiles.

### *2.4 Strategy profiles as finite automata*

By definition, a player's strategy is a mapping from an infinite set of histories into the set of player's actions. In order to construct a strategy for an artificial agent (which is usually bounded in terms of memory and performance), one needs a way to specify strategies by means of finite representations.

Intuitively, one can see that, given a strategy profile $\sigma$, two different histories $h^t$ and $h^\tau$ can induce identical continuation strategy profiles, that is, $\sigma|_{h^t} = \sigma|_{h^\tau}$. For example, in the case of TFT strategy, agents will have the same continuation strategy both after the history $((C, C), (C, C))$ and after the history $((D, C), (C, D), (C, C))$. One can put all such histories into the same equivalence class. If one views these equivalence classes of histories as players' states, then a strategy profile can be viewed as an automaton.

Let $M \equiv (Q, q^0, f, \tau)$ be an *automaton implementation of a strategy profile* $\sigma$. $M$ consists of (i) a set of states $Q$, with the initial state $q^0 \in Q$; (ii) a profile of decision functions $f \equiv \times_{i\in N} f_i$, where the decision function of player $i$, $f_i : Q \mapsto \Delta(A_i)$, associates mixed actions with states, and; (iii) a transition function $\tau : Q \times A \mapsto Q$, which identifies the next state of the automaton given the current state and the action profile played in the current state.

Let $M$ be an automaton. In order to demonstrate how $M$ induces a strategy profile, one can first recursively define $\tau(q, h^t)$, the transition function specifying the next state of the automaton given its initial state $q$ and a history $h^t$ that starts in $q$, as

$$\begin{cases} \tau(q,h^t) & \equiv \tau\big(\tau(q,h^{t-1}),a^{t-1}\big), \\ \tau(q,h^1) & \equiv \tau\big(q,a^0\big) \end{cases}$$

With the above definition in hand, one can define $\sigma_i^M$, the strategy of player $i$ induced by the automaton $M$, as

$$\begin{cases} \sigma_i^M(\varnothing) & \equiv f_i\big(q^0\big), \\ \sigma_i^M(h^t) & \equiv f_i\big(\tau\big(q^0,h^t\big)\big) \end{cases}$$

An example of a strategy profile implemented as an automaton is shown in Figure 4. This automaton implements the profile of two grim trigger strategies. The circles are the states of the automaton. The arrows are the transitions between the corresponding states; they are labeled with outcomes. The states are labeled with the action profiles prescribed by the profiles of decision functions.

Since any automaton induces a strategy profile, any two automata can be compared in terms of the utility they bring to the players. Let an automaton $M$ induce a strategy profile $\sigma^M$. The utility $u_i$ $(M)$ of the automaton $M$ for player $i$ is then equal to $u_i$ $(\sigma^M)$, where $u_i$ $(\sigma^M)$ is given by Equation (3).

Let $|M|$ denote the number of states of automaton $M$. If $|M|$ is finite, then $M$ is called a *finite automaton*; otherwise the automaton is called *infinite*. In MAS, most of the time, we are interested in finite automata, because artificial agents always have a finite memory to stock their strategies and a finite processing power to construct them.

Note that any finite automaton induces a strategy profile, however, not any strategy profile can be represented using finite automata. Kalai and Stanford (1988) demonstrated that any SPE can be *approximated* with a finite automaton.

Note also that an automaton implementation of a strategy profile can be naturally split into a set of individual player automatons. Each such automaton will share the same set of states and the same transition function, and only decision functions will be individual. Below we will sometimes consider players' individual automata as well.

Before presenting important theoretical results about repeated games, let us first outline a taxonomy of different settings, in which the model of repeated games can be applied.

## 3 Taxonomy

To construct a taxonomy of repeated game settings, let us enumerate a number of different assumptions that can affect the decision making in the repeated games. These assumptions are due to various conditions of the real-world environment, which we model as a repeated game. Players' patience, their payoff criteria, ability to execute mixed strategies, observability of the opponents' actions, and knowledge of their payoff functions are all examples of different conditions of the real-world environment. To reflect those conditions, there exists a set of formal assumptions that can be integrated into the model of repeated games. Approaches to the analysis of repeated games under different assumptions can also be quite different. The present overview paper covers only a small subset of all possible repeated game settings considered in the literature.

*Number of iterations*: Repeated games differ by the expected number of game iterations. One can distinguish *finitely repeated games* and *infinitely repeated games*. In the finitely repeated games (Benoit & Krishna, 1985), the horizon $T$ of the repeated game (i.e. the number of stage-game repetitions) is supposed to be finite, fixed, and known by the players before the repeated game starts. The analysis and the solutions in the finitely repeated games have quite specific properties due to the fact that one can use backward induction to compute game solutions.

*Payoff criteria*: Infinitely repeated games can be distinguished from the point of view of long-term payoff criteria adopted by the players. In previous section, we have already seen that players can have the AP criterion or the DAP criterion. In addition to the above two widely used criteria, Rubinstein (1979) studies equilibrium properties in repeated games with another payoff criterion, called an *overtaking criterion*.

*Player patience*: In infinitely repeated games with the DAP criterion, players are considered *patient* if their discount factor is close to 1. On the other hand, players are said to be *impatient* if $0 < \gamma \ll 1$. Fudenberg *et al.* (1990) study a case where patient and impatient players are playing together.

*Game information*: Players can have different knowledge of the game. For instance, if all players know their own payoff functions and those of the other players, such game is said to be of *complete information* (Sorin, 1986). On the other hand, if all players are uncertain about either of game properties, such game is said to be of *incomplete information* (Rasmusen, 1994; Aumann *et al*., 1995). One can also distinguish an intermediate situation, in which players having a complete information about the game they play are playing together with players having only a partial information about certain game properties (Lehrer & Yariv, 1999; Laraki, 2002).

*Game monitoring*: Players can be supposed to either perfectly observe the actions executed by the other players, observe them with an uncertainty, or not to observe them at all. When the actions of the other players are observable, a repeated game is said to be of *perfect monitoring*. A special case of perfect monitoring is the observability of the opponent players' *mixed actions*.

When the players receive only a *stochastic signal* giving only an imprecise idea of the actions taken by the opponents, a repeated game is said to be of *imperfect monitoring*. One can also distinguish between the games of *imperfect public monitoring*, where all players receive the same signal characterizing the executed action profile (Abreu *et al.*, 1990; Fudenberg *et al.*, 1994); and the games of *imperfect private monitoring*, where the signals received by the different players are different but correlated (Piccione, 2002; Matsushima, 2004).

*Strategic complexity*: One can distinguish between the basic properties of the strategies that can be implemented by players. We have already seen that strategies can be defined using either *finite* or *infinite representations*. An example of a finite strategy representation is a finite automaton (Neyman, 1985; Abreu & Rubinstein, 1988; Kalai & Stanford, 1988). An example of a non-finite representation of a strategy profile can be found in Judd *et al.* (2003), where, in order to choose an action at a given iteration, each player has to solve a linear program corresponding to a certain payoff profile.

Players' strategies also generally have a less complex structure if the players are only allowed to adopt pure strategies (Abreu, 1988); on the other hand, strategy profiles usually have a more complex structure if mixed actions are allowed (Burkov & Chaib-draa, 2010).

*Public correlation*: An important special case includes repeated games with public correlation. In this case, an independent correlating device is available and it is capable of transmitting a random signal to the players at the beginning of each game period. It permits convexifying the set of payoff profiles without having recourse to complex sequence of outcomes (Mailath & Samuelson, 2006, p. 17). Possibility to convexify the set of payoff profiles is an important property permitting to considerably simplify both the study of subgame-perfect equilibrium payoffs of repeated game and the construction of the corresponding strategy profiles (Judd *et al.*, 2003). A correlating device can be viewed as a special player whose rewards are independent of the actions of the other players. Alternatively in two-player games, a special communication protocol can simulate a correlating device (Aumann *et al.*, 1995).

In this overview, we are primarily focused at the setting characterized by the infinitely repeated games with complete information and perfect monitoring. The players are supposed to be sufficiently patient ($\gamma$ is close enough to 1), capable of using only finite representations for their strategies and having either the average or the discounted average long-term payoff criterion. Our results presented in Section 4 are all based on these assumptions. In Section 5, we give a brief overview of important results for a number of other repeated game settings. For instance, we consider a setting described by the imperfect monitoring. Contrary to repeated games of perfect monitoring, which are relatively well studied, there exist considerably less results about games of imperfect monitoring, especially for the most complicated, private monitoring case. In particular, we do not know about any result concerning a finite strategic complexity in repeated games of imperfect monitoring.

Now let us first present a number of important results characterizing the repeated games of perfect monitoring.

## 4 Perfect monitoring

### 4.1 The folk theorems

There is a cluster of results, under the common name of *folk theorems* (FT), characterizing the set of payoffs of Nash and subgame-perfect equilibria in a repeated game. In this section, we will present only a few of these results. The reason is that the proofs of all folk theorems are conceptually similar, while certain are too space consuming for an overview paper.

As we have mentioned in the beginning of this paper, in order to use a strategy profile in practice, two notions are important: individual rationality and subgame-perfection. As we have seen, any Nash equilibrium is individually rational but not necessarily subgame-perfect. Recall that the property of subgame-perfection guarantees that when any player *does actually* deviate from the long-term equilibrium strategy in favor of a short-term gain, the other players *will indeed* prefer to continue following their prescribed strategies. The two following *perfect folk theorems* (PFT) characterize the set of payoff profiles of subgame-perfect equilibria.

THEOREM 1 **(PFT with the AP criterion)** *Every payoff profile $v \in F^{\dagger+}$ is a payoff profile of a certain subgame-perfect equilibrium in the repeated game.*

*Proof.* See Osborne and Rubinstein (1994, p. 147) for a proof for the case of pure strategies, that is, $v \in F^{\dagger+p}$ That proof can easily be extended to mixed strategies. Here, let us briefly outline the intuition behind the proof. The proof is by construction. We first construct an automaton in which there are basically two types of states: cooperative states and punishment states. The cooperative states form a cycle. When all players follow this cycle, they obtain the AP $v$. If one player deviates, the other players punish the deviator by passing through a finite sequence of punishment states in which the minmax action profile for deviator is played. It is remaining to compute the length of the sequence of punishment states so that the maximal gain the deviator can obtain by deviating from the cooperative cycle is vanished by the punishment. As soon as the final punishment state is reached and the corresponding minmax profile is played, the automaton transits to the beginning of the cooperative cycle. Since the punishment sequence is finite, any loss one player endures when it punishes the other player becomes infinitesimal as the number of repeated game periods tends to infinity. This justifies that the deviator will indeed be punished by the remaining players, which is, in turn, a property of subgame-perfection. □

For repeated games with the DAP criterion, there is no corresponding *exact* folk theorem with the players' strategies being represented with *finite* automata. In other words, there does not exist a theorem specifying the equilibrium payoffs laying in $F^{\dagger+}$. However, there are three mutually complementary results that can be satisfactory in many situations.

