

# Qualitative Determinacy and Decidability of Stochastic Games with Signals \*

Nathalie Bertrand<sup>1</sup>, Blaise Genest<sup>2</sup>, Hugo Gimbert<sup>3</sup>

<sup>1</sup> INRIA, IRISA  
Rennes, France

nathalie.bertrand@irisa.fr

<sup>2</sup> CNRS, IRISA  
Rennes, France

blaise.genest@irisa.fr

<sup>3</sup> CNRS, LaBRI  
Bordeaux, France

hugo.gimbert@labri.fr

## Abstract

We consider the standard model of finite two-person zero-sum stochastic games with signals. We are interested in the existence of almost-surely winning or positively winning strategies, under reachability, safety, Büchi or co-Büchi winning objectives. We prove two qualitative determinacy results. First, in a reachability game either player 1 can achieve almost-surely the reachability objective, or player 2 can ensure surely the complementary safety objective, or both players have positively winning strategies. Second, in a Büchi game if player 1 cannot achieve almost-surely the Büchi objective, then player 2 can ensure positively the complementary co-Büchi objective. We prove that players only need strategies with finite-memory, whose sizes range from no memory at all to doubly-exponential number of states, with matching lower bounds. Together with the qualitative determinacy results, we also provide fix-point algorithms for deciding which player has an almost-surely winning or a positively winning strategy and for computing the finite memory strategy. Complexity ranges from EXPTIME to 2EXPTIME with matching lower bounds, and better complexity can be achieved for some special cases where one of the players is better informed than her opponent.

## Introduction

Numerous advances in algorithmics of stochastic games have recently been made [10, 9, 7, 5, 12, 14], motivated in part by application in controller synthesis and verification of open systems. Open systems can be viewed as two-players games between the system and its environment. At each round of the game, both players independently and simultaneously choose actions and the two choices together with the current state of the game determine transition probabilities to the next state of the game. Properties of open sys-

tems are modeled as objectives of the games [9, 13], and strategies in these games represent either controllers of the system or behaviors of the environment.

Most algorithms for stochastic games suffer from the same restriction: they are designed for games where players can fully observe the state of the system (e.g. concurrent games [10, 9] and stochastic games with perfect information [8, 14]). The full observation hypothesis can hinder interesting applications in controller synthesis. In many systems in interaction with both a user and a controller, full monitoring for the controller is hardly implementable in practice and the user has very partial information about the system. Recently, algorithms for games where one of the players has partial observation and her opponent is fully informed have been proposed [17, 6]. Here we consider the general case where both players have partial observations.

In the present paper, we consider *stochastic games with signals*, that are a standard tool in game theory to model partial observation [23, 20, 18]. When playing a stochastic game with signals, players cannot observe the actual state of the game, nor the actions played by their opponent, but are only informed via private signals they receive throughout the play. Stochastic games with signals subsume standard stochastic games [22], repeated games with incomplete information [1], games with imperfect monitoring [20], concurrent games [9] and deterministic games with imperfect information on one side [17, 6]. Players make their decisions based upon the sequence of signals they receive: a strategy is hence a mapping from finite sequences of private signals to probability distributions over actions.

From the algorithmic point of view, stochastic games with signals are considerably harder to deal with than stochastic games with full observation. While *values* of the latter games are computable [9, 5], simple questions like ‘is there a strategy for player 1 which guarantees winning with probability more than  $\frac{1}{2}$ ?’ are *undecidable* even for restricted classes of stochastic games with signals [16]. For this reason, rather than *quantitative* properties (i.e. questions about values), we focus in the present paper on *qualitative* properties of stochastic games with signals.

\*This work is supported by ANR-06-SETI DOTS. An earlier version with complete proofs is available as technical report [3].

We study the following qualitative questions about stochastic games with signals, equipped with reachability, safety, Büchi or co-Büchi objectives:

- (i) Does player 1 have an *almost-surely winning strategy*, i.e. a strategy which guarantees the objective to be achieved with probability 1, whatever the strategy of player 2?
- (ii) Does player 2 have a *positively winning strategy*, i.e. a strategy which guarantees the opposite objective to be achieved with strictly positive probability, whatever the strategy of player 1?

