

# NASH EQUILIBRIA IN PARTIAL-INFORMATION GAMES ON MARKOV CHAINS\*

João P. Hespanha

hespanha@usc.edu

Electrical Engineering-Systems  
Univ. of Southern California, USA

Maria Prandini

prandini@ing.unibs.it

Electronics for Automation  
Univ. of Brescia, Italy

## Abstract

We consider a two-player partial-information game on a Markov chain, where each player attempts to minimize its own cost over a finite time horizon. We show that this game has always a Nash equilibrium in stochastic behavioral policies. The technique used to prove this result is constructive but has severe limitations because it involves solving an extremely large bi-matrix game. To alleviate this problem, we derive a dynamic-programming-like condition that is necessary and sufficient for a pair of policies to be a Nash equilibrium. This condition automatically gives Nash equilibria when a pair of “cost-to-go” functions can be found that satisfy certain inequalities.

## 1 Introduction

Competitive games are usually classified as either having full or partial-information. In *full-information* games both players know the whole state of the game when they have to make decisions. By state, we mean all information that is needed to completely describe the future evolution of the game, when the decision rules used by both players are known. Examples of full-information games include Chess, Checkers, and Go. *Partial-information* games differ from these in that at least one of the players does not know the whole state of the game. Poker, Bridge, and Hearts are examples of such games. In full-information games, as a player is planning its next move, it only needs to hypothesize over its and the opponent’s future moves to predict the possible outcomes of the game. Partial-information games are especially challenging because this reasoning may fail. In many partial-information games, a player must hypothesize not only on the future moves of both players, but also on the past moves of the opponent to predict the possible outcomes of the game. This often leads to a tremendous increase in the complexity of the games. In general, partial-information stochastic games are poorly understood and the literature is relatively sparse. Notable exceptions are games with lack of information for one of the players [1, 2], games with private information where each

player can estimate the opponent’s state based only on its current private state [3], single-stage games [4, 5], or games with particular structures such as the Duel game [6], the Rabbit and Hunter game [7], and the Searchlight game [8]. Linear, quadratic, Gaussian differential games have also been studied but the results available are mostly restricted to cases where the information available to one of the players is a subset of the information available to the other player [9, 10] or when both players share the same observations (although they may not know the controls used by the other player) [11]. Another relevant class of games that have been analyzed to some extent are the so-called repeated games with *incomplete information* [12, 13]. These correspond to iterated plays of the same static game, in which the payoff matrix is randomly chosen by “nature” and not known to the players (who can try to estimate it based on the information collected at each play).

In this paper, we address fairly general partial-information games on Markov chains. These are games in which two players are able to influence the state transitions in a Markov chain by taking appropriate actions [14]. Each player attempts to minimize its own cost that is additive over time with the incremental costs depending on the state of the Markov chain and the actions taken by the players. We deviate from most of the literature on Markov games in that we do not assume full-information. In fact, each player only has available stochastic measurements that, in general, do not allow it to determine the current state of the Markov chain. We consider here finite games played over a finite time horizon and with possibly nonzero-sum costs.

This paper contains two main contributions. We first show that there always exists a Nash equilibrium in the set of stochastic behavioral policies, thus extending Kuhn’s result for extensive games with perfect recall [15] to the setting of partial-information Markov games. The second contribution consists in deriving a dynamic programming-like condition that, when satisfied by a pair of “cost-to-go” functions, provides a Nash equilibrium in stochastic behavioral policies. We show that this condition is non-conservative because it is also necessary for the existence of a Nash equilibrium. This paper falls short of actually providing an efficient algorithm to determine the cost-to-go functions for a generic

---

\*Research supported by the Office of Naval Research, the Defense Advanced Research Projects Agency, the National Science Foundation, and the Ministero dell’Università e della Ricerca Scientifica e Tecnologica.

In Proc. of the 40th IEEE Conf. on Dec. and Control.

partial information game. This is an important topic for future research.

**Notation:** Let  $(\Omega, \mathcal{F})$  be the relevant *measurable space*. Bold face symbols are used to denote random variables. Consider a probability measure  $P : \mathcal{F} \rightarrow [0, 1]$  and a random variable  $\xi : \Omega \rightarrow \mathcal{C}$ . Given two events  $A, B \in \mathcal{F}$  with  $P(B) \neq 0$ , we write  $P(A|B)$  for the *conditional probability of A given B*, i.e.,  $P(A|B) = P(A \cap B) / P(B)$ . Given  $c \in \mathcal{C}$ ,  $P(\xi = c)$  is the probability of  $\xi$  to take the value  $c$ . A similar notation is used for conditional probabilities. Moreover, we write  $E[\xi]$  for the expected value of  $\xi$  and  $E[\xi|A]$  for the expected value of  $\xi$  conditioned to the event  $A \in \mathcal{F}$ . Given a finite set  $\mathcal{G}$ , we denote its cardinality by  $|\mathcal{G}|$  and by  $[0, 1]^\mathcal{G}$ , the set of all probability distributions over  $\mathcal{G}$ . Finally, given a distribution  $p \in [0, 1]^\mathcal{G}$ ,  $p_g$  denotes the probability of  $g \in \mathcal{G}$  in  $p$ .

## 2 Two-player Markov Game

We consider a game between two players U and D. The game takes place on a controlled Markov chain with finite *state space*  $\mathcal{S}$ . We denote by  $\mathbf{s}_t$  the random variable representing the *state* of the chain at time  $t \in \mathcal{T} := \{0, 1, 2, \dots, T\}$ ,  $T \leq +\infty$ .

**Transition probability.** Players U and D can affect the evolution of the game by applying *control actions*  $\mathbf{u}_t$  and  $\mathbf{d}_t$ , respectively. We assume that  $\mathbf{u}_t$  and  $\mathbf{d}_t$ ,  $t \in \mathcal{T}$ , take values in finite *action spaces*  $\mathcal{U}$  and  $\mathcal{D}$ , respectively. The probability of transition from a given state  $s \in \mathcal{S}$  at time  $t$  to another state  $s' \in \mathcal{S}$  at time  $t + 1$  when actions  $u \in \mathcal{U}$  and  $d \in \mathcal{D}$  are applied is supposed to be stationary, i.e.,

$$P(\mathbf{s}_{t+1} = s' \mid \mathbf{s}_t = s, \mathbf{u}_t = u, \mathbf{d}_t = d) = p(s, s', u, d),$$

for all  $s, s' \in \mathcal{S}$ ,  $u \in \mathcal{U}$ ,  $d \in \mathcal{D}$ ,  $t \in \mathcal{T}$ , where  $p : \mathcal{S} \times \mathcal{U} \times \mathcal{D} \rightarrow [0, 1]$  is the *transition probability function*.  $\mathbf{s}_{t+1}$  is assumed to be conditionally independent of all other random variables at times smaller or equal to  $t$ , given  $\mathbf{s}_t$ ,  $\mathbf{u}_t$  and  $\mathbf{d}_t$ . Moreover, the initial state  $\mathbf{s}(0)$  is supposed to be independent of all the other random variables at time  $t = 0$ . Its probability distribution is denoted by  $\bar{p} \in [0, 1]^\mathcal{S}$ .