The first such result (Theorem 2) establishes an exact folk theorem for the case of feasible and *pure* strictly individually rational payoffs (i.e. $v \in F^{\dagger+p}$). This result is suitable for the case when the players are limited to use pure strategies. The two remaining results (Theorems 4 and 5) state that (i) any feasible and strictly individually rational payoff profile in repeated games with the DAP criterion is a payoff profile of a certain subgame-perfect equilibrium whenever the strategies are representable by *infinite* automata; and (ii) any subgame-perfect equilibrium strategy profile can be approximated to an arbitrary precision using finite automata.

A payoff profile $v$ is said to be an *interior* feasible payoff profile if $v$ is an interior point of $F^{\dagger}$, that is, $v \in \operatorname{int} F^{\dagger}$.

THEOREM 2 **(PFT with the DAP criterion)** *Let v be an interior feasible and strictly pure individually rational payoff profile in a repeated game with the DAP criterion. For all $\varepsilon > 0$ there exist a discount factor $\underline{\gamma} \in (0, 1)$ and a payoff profile $v'$ for which $\forall i \left| v'_i - v_i \right| < \varepsilon$, such that for any $\gamma \in (\underline{\gamma}, 1)$, $v'$ is a payoff profile of a certain subgame-perfect equilibrium.*

*Proof.* For the proof of this theorem formulated in a slightly different form, see Ben-Porath and Peleg (1987). In this paper, we only present the proof of Theorem 3, a simplified and restricted version of Theorem 2. Theorem 3 focuses on the payoff profiles from the set $F^{+p}$, that is, all strictly pure individually rational payoff profiles that can be generated by pure action profiles. □

Before introducing perfect folk theorems for DAP criterion, let us first introduce the *one-shot deviation principle* (Fudenberg & Tirole, 1991).

DEFINITION 6 *Given a strategy profile $\sigma = (\sigma_i, \sigma_{-i})$, a one-shot deviation strategy for player i from strategy $\sigma_i$ is a strategy $\hat{\sigma}_i$ with the property that there exists a unique history $\hat{h}^t$ such that*

$$\begin{cases} \hat{\sigma}_i(h^\tau) \neq \sigma_i(h^\tau), & \text{if } h^\tau = \hat{h}^t, \\ \hat{\sigma}_i(h^\tau) = \sigma_i(h^\tau), & \text{otherwise} \end{cases}$$

*A one-shot deviation for player i is called* profitable *if $u_i(\hat{\sigma}_i|_{\hat{h}^t}, \sigma_{-i}|_{\hat{h}^t}) > u_i(\sigma|_{\hat{h}^t})$*

Recall that $\sigma_i(h^t)$ denotes the (possibly mixed) action player $i$ should do according to the strategy $\sigma$ after the history $h^t$, while $\sigma_i|_{h^t}$ is the complete description of strategy $\sigma_i$ in the subgame induced by the history $h^t$.

At first glance, it may appear that in order to verify whether a given strategy profile $\sigma$ is a subgame-perfect equilibrium, one has to check, for every subgame induced by each possible history $h^t$, whether $\sigma|_{h^t}$ remains Nash equilibrium in this subgame. To do this, one would have to check all possible deviations (typically, an infinity of them) of each player and after each history. However, in many situations the task can be drastically simplified. For instance, for repeated games with the DAP criterion it can be demonstrated that it is sufficient to check only the one-shot deviation.

PROPOSITION 1 **(The one-shot deviation principle)** *In any repeated game with the DAP criterion, a strategy profile is a subgame-perfect equilibrium if and only if for all players and after all histories there is no profitable one-shot deviation.*

*Proof.* For the proof, see Mailath and Samuelson (2006). The principal idea behind the proof is that since the payoffs are discounted, any strategy that offers a higher payoff than an equilibrium strategy, must do so in a finite number of iterations. Then, by using dynamic programming, one can show that when there is a profitable deviation, there should be a profitable one-shot deviation. □

THEOREM 3 **(PFT with DAP and pure action payoff profiles)** *If, in a repeated game with the DAP criterion, there is a payoff profile $v \in F^{+p}$ for which there exists a collection $\{v^j\}_{j \in N}$ of payoff profiles such that (i) $\forall j \in N$, $v^j \in F^{+p}$ and $v^j_j < v_j$, and (ii) $\forall i \neq j, v^j_j < v^i_j$, then there exists a discount factor $\underline{\gamma} \in (0, 1)$ such that for every $\gamma \in (\underline{\gamma}, 1)$, there exists a subgame-perfect equilibrium yielding the payoff profile $v$.*

*Proof.* The complete proof of this theorem can be found in Osborne & Rubinstein (1994, p. 151). As previously, here we will only outline the intuition behind the proof. Let us first explain why the existence of the collection $\{v^i\}_{i \in N}$ with the aforementioned properties is necessary. This collection represents the so-called *player-specific punishments* (Mailath & Samuelson, 2006, p. 82). The basic idea behind the proof of any perfect folk theorem is that one needs to construct a profile of strategies with the property that when any player $i$ deviates from the prescribed 'cooperative' behavior (yielding the payoff $v_i$ to player $i$) then two conditions are satisfied: (i) player $i$ can be successfully punished by the other players (by vanishing its deviation gain) and (ii) any player $j \neq i$ has to prefer to punish $i$ on penalty of being punished itself by the other players (including player $i$, which, thanks to the one-shot deviation principle, can be assumed to return to the equilibrium strategy right after its own deviation). Consequently, in the case of the DAP criterion one not only needs to find the length of the sequence of punishment states for each player, but also the condition under which any deviation from this punishment sequence by any punisher can be successfully punished by the remaining players. □

Now let us for a while consider strategies implementable with infinite automata. In this case, for repeated games with the DAP criterion there exists an exact folk theorem (Theorem 4) having the desired properties. That is, Theorem 4 considers all payoff profiles in the set $F^{\dagger+}$ and not only those that belong to $F^{\dagger+p}$. We will then show how this result can be approximated by reducing the space of player strategies to those implementable by finite automata.

THEOREM 4 **(PFT with DAP and infinite automata)** *For any payoff profile $v \in \{\tilde{v} \in F^{\dagger+} : \exists v' \in F^{\dagger+}, v'_i < \tilde{v}_i \ \forall_i\}$ in a repeated game with the DAP criterion, there exists a discount factor $\underline{\gamma} \in (0,\ 1)$ such that for every $\gamma \in (\underline{\gamma},\ 1)$, there exists a subgame-perfect equilibrium yielding the payoff profile $v$.*

*Proof.* For the complete proof, see Mailath and Samuelson (2006, p. 101). Here, we will only draw the reader's attention to the fact that the formulation of this theorem is similar to that of

Theorem 3. The ideas behind the proof are also similar. To prove this theorem, one needs to recall that any $v \in F^{\dagger}$ can be obtained as a payoff of a (possibly infinite) sequence of pure action profiles (see Mailath & Samuelson, 2006, pp. 97–99). Fix $v \in \{\tilde{v} \in F^{\dagger +} : \exists v' \in F^{\dagger +}, v'_i < \tilde{v}_i \ \forall i\}$. Because $F^{\dagger +}$ is convex, one can always find a vector $v'$ of payoff profiles with the property that for all players $i$, $v'_i < v_i$. Furthermore, one can find an $\varepsilon > 0$ such that for each player $i$ there exists a payoff profile

$$v'^{i} = (v'_1 + \varepsilon, v'_2 + \varepsilon, \ldots, v'_{i-1} + \varepsilon, v'_i, v'_{i+1} + \varepsilon, \ldots, v'_n + \varepsilon)$$

Observe that the collection $\{v'^{i}\}_{i \in N}$ of such payoff profiles specifies player-specific punishments analogous to those required for Theorem 3 with the difference that in the latter theorem, the punishment actions are pure and, therefore, any deviation is detectable. When the punishment actions are allowed to be mixed, deviations are only detectable when they are outside the support of the mixed action minmax[5]. Otherwise, short-term deviations of one player cannot be instantly detected by the other players. To tackle this problem, it is required that each player $i$ were indifferent (in terms of expected payoff) between all its pure actions in the support of its mixed action $\underline{\alpha}^j_i$ in the mixed minmax profile $\underline{\alpha}^j = (\underline{\alpha}^j_i, \underline{\alpha}^j_{-i})$ for any player $j \neq i$. The technique is to specify the continuation play (after punishment is over) so as to obtain the required indifference during punishment. See Mailath and Samuelson (2006, p. 102) for more details. □

Before presenting the approximation result, let us first define the notion of an approximately subgame-perfect equilibrium. A strategy profile $\sigma = (\sigma_i, \sigma_{-i})$ is a Nash $\varepsilon$-equilibrium of a repeated game if $u_i(\sigma) \geq u_i(\sigma'_i, \sigma_{-i}) - \varepsilon$ for each player $i \in N$ and every strategy $\sigma'_i$. Similarly, a strategy profile $\sigma = (\sigma_i, \sigma_{-i})$ is a subgame-perfect $\varepsilon$-equilibrium of a repeated game, if for any $h^t \in H$, $\sigma|_{h^t}$ is a Nash $\varepsilon$-equilibrium of the subgame induced by $h^t$.

THEOREM 5 *Consider a repeated game with the DAP criterion, and let* $\varepsilon > 0$. *Then for any subgame-perfect equilibrium* $\sigma$, *there exists a finite automaton* $M \equiv (M_i)_{i \in N}$ *with the property that* $|u_i(\sigma) - u_i(M)| < \varepsilon$ *for all* $i \in N$, *and such that* $M$ *induces a subgame-perfect* $\varepsilon$*-equilibrium.*

*Proof.* For the detailed proof, see Kalai and Stanford (1988). The key to the proof is the observation that one can partition the convex and, therefore, bounded set of feasible and individually rational payoff profiles (in $\mathbb{R}^n$) into disjoint adjacent $n$-cubes. There will necessarily be a finite set $C$ of such $n$-cubes. Let $\sigma$ be a subgame-equilibrium strategy, which we want to approximate with precision $\varepsilon$. In each $n$-cube $c \in C$, choose a payoff profile, and find the corresponding subgame-perfect equilibrium strategy profile $\sigma^c$. Recall that by definition, for each history $h^t$, $\sigma|_{h^t}$ is a subgame-perfect equilibrium. The approximate joint strategy $g$ is implemented by the automaton $M \equiv (C, c^0, f, \tau)$, where the set $C$ of $n$-cubes plays the role of the automaton states; $c^0$ is the $n$-cube, which the payoff profile $u(\sigma)$ belongs to; $f(\bar{\sigma}, a)$ is the transition function that takes a subgame-perfect equilibrium strategy profile $\bar{\sigma}$ and an action profile $a$, finds $v \equiv u(\bar{\sigma}|_a)$ and then returns the $n$-cube $\bar{c}$, which $v$ belongs to; $\tau(\bar{\sigma}) \equiv \bar{\sigma}(\varnothing)$. Kalai and Stanford (1988) show that when the side of each $n$-cube is of length $(1-\gamma)^2 \varepsilon/2$, the automaton $M$ implements a subgame-perfect $\varepsilon$-equilibrium. □

### 4.2 Constructing an equilibrium strategy

Roughly speaking, the folk theorems tell us that in repeated games ‘everything is possible’ if the *players are sufficiently patient*. For instance, to any feasible and individually rational payoff profile there corresponds a strategy profile which every player will want to follow given that the other players do the same. However, the subject matter of the research in MAS has its inherent practical

[5] The support of a mixed action $\alpha_i \in \Delta(A_i)$ are those pure actions $a_i \in A_i$ that have a non-zero probability in $\alpha_i$.

questions that are not directly addressed by the folk theorems. Here are two examples of such questions. The first question is: does a given collection of players' strategies induce an equilibrium strategy profile? The second question is: how to construct a strategy of player $\underline{i}$, such that the collection of all players' strategies induces an equilibrium strategy profile? In this subsection, we explore several approaches to answering these two questions.