Obviously, given an objective, properties (i) and (ii) cannot hold simultaneously. For games with a reachability, safety or Büchi objective, we obtain the following results:

- (1) Either property (i) holds or property (ii) holds; in other words these games are *qualitatively determined*.
- (2) Players only need strategies with *finite-memory*, whose memory sizes range from no memory at all to doubly-exponential number of states.
- (3) Questions (i) and (ii) are decidable. We provide fix-point algorithms for computing uniformly all initial states that satisfy (i) or (ii), together with the corresponding finite-memory strategies. The complexity of the algorithms ranges from EXPTIME to 2EXPTIME.

These three results are detailed in Theorems 1, 2, 3 and 4. We prove that these results are tight and robust in several aspects. Games with co-Büchi objectives are absent from these results, since they are neither qualitatively determined (see Fig. 3) nor decidable (as proven in [2]).

Our *main result*, and the element of surprise, is that for winning positively a safety or co-Büchi objective, a player *needs* a memory with a doubly-exponential number of states, and the corresponding decision problem is 2EXPTIME-complete. This result departs from what was previously known [17, 6], where both the number of memory states and the complexity are simply exponential. These results also reveal a nice property of *reachability* games, that Büchi games do not enjoy: Every initial state is either *almost-surely winning* for player 1, *surely winning* for player 2 or *positively winning* for both.

Our results strengthen and generalize in several ways results that were previously known for concurrent games [10, 9] and deterministic games with imperfect information on one side [17, 6]. First, the framework of stochastic games with signals strictly encompasses all the settings of [17, 10, 9, 6]. In concurrent games there is no signaling structure at all, and in deterministic games with imperfect information on one side [6] transitions are deterministic and player 2 observes everything that happens in the game, including results of random choices of her opponent.

No determinacy result was known for deterministic games with imperfect information on one side. In [17, 6], algorithms are given for deciding whether the imperfectly informed player has an almost-surely winning strategy for a Büchi (or reachability) objective but nothing can be inferred in case she has no such strategy. This open question is solved in the present paper, in the broader framework of stochastic games with signals.

Our qualitative determinacy result (1) is a radical generalization of the same result for concurrent games [9, Th.2], while proofs are very different. Interestingly, for concurrent games, qualitative determinacy holds for every omega-regular objectives [9], while for games with signals we show that it fails already for co-Büchi objectives. Interestingly also, stochastic games with signals and a reachability objective have a value [19] but this value is not computable [16], whereas it is computable for concurrent games with omega-regular objectives [11]. The use of randomized strategies is mandatory for achieving determinacy results, this also holds for stochastic games without signals [22, 10] and even matrix games [24], which contrasts with [4, 17] where only deterministic strategies are considered.

Our results about randomized finite-memory strategies (2), stated in Theorem 2, are either brand new or generalize previous work. It was shown in [6] that for deterministic games where player 2 is perfectly informed, strategies with a finite memory of exponential size are sufficient for player 1 to achieve a Büchi objective almost-surely. We prove the same result holds for the whole class of stochastic games with signals. Moreover we prove that for player 2 a doubly-exponential number of memory states is necessary and sufficient for achieving positively the complementary co-Büchi objective.

Concerning algorithmic results (3) (see details in Theorem 3 and 4) we show that our algorithms are optimal in the following meaning. First, we give a fix-point based algorithm for deciding whether a player has an almost-surely winning strategy for a Büchi objective. In general, this algorithm is 2EXPTIME. We show in Theorem 5 that this problem is indeed 2EXPTIME-hard. However, in the restricted setting of [6], it is already known that this problem is only EXPTIME-complete. We show that our algorithm is also optimal with an EXPTIME complexity not only in the setting of [6] where player 2 has perfect information but also under weaker hypothesis: it is sufficient that player 2 has *more information* than player 1. Our algorithm is also EXPTIME when player 1 has full information (Proposition 2). In both subcases, player 2 needs only exponential memory.

Part of our results have been concurrently obtained in [21] whose contribution is weaker than our: no determinacy result is provided, nothing is said about strategies used by player 2 nor the memory she needs, and the algorithm provided is enumerative rather than fix-point based.

The paper is organized as follows. In Section 1 we introduce partial observation games, in Section 2 we define the notion of qualitative determinacy and we state our determinacy result, in Section 3 we discuss the memory needed by strategies. Section 4 is devoted to decidability questions and Section 5 investigates the precise complexity of the general problem as well as special cases. Due to lack of space most proofs are missing, they can be found in [3].