**Observation probability.** To choose their actions, *measurements*  $\mathbf{y}_t$  and  $\mathbf{z}_t$  are available to players U and D, respectively, at time  $t \in \mathcal{T}$ . Here, the random variables  $\mathbf{y}_t$  and  $\mathbf{z}_t$ ,  $t \in \mathcal{T}$ , take values in finite *measurement spaces*  $\mathcal{Y}$  and  $\mathcal{Z}$ , respectively.  $\mathbf{y}_t$  is assumed to be conditionally independent of all the other random variables at times smaller or equal to  $t$ , given  $\mathbf{s}_t$ . Similarly for  $\mathbf{z}_t$ . Moreover, the conditional distributions of  $\mathbf{y}_t$  and  $\mathbf{z}_t$ , given the current value of the state  $\mathbf{s}_t$ , are assumed to be stationary, i.e.,  $P(\mathbf{y}_t = y \mid \mathbf{s}_t = s) = p_Y(y, s)$  and  $P(\mathbf{z}_t = z \mid \mathbf{s}_t = s) = p_Z(z, s)$ , for all  $s \in \mathcal{S}$ ,  $y \in \mathcal{Y}$ ,  $z \in \mathcal{Z}$ ,  $t \in \mathcal{T}$ , where  $p_Y : \mathcal{Y} \times \mathcal{S} \rightarrow [0, 1]$  and  $p_Z : \mathcal{Z} \times \mathcal{S} \rightarrow [0, 1]$  are the *observation probability functions* for players U and D, respectively. The information available to players U and D to decide which action to take at time  $t \in \mathcal{T}$

is respectively given by the sequence of measurements and past controls

$$\begin{aligned} \mathbf{Y}_t &:= \{\mathbf{y}_0, \mathbf{u}_0, \mathbf{y}_1, \mathbf{u}_1, \dots, \mathbf{y}_{t-1}, \mathbf{u}_{t-1}, \mathbf{y}_t\}, \\ \mathbf{Z}_t &:= \{\mathbf{z}_0, \mathbf{d}_0, \mathbf{z}_1, \mathbf{d}_1, \dots, \mathbf{z}_{t-1}, \mathbf{d}_{t-1}, \mathbf{z}_t\}. \end{aligned}$$

These sequences are said to be of *length t*. The sets of all possible outcomes for  $\mathbf{Y}_t$  and  $\mathbf{Z}_t$ ,  $t \in \mathcal{T}$ , are denoted by  $\mathcal{Y}^*$  and  $\mathcal{Z}^*$ , respectively. In the sequel,  $\ell(Y)$  and  $\ell(Z)$  denote the lengths of  $Y \in \mathcal{Y}^*$  and  $Z \in \mathcal{Z}^*$ , respectively. When the measurements available to the players are sufficient to let them both know at each time instant  $t \in \mathcal{T}$  which is the current value realized by the state  $\mathbf{s}_t$  with probability one, the game is said to be of *full-information*. Games for which this does not happen are said to be of *partial-information*.

**Stochastic Policies.** Informally, a “policy” for one of the players is a rule the player uses to select which actions to take over the time horizon  $\mathcal{T}$ . We consider here policies that are *stochastic* in that, at every time  $t \in \mathcal{T}$ , each player selects an action over the action set according to some probability distribution. The policies considered are also *behavioral* in that the specific probability distribution depends on the information collected up to time  $t$ . Specifically, a *stochastic behavioral policy*  $\mu$  of player U is a function  $\mu : \mathcal{Y}^* \rightarrow [0, 1]^\mathcal{U}$ . We denote by  $\Pi_U$  the set of all such policies. For each  $Y \in \mathcal{Y}^*$ ,  $\mu(Y)$  is called a *stochastic action* of player U. We say that  $\mu \in \Pi_U$  is a *pure behavioral policy* for player U if, for every  $Y \in \mathcal{Y}^*$ , the entries of the vector  $\mu(Y)$  are in the set  $\{0, 1\}$ . The set of all pure policies for player U is denoted by  $\bar{\Pi}_U$ . Similarly, a *stochastic behavioral policy*  $\delta$  of player D is a function  $\delta : \mathcal{Z}^* \rightarrow [0, 1]^\mathcal{D}$ . For every  $Z \in \mathcal{Z}^*$ ,  $\delta(Z)$  is a *stochastic action* of player D. The sets of stochastic and pure behavioral policies for player D are respectively denoted by  $\Pi_D$  and  $\bar{\Pi}_D$ .

Stochastic behavioral policies are a generalization to partial-information Markov games of the behavioral policies for multi-act extensive games [16]. The fact that the stochastic actions of the players are allowed to depend on their past actions is key to guaranteeing that a Nash equilibrium always exists for the game presented here, and is the analogous to the “perfect recall” property in the extensive games setting [15].

Opting for specific stochastic policies corresponds to selecting a particular probability measure that is consistent with the distributions chosen for the control actions. In the following we use the subscript  $\mu\delta$  in the probability measure  $P$  to denote the probability measure associated with  $\mu \in \Pi_U$  and  $\delta \in \Pi_D$ . When an assertion holds true with respect to  $P_{\mu\delta}$  independently of  $\mu \in \Pi_U$ , or of  $\delta \in \Pi_D$ , or of both  $\mu \in \Pi_U$  and  $\delta \in \Pi_D$ , we use the notation  $P_\delta$ ,  $P_\mu$ , or  $P$ , respectively. Similarly for the expected value operator  $E$ . According to this notation, the transition and observation probabilities, and the initial state distribution introduced earlier are independent of  $\mu$  and  $\delta$ .