The case of AP is trivial: the players are supposed to be extremely patient (i.e. $\gamma = 1$). For example, a way to construct the player $i$'s automaton yielding any payoff profile in $F^{\dagger +}$ is described in the proof of Theorem 1. Using the same principle, Littman and Stone (2005) describe an efficient algorithm for constructing equilibrium strategy profiles in any two-player repeated game. The opposite case is also trivial: when all players are extremely impatient (i.e. $\gamma = 0$), the set of subgame-perfect equilibrium strategies in the repeated game coincides with the set of stage-game equilibria (recall the proof of Lemma 2). In MAS, most of the time one deals with an intermediate case: the players are neither extremely patient, nor extremely impatient, that is, their discount factor is an interior point of the set (0, 1). The reader interested in a situation where patient players play a repeated game with a set of extremely impatient players is referred to Fudenberg *et al.* (1990).

### *4.2.1 Nash reversion*

As we already mentioned above, any strategy profile prescribing playing a stage-game Nash equilibrium at every iteration of repeated game is a subgame-perfect equilibrium. Such SPE are trivial and can be constructed for any game regardless of the discount factor.

The simplest non-trivial subgame-perfect equilibria can be constructed based on a similar principle, called *Nash reversion* (Friedman, 1971). We have already seen an example of Nash reversion when we considered the grim trigger strategy in Section 2. As in grim trigger, any Nash reversion-based strategy profile $\sigma$ prescribes to start by playing a certain 'cooperative' sequence of action profiles. If either player deviates from the prescribed sequence, all players revert to permanently playing a certain stage-game Nash equilibrium. Therefore, stage-game Nash equilibrium is viewed as a punishment that supports the payoff profile corresponding to the cooperative sequence. Because such punishment is itself a subgame-perfect equilibrium, one can state that $\sigma$ is a subgame-perfect equilibrium whenever no one-shot deviation from the cooperative sequence is profitable (due to the threat of the continuation strategy that prescribes to permanently play a stage-game Nash equilibrium).

Let us develop this argument more formally. As previously, let us limit ourselves to pure strategies. Given a strategy profile $\sigma$, one can rewrite Equation (3) as follows:

$$\begin{aligned} u_i(\sigma) &\equiv (1-\gamma)\sum_{t=0}^{\infty}\gamma^t r_i(a^t(\sigma)) \\ &= (1-\gamma)r_i(a^0(\sigma)) + \gamma\left[\sum_{t=1}^{\infty}\gamma^{t-1}r_i(a^t(\sigma))\right] \\ &= (1-\gamma)r_i(a^0(\sigma)) + \gamma u_i(\sigma|_{a^0(\sigma)}) \end{aligned}$$

Let $v_i(a_i, \sigma|_{h^t})$ denote player $i$'s long-term payoff for playing action $a_i$ after history $h^t$ given the strategy profile $\sigma$. Let $\bar{a} \equiv (\bar{a}_i, \bar{a}_{-i})$ be the action profile prescribed by strategy $\sigma$ after the history $h^t$, that is, $\bar{a} \equiv \sigma(h^t) \equiv \sigma|_{h^t}(\varnothing)$. For all $a_i \in A_i$ one can write:

$$v_i(a_i, \sigma|_{h^t}) = (1-\gamma)r_i(a_i, \bar{a}_{-i}) + \gamma u_i(\sigma|_{h^{t+1}}) \tag{4}$$

where $h^{t+1} = h^t \cdot a$ is obtained as a concatenation of the history $h^t$ and the action profile $a \equiv (a_{i,}\bar{a}_{-i})$ and $u_i(\sigma|_{h^{t+1}})$ represents the so-called *continuation promise* of the strategy $\sigma$ after the history $h^t \cdot a$.

DEFINITION 7 **(Continuation promise)** *The continuation promise of the strategy profile* $\sigma$ *to player* $i$, $u_i(\sigma|_{h^{t+1}})$, *is the utility of the strategy profile* $\sigma$ *to player* $i$ *if the history at the next period is* $h^{t+1}$. *The continuation promise for period* $t+1$ *is computed at period* $t$.

At each iteration of the repeated game, player $i$ has a choice between different actions $a_i$, each proposing to that player a particular long-term payoff $v_i$. Consequently, each iteration of the repeated game can be represented as a certain normal form game whose payoffs equal to the stage-game payoffs augmented by the corresponding continuation promises. Let us call such game an *augmented game*.

Let the stage-game of a repeated game be as shown in Figure 5.

Given a strategy profile $\sigma$ and a history $h^t$, the augmented game corresponding to this stage-game is shown in Figure 6.

We can now reformulate the definition of subgame-perfect equilibrium by saying that a strategy profile $\sigma$ is a subgame-perfect equilibrium if and only if it induces a stage-game Nash equilibrium in augmented games after any history.

DEFINITION 8 *A Nash reversion-based strategy profile is such that the players execute a certain infinite sequence of action profiles unless one agent deviates. Following the deviation, a certain stage-game Nash equilibrium is played at every subsequent period.*

Grim trigger is an example of a Nash reversion strategy. Consider the repeated Prisoner's Dilemma from Figure 1. Let the strategy profile $\sigma$ be a profile of two grim trigger strategies. We have already seen an automaton representation of a profile of two grim trigger strategies in Figure 4.

Consider a history $h^t$ in which all players played $C$ at each iteration. Now, each player has to take a decision whether to play $C$, as prescribed by the strategy, or to play $D$ instead. In particular, we have: $u(\sigma|_{h^t\cdot(C,C)}) = (2,2)$. Because, in the case of deviation, the unique stage-game Nash equilibrium $(C, C)$, whose payoff profile is (0, 0), should be played, we have $u(\sigma|_{h^t\cdot(D,C)}) = u(\sigma|_{h^t\cdot(C,D)}) = u(\sigma|_{h^t\cdot(D,D)}) = (0,\ 0)$. The augmented game corresponding to this situation is shown in Figure 7.

When $\gamma \geq 1/3$, both $(C, C)$ and $(D, D)$ are stage-game Nash equilibria in this augmented game. Observe that the payoff profile corresponding to the outcome $(D, D)$ is inefficient.

Now, consider a history $h^\tau$ in which either of two players has deviated. For this case, grim trigger prescribes permanently playing the action $D$ regardless of the action played by the other player. The augmented game corresponding to this situation is shown in Figure 8. For any value of

| | | Player 2 | |
|---|---|---|---|
| | | $C$ | $D$ |
| Player 1 | $C$ | $r(C,C)$ | $r(C,D)$ |
| | $D$ | $r(D,C)$ | $r(D,D)$ |

**Figure 5** A generic stage-game

| | | Player 2 | |
|---|---|---|---|
| | | $C$ | $D$ |
| Player 1 | $C$ | $(1-\gamma)r(C,C) + \gamma u(\sigma\vert_{h^t\cdot(C,C)})$ | $(1-\gamma)r(C,D) + \gamma u(\sigma\vert_{h^t\cdot(C,D)})$ |
| | $D$ | $(1-\gamma)r(D,C) + \gamma u(\sigma\vert_{h^t\cdot(D,C)})$ | $(1-\gamma)r(D,D) + \gamma u(\sigma\vert_{h^t\cdot(D,D)})$ |

**Figure 6** An augmented game for the generic stage-game from Figure 5

| | | Player 2 | |
|---|---|---|---|
| | | $C$ | $D$ |
| Player 1 | $C$ | $2, 2$ | $-(1-\gamma), 3(1-\gamma)$ |
| | $D$ | $3(1-\gamma), -(1-\gamma)$ | $0, 0$ |

**Figure 7** An augmented game for Prisoner's Dilemma from Figure 1

| | | Player 2 | |
|---|---|---|---|
| | | $C$ | $D$ |
| Player 1 | $C$ | $2(1-\gamma), 2(1-\gamma)$ | $-(1-\gamma), 3(1-\gamma)$ |
| | $D$ | $3(1-\gamma), -(1-\gamma)$ | $0, 0$ |

**Figure 8** An augmented game for Prisoner's Dilemma from Figure 1

| | | Player 2 | | |
|---|---|---|---|---|
| | | $L$ | $M$ | $H$ |
| | $L$ | $10, 10$ | $3, 15$ | $0, 7$ |
| Player 1 | $M$ | $15, 3$ | $7, 7$ | $-4, 5$ |
| | $H$ | $7, 0$ | $5, -4$ | $-15, -15$ |

**Figure 9** Payoff matrix of both players in the Duopoly

$\gamma \in [0, 1)$, the only stage-game Nash equilibrium in this augmented game is $(D, D)$. Putting the above two results together, the profile of two grim trigger strategies is a subgame-perfect equilibrium in the repeated Prisoner's Dilemma if and only if $\gamma \geq 1/3$.

In Prisoner's Dilemma, it is sufficient to study Nash reversion-based strategies, such as grim trigger. This is due to the fact that in this game the only stage-game equilibrium payoff profile coincides with the minmax profile for both players. Therefore, the strategy profile that prescribes playing $(D, D)$ after any history is the most severe subgame-perfect equilibrium punishment available in this game. In other words, any playoff profile that can be supported by any subgame-perfect equilibrium continuation promise can be supported by Nash reversion. However, not all games have such a property. In many games, Nash reversion is not the most severe subgame-perfect equilibrium punishment, and the set of Nash reversion-based equilibria is either empty or excludes some equilibrium payoff profiles.