## 1 Stochastic games with signals.

We consider the standard model of finite two-person zero-sum stochastic games with signals [23, 20, 18]. These are stochastic games where players cannot observe the actual state of the game, nor the actions played by their opponent, their only source of information are private signals they receive throughout the play. Stochastic games with signals subsume standard stochastic games [22], repeated games with incomplete information [1], games with imperfect monitoring [20] and games with imperfect information [6].

**Notations.** Given a finite set  $K$ , we denote by  $\mathcal{D}(K) = \{\delta : K \rightarrow [0, 1] \mid \sum_k \delta(k) = 1\}$  the set of probability distributions on  $K$  and for a distribution  $\delta \in \mathcal{D}(K)$ , we denote  $\text{supp}(\delta) = \{k \in K \mid \delta(k) > 0\}$  its support.

**States, actions and signals.** Two players called 1 and 2 have opposite goals and play for an infinite sequence of steps, choosing actions and receiving signals. Players observe their own actions and signals but they cannot observe the actual state of the game, nor the actions played and the signals received by their opponent. We borrow notations from [18]. Initially, the game is in a state  $k_0 \in K$  chosen according to an initial distribution  $\delta \in \mathcal{D}(K)$  known by both players; the initial state is  $k_0$  with probability  $\delta(k_0)$ . At each step  $n \in \mathbb{N}$ , players 1 and 2 choose some actions  $i_n \in I$  and  $j_n \in J$ . They respectively receive signals  $c_n \in C$  and  $d_n \in D$ , and the game moves to a new state  $k_{n+1}$ . This happens with probability  $p(k_{n+1}, c_n, d_n \mid k_n, i_n, j_n)$  given by fixed transition probabilities  $p : K \times I \times J \rightarrow \mathcal{D}(K \times C \times D)$ , known by both players. Formally a game is a tuple  $(K, I, J, C, D, p)$ .

**Plays and strategies.** Players observe their own actions and the signals they receive. It is convenient to assume that the action  $i$  player 1 plays is encoded in the signal  $c$  she receives, with the notation  $i = i(c)$  (and symmetrically for player 2). This way, plays can be described by sequences of states and signals for both players, without mentioning which actions were played. A finite play is a sequence  $(k_0, c_1, d_1, \dots, c_n, d_n, k_n) \in (KCD)^*K$

such that for every  $0 \leq m < n$ ,  $p(k_{m+1}, c_{m+1}, d_{m+1} \mid k_m, i(c_{m+1}), j(d_{m+1})) > 0$ . An infinite play is a sequence in  $(KCD)^\omega$  whose prefixes are finite plays.

A (behavioral) strategy of player 1 is a mapping  $\sigma : \mathcal{D}(K) \times C^* \rightarrow \mathcal{D}(I)$ . If the initial distribution is  $\delta$  and player 1 has seen signals  $c_1, \dots, c_n$  then she plays action  $i$  with probability  $\sigma(\delta, c_1, \dots, c_n)(i)$ . Strategies for player 2 are defined symmetrically. In the usual way, an initial distribution  $\delta$  and two strategies  $\sigma$  and  $\tau$  define a probability measure  $\mathbb{P}_\delta^{\sigma, \tau}$  on the set of infinite plays, equipped with the  $\sigma$ -algebra generated by cylinders.

We use random variables  $K_n, I_n, J_n, C_n$  and  $D_n$  to denote respectively the  $n$ -th state, action of player 1, action of player 2, signal of player 1 and signal of player 2.

**Winning conditions.** The goal of player 1 is described by a measurable event Win called the *winning condition*. Motivated by applications in logic and controller synthesis [13], we are especially interested in *reachability*, *safety*, *Büchi* and *co-Büchi conditions*. These four winning conditions use a subset  $T \subseteq K$  of *target states* in their definition. The reachability condition stipulates that  $T$  should be visited at least once,  $\text{Win} = \{\exists n \in \mathbb{N}, K_n \in T\}$ , the safety condition is complementary  $\text{Win} = \{\forall n \in \mathbb{N}, K_n \notin T\}$ . For the Büchi condition the set of target states has to be visited infinitely often,  $\text{Win} = \{\forall m \in \mathbb{N}, \exists n \geq m, K_n \in T\}$ , and the co-Büchi condition is complementary  $\text{Win} = \{\exists m \in \mathbb{N}, \forall n \geq m, K_n \notin T\}$ .