When player U selects a stochastic behavioral policy  $\mu \in \Pi_U$  and player D a policy  $\delta \in \Pi_D$ , the two players are jointly selecting a probability measure in the family  $\{P_{\mu\delta} : \mu \in \Pi_U, \delta \in \Pi_D\}$ . This family of probability measures has the property that  $P_{\mu}(\mathbf{u}_t = u \mid \mathbf{Y}_t = Y) = \mu_u(Y)$ ,  $P_{\delta}(\mathbf{d}_\tau = d \mid \mathbf{Z}_\tau = Z) = \delta_d(Z)$ ,  $t := \ell(Y)$ ,  $u \in \mathcal{U}$ ,  $Y \in \mathcal{Y}^*$ ,  $\tau := \ell(Z)$ ,  $d \in \mathcal{D}$ ,  $Z \in \mathcal{Z}^*$ . Moreover,  $\mathbf{u}_t(\mathbf{d}_t)$  is conditionally independent of all other random variables for times smaller or equal to  $t$ , given  $\mathbf{Y}_t(\mathbf{Z}_t)$ .

**Cost Structure.** In this paper we consider non-cooperative games over a finite horizon  $T < \infty$ , in which each player  $K \in \{U, D\}$  chooses the actions so as to minimize its own cost

$$J_{\mu\delta}^K := \mathbb{E}_{\mu\delta} \left[ \sum_{\tau=0}^T l_K(\mathbf{s}_\tau, \mathbf{u}_\tau, \mathbf{d}_\tau, \tau) \right],$$

Typically,  $l_K : \mathcal{S} \times \mathcal{U} \times \mathcal{D} \times \mathcal{T} \rightarrow \mathbb{R}$  takes the form

$$l_K(s, u, d, \tau) = \begin{cases} c_K(s, u, d), & \tau < T \\ r_K(s), & \tau = T \end{cases},$$

where  $r_K : \mathcal{S} \rightarrow \mathbb{R}$  assigns to each state  $s$  the cost for player  $K$  of finishing the game at  $s$ , whereas  $c_K : \mathcal{S} \times \mathcal{U} \times \mathcal{D} \rightarrow \mathbb{R}$  assigns to each state  $s$  and pair of actions  $u, d$  the marginal cost for player  $K$  for continuing the game at state  $s$  when the applied actions are  $u$  and  $d$ , respectively. We say that the game is zero-sum if  $l_D = -l_U$ , since under this condition  $J_{\mu\delta}^U + J_{\mu\delta}^D = 0$ . The coupling between the players' costs models situations where they are sharing a common environment and competing for the same resources.

We suppose that each player tries to best counteract the other player's action so as to achieve a certain performance level irrespectively of the other player's choice. In the game of interest, this translates into the players selecting a pair of stochastic behavioral policies  $(\mu^*, \delta^*) \in \Pi_U \times \Pi_D$  for which

$$J_{\mu^*\delta^*}^U \leq J_{\mu\delta^*}^U, \quad \forall \mu \in \Pi_U, \quad J_{\mu^*\delta^*}^D \leq J_{\mu\delta^*}^D, \quad \forall \delta \in \Pi_D. \quad (1)$$

The policies  $(\mu^*, \delta^*)$  satisfying (1) are said to constitute a *Nash equilibrium (in stochastic behavioral policies)*.

### 3 Existence of Nash equilibria

It turns out that, if we restrict our attention to pure policies, a Nash equilibrium in the sense of (1) may not exist. This is actually the reason why we consider the stochastic behavioral policies. In order to prove that Nash equilibria always exist in stochastic behavioral policies, we introduce another type of policies, called mixed policies, for which one can use standard arguments to show that an equilibrium exists. Mixed policies can be thought of as another method to enlarge the set of pure policies which is well suited for games that are played repeatedly.

Suppose that both players do restrict their attention to pure behavioral policies but they independently extract at random which policy to use according to some probability distribution over the sets of pure policies. This extraction is done before the game starts and the resulting game is therefore known as a *prior commitment game*. Denoting by  $\rho := \{\rho_\mu : \mu \in \bar{\Pi}_U\} \in [0, 1]^{\bar{\Pi}_U}$  and  $\sigma := \{\sigma_\delta : \delta \in \bar{\Pi}_D\} \in [0, 1]^{\bar{\Pi}_D}$  the distributions used by players U and D, respectively, to choose among their pure policies, the expected cost for player  $K \in \{U, D\}$  is given by

$$\bar{J}_{\rho\sigma}^K := \sum_{\mu \in \bar{\Pi}_U, \delta \in \bar{\Pi}_D} \rho_\mu \sigma_\delta J_{\mu\delta}^K.$$

The distributions  $\rho$  and  $\sigma$  are called *mixed behavioral policies* for player U and D, respectively ([17]). The cost  $\bar{J}_{\rho\sigma}^K$  can also be expressed in matrix form as

$$\bar{J}_{\rho\sigma}^K = \rho' A^K \sigma,$$

where  $A^K$  is a  $|\bar{\Pi}_U| \times |\bar{\Pi}_D|$  matrix defined by

$$[A^K]_{(\mu, \delta) \in \bar{\Pi}_U \times \bar{\Pi}_D} := J_{\mu\delta}^K,$$

with one row for each pure policy for player U and one column for each pure policy for player D, and  $\rho$  and  $\mu$  are interpreted as column vectors.

It is well known that at least one Nash equilibrium always exists in mixed policies [17], i.e., that there exists  $(\rho^*, \sigma^*) \in [0, 1]^{\bar{\Pi}_U} \times [0, 1]^{\bar{\Pi}_D}$  for which

$$\rho^{*'} A^U \sigma^* \leq \rho' A^U \sigma^*, \quad \forall \rho \in [0, 1]^{\bar{\Pi}_U}, \quad (2)$$

$$\rho^{*'} A^D \sigma^* \leq \rho^{*'} A^D \sigma, \quad \forall \sigma \in [0, 1]^{\bar{\Pi}_D}. \quad (3)$$

We proceed now to prove the existence of Nash equilibria in stochastic behavioral policies. To achieve this, we need the following result:

**Lemma 1** *There exist surjective functions  $L^U : [0, 1]^{\bar{\Pi}_U} \rightarrow \Pi_U$  and  $L^D : [0, 1]^{\bar{\Pi}_D} \rightarrow \Pi_D$  such that, for every pair of mixed policies  $(\rho, \sigma) \in [0, 1]^{\bar{\Pi}_U} \times [0, 1]^{\bar{\Pi}_D}$ ,*

$$\bar{J}_{\rho\sigma}^U = J_{\mu\delta}^U, \quad \bar{J}_{\rho\sigma}^D = J_{\mu\delta}^D, \quad (4)$$

where  $(\mu, \delta) \in \Pi_U \times \Pi_D$  are stochastic policies given by  $\mu := L^U(\rho)$ ,  $\delta := L^D(\sigma)$ .