Another way to construct subgame-perfect equilibrium strategy profiles, appropriate in certain cases, is based on the notion of *simple strategies*. We now consider such strategies.

### 4.2.2 *Simple strategies and penal codes*

For the illustrative purposes, in this subsection we continue restricting ourselves to the case of two players.

If we do not apply restrictions on the size of player's automaton and only suppose that it is finite, the number of different pairs of automata capable of inducing an equilibrium strategy profile grows combinatorially with the number of automaton states. For example, we already mentioned above that any convex combination of payoffs can be attained using an infinite sequence of pure outcomes. Given a payoff profile $v \in F^{\dagger}$, such sequence need not be unique. This results in an overwhelming number of only cooperative states that can make part of player $i$'s automaton. The same is true for punishment states as well.

Thereby, we need to narrow the choice of automata. This can be done in different ways. For example: (1) by applying restrictions on the set of players' payoffs in equilibrium, (2) by focusing on a particular structure of players' strategies, or (3) by restricting the size of players' automata. A natural environment where one can restrict one's attention to pure action payoff profiles, without missing interesting opportunities, is an environment that can be modeled as a symmetric game (Cheng *et al.*, 2004). To illustrate the construction of equilibrium in such setting, we follow an example taken from Abreu (1988). Consider the Duopoly game in Figure 9. The payoff profile $v = (10, 10)$ corresponding to the pure action profile $(L, L)$ is clearly the only Pareto efficient payoff profile in this game, yielding the equal (symmetric) per player payoffs. Let suppose that we want to construct an automaton profile inducing a subgame-perfect equilibrium strategy profile with payoff profile $v$. One of possibilities is to have recourse to *penal codes* (Abreu, 1988). We begin with the concept of a *simple strategy profile*:

DEFINITION 9 *Given a collection of* $(n+1)$ *outcome paths*[6] $(\mathbf{a}(0), \mathbf{a}(1)), \ldots, \mathbf{a}(n))$ *the associated simple strategy profile* $\sigma(\mathbf{a}(0), \mathbf{a}(1), \ldots, \mathbf{a}(n))$ *is given by the following automaton for each player* $i$:

- *Set of states*: $Q_i = \{q(j,t) : j \in \{0\} \cup N, t = 0, 1, \ldots\}$.
- *Initial state*: $q^0 = q(0, 0)$.
- *Output function*: $f(q(j, t)) = a^t(j)$, *and*
- *Transition function*: *for an outcome* $a \in A$.

$$\tau(q(j, t), a) = \begin{cases} q(i, 0), & \textit{if } a_i \neq a_i^t(j) \wedge a_{-i} = a_{-i}^t(j), \\ q(j, t+1), & \textit{otherwise.} \end{cases}$$

Therefore, a simple strategy profile consists of a cooperative outcome path $\mathbf{a}(0)$ and punishment outcome paths $\mathbf{a}(i)$ for each player $i$. A collection $\{\mathbf{a}(i)\}_{i \in N}$ is called a *penal code* and embodies player-specific punishments in the sense of Theorem 3. Using a one-shot deviation principle, one can show that a simple strategy profile $\sigma(\mathbf{a}(0), \mathbf{a}(1), \ldots, \mathbf{a}(n))$ induces a subgame-perfect equilibrium if and only if for any $t$ (Mailath & Samuelson, 2006, p. 52):

$$u_i^t(\mathbf{a}(j)) \geq \max_{a_i \in A_i} (1-\gamma) r_i(a_i, a_{-i}^t(j)) + \gamma u_i^0(\mathbf{a}(i))$$

for all $i \in N$, and $j \in \{0\} \cup N$, and $t = 0, 1, \ldots$, where

$$u_i^t(\mathbf{a}) \equiv (1-\gamma) \sum_{\tau=t}^{\infty} \gamma^{\tau-t} r_i(a^\tau)$$

In this context, Abreu (1988, Proposition 5) shows that any feasible and pure individually rational subgame-perfect equilibrium payoff profile is supported by a certain simple strategy profile.

Therefore, there exists an *optimal penal code* $\{\mathbf{a}(i)\}_{i \in N}$, such that there exists a collection of strategy profiles $\{\sigma(i)\}_{i \in N}$ in which each $\sigma(i) \equiv \sigma(\mathbf{a}(i), \mathbf{a}(1), \ldots, \mathbf{a}(n))$ is a subgame-perfect equilibrium. In other words, an optimal penal code is a collection of outcome paths, embodying player-specific punishments, with the property that the punishment outcome path $\mathbf{a}(i)$ for player $i$ is supported (in the sense of subgame-perfection) by the threat of restarting $\mathbf{a}(i)$ from the beginning.

While, by focusing on simple strategy profiles, we have already considerably narrowed down the choice of strategies, there still remains an infinity of outcome paths that could induce a subgame-perfect equilibrium in simple strategies. Besides, verifying the conditions of Equation (5) for any arbitrary structure of the simple strategy profile and for any $t$ can be extremely difficult. For our example of Duopoly games and with our preference for symmetric payoffs, we can narrow the choice further, by restricting our attention to a particularly simple structure of penal codes, called *carrot-and-stick*.

Carrot-and-stick (Abreu, 1988) is a punishment defined for each player $i$ by two outcomes, $\bar{a}(i)$ and $\tilde{a}(i)$. The punishment outcome path for player $i$, $\mathbf{a}(i)$, looks as follows:

$$\mathbf{a}(i) = (\tilde{a}(i), \bar{a}(i), \bar{a}(i), \ldots)$$

where the outcomes $\tilde{a}(i)$ and $\bar{a}(i)$ play, respectively, the roles of stick and carrot. Abreu (1986) has shown that in repeated Oligopoly games (that include our Duopoly example as a special case), the use of penal codes having a structure of carrot-and-stick, is sufficient to support any symmetric subgame-perfect equilibrium payoff. In other words, in such cases carrot-and-stick-based penal codes are optimal. In our example from Figure 9, let $\gamma = 4/7$. Choose stick and carrot

[6] Recall that an outcome path is defined as $\mathbf{a} \equiv (a^0, a^1, \ldots)$ with $a^t \in A$.

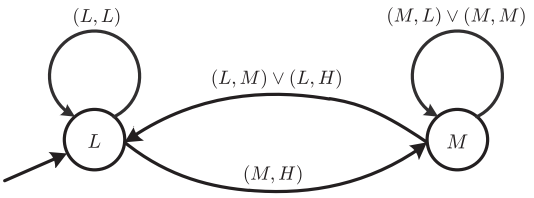


**Figure 10** A finite automaton representing a carrot-and-stick based subgame-perfect equilibrium strategy for Player 1 in the Duopoly game from Figure 9

for Player 1 to be, respectively, $\tilde{a}(1) = (M, H)$ and $\bar{a}(1) = (L, M)$. It is easy to see that two outcome paths

$$\mathbf{a}(1) = ((M,\ H), (L,\ M), (L,\ M), \ldots)$$
$$\mathbf{a}(2) = ((H,\ M), (M,\ L), (M,\ L), \ldots)$$

define an optimal penal code. For instance, let $\mathbf{a}(1)$ be in force. Both Player 1 and Player 2 will follow this outcome path because:

1. By following strategy $\sigma(1) \equiv \sigma(\mathbf{a}(1), \mathbf{a}(1), \mathbf{a}(2))$, Player 1 obtains its pure minmax payoff $\underline{v}_1^p = 0$, because $-4(1-\gamma) + 3\gamma = 0$. Therefore, this payoff is individually rational. Make sure that the payoff of Player 2 exceeds its minmax payoff.
2. Player 1 cannot 'mitigate' the punishment by deviating to playing $L$ instead of $M$ at the very beginning of $\mathbf{a}(1)$. This is because by deviating from it still obtains the same long-term payoff: $0(1-\gamma) + 0\gamma = 0$. Moreover, it cannot deviate afterwards, because playing $M$ instead of $L$, whenever Player 2 plays $M$, yields the payoff of $7(1-\gamma) + 0\gamma = 3$, which equals the payoff of playing the prescribed action $L$: $3(1-\gamma) + 3\gamma = 3$.
3. Player 2 will play $H$ whenever Player 1 is supposed to play $M$, that is, at the very beginning of $\mathbf{a}(1)$. This is because if Player 2 does not do so and deviates to playing $M$, it obtains the long-term payoff of $7(1-\gamma) + 0\gamma = 3$ due to the subsequent punishment[7] on the part of Player 1 (i.e. $\mathbf{a}(2)$ is played instead of $\mathbf{a}(1)$). On the other hand, if Player 2 follows $\mathbf{a}(1)$, it obtains $5(1-\gamma) + 15\gamma = 75/7 > 3$.

The corresponding subgame-perfect equilibrium strategy for Player 1 in the form of a finite automaton is shown in Figure 10.

### *4.2.3 Self-generation*

There exist a number of numerical methods (Cronshaw & Luenberger, 1994; Cronshaw, 1997; Judd *et al.*, 2003) for computing the set of subgame-perfect equilibrium payoffs for a given stage-game and a given discount factor. Judd *et al.* (2003) describe an algorithmic approach permitting to first identify the set of subgame-perfect equilibrium payoff profiles, and then to extract players' strategies yielding a certain payoff profile belonging to this set.

Let $V^p$ denote the set of subgame-perfect equilibrium payoffs that we want to identify. Recall Equation (4): after history $h^t$, to make part of a subgame-perfect equilibrium strategy, action $a_i$ has to be supported by a certain continuation promise $u_i(\sigma|_{h^{t+1}})$. By the definition of subgame-perfection, this must hold after any history. Therefore, if $a_i$ makes part of a subgame-perfect equilibrium $\sigma$, then $u_i(\sigma|_{h^{t+1}})$ has to belong to $V^p$ as well as $v_i(a_i, \sigma|_{h^t})$. This self-referential property of subgame-perfect equilibrium suggests a way by which one can find $V^p$. The key to finding $V^p$ is a

[7] The assumption of the subsequent punishment on the part of Player 1 is a corollary of the one-shot deviation principle. We have already mentioned in the proof of Theorem 3 that we can assume that the deviator (say Player 1) deviates only once and then it follows the prescribed strategy. In practice, this means that it punishes Player 2 for not being punished by Player 2.

construction of *self-generating sets* (Abreu *et al.*, 1990). The analysis focuses on the map $B^p$ defined on a set $W^p \subset \mathbb{R}^n$:

$$B^p(W^p) = \bigcup_{(a,w)\in A \times W^p} (1-\gamma)r(a) + \gamma w$$

where $w \in \mathbb{R}^n$ has to verify:

$$(1-\gamma)r_i(a) + \gamma w_i - (1-\gamma)BR_i(a) + \gamma \underline{w}_i \geq 0$$

$BR_i(a)$ denotes a best response of player $i$ to the action profile $a$, and $\underline{w}_i \equiv \inf_{w \in w^p} w_i$. Abreu *et al.* (1990) show that the largest fixed point of $B^p(W^p)$ is $V^p$, that is, $B^p(V^p) = V^p$ and $V^p$ is the largest such set.