**Almost-surely and positively winning strategies.** When player 1 and 2 use strategies  $\sigma$  and  $\tau$  and the initial distribution is  $\delta$ , then player 1 wins the game with probability:

$$\mathbb{P}_\delta^{\sigma, \tau}(\text{Win}) .$$

Player 1 wants to maximize this probability, while player 2 wants to minimize it. The best situation for player 1 is when she has an almost-surely winning strategy.

**Definition 1** (Almost-surely winning strategy). *A strategy  $\sigma$  for player 1 is almost-surely winning from an initial distribution  $\delta$  if*

$$\forall \tau, \mathbb{P}_\delta^{\sigma, \tau}(\text{Win}) = 1 . \quad (1)$$

*When such a strategy  $\sigma$  exists, both  $\delta$  and its support  $\text{supp}(\delta)$  are said to be almost-surely winning as well.*

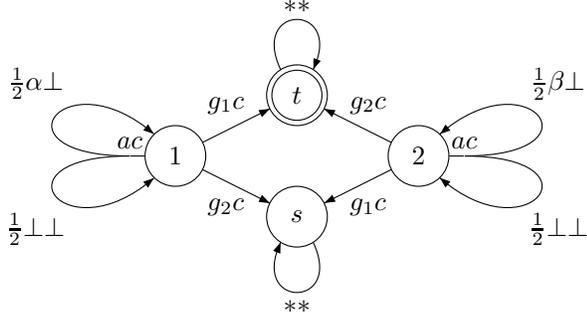
A less enjoyable situation for player 1 is when she only has a positively winning strategy.

**Definition 2** (Positively winning strategy). *A strategy  $\sigma$  for player 1 is positively winning from an initial distribution  $\delta$  if*

$$\forall \tau, \mathbb{P}_\delta^{\sigma, \tau}(\text{Win}) > 0 . \quad (2)$$

*When such a strategy  $\sigma$  exists, both  $\delta$  and its support  $\text{supp}(\delta)$  are said to be positively winning as well.*

The worst situation for player 1 is when her opponent has an almost-surely winning strategy  $\tau$ , which ensures  $\mathbb{P}_\delta^{\sigma, \tau}(\text{Win}) = 0$  for all strategies  $\sigma$  chosen by player 1. Symmetrically, a strategy  $\tau$  for player 2 is positively winning if it guarantees  $\forall \sigma, \mathbb{P}_\delta^{\sigma, \tau}(\text{Win}) < 1$ . These notions only depend on the support of  $\delta$  since  $\mathbb{P}_\delta^{\sigma, \tau}(\text{Win}) = \sum_{k \in K} \delta(k) \cdot \mathbb{P}_{1_k}^{\sigma, \tau}(\text{Win})$ .



**Figure 1.** When the initial state is chosen at random between states 1 and 2, player 1 has a strategy to reach  $t$  almost surely.

Consider the one-player game depicted on Fig. 1. The objective of player 1 is to reach state  $t$ . The initial distribution is  $\delta(1) = \delta(2) = \frac{1}{2}$  and  $\delta(t) = \delta(s) = 0$ . Player 1 plays with actions  $I = \{a, g_1, g_2\}$ , where  $g_1$  and  $g_2$  mean respectively ‘guess 1’ and ‘guess 2’, while player 2 plays with actions  $J = \{c\}$  (that is, player 2 has no choice). Player 1 receives signals  $C = \{\alpha, \beta, \perp\}$  and player 2 is ‘blind’, she always receives the same signal  $D = \{\perp\}$ . Transitions probabilities are represented in a quite natural way. When the game is in state 1, player 1 plays  $a$  and player 2 plays  $c$ , then player 1 receives signal  $\alpha$  or  $\perp$  with probability  $\frac{1}{2}$ , player 2 receives signal  $\perp$  and the game stays in state 1. In state 2 when action of player 1 is  $a$  and action of player 2 is  $c$ , player 1 cannot receive signal  $\alpha$  but instead she may receive signal  $\beta$ . When ‘guessing the state’ i.e. playing action  $g_i$  in state  $j \in \{1, 2\}$ , player 1 wins the game if  $i = j$  (she guesses the correct state) and loses the game if  $i \neq j$ . The star symbol  $*$  stands for any action. In this game, player 1 has a strategy to reach  $t$  almost surely. Her strategy is to keep playing action  $a$  as long as she keeps receiving signal  $\perp$ . The day player 1 receives signal  $\alpha$  or  $\beta$ , she plays respectively action  $g_1$  or  $g_2$ . This strategy is almost-surely winning because the probability for player 1 to receive signal  $\perp$  forever is 0.