Lemma 1 extends Kuhn's equivalence result between mixed and stochastic behavioral policies in extensive games with perfect recall to the Markov games setting. To prove it we need to introduce the following notation: Fix  $Y \in \mathcal{Y}^*$  and  $u \in \mathcal{U}$ . For any  $\tau < \ell(Y)$ ,  $Y_\tau$  and  $u_\tau$  respectively denote the truncation of  $Y$  to length  $\tau$  and the action at time  $\tau$  in the sequence  $Y$ . Moreover,  $\bar{\Pi}_{U|\{Y, u\}}$  denotes the set of pure policies that are *compatible* with  $Y$  and  $u$ , i.e., the set of policies  $\mu$  for which  $\mu_u(Y) = 1$ , and  $\mu_{u_\tau}(Y_\tau) = 1$ ,  $\forall \tau < \ell(Y)$ . Given  $Z \in \mathcal{Z}$  and  $d \in \mathcal{D}$ , the notations  $Z_\tau$  and  $d_\tau$ ,  $\tau < \ell(Z)$ , and  $\bar{\Pi}_{D|\{Z, d\}}$  are defined similarly. The following Lemma (see [18] for a proof) is also needed to prove Lemma 1.

**Lemma 2** Given any  $t \in \mathcal{T}$ ,  $s_0, s_1, \dots, s_t \in \mathcal{S}$ ,  $u_t \in \mathcal{U}$ ,  $d_t \in \mathcal{D}$ , and  $Y_t \in \mathcal{Y}^*$ ,  $Z_t \in \mathcal{Z}^*$ , with  $\ell(Y_t) = \ell(Z_t) = t$ , for every pair of stochastic policies  $(\mu, \delta) \in \Pi_U \times \Pi_D$ ,

$$\begin{aligned} P_{\mu\delta}(s_0 = s_0, s_1 = s_1, \dots, s_t = s_t, u_t = u_t, d_t = d_t, \\ Y_t = Y_t, Z_t = Z_t) = k_t(s_0, s_1, \dots, s_t, u_t, d_t, Y_t, Z_t) \\ \prod_{\tau=0}^t \mu_{u_\tau}(Y_\tau) \delta_{d_\tau}(Z_\tau), \end{aligned}$$

where  $k_t$  is a function that does not depends on  $(\mu, \delta)$ .

For ease of notation, in the sequel we shall use  $k_t(\cdot)$  for  $k_t(s_0, s_1, \dots, s_t, u_t, d_t, Y_t, Z_t)$ ,  $t \in \mathcal{T}$ .

**Proof of Lemma 1.**  $L^U$  can be defined as follows: for a given  $\rho \in [0, 1]^{\bar{\Pi}_U}$ ,  $L^U(\rho) = \mu$ , with

$$\mu_u(Y) := \frac{\sum_{\bar{\mu} \in \bar{\Pi}_{U|\{Y, u\}}} \rho_{\bar{\mu}}}{\sum_{\hat{u} \in \mathcal{U}} \sum_{\bar{\mu} \in \bar{\Pi}_{U|\{Y, \hat{u}\}}} \rho_{\bar{\mu}}}, u \in \mathcal{U}, Y \in \mathcal{Y}^*. \quad (5)$$

To verify that this function is surjective, we show next that  $\bar{L}^U : \Pi_U \rightarrow [0, 1]^{\bar{\Pi}_U}$ , defined by  $\bar{L}^U(\mu) := \rho$ , with

$$\rho_{\bar{\mu}} := \prod_{Y \in \mathcal{Y}^*} \sum_{u \in \mathcal{U}} \bar{\mu}_u(Y) \mu_u(Y), \quad \bar{\mu} \in \bar{\Pi}_U,$$

is right-inverse of  $L^U$ . To verify that this is true, let  $\tilde{\mu} := L^U(\bar{L}^U(\mu))$  for some  $\mu \in \Pi_U$ . Fix an arbitrary  $Y \in \mathcal{Y}^*$ . From the definitions of  $L^U$  and  $\bar{L}^U$ ,

$$\begin{aligned} \tilde{\mu}_u(Y) &= \frac{\sum_{\bar{\mu} \in \bar{\Pi}_{U|\{Y, u\}}} \prod_{\bar{Y}} \sum_{\bar{u}} \bar{\mu}_{\bar{u}}(\bar{Y}) \mu_{\bar{u}}(\bar{Y})}{\sum_{\hat{u} \in \mathcal{U}} \sum_{\bar{\mu} \in \bar{\Pi}_{U|\{Y, \hat{u}\}}} \prod_{\bar{Y}} \sum_{\bar{u}} \hat{\mu}_{\bar{u}}(\bar{Y}) \mu_{\bar{u}}(\bar{Y})} \\ &= \frac{\sum_{\bar{\mu} \in \bar{\Pi}_{U|\{Y, u\}}} \mu_u(Y) \prod_{\bar{Y} \neq Y} \sum_{\bar{u}} \bar{\mu}_{\bar{u}}(\bar{Y}) \mu_{\bar{u}}(\bar{Y})}{\sum_{\hat{u} \in \mathcal{U}} \sum_{\bar{\mu} \in \bar{\Pi}_{U|\{Y, \hat{u}\}}} \mu_{\hat{u}}(Y) \prod_{\bar{Y} \neq Y} \sum_{\bar{u}} \hat{\mu}_{\bar{u}}(\bar{Y}) \mu_{\bar{u}}(\bar{Y})} \\ &= \frac{\mu_u(Y)}{\sum_{\hat{u} \in \mathcal{U}} \mu_{\hat{u}}(Y)} = \mu_u(Y) \end{aligned} \quad (6)$$