Any numerical implementation of $B^p(W^p)$ requires an efficient representation of the set $W^p$ in a machine. Judd *et al.* (2003) propose to use convex sets in order to approximate both $W^p$ and $B^p(W^p)$ as an intersection of hyperplanes. With this in hand, each application of a $B(W)$ (a convexified version of map $B^p(W^p)$) is reduced to solving a convex optimization problem. We omit further details: the interested reader can refer to Judd *et al.* (2003); an extended version of the paper with useful illustrations and examples is also available. The algorithm of Judd *et al.* (2003) permits computing only pure action subgame-perfect equilibria in repeated games. Burkov and Chaib-draa (2010) leverage self-generation to approximate all (pure and mixed) action equilibria.

As we already mentioned in the beginning of this section, for any two-player repeated game with the AP criterion, there exists an efficient algorithm returning a pair of automata inducing a subgame-perfect equilibrium strategy profile (Littman & Stone, 2005). For more than two players, however, it has recently been demonstrated that an efficient algorithm for computing subgame-perfect equilibria is unlikely to exist (Borgs *et al.*, 2008). In Section 4.3, we will explore the question of strategy computability and implementability.

### *4.3 Complexity in repeated games*

The word 'complexity' in the context of repeated games can have two different meanings: design complexity and implementation complexity. Kalai and Stanford (1988) talk about *implementation complexity* as the complexity of player's strategy. More precisely, they define implementation complexity as the cardinality of the set $\sum_i(\sigma_i) = \{\sigma_i|_{h^t} : h^t \in H\}$. This reflects the number of continuation game strategies, which the player's strategy induces after different histories. The authors then establish that this measure of strategic complexity is equal to the number of the states of the smallest automaton that implements the strategy (Kalai & Stanford, 1988, Theorem 3.1). A similar theory is due to Abreu and Rubinstein (1988).

Another meaning of complexity, *design complexity*, can be described as the measure of computational resources of players required to compute (or design) a strategy having the desired properties (Gilboa, 1988; Ben-Porath, 1990; Papadimitriou, 1992).

#### *4.3.1 Implementation complexity*

Abreu and Rubinstein (1988) modified the subgame-perfect equilibrium concept by incorporating into it the notion of implementation complexity. They first defined *machine game* as a two-player normal form game build upon an original repeated game. In a machine game, the set $\mathcal{M}_i$ of player $i$'s actions, $i \in \{1, 2\}$, is a finite set containing finite automata $M_i \in \mathcal{M}_i$. Every such automaton induces a different player $i$'s strategy in the original repeated game. In the machine game, player $i$ prefers a strategy profile induced by an automaton profile $M \equiv (M_1, M_2)$ to another strategy profile induced by an automaton profile $M' \equiv (M'_1, M'_2)$ (we then write $M \succ_i M'$) if either $u_i(M) > u_i(M') \wedge |M_i| = |M'_i|$ or $u_i(M) = u_i(M') \wedge |M_i| < |M'_i|$. In other words, the players playing a machine game have *lexicographic preferences* over the automata inducing equal payoffs. Abreu and Rubinstein (1988) then define a Nash equilibrium of the machine game as a pair of machines $(M_1, M_2)$, $M_1 \in \mathcal{M}_1$, $M_2 \in \mathcal{M}_2$, with the property that there is no another automaton $M'_1 \in \mathcal{M}_1$ or $M'_2 \in \mathcal{M}_2$ such that

$$(M'_1,\, M_2) \succ_1 (M_1,\, M_2)\, or\, (M_1,\, M'_2) \succ_2 (M_1,\, M_2)$$

Abreu and Rubinstein (1988) described the structure of an equilibrium strategy profile in the machine game: similarly to the subgame-perfect equilibrium strategies, which we constructed in the proofs of folk theorems, in a Nash equilibrium of the machine game the players' automata must contain a non-cyclic punishment part and a cyclic cooperative part respecting a set of interdependent conditions.

An important result for the DAP criterion (a similar one exists for the AP criterion as well) is formulated as follows. Let $(M_1, M_2)$ be a Nash equilibrium of the machine game. This induces two properties. First, the players' automata $M_1$ and $M_2$ have an equal implementation complexity, and maximize the repeated game payoff against one another. Second, if in any two periods, $M_1$ (respectively, $M_2$) plays the same stage-game action, this must be true for the other automaton as well. The latter property permits restricting to a great extent the space of strategies the players can choose from. Also, this permits considerably reducing the set of payoffs that can arise in an equilibrium (see Abreu & Rubinstein, 1988 for more details). Kalai and Stanford (1988) established similar results (while under distinct assumptions) about the implementation complexity for $n$-player ($n > 2$) machine games.

In order to demonstrate that not every subgame-perfect equilibrium strategy profile in a repeated game is also a Nash equilibrium in the corresponding machine game, let us consider the following example. Let us suppose there are two players playing a repeated Prisoner's Dilemma from Figure 1 using the profile of two TFT strategies. With the AP criterion, this strategy profile is clearly a subgame-perfect equilibrium. However, this is not a Nash equilibrium of the corresponding machine game. The reason is that when Player 2 uses TFT, Player 1 can be better of by deviating (in the machine game) to a one-state automaton strategy in which it always plays $C$. Such a strategy has a lower complexity while bringing the same AP to Player 1. On the other hand, if Player 1 chooses this one-state automaton strategy, Player 2 will prefer to choose a one-state automaton that always plays $D$. This process of strategy changes terminates when each player chooses a one-state automaton that always plays $D$. Observe that such a strategy profile now constitues a Nash equilibrium of the machine game and also a subgame-perfect equilibrium in the repeated Prisoner's Dilemma.

Notice the following implicit feature of the approach of Abreu and Rubinstein (1988). Assuming that the players playing a machine game have lexicographic preferences over the automata inducing equal payoffs is equivalent to assigning an infinitesimal cost to each state of the automaton. Neme and Quintas (1995) followed the direction proposed by Abreu and Rubinstein. They considered the case when the complexity enters the utility function as a non-infinitesimal real number defining the cost of using each additional automaton state:

$$u_i^k(\sigma) \equiv u_i(\sigma) - k(comp(\sigma_i))$$

In the above definition, $u_i^k$ is called the *utility function with cost*, $k$ is the *cost function* defined as a mapping $k : \mathbb{N} \mapsto \mathbb{R}^+$ and $comp(\sigma_i)$ denotes the number of states of the minimal automaton inducing the strategy $\sigma_i$. Neme and Quintas studied the structure of Nash equilibrium in repeated games with complexity costs. They provided a corresponding folk theorem for infinite automata, as well as a finite approximation result (Neme & Quintas, 1995). At their turn, Ben-Sasson *et al.* (2007) studied repeated zero-sum games with costs and the properties of certain game theoretic algorithms applied to such games. Notice that the notion of strategy cost naturally reflects a fundamental property of any real-world application: one should pay for every additional resource artificial agents can use in order to become more effective.

For their part, Lipman and Wang (2000, 2009) introduced the concept of *switching costs*. In particular, the authors modify the standard repeated game model by adding a small cost endured by the players when they change their actions between two subsequent periods. The authors show that the addition of such costs changes the properties of equilibria. For example, a multiplicity of equilibria arise in certain games that have a unique subgame-perfect equilibrium without switching costs. On the other hand, in the coordination games, which have multiple equilibria without switching costs, it appears that with small switching costs one can have a unique subgame-perfect equilibrium.

Banks and Sundaram (1990) studied the structure of Nash equilibria in two-player repeated games played by finite automata. Their complexity criterion takes into account not only the size of the automaton but also its transitional structure. In this context, the authors show that the only Nash equilibria in machine games are the pairs of automata that every period recommend actions that are stage-game Nash equilibria.

Another approach to implementation complexity in repeated games is due to Neyman (1985, 1995), Ben-Porath (1993), and others (Zemel, 1989; Papadimitriou & Yannakakis, 1994). According to this approach, players do not include complexity cost into their strategies, but each player $i$ is *exogenously* restricted to choose among the automata having the number of states that does not exceed $m_i$. In contrast to most of the results presented in this paper (i.e. applicable to infinitely repeated games) the following results have only been established for finitely repeated games.

As we already mentioned in Section 3, in finitely repeated games, the horizon $T$ of the repeated game (i.e. the number of game repetitions) is supposed to be finite, fixed and known by all players before the game starts. The analysis and the solutions in finitely repeated games have quite different properties due to the fact that one can use backward induction to compute game solutions. For example, when the players are not given with an upper bound on the complexity of their strategies, it can be shown that cooperation cannot be an equilibrium solution of the finitely repeated Prisoner's Dilemma. Indeed, when there exists the last period $T$, in the absence of the threat of future punishments, the action profile of players at iteration $t = T$ should constitute a stage-game Nash equilibrium $(D, D)$. Therefore, at iteration $t = T-1$, because the future play is known and independent of the present action profile, there is also no way to impose the cooperation, and so on. The similar considerations can apply to all other repeated games having the property that all Nash equilibria of the stage-game yield a payoff profile $v$, such that, $\forall_i \in N$, $v_i = \underline{v}_i$ (Osborne & Rubinstein, 1994, p. 155). Notice that there exists a notion of subgame-perfection in finitely repeated games. There also exists a number of folk theorems for such games (Benoit & Krishna, 1985).

Despite the absence, in finitely repeated Prisoner's Dilemma, of cooperative outcomes sustained by an equilibrium, it is known that a certain degree of cooperation is possible when the players' strategic complexities are bounded. For example, this is true when the minimal of the players' automata has the size that is less than exponential in the value of the repeated game horizon $T$. This result is due to Papadimitriou and Yannakakis (1994). The seminal work providing a weaker result is due to Neyman (1985). The main idea behind this phenomenon is that in the equilibrium, player $i$ adopts a strategy that realizes a complex sequence of $C$ and $D$. From the part of player $-i$, in order to have a profitable deviation, this would require to keep track of this sequence. The latter fact, in turn, would imply that player $-i$ must have a number of states that exceeds the given bound. See Papadimitriou and Yannakakis (1994) and Neyman (1998) for the details of the proof.

#### 4.3.2 Design complexity

In addition to characterizing the equilibrium behavior of automata under the condition of limited or paid resources, certain researchers also investigated the computational complexity of the task of designing a best-response automaton. Gilboa (1988) considered the problem of computing a best-response automaton $M_i$ for player $i$ in a repeated game with $n$ players when the other players' *pure* strategies are induced by finite automata. He demonstrated that both (1) the problem of determining whether $M_i$ induces a strategy that is a best response to the strategy profile induced by the automata of the remaining players, and (2) the problem of finding an automaton $M_i$ inducing such a best-response strategy, can be solved in a time polynomial in the size of the automata.