## 2 Qualitative Determinacy.

If an initial distribution is positively winning for player 1 then by definition it is *not* almost-surely winning for his op-

ponent player 2. A natural question is whether the converse implication holds.

**Definition 3** (Qualitative determinacy). *A winning condition Win is qualitatively determined if for every game equipped with Win, every initial distribution is either almost-surely winning for player 1 or positively winning for player 2.*

**Comparison with value determinacy.** *Qualitative determinacy is similar to but different from the usual notion of (value) determinacy which refers to the existence of a value. Actually both qualitative determinacy and value determinacy are formally expressed by a quantifier inversion. On one hand, qualitative determinacy rewrites as:*

$$(\forall \sigma \exists \tau \mathbb{P}_\delta^{\sigma, \tau}(\text{Win}) < 1) \implies (\exists \tau \forall \sigma \mathbb{P}_\delta^{\sigma, \tau}(\text{Win}) < 1) .$$

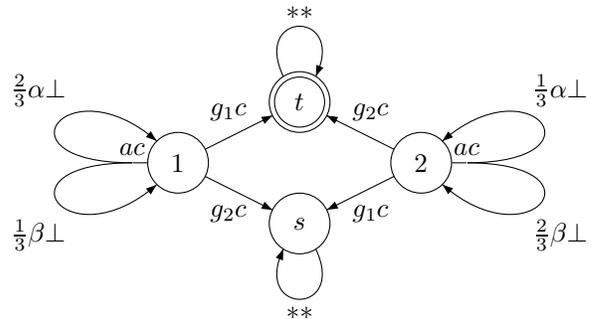
On the other hand, the game has a value if:

$$\sup_\sigma \inf_\tau \mathbb{P}_\delta^{\sigma, \tau}(\text{Win}) \geq \inf_\tau \sup_\sigma \mathbb{P}_\delta^{\sigma, \tau}(\text{Win}) .$$

Both the converse implication of the first equation and the converse inequality of the second equation are obvious.

While *value determinacy* is a classical notion in game theory [15], to our knowledge the notion of *qualitative determinacy* appeared only in the context of omega-regular concurrent games [10, 9] and stochastic games with perfect information [14].

Existence of an almost-surely winning strategy ensures that the value of the game is 1, but the converse is not true. Actually it can even hold that player 2 has a positively winning strategy while at the same time the value of the game is 1. For example, consider the game depicted on Fig. 2, which is a slight modification of Fig. 1 (only signals of player 1 and transitions probabilities differ). Player 1 has



**Figure 2.** A reachability game with value 1 where player 2 has a positively winning strategy.

signals  $\{\alpha, \beta\}$  and similarly to the game on Fig 1, her goal is to reach the target state  $t$  by guessing correctly whether the initial state is 1 or 2. On one hand, player 1 can guarantee a winning probability as close to 1 as she wants: she plays  $a$  for a long time and compares how often she received signals  $\alpha$  and  $\beta$ . If signals  $\alpha$  were more frequent, then she plays action  $g_1$ , otherwise she plays action  $g_2$ . Of course, the longer player 1 plays  $a$ 's the more accurate the prediction will be. On the other hand, the only strategy available to player 2 (always playing  $c$ ) is positively winning, because any sequence of signals in  $\{\alpha, \beta\}^*$  can be generated with positive probability from both states 1 and 2.

**Qualitative determinacy results.** The first main result of this paper is the qualitative determinacy of stochastic games with signals for the following winning objectives.

**Theorem 1.** *Reachability, safety and Büchi games are qualitatively determined.*

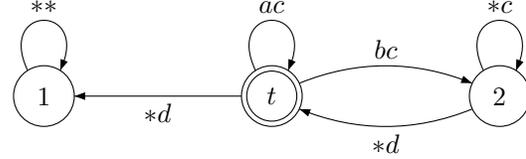
While qualitative determinacy of safety games is not too hard to establish, proving determinacy of Büchi games is harder. Notice that the qualitative determinacy of Büchi games implies the qualitative determinacy of reachability games, since any reachability game can be turned into an equivalent Büchi one by making all target states absorbing.