Here, we used the fact that

$$\begin{aligned} \sum_{\bar{\mu} \in \bar{\Pi}_{U|\{Y, u\}}} \prod_{\bar{Y} \neq Y} \sum_{\bar{u}} \bar{\mu}_{\bar{u}}(\bar{Y}) \mu_{\bar{u}}(\bar{Y}) \\ = \sum_{\hat{\mu} \in \bar{\Pi}_{U|\{Y, \hat{u}\}}} \prod_{\bar{Y} \neq Y} \sum_{\bar{u}} \hat{\mu}_{\bar{u}}(\bar{Y}) \mu_{\bar{u}}(\bar{Y}), \quad \forall \hat{u} \in \mathcal{U}. \end{aligned} \quad (7)$$

This equality holds because for each  $\bar{\mu} \in \bar{\Pi}_{U|\{Y, u\}}$  there is exactly one  $\hat{\mu} \in \bar{\Pi}_{U|\{Y, \hat{u}\}}$  such that  $\hat{\mu}(\bar{Y}) = \bar{\mu}(\bar{Y})$ ,  $\bar{Y} \neq Y$ . This means that each term in the summation on the right-hand side of (7) equals exactly one term in the summation in the left-hand side of the same equation (and vice-versa). Equation (6) proves that  $\bar{L}^U$  is a right-inverse of  $L^U$ , and hence that  $L^U$  is surjective.  $L^D$  can be defined similarly to  $L^U$  in (5) and can be proven to be surjective.

We are now ready to prove that (4) holds. To accomplish this let  $\mu := L^U(\rho)$ ,  $\rho \in [0, 1]^{\bar{\Pi}_U}$ , and  $\delta := L^D(\sigma)$ ,

$\sigma \in [0, 1]^{\bar{\Pi}_D}$ . By definition of  $J_{\mu\delta}^U$  and Lemma 2,

$$J_{\mu\delta}^U = \sum_{\substack{s_0, s_1, \dots, s_T, \\ u_T, d_T, Y_T, Z_T}} \sum_{t=0}^T l_U(s_t, u_t, d_t, t) k_T(\dots) \prod_{\tau=0}^T \mu_{u_\tau}(Y_\tau) \delta_{d_\tau}(Z_\tau). \quad (8)$$

By induction on  $t$ , it is straightforward to show that

$$\prod_{\tau=0}^t \mu_{u_\tau}(Y_\tau) \delta_{d_\tau}(Z_\tau) = \sum_{\bar{\mu} \in \bar{\Pi}_{U|\{Y_t, u_t\}}} \rho_{\bar{\mu}} \sum_{\bar{\delta} \in \bar{\Pi}_{D|\{Z_t, d_t\}}} \sigma_{\bar{\delta}},$$

$t \in \mathcal{T}$  (cf. [18]). Using this equation in (8), we get

$$J_{\mu\delta}^U = \sum_{\substack{s_0, s_1, \dots, s_T, \\ u_T, d_T, Y_T, Z_T}} \sum_{t=0}^T l_U(s_t, u_t, d_t, t) k_T(\dots) \sum_{\bar{\mu} \in \bar{\Pi}_{U|\{Y_T, u_T\}}} \rho_{\bar{\mu}} \sum_{\bar{\delta} \in \bar{\Pi}_{D|\{Z_T, d_T\}}} \sigma_{\bar{\delta}}. \quad (9)$$

On the other hand,

$$\bar{J}_{\rho\sigma}^U := \sum_{\bar{\mu} \in \bar{\Pi}_U} \sum_{\bar{\delta} \in \bar{\Pi}_D} J_{\bar{\mu}\bar{\delta}}^U \rho_{\bar{\mu}} \sigma_{\bar{\delta}}. \quad (10)$$

By specializing (8) to pure policies, we have

$$J_{\bar{\mu}\bar{\delta}}^U = \begin{cases} \sum_{\substack{s_0, s_1, \dots, s_T, \\ u_T, d_T, Y_T, Z_T}} \sum_{t=0}^T l_U(s_t, u_t, d_t, t) k_T(\dots), \\ \text{if } \bar{\mu} \in \bar{\Pi}_{U|\{Y_T, u_T\}}, \bar{\delta} \in \bar{\Pi}_{D|\{Z_T, d_T\}}, \\ 0, \text{ otherwise.} \end{cases}$$

This, together with (10), show that both  $\bar{J}_{\rho\sigma}^U$  and  $J_{\mu\delta}^U$  are equal to the right-hand side of (9). The proof that  $\bar{J}_{\rho\sigma}^D = J_{\mu\delta}^D$  is analogous. ■

We next state and prove the main result of this section:

**Theorem 1** Let  $(\rho^*, \sigma^*) \in [0, 1]^{\bar{\Pi}_U} \times [0, 1]^{\bar{\Pi}_D}$  be a Nash equilibrium in mixed policies. Then  $(\mu^*, \delta^*) \in \Pi_U \times \Pi_D$ , with  $\mu^* := L^U(\rho^*)$ ,  $\delta^* := L^D(\sigma^*)$ , is a Nash equilibrium in stochastic policies.

**Proof of Theorem 1.** We start by proving the first inequality in (1). By contradiction assume that there is a policy  $\mu \in \Pi_U$  for which  $J_{\mu^*\delta^*}^U > J_{\mu\delta^*}^U$ . Since the map  $L^U$  is surjective, there must exist some  $\rho \in [0, 1]^{\bar{\Pi}_U}$  such that  $\mu = L^U(\rho)$ . From  $J_{\mu^*\delta^*}^U > J_{\mu\delta^*}^U$  and Lemma 1, one then concludes that  $\bar{J}_{\rho^*\sigma^*}^U > \bar{J}_{\rho\sigma^*}^U$ , which violates (2). The second inequality in (1) is proved similarly. ■

Since there always exists one Nash equilibrium in mixed policies, from Theorem 1 it follows that there always exists at least one Nash equilibrium in stochastic policies. Moreover, Theorem 1 gives a procedure to actually compute the corresponding stochastic policies, though the bi-matrix game to be solved can be extremely large.

## 4 Dynamic Programming Approach

In this section we look for a necessary and sufficient condition for a pair of stochastic behavioral policies  $(\mu^*, \delta^*) \in \Pi_U \times \Pi_D$  to be Nash equilibrium, based on a dynamic programming approach. With this in mind we first treat the case where one player computes its optimal policy assuming that its opponent is following a known policy. In particular, in Section 4.1 we deal with the following optimization problem: Given  $\delta \in \Pi_D$ , determine  $\mu^* \in \Pi_U$  such that

$$J_{\mu^* \delta}^U = \inf_{\mu \in \Pi_U} J_{\mu \delta}^U. \quad (11)$$

This can be viewed as the problem of determining a Stackelberg equilibrium with player D being the leader and player U the follower ([16]). Analogous results are valid for the case when the players' roles are inverted. The interested reader is referred to [18] for a proof of the results stated in Section 4.1.