Ben-Porath (1990) demonstrated, in turn, that for a repeated two-player game where player 2 plays a mixed strategy by sampling pure finite automata from a distribution with a finite support, both (1) the problem of determining whether a given automaton of player 1 induces a best-response strategy to the strategy of player 2, and (2) the problem of finding such a best-response automaton, cannot be solved efficiently.

Papadimitriou (1992) explored the relationship between the computational complexity of finding a best-response strategy and the implementation complexity in the repeated Prisoner's Dilemma. In particular, he showed that if an upper bound $M_1$ is placed on the number of states of the automaton $M_1$ of Player 1, the problem of determining whether $M_1$ induces a best response to the strategy of Player 2, represented by the automaton $M_2$, can only be solved efficiently, when $m_1 \geq |M_2|$.

The complexity of *constructing a profile* of strategies inducing a Nash equilibrium in a repeated game (and not only finding or verifying a best-response strategy) has been studied by Littman and Stone (2005) and Borgs *et al.* (2008). The former have shown that a subgame-perfect equilibrium strategy profile for a two-player repeated game with the AP criterion can be computed in a time polynomial in a number of input parameters. Borgs *et al.* (2008), in turn, have shown that for more than two players a polynomial algorithm cannot probably[8] exist.

There exists another interesting complexity measure: the *communication complexity*. This complexity measure counts the amount of information exchanged between the participants of a distributed optimization problem. This measure is useful for establishing the results in the form of lower bounds on the number of bits of communication needed to solve the problem (Kushilevitz & Nisan, 1997). The applications include equilibrium computation in games with a large number of players. For example, Hart and Mansour (2007) established that the worst-case amount of payoff information that must be exchanged between $n$ players to reach a Nash equilibrium (when initially each player knows only its own payoff function) is exponential in $n$.

In Section 5, we pass from the setting of perfect monitoring to that of imperfect monitoring. We first adjust the model of repeated games so as to take into account the imperfection of the observability model. Then, we briefly overview the most important results obtained in this setting.

## 5 Imperfect monitoring

The setting of perfect monitoring in repeated games roughly corresponds to observability without noise in the algorithmic decision theory (Berry & Fristedt, 1985; Sutton & Barto, 1998). Imperfect private monitoring can be compared with the notion of partial observability well known in the computer science (Kaelbling *et al.*, 1998; Bernstein *et al.*, 2003). Indeed, the imperfect monitoring setting is closely related to general MAS, because the model of artificial agent assumes an uncertain observability of the environment by the agent due to its noised sensors (Russell & Norvig, 2009).

The study of repeated games of perfect monitoring is considerably simplified by the fact that often the structure of an equilibrium strategy is quite simple: a cycle of cooperative actions is followed; any deviation of one player is immediately detected and jointly punished by the other players. When monitoring is imperfect, not every deviation can be immediately and with certainty detected, as well as the identity of deviator cannot always be precisely determined. Therefore, an important part of the analysis in games of imperfect monitoring is devoted to specifying conditions, under which the deviations are detectable and the deviators are identifiable.

In Section 4, we have seen that equilibria in repeated games of perfect monitoring have a recursive structure: for any action, to make part of an equilibrium strategy, it should induce an

[8] More precisely, the problem of computing a Nash equilibrium profile for an $n$-player repeated game with $n > 2$ is at least as complex as the problem of computing a Nash equilibrium in an $(n-1)$-player stage-game. The latter problem does not have efficient algorithms solving it. Furthermore, this problem has recently been shown to be PPAD-complete (see Papadimitriou & Yannakakis, 1988; Chen & Deng, 2006; Daskalakis *et al.*, 2006 for more details on this subject).

equilibrium continuation strategy. This recursive structure can also be preserved in discounted repeated games of imperfect public monitoring. The latter fact considerably simplifies the study of such games. Furthermore, it can be asserted that the games of public monitoring are relatively well studied (Abreu *et al.*, 1990; Fudenberg *et al.*, 2007; Hörner & Olszewski, 2007). On the other hand, repeated games of private monitoring still remain considerably less explored in the literature.

In the next subsections, we survey the literature studying the games of imperfect monitoring. To do that, we first need to adjust our model of repeated games so as to take into account the imperfection of monitoring.

### 5.1 *Repeated game: revisited*

The model of repeated games of imperfect monitoring differs from the perfect monitoring model, which we defined in Section 2. As previously, there is a finite set of players $N$, a finite set of actions $A_i$ for each player $i \in N$, and a collection $\{r_i\}_{i \in N}$ of player-specific payoff functions $r_i : A \mapsto \mathbb{R}$, where $A \equiv \times_{i \in N} A_i$. The difference induced by the particular structure of imperfect game monitoring can be described as follows. At the end of each stage-game, each player observes a *signal* $y_i$ drawn from a finite *signal space* $Y_i$. Let $Y \equiv \times_{i \in N} Y_i$. The function $\rho : Y \times A \mapsto [0, 1]$ defines the probability with which each signal profile $y \in Y$ is drawn given an action profile $a \in A$.

Notice that in this model, a repeated game of perfect monitoring is a special case, in which for every player $i$, $Y_i \equiv A$, and,

$$\begin{cases} \rho((y_i)_{i \in N}, a) = 1, & \text{if } y_i = a \ \forall i; \\ \rho((y_i)_{i \in N}, a) = 0, & \text{otherwise.} \end{cases}$$

A repeated game is said to be a game of *public monitoring*, if for all signal profiles $y \in Y$ and for each pair of players $i, j \in N$, we have $y_i = y_j$. Otherwise, the repeated game is said to be a game of *private monitoring*.

When thinking about decision making in repeated games with imperfect monitoring, two types of finite histories have to be considered. The first set of histories up to period $t$, $H_i^t \equiv \times_t (A_i \times Y_i)$, contains player-specific histories of length $t$. Those histories are uniquely based on the information available to player $i$: the observed signals and the own played actions. The other set of histories, $H^t \equiv \times_t (A \times Y)$, contains the histories based on the full information generated in the repeated game up to period $t$. Any history belonging to the set $H^t$ can be viewed as a sequence of action profiles observed by an omniscient observer capable of perceiving an imperfect monitoring game as if it was a game of perfect monitoring.

Let $H_i \equiv \bigcup_{t=0}^{\infty} H_i^t$ denote the set of *player i's personal histories*, and let $H \equiv \bigcup_{t=0}^{\infty} H^t$ be the set of *game histories*. A pure private strategy $\sigma_i$ of player $i$ is a mapping $\sigma_i : H_i \mapsto A_i$ from the set of player $i$'s personal histories to the set of player $i$'s actions. Similarly, a mixed private strategy is a mapping $\sigma_i : H_i \mapsto \Delta(A_i)$.

In repeated games of imperfect monitoring, Nash equilibrium is defined similarly to the definition of this concept for games of perfect monitoring. On the other hand, we cannot simply define and study the subgame-perfect equilibria, because in the case of personal histories, there are no non-trivial subgames. The appropriate notion of sequential rationality in the imperfect information setting is *sequential equilibrium* (Kreps & Wilson, 1982). In words, a strategy profile $\sigma$ is a sequential equilibrium if, for each player $i$, after any personal history, player $i$'s strategy is a best response to its beliefs over the strategies of the other players. The beliefs of player $i$ are conditioned on its personal history.

### 5.2 *Public monitoring*

Studying equilibria in private strategies, such as sequential equilibria, is complicated by a potential need of keep track of player $i$'s beliefs about the other players' personal histories. Equilibrium actions will then depend on the infinitely nested beliefs of players, about the beliefs of other players about their beliefs, and so on. Reducing attention to *public strategies* permits simplifying the analysis of equilibria in repeated games of public monitoring (Abreu *et al.*, 1990; Fudenberg *et al.*, 2007).

### *5.2.1 Equilibria in public strategies*

The set of *public histories* up to period $t$ is denoted by $H^t_{pub} \equiv \times_t Y$. The set of all public histories, H$_{pub}$, can then be defined as $H_{pub} \equiv \bigcup_{t=0}^{\infty} H^t_{pub}$. A *pure public strategy* $\sigma_i$ of player $i$ is a mapping from the set of public histories to the set of player $i$'s actions, $\sigma_i : H_{pub} \mapsto A_i$. A similar definition can be obtained for the *mixed public strategies.*

When we only consider public strategies, the notions of subgame and subgame strategy can easily be defined, because now after each game history all players use the same information to base their continuation play on.

DEFINITION 10 A perfect public equilibrium *is a profile of public strategies such that after every public history, each player's continuation strategy is a best response to the opponent's continuation strategy.*

As we have already pointed out, in order to construct an equilibrium strategy in the conditions of imperfect monitoring, one needs to have all deviations detectable and all deviators identifiable.

DEFINITION 11 *A mixed action profile* $\alpha \equiv (\alpha_i, \alpha_{-i})$ *has*individual full rank for playe*r* $i$ *if the collection of probability distributions* $\{\rho(\cdot, a_i, \alpha_{-i})\}_{a_i \in A_i}$ *is linearly independent. If this holds for every player* $i \in N$, *then* $\alpha$ *has individual full rank.*

In other words, if a mixed action profile has individual full rank, no player can change the distribution of its actions without affecting the distribution of public signals. Therefore, individual full rank is a condition of detectability of deviations.

DEFINITION 12 *Let* $\alpha \equiv (\alpha_i, \alpha_{-i})$ *be a mixed action profile. Let the* $(|A_i| \times |Y|)$ *matrix* $R_i(\alpha_{-i})$ *have its element* $[R_i(\alpha_{-i})]_{a_i y} \equiv \rho(y, a_i, \alpha_{-i})$. *The profile* $\alpha$ *has pairwise* full rank for players *i and j if the* $(|A_i| + |A_j| \times |Y|)$ *matrix,*

$$R_{ij}(\alpha) = \begin{pmatrix} R_i(a_{-i}) \\ R_j(a_{-j}) \end{pmatrix}$$

*has rank* $|A_i| + |A_j| - 1$. *If this holds for every pair of players, then* $\alpha$ *has* pairwise full rank.

Under the condition of pairwise full rank, deviations from two different players induce different distributions of public signals. Therefore, pairwise full rank is a condition of identifiability of deviators.

Several folk theorems for public monitoring have originally been established by Fudenberg *et al.* (1994). Here, we present a generalized and aggregated formulation from Mailath and Samuelson (2006, p. 301).