The proof of Theorem 1 is postponed to Section 4, where the determinacy result will be completed by a decidability result: there are algorithms for computing which initial distributions are almost-surely winning for player 1 or positively winning for player 2. This is stated precisely in Theorems 3 and 4.

A consequence of Theorem 1 is that in a reachability game, every initial distribution is either almost-surely winning for player 1, *surely* winning for player 2, or positively winning for both players. Surely winning means that player 2 has a strategy  $\tau$  for preventing every finite play consistent with  $\tau$  from visiting target states.

Büchi games do not share this nice feature because co-Büchi games are not qualitatively determined. An example of a co-Büchi game which is not determined is represented in Fig. 3. In this game, player 1 observes everything, player 2 is blind (she only observes her own actions), and player 1's objective is to avoid state  $t$  from some moment on. The initial state is  $t$ .

On one hand, player 1 does not have an almost-surely winning strategy for the co-Büchi objective. Fix a strategy  $\sigma$  for player 1 and suppose it is almost-surely winning. To win against the strategy where player 2 plays  $c$  forever, with probability  $1 - \sigma$  should eventually play a  $b$ . Otherwise, the probability that the play stays in state  $t$  is positive, and  $\sigma$  is not almost-surely winning, a contradiction. Since  $\sigma$  is fixed there exists a date after which player 1 has played  $b$  with probability arbitrarily close to 1. Consider the strategy



**Figure 3. Co-Büchi games are not qualitatively determined.**

of player 2 which plays  $d$  at that date. Although player 2 is blind, obviously she can play such a strategy which requires only counting time elapsed since the beginning of the play. With probability arbitrarily close to 1, the game is in state 2 and playing a  $d$  puts the game back in state  $t$ . Playing long sequences of  $c$ 's followed by a  $d$ , player 2 can ensure with probability arbitrarily close to 1 that if player 1 plays according to  $\sigma$ , the play will visit states  $t$  and 2 infinitely often, hence will be lost by player 1. This contradicts the existence of an almost-surely winning strategy for player 1.

On the other hand, player 2 does not have a positively winning strategy either. Fix a strategy  $\tau$  for player 2 and suppose it is positively winning. Once  $\tau$  is fixed, player 1 knows how long she should wait so that if action  $d$  was never played by player 2 then there is arbitrarily small probability that player 2 will play  $d$  in the future. Player 1 plays  $a$  for that duration. If player 2 plays a  $d$  then the play reaches state 1 and player 1 wins, otherwise the play stays in state  $t$ . In the latter case, player 1 plays action  $b$ . Player 1 knows that with very high probability player 2 will play  $c$  forever in the future, in that case the play stays in state 2 and player 1 wins. If player 1 is very unlucky then player 2 will play  $d$  again, but this occurs with small probability and then player 1 can repeat the same process again and again. Similar examples can be used to prove that stochastic Büchi games with signals do not have a value [19].

### 3 Memory needed by strategies.

#### 3.1 Finite-memory strategies.

Since our ultimate goal are algorithmic results and controller synthesis, we are especially interested in strategies that can be finitely described, like finite-memory strategies.

**Definition 4** (Finite-memory strategy). *A finite-memory strategy for player 1 is given by a finite set  $M$  called the memory together with a strategic function  $\sigma_M : M \rightarrow \mathcal{D}(I)$ , an update function  $\text{upd}_M : M \times C \rightarrow \mathcal{D}(M)$ , and an initialization function  $\text{init}_M : \mathcal{P}(K) \rightarrow \mathcal{D}(M)$ . The memory size is the cardinal of  $M$ .*

In order to play with a finite-memory strategy, a player proceeds as follows. She initializes the memory of  $\sigma$  to  $\text{init}_M(L)$ , where  $L = \text{supp}(\delta)$  is the support of the initial distribution  $\delta$ . When the memory is in state  $m \in M$ , she plays action  $i$  with probability  $\sigma_M(m)(i)$  and after receiving signal  $c$ , the new memory state is  $m'$  with probability  $\text{upd}_M(m, c)(m')$ .

On one hand it is intuitively clear how to play with a finite-memory strategy, on the other hand the behavioral strategy associated with a finite-memory strategy<sup>1</sup> can be quite complicated and requires the player to use infinitely many different probability distributions to make random choices (see discussions in [10, 9, 14]).

In the games we consider, the construction of