### 4.1 Solution to the optimization problem

For given policies  $\mu \in \Pi_U, \delta \in \Pi_D$ , we define  $V_{\mu \delta}^U(Y)$ ,  $Y \in \mathcal{Y}^*$ , to be *player U's cost-to-go from Y* associated with the policies  $\mu$  and  $\delta$ , after having collected a sequence  $Y$  of length  $t := \ell(Y) \in \mathcal{T}$  of observations and controls, i.e.,  $E_{\mu \delta}[\sum_{\tau=t}^T l_U(\mathbf{s}_\tau, \mathbf{u}_\tau, \mathbf{d}_\tau, \tau) \mid \mathbf{Y}_t = Y]$ . The expected value above is only well defined when  $P_{\mu \delta}(\mathbf{Y}_t = Y) \neq 0$  but it is actually convenient to define cost-to-go for any  $Y \in \mathcal{Y}^*$  such that there is some policy  $\hat{\mu}_Y$  for which  $P_{\hat{\mu}_Y \delta}(\mathbf{Y}_t = Y) \neq 0$ . To do this we formally define  $V_{\mu \delta}^U(Y) := E_{\hat{\mu}_Y \delta}[\sum_{\tau=t}^T l_U(\mathbf{s}_\tau, \mathbf{u}_\tau, \mathbf{d}_\tau, \tau) \mid \mathbf{Y}_t = Y]$ , where the policy  $\hat{\mu}_Y$  is given by

$$\tilde{\mu}_Y(\bar{Y}) := \begin{cases} \mu(\bar{Y}), & \ell(\bar{Y}) \geq \ell(Y) \\ \hat{\mu}_Y(\bar{Y}), & \ell(\bar{Y}) < \ell(Y) \end{cases}. \quad (12)$$

The cost  $V_{\mu \delta}^U$  is always well defined because, for all  $Y \in \mathcal{Y}^*$ , 1)  $P_{\tilde{\mu}_Y \delta}(\mathbf{Y}_t = Y) = P_{\hat{\mu}_Y \delta}(\mathbf{Y}_t = Y) \neq 0$ , and 2) the value of  $V_{\mu \delta}^U(Y)$  is independent of the policy  $\hat{\mu}_Y$  chosen to define  $\tilde{\mu}_Y$ . The intuition behind this is that once  $\mathbf{Y}_t = Y$ , it does not really matter what was the value of the policy before time  $t$ .

The cost  $J_{\mu \delta}^U$  associated with a pair of policies  $\mu \in \Pi_U, \delta \in \Pi_D$  can be easily computed from  $V_{\mu \delta}^U$ . Indeed, from the fact that the probability distribution of  $\mathbf{y}_0$  is independent of the policies  $\mu$  and  $\delta$ , we conclude that

$$J_{\mu \delta}^U = E_{\mu \delta}[V_{\mu \delta}^U(\{\mathbf{y}_0\})] = E[V_{\mu \delta}^U(\{\mathbf{y}_0\})]. \quad (13)$$

We shall see next that it is possible to compute  $V_{\mu \delta}^U$  using the operator  $T_{\mu \delta}^U$  from the set of functionals  $\mathcal{V}^U := \{V : \mathcal{Y}^* \rightarrow \mathbb{R}\}$  into itself, defined for each  $Y \in \mathcal{Y}^*$  by

$$T_{\mu \delta}^U V(Y) := E_{\tilde{\mu}_Y \delta}[l_U(\mathbf{s}_t, \mathbf{u}_t, \mathbf{d}_t, t) + V(\mathbf{Y}_{t+1}) \mid \mathbf{Y}_t = Y],$$

where  $t := \ell(Y)$ ,  $V(\mathbf{Y}_{T+1}) := 0$  and  $\tilde{\mu}_Y$  is defined in (12). The following theorem summarizes the relation

between  $V_{\mu \delta}^U$  and  $T_{\mu \delta}^U$ . To state this theorem we need to introduce the following notation: Given policies  $\mu \in \Pi_U$  and  $\delta \in \Pi_D$ , we use  $\mathcal{Y}_{\mu \delta}^*$  to denote the set of values  $Y \in \mathcal{Y}^*$  for which  $P_{\mu \delta}(\mathbf{Y}_\tau = Y) > 0$ ,  $\tau := \ell(Y)$ .

**Theorem 2** *Given  $\mu \in \Pi_U$  and  $\delta \in \Pi_D$ ,*

- i)  $V_{\mu \delta}^U = T_{\mu \delta}^U V_{\mu \delta}^U$ .
- ii) *Any function  $V \in \mathcal{V}^U$  satisfying  $V = T_{\mu \delta}^U V$  on  $\mathcal{Y}_{\mu \delta}^*$  is equal to  $V_{\mu \delta}^U$  on  $\mathcal{Y}_{\mu \delta}^*$ .*

We next show how to actually compute the function in  $\mathcal{V}^U$  that results from applying  $T_{\mu \delta}^U$  to some function  $V \in \mathcal{V}^U$ . To this effect let  $\mathcal{P} := [0, 1]^{\mathcal{U}}$ . For each  $p := \{p_u : u \in \mathcal{U}\} \in \mathcal{P}$ , and  $\delta \in \Pi_D$  define an operator  $H_{p \delta}^U$  from  $\mathcal{V}^U$  into itself by setting

$$H_{p \delta}^U V(Y) := \sum_{\substack{u, s, s' \\ y, d, Z}} p_u \left( l_U(s, u, d, \tau) + V(\{Y, u, y\}) \right) p_Y(y, s') p(s, s', u, d) \delta_d(Z) I_\delta(s, Z, Y),$$

for every  $Y \in \mathcal{Y}^*$ , where  $\tau := \ell(Y)$  and  $V(\{Y, u, y\}) = 0$  for  $\tau = T$ . The function  $I_\delta : \mathcal{S} \times \mathcal{Z}^* \times \mathcal{Y}^* \rightarrow \mathbb{R}$  is defined as follows: For given  $Y \in \mathcal{Y}^*$  and  $Z \in \mathcal{Z}^*$ ,  $I_\delta(s, Z, Y) = 0$ ,  $s \in \mathcal{S}$  when  $\ell(Y) \neq \ell(Z)$ . Otherwise,  $I_\delta$  is defined recursively by

$$I_\delta(s', \{Z, d, z\}, \{Y, u, y\}) = \frac{\sum_s p_Y(y, s') p_Z(z, s') p(s, s', u, d) \delta_d(Z) I_\delta(s, Z, Y)}{\sum_{\substack{\bar{s}, \bar{z}, \bar{d} \\ \bar{s}', \bar{Z}}} p_Y(y, \bar{s}') p_Z(\bar{z}, \bar{s}') p(\bar{s}, \bar{s}', u, \bar{d}) \delta_{\bar{d}}(\bar{Z}) I_\delta(\bar{s}, \bar{Z}, Y)}, \quad (14)$$