THEOREM 6 **(The public monitoring FT with DAP)** *Let all pure action profiles yielding the extreme points of* $F^\dagger$ *have pairwise full rank.*

1. *If* $\tilde{\alpha}$ *is an inefficient stage-game Nash equilibrium, then for all* $v \in \mathrm{int}\{v' \in F^\dagger : v'_i \geq r_i(\tilde{\alpha}) \forall i\}$, *there exists* $\underline{\gamma} < 1$ *such that for all* $\gamma \in (\underline{\gamma}, 1)$, $v$ *is a payoff of a certain perfect public equilibrium.*
2. *If* $\underline{v} \equiv (\underline{v}_i)_{i \in N}$ *is inefficient and each player's minmax profile* $\hat{\alpha}^i$ *has individual full rank, then, for all* $v \in \mathrm{int}\, F^{\dagger +}$, *there exists* $\underline{\gamma} < 1$ *such that for all* $\gamma \in (\underline{\gamma}, 1)$, $v$ *is a payoff of a certain perfect public equilibrium.*

*Proof.* For the complete proof, see Mailath and Samuelson (2006, p. 301). Intuitively, in the first case, the existence of an inefficient stage-game Nash equilibrium implies that it can be used as a threat to prevent deviations and to support the payoff profiles that Pareto dominate it.

The punishment will happen because $\tilde{\alpha}$ is a stage-game equilibrium and therefore no one-shot deviation is profitable. In the second case, the fact that the profile of minmax payoffs is inefficient and that each player's minmax profile has individual full rank implies that punishment will happen since the deviations from it are detectable. □

### 5.2.2 *Equilibria in private strategies*

The payoffs generated by public perfect equilibrium strategy profiles do not cover all sequential equilibrium payoffs that can arise in repeated games of public monitoring. Certain sequential equilibrium payoffs can only be generated by the profiles of private strategies (Kandori & Obara, 2006). Due to the complexity of the setting, only few results exist that characterize the set of sequential equilibrium payoff in games of public monitoring with private strategies. The examples include Mailath *et al.* (2002) and Renault *et al.* (2005, 2008).

## 5.3 *Private monitoring*

The main reason of difficulties in studying games of private monitoring lies in the fact that players do not have any common information to base their continuation play on. In games of perfect monitoring, this common information is the history of the repeated game. In games of imperfect public monitoring, players can maintain a common public history. The signals received by two different players in a repeated game of private monitoring can be different. Therefore, their private histories can differ. Consequently, the notion of subgame cannot easily be defined. There exist several approaches to bypass this problem. They consist in considering different important special cases of a more generally stated problem:

*Almost perfect and almost public monitoring*: These are two special cases of private monitoring (Mailath & Morris, 2002). Almost perfect monitoring is a private monitoring with the property that each player can identify the action taken by its opponents with an error $\varepsilon$. Almost public monitoring, in turn, is a private monitoring when all players receive the same public signal with an error $\varepsilon$. In both cases, the error $\varepsilon$ is supposed to be small, that is, $0 < \varepsilon \ll 1$.

The first significant result has been obtained by Sekiguchi (1997) in the repeated Prisoner's Dilemma with almost perfect monitoring. In this seminal work, he showed that the payoff profile corresponding to action profile $(C, C)$ can be approximated by an equilibrium strategy profile when the monitoring is almost perfect. Bhaskar and Obara (2002) then extended Sekiguchi's approach to support in equilibrium any payoff profile that Pareto dominates the minmax payoff profile. Ely and Valimaki (2002) then proved the folk theorem for the repeated Prisoner's Dilemma of almost perfect monitoring. Recently, Hörner and Olszewski (2006) proved a folk theorem for general repeated games of almost perfect monitoring.

In their turn, Mailath and Morris (2002) considered the repeated games of almost public monitoring and prove a folk theorem. Their results only apply when the strategies are of a *finite memory* (Mailath & Morris, 2002). Under similar conditions, Hörner and Olszewski (2007) obtained a strengthened result.

*Belief-free equilibria*: Ely *et al.* (2005) considered so-called *belief-free* equilibrium strategies. A sequential equilibrium is belief-free if, after every private history, each player's continuation strategy is optimal independently of its belief about the opponents' private histories. The authors provided a characterization of equilibrium payoffs generated by those strategies. They showed that belief-free strategies are not rich enough to generate a folk theorem in most games besides the repeated Prisoner's Dilemma. In their turn, Hörner and Lovo (2009) characterized a set of payoffs that includes all belief-free equilibrium payoffs.

*Other approaches*: Efficient equilibria under private monitoring have been obtained in the repeated games with the AP criterion (Radner, 1986). An approximated folk theorem for games of private monitoring has been established by Fudenberg and Levine (1991). The idea behind those papers is to construct a strategy profile, in which a player has to deviate infinitely often in order to get a payoff improvement.

Another approach to bypass the aforementioned inherent difficulties of private monitoring, is to introduce a possibility of public communication between players. Compte (1998) proved a folk theorem with communication for more than two players and showed that an approximated result can be obtained for two-player case as well. There are two main ideas behind Compte's approach. The first one is that public communication histories are analogous to public histories defined in repeated games of public monitoring. Therefore, they permit establishing a notion of subgame. The second idea is the notion of delayed communication: the players reveal their private signals every $T$ periods. The intuition is that deviating a small fraction of those $T$ periods yields a small gain, while deviating a larger fraction of periods can be detected by the other players. Obara (2009) extends the idea of delayed communication to the case where private signals are correlated. He then proves a new folk theorem for repeated games with private monitoring and communication.

## 6 Conclusion

Repeated games are proven to be useful for MAS as a mathematical abstraction used to represent the interactive aspects of MAS in a compact form. The model of repeated games has been adopted by different researchers to represent complex multiagent interaction schemas. Computer scientists, however, mainly focused on the stationary solutions of repeated games, by only using its repetitive property as a mean to implement an iterative algorithm searching for a stationary equilibrium solution (Bowling & Veloso, 2002; Banerjee & Peng, 2003; Conitzer & Sandholm, 2007). While stationary equilibria are the appropriate solutions in the repeated games, their corresponding set of players' payoffs is typically very restricted. When a repeated game is supposed to be played only once by the same set of players, stationary equilibrium can be considered as the only possible solution. However, when multiagent interactions are extended in time and have a repeated nature, there can exist non-stationary strategy profiles whose payoffs are more attractive for the whole set of players. For instance, the payoffs of such non-stationary strategy profiles can Pareto dominate any stationary equilibrium payoff. The most important (and complex) question is to determine the conditions under which these strategy profiles constitute an equilibrium.

Research on repeated games in economics considered the repetitive property of repeated games as an important distinction from the other interaction models (Abreu, 1988; Aumann & Shapley, 1994; Osborne & Rubinstein, 1994; Mailath & Samuelson, 2006). Economists used the discounted factor to compare player's preferences between its present and future payoffs. The proper definition of the player's long-term payoff function gave rise to the variety of new solutions in the existing games. Thus, for example, the repeated Prisoner's Dilemma, which has the unique non-cooperative stationary equilibrium with low payoffs, obtains a variety of Pareto efficient equilibrium solutions in the form of the TFT strategy profile.

The basic model of repeated games then gives rise to a variety of different settings that can affect the preferences and therefore the decisions of players. These settings differ by the number of iterations of the repeated game, payoff criteria chosen by the players, their patience, information they dispose about the properties of the game, and the perfection (or imperfection) of game monitoring. A collection of folk theorems establish, for different such settings, the bounds on the long-term payoffs that can be obtained by the players in equilibrium when the discount factor tends to one.

After determining the set of all possible equilibrium payoffs, the next important question is which of these payoffs are attainable given a particular value of the discount factor. For instance, a low discount factor means that the players are impatient and, therefore, the future punishments for the present deviations cannot be severe enough to support all the payoffs promised by the corresponding folk theorem. As a consequence, a number of works exist on finding a set of equilibrium payoffs supported by a given discount factor (Abreu, 1988; Cronshaw & Luenberger, 1994; Cronshaw, 1997; Judd *et al.*, 2003; Littman & Stone, 2005).

When the set of payoff profiles supported by a certain discount factor is found, the next step is the extraction of strategy profiles corresponding to a certain payoff from this set. In MAS, populated by artificial agents, this task is complicated by the fact that this extraction has to be done efficiently; furthermore, the extracted strategies have to be representable in a finite way. A number of works have been devoted to the study of computability of strategies in repeated games, and their representability as finite automata (Neyman, 1985; Abreu & Rubinstein, 1988; Kalai & Stanford, 1988; Papadimitriou, 1992; Papadimitriou & Yannakakis, 1994; Littman & Stone, 2005).

While the setting characterized by complete information and monitoring in repeated games is relatively well studied, a number of important questions are still remaining in the other settings. The most difficult one is the setting of imperfect private monitoring characterizing many important practical problems, including multirobot problem solving, electronic commerce, and others. The main difficulty of this setting is that when the observations of different players are different, there is no non-trivial notion of subgame. Therefore, in order to compute the set of equilibrium payoffs, one cannot generally leverage the self-referential structure of an equilibrium payoff, as, for example, was done for the settings characterized by perfect or public monitoring.

Even in the perfect monitoring case, a number of important questions is remaining. Characterization of payoffs in mixed strategies is one example; finding fast and efficient approximate algorithms for computing equilibrium strategies is another one. Finally, extending the existing theoretical results and algorithms for repeated games, that go beyond stationary equilibria, to more complex multistate environments, such as dynamic or stochastic games, is an exciting and challenging future research direction.

## References

Abreu, D. 1986. Extremal equilibria of oligopolistic supergames. *Journal of Economic Theory* **39**(1), 191–225.

Abreu, D. 1988. On the theory of infinitely repeated games with discounting. *Econometrica* **56**, 383–396.

Abreu, D., Pearce, D. & Stacchetti, E. 1990. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica* **58**(5), 1041–1063.

Abreu, D. & Rubinstein, A. 1988. The structure of Nash equilibrium in repeated games with finite automata. *Econometrica* **56**(6), 1259–1281.

Aumann, R. 1981. Survey of repeated games. *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, 11–42.

Aumann, R., Maschler, M. & Stearns, R. 1995. *Repeated Games With Incomplete Information*. The MIT press.

Aumann, R. & Shapley, L. 1994. Long term competition: a game theoretic analysis. *Essays in Game Theory in Honor of Michael Maschler*, 1–15.

Banerjee, B. & Peng, J. 2003. Adaptive policy gradient in multiagent learning. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'03)*. ACM Press, 686–692.

Banks, J. & Sundaram, R. 1990. Repeated games, finite automata, and complexity. *Games and Economic Behavior* **2**(2), 97–117.

Ben-Porath, E. 1990. The complexity of computing a best response automaton in repeated games with mixed strategies. *Games and Economic Behavior* **2**(1), 1–12.

Ben-Porath, E. 1993. Repeated games with finite automata. *Journal of Economic Theory* **59**, 17–39.