$s' \in \mathcal{S}, u \in \mathcal{U}, d \in \mathcal{D}, y \in \mathcal{Y}, z \in \mathcal{Z}$ , and initialized with

$$I_\delta(s', \{z\}, \{y\}) = \frac{p_Y(y, s') p_Z(z, s') \bar{p}_{s'}}{\sum_{\bar{s}} p_Y(y, \bar{s}) \bar{p}_{\bar{s}}}.$$

It can be shown that, for each  $\mu \in \Pi_U$  and  $\delta \in \Pi_D$  such that  $P_{\mu \delta}(\mathbf{Y}_\tau = Y) > 0$ ,  $I_\delta(s, Z, Y) = P_{\mu \delta}(\mathbf{s}(\tau) = s, \mathbf{Z}_\tau = Z \mid \mathbf{Y}_\tau = Y)$ ,  $s \in \mathcal{S}, Z \in \mathcal{Z}^*$ , i.e.,  $I_\delta$  is the so-called *information state*. Moreover,

$$T_{\mu \delta}^U V(Y) = H_{\tilde{\mu}_Y \delta}^U V(Y), \quad Y \in \mathcal{Y}^*. \quad (15)$$

This is actually the key equation for showing that the multi-step optimization problem (11) can be reduced to multiple single-step optimization problems. This is the subject of the developments that follow.

**Optimal Cost-to-Go.** For a given policy  $\delta \in \Pi_D$  we define *player U's optimal cost-to-go function*  $V_\delta^{U*}$  associated with the policy  $\delta$  as  $V_\delta^{U*}(Y) := \inf_{\mu \in \Pi_U} V_{\mu \delta}^U(Y)$ ,  $Y \in \mathcal{Y}^*$ . The optimal cost  $J_\delta^{U*} := \inf_{\mu \in \Pi_U} J_{\mu \delta}^U$  can be easily computed from  $V_\delta^{U*}$ . Indeed,

$$J_\delta^{U*} = E[V_\delta^{U*}(\{\mathbf{y}_0\})]. \quad (16)$$

It turns out that one can compute  $V_\delta^{U*}$  using the operator  $T_\delta^U : \mathcal{V}^U \rightarrow \mathcal{V}^U$  defined by

$$T_\delta^U V(Y) = \inf_{p \in \mathcal{P}} H_{p \delta}^U V(Y), \quad Y \in \mathcal{Y}^*. \quad (17)$$

Due to the linearity of the map  $p \mapsto H_{p\delta}^U V(Y)$  and the particular structure of  $\mathcal{P}$ , the infimum is actually a minimum and can be achieved at some vector in  $\mathcal{P}$  with all entries in the set  $\{0, 1\}$ . The following Theorem summarizes the relation between  $V_\delta^{U*}$  and  $T_\delta^U$ :

**Theorem 3** *Given an arbitrary policy  $\delta \in \Pi_D$ ,*

- i)  $V_\delta^{U*} = T_\delta^U V_\delta^{U*}$ .
- ii) *For any policy  $\mu \in \Pi_U$  satisfying  $V_{\mu\delta}^U = T_\delta^U V_{\mu\delta}^U$  on  $\mathcal{Y}_{\mu\delta}^*$ , we have that  $V_{\mu\delta}^U = V_\delta^{U*}$  on  $\mathcal{Y}_{\mu\delta}^*$ .*
- iii) *If  $\mu^* \in \Pi_U$  is an optimal policy, then  $T_{\mu^*\delta}^U V_\delta^{U*} = T_\delta^U V_\delta^{U*}$  on  $\mathcal{Y}_{\mu^*\delta}^*$ .*

## 4.2 Characterization of Nash equilibria

Using the results in Section 4.1, we finally derive a necessary and sufficient condition for  $(\mu^*, \delta^*) \in \Pi_U \times \Pi_D$  to be a Nash equilibrium. In the sequel, we shall use  $\mathcal{V}^D$ ,  $V_{\mu\delta}^D$ ,  $H_{\mu q}^D$ ,  $V_\mu^{D*}$ ,  $T_\mu^D$ , and  $\mathcal{Q}$  to denote the duals of  $\mathcal{V}^U$ ,  $V_{\mu\delta}^U$ ,  $H_{p\delta}^U$ ,  $V_\delta^{U*}$ ,  $T_\delta^U$ , and  $\mathcal{P}$ , respectively.

**Theorem 4**  $(\mu^*, \delta^*) \in \Pi_U \times \Pi_D$  *constitute a Nash equilibrium if and only if there exists two functionals  $V^U \in \mathcal{V}^U$  and  $V^D \in \mathcal{V}^D$  satisfying*

$$H_{p\delta^*}^U V^U(Y) \geq H_{\mu^*(Y)\delta^*}^U V^U(Y) = V^U(Y), \quad (18)$$

$$H_{\mu^*q}^D V^D(Z) \geq H_{\mu^*\delta^*(Z)}^D V^D(Z) = V^D(Z), \quad (19)$$

$$p \in \mathcal{P}, Y \in \mathcal{Y}_{\mu^*\delta^*}^*, q \in \mathcal{Q}, Z \in \mathcal{Z}_{\mu^*\delta^*}^*.$$

Before proving Theorem 4, note that we actually know from Theorem 1 that there always exists at least one Nash equilibrium in stochastic policies. This means that there must always exist functionals  $V^U \in \mathcal{V}^U$  and  $V^D \in \mathcal{V}^D$  satisfying (18)–(19).

**Proof of Theorem 4.** Suppose that there exist functionals  $V^U \in \mathcal{V}^U$ ,  $V^D \in \mathcal{V}^D$  satisfying (18)–(19). We prove next that  $(\mu^*, \delta^*)$  is a Nash equilibrium because these policies satisfy (1). Because of (15) and the definition of  $T_\delta^U$ , (18) can be rewritten as

$$V^U(Y) = T_{\delta^*}^U V^U(Y) = T_{\mu^*\delta^*}^U V^U(Y), Y \in \mathcal{Y}_{\mu^*\delta^*}^*. \quad (20)$$

Then, by ii) in Theorem 2,  $V^U(Y) = V_{\mu^*\delta^*}^U(Y)$  on  $\mathcal{Y}_{\mu^*\delta^*}^*$ . Based on this, (20) becomes  $V_{\mu^*\delta^*}^U(Y) = T_{\delta^*}^U V_{\mu^*\delta^*}^U(Y) = T_{\mu^*\delta^*}^U V_{\mu^*\delta^*}^U(Y)$ ,  $Y \in \mathcal{Y}_{\mu^*\delta^*}^*$ , which, by ii) in Theorem 3, leads to  $V^U(Y) = V_{\mu^*\delta^*}^U(Y) = V_{\delta^*}^{U*}(Y)$ ,  $Y \in \mathcal{Y}_{\mu^*\delta^*}^*$ . This equation specializes to  $V_{\delta^*}^{U*}(\{y\}) = V_{\mu^*\delta^*}^U(\{y\})$ ,  $y \in \mathcal{Y}$ ,  $P_{\mu^*\delta^*}(\mathbf{y}_0 = y) > 0$ , for sequences of length 0, from which  $J_{\delta^*}^{U*} = J_{\mu^*\delta^*}^U$  follows because of (16) and (13). This concludes the proof of the first inequality in (1). The second inequality can be proved similarly using (19) and reversing the roles of the players.

To prove the converse statement, suppose that  $(\mu^*, \delta^*)$  is a Nash equilibrium. This means, in particular, that  $\mu^*$  minimizes the cost  $J_{\mu\delta^*}^U$ . From i) and

iii) in Theorem 3, it then follows that  $V_{\delta^*}^{U*}(Y) = T_{\delta^*}^U V_{\delta^*}^{U*}(Y) = T_{\mu^*\delta^*}^U V_{\delta^*}^{U*}(Y)$ ,  $Y \in \mathcal{Y}_{\mu^*\delta^*}^*$ . Similarly,  $V_{\mu^*}^{D*}(Z) = T_{\mu^*}^D V_{\mu^*}^{D*}(Z) = T_{\mu^*\delta^*}^D V_{\mu^*}^{D*}(Z)$ ,  $Z \in \mathcal{Z}_{\mu^*\delta^*}^*$ . Equations (18)–(19) are then satisfied by  $V^U = V_{\delta^*}^{U*}$  and  $V^D = V_{\mu^*}^{D*}$ .

## References

- [1] S. Sorin and S. Zamir, ““Big Match” with lack of information on one side (III),” in *Raghavan et al.* [19], pp. 101–112.
- [2] C. Melolidakis, “Stochastic games with lack of information on one side and positive stop probabilities,” in *Raghavan et al.* [19], pp. 113–126.
- [3] H. Cole and N. Kocherlakota, “Dynamic games with hidden actions and hidden states,” *J. Economic Theory*, vol. 98, no. 1, pp. 114–126, 2001.
- [4] J. P. Hespanha, M. Prandini, and S. Sastry, “Probabilistic pursuit-evasion games: A one-step Nash approach,” in *Proc. of the 39th CDC Conf.*, Dec. 2000.
- [5] J. P. Hespanha, Y. S. Ateşkan, and H. H. Kızıloçak, “Deception in non-cooperative games with partial information,” in *Proc. of the 2nd DARPA-JFACC Symp. on Advances in Enterprise Control*, July 2000.
- [6] G. Kimeldorf, “Duels: An overview,” in *Math. of Conflict* (M. Shubik, ed.), pp. 55–72, Amsterdam: North-Holland, 1983.
- [7] P. Bernhard, A.-L. Colomb, and G. P. Papavasiliopoulos, “Rabbit and hunter game: Two discrete stochastic formulations,” *Comput. Math. Applic.*, vol. 13, no. 1–3, pp. 205–225, 1987.
- [8] G. J. Olsder and G. P. Papavasiliopoulos, “About when to use a searchlight,” *J. Math. Analysis and Applic.*, vol. 136, pp. 466–478, 1988.
- [9] I. B. Rhodes and D. G. Luenberger, “Differential games with imperfect state information,” *IEEE Trans. Automat. Contr.*, vol. AC-14, pp. 29–38, Feb. 1969.
- [10] P. Kumar and J. V. Schuppen, “On Nash equilibrium solutions in stochastic dynamic games,” *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 1146–1149, Dec. 1980.
- [11] P. R. Kumar, “A differential game with jump process observations,” *J. Optim. Theory and Applic.*, vol. 31, June 1980.
- [12] J. C. Harsanyi, “Games of incomplete information played by Bayesian players. Part I, II, III,” *Management Science*, vol. 14, no. 3, 5, 7, pp. 159–182, 320–334, 486–502, 1967–68.
- [13] S. Zamir, “Repeated games of incomplete information,” in *Handbook of Game Theory* (R. J. Aumann and S. Hart, eds.), vol. 1, pp. 109–154, Elsevier Science Publisher B.V., 1992.
- [14] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*. New York: Springer-Verlag, 1997.
- [15] H. W. Kuhn, “Extensive games and the problem of information,” in *Contributions to the Theory of Games* (H. W. Kuhn and A. W. Tucker, eds.), vol. II, pp. 193–216, Princeton: Princeton University Press, 1953.
- [16] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. No. 23 in Classics in Applied Mathematics, Philadelphia: SIAM, 2nd ed., 1999.
- [17] J. Nash, “Non-cooperative games,” *Annals of Mathematics*, vol. 54, pp. 286–295, 1951.
- [18] J. P. Hespanha and M. Prandini, “Nash equilibria in partial-information games on Markov chains,” tech. rep., Univ. of Southern California, Los Angeles, CA, Mar. 2001.
- [19] T. E. S. Raghavan, T. S. Ferguson, and T. Parthasarathy, eds., *Stochastic Games and Related Topics: In Honor of Professor L. S. Shapley*, vol. 7 of *Theory and Decision Library, Series C, Game Theory, Mathematical Programming and Operations Research*. Dordrecht: Kluwer Academic Publishers, 1991.