Ben-Porath, E. & Peleg, B. 1987. *On the Folk Theorem and Finite Automata*. Mimeo, Hebrew University of Jerusalim.

Ben-Sasson, E., Kalai, A. T. & Kalai, E. 2007. An approach to bounded rationality. In *Advances in Neural Information Processing Systems 19*, Schisölkopf, B., Platt J. & Hoffman, T. (eds). MIT Press, 145–152.

Benoit, J.-P. & Krishna, V. 1985. Finitely repeated games. *Econometrica* **53**(4), 905–922.

Benoit, J.-P. & Krishna, V. 1999. *The Folk Theorems for Repeated Games: A Synthesis*. Mimeo, Pennsylvania State University.

Bernstein, D., Givan, R., Immerman, N. & Zilberstein, S. 2003. The complexity of decentralized control of Markov decision processes. *Mathematics of Operations Research* **27**(4), 819–840.

Berry, D. & Fristedt, B. 1985. *Bandit Problems*. Chapman and Hall London.

Bhaskar, V. & Obara, I. 2002. Belief-based equilibria in the repeated Prisoners' Dilemma with private monitoring. *Journal of Economic Theory* **102**(1), 40–69.

Borgs, C., Chayes, J., Immorlica, N., Kalai, A., Mirrokni, V. & Papadimitriou, C. 2008. The myth of the folk theorem. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC'08)*. ACM Press, 365–372.

Bowling, M. & Veloso, M. 2002. Multiagent learning using a variable learning rate. *Artificial Intelligence* **136**(2), 215–250.

Burkov, A. & Chaib-draa, B. 2009. Effective learning in the presence of adaptive counterparts. *Journal of Algorithms* **64**(4), 127–138.

Burkov, A. & Chaib-draa, B. 2010. An approximate subgame-perfect equilibrium computation technique for repeated games. In *Proceedings of Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI'10)*. AAAI Press, 729–736.

Chen, X. & Deng, X. 2006. Settling the complexity of two-player Nash equilibrium. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*. IEEE Computer Society, 261–272.

Cheng, S., Reeves, D., Vorobeychik, Y. & Wellman, M. 2004. Notes on equilibria in symmetric games. In *AAMAS-04 Workshop on Game-Theoretic and Decision-Theoretic Agents*.

Compte, O. 1998. Communication in repeated games with imperfect private monitoring. *Econometrica* **66**(3), 597–626.

Conitzer, V. & Sandholm, T. 2007. AWESOME: a general multiagent learning algorithm that converges in self-play and learns a best response against stationary opponents. *Machine Learning* **67**(1), 23–43.

Cronshaw, M. 1997. Algorithms for finding repeated game equilibria. *Computational Economics* **10**(2), 139–168.

Cronshaw, M. & Luenberger, D. 1994. Strongly symmetric subgame perfect equilibria in infinitely repeated games with perfect monitoring and discounting. *Games and Economic Behavior* **6**(2), 220–237.

Daskalakis, C., Goldberg, P. & Papadimitriou, C. 2006. The complexity of computing a Nash equilibrium. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing (STOC'06)*. ACM Press, 71–78.

Ely, J., Hörner, J. & Olszewski, W. 2005. Belief-free equilibria in repeated games. *Econometrica* **73**(2), 377–415.

Ely, J. & Valimaki, J. 2002. A robust folk theorem for the prisoner's dilemma. *Journal of Economic Theory* **102**(1), 84–105.

Friedman, J. 1971. A non-cooperative equilibrium for supergames. *The Review of Economic Studies*, 1–12.

Fudenberg, D., Kreps, D. & Maskin, E. 1990. Repeated games with long-run and short-run players. *The Review of Economic Studies* **57**(4), 555–573.

Fudenberg, D. & Levine, D. 1991. An approximate folk theorem with imperfect private information. *Journal of Economic Theory* **54**(1), 26–47.

Fudenberg, D., Levine, D. & Maskin, E. 1994. The folk theorem with imperfect public information. *Econometrica* **62**(5), 997–1039.

Fudenberg, D., Levine, D. & Takahashi, S. 2007. Perfect public equilibrium when players are patient. *Games and Economic Behavior* **61**(1), 27–49.

Fudenberg, D. & Tirole, J. 1991. *Game Theory*. MIT Press.

Gilboa, I. 1988. The complexity of computing best-response automata in repeated games. *Journal of economic theory* **45**(2), 342–352.

Gossner, O. & Tomala, T. 2009. Repeated games. *Encyclopedia of Complexity and Systems Science*, forthcoming.

Hart, S. & Mansour, Y. 2007. The communication complexity of uncoupled Nash equilibrium procedures. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing (STOC'07)*. ACM Press, 345–353.

Hörner, J. & Lovo, S. 2009. Belief-free equilibria in games with incomplete information. *Econometrica* **77**(2), 453–487.

Hörner, J. & Olszewski, W. 2006. The folk theorem for games with private almost-perfect monitoring. *Econometrica* **74**(6), 1499–1544.

Hörner, J. & Olszewski, W. 2007. How robust is the folk theorem with imperfect public monitoring. *Northwestern University*.

Jong, S., Tuyls, K. & Verbeeck, K. 2008. Fairness in multi-agent systems. *The Knowledge Engineering Review* **23**(2), 153–180.

Judd, K., Yeltekin, S. & Conklin, J. 2003. Computing supergame equilibria. *Econometrica* **71**(4), 1239–1254.

Kaelbling, L., Littman, M. & Cassandra, A. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence* **101**(1–2), 99–134.

Kalai, E. & Stanford, W. 1988. Finite rationality and interpersonal complexity in repeated games. *Econometrica* **56**(2), 397–410.

Kandori, M. & Obara, I. 2006. Efficiency in repeated games revisited: the role of private strategies. *Econometrica* **74**(2), 499–519.

Kreps, D. & Wilson, R. 1982. Sequential equilibria. *Econometrica: Journal of the Econometric Society*, 863–894.

Kushilevitz, E. & Nisan, N. 1997. *Communication Complexity*. Cambridge University Press.

Laraki, R. 2002. Repeated games with lack of information on one side: the dual differential approach. *Mathematics of Operations Research* **27**(2), 419–440.

Lehrer, E. & Pauzner, A. 1999. Repeated games with differential time preferences. *Econmetrica* **67**(2), 393–412.

Lehrer, E. & Yariv, L. 1999. Repeated games with incomplete information on one side: the case of different discount factors. *Mathematics of Operations Research* **24**(1), 204–218.

Lipman, B. & Wang, R. 2000. Switching costs in frequently repeated games. *Journal of Economic Theory* **93**(2), 149–190.

Lipman, B. & Wang, R. 2009. Switching costs in infinitely repeated games. *Games and Economic Behavior* **66**(1), 292–314.

Littman, M. & Stone, P. 2005. A polynomial-time Nash equilibrium algorithm for repeated games. *Decision Support Systems* **39**(1), 55–66.

Mailath, G., Matthews, S. & Sekiguchi, T. 2002. Private strategies in finitely repeated games with imperfect public monitoring. *Contributions to Theoretical Economics* **2**(1), 1046.

Mailath, G. & Morris, S. 2002. Repeated games with almost-public monitoring. *Journal of Economic Theory* **102**(1), 189–228.

Mailath, G. & Samuelson, L. 2006. *Repeated Games and Reputations: Long-run Relationships*. Oxford University Press.

Matsushima, H. 2004. Repeated games with private monitoring: two players. *Econometrica* **72**(3), 823–852.

Mertens, J., Sorin, S. & Zamir, S. 1994. Repeated games, Part A: background material. *CORE Discussion Papers*, 9420.

Myerson, R. 1991. *Game Theory: Analysis of Conflict*. Harvard University Press.

Nash, J. 1950. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America* **36**(1), 48–49.

Neme, A. & Quintas, L. 1995. Subgame perfect equilibrium of repeated games with implementation costs. *Journal of Economic Theory* **66**(2), 599–608.

Neyman, A. 1985. Bounded complexity justifies cooperation in the finitely repeated Prisoner's Dilemma. *Economics Letters* **19**(3), 227–229.

Neyman, A. 1995. Cooperation, repetition, and automata. In *Cooperation: Game Theoretic Approaches*, volume 155 *of NATO ASI Series F*. Springer-Verlag, 233–255.

Neyman, A. 1998. Finitely repeated games with finite automata. *Mathematics of Operations Research* **23**(3), 513–552.

Obara, I. 2009. Folk theorem with communication. *Journal of Economic Theory* **144**(1), 120–134.

Osborne, M. & Rubinstein, A. 1994. *A Course in Game Theory*. MIT Press.

Papadimitriou, C. 1992. On players with a bounded number of states. *Games and Economic Behavior* **4**(1), 122–131.

Papadimitriou, C. & Yannakakis, M. 1988. Optimization, approximation, and complexity classes. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*. ACM Press, 229–234.

Papadimitriou, C. & Yannakakis, M. 1994. On complexity as bounded rationality (extended abstract). In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing*. ACM Press, 726–733.

Pearce, D. 1992. Repeated games: cooperation and rationality. In *Advances in Economic Theory: Sixth World Congress*, vol. 1. Cambridge University Press, 132–174.

Piccione, M. 2002. The repeated prisoner's dilemma with imperfect private monitoring. *Journal of Economic Theory* **102**(1), 70–83.

Radner, R. 1986. Repeated partnership games with imperfect monitoring and no discounting. *The Review of Economic Studies* **53**(1), 43–57.

Ramchurn, S., Huynh, D. & Jennings, N. 2004. Trust in multi-agent systems. *The Knowledge Engineering Review* **19**(1), 1–25.

Rasmusen, E. 1994. *Games and Information*. Blackwell Cambridge.

Renault, J., Scarlatti, S. & Scarsini, M. 2005. A folk theorem for minority games. *Games and Economic Behavior* **53**(2), 208–230.

Renault, J., Scarlatti, S. & Scarsini, M. 2008. Discounted and finitely repeated minority games with public signals. *Mathematical Social Sciences* **56**(1), 44–74.

Rubinstein, A. 1979. Equilibrium in supergames with the overtaking criterion. *Journal of Economic Theory* **21**(1), 1–9.

Russell, S. & Norvig, P. 2009. *Artificial Intelligence: A Modern Approach*, 3rd edn. Prentice Hall.

Sekiguchi, T. 1997. Efficiency in repeated Prisoner's Dilemma with private monitoring. *Journal of Economic Theory* **76**(2), 345–361.

Sorin, S. 1986. On repeated games with complete information. *Mathematics of Operations Research* **11**(1), 147–160.

Sutton, R. S. & Barto, A. G. 1998. *Reinforcement Learning: An Introduction*. The MIT Press.

Zemel, E. 1989. Small talk and cooperation: a note on bounded rationality. *Journal of Economic Theory* **49**(1), 1–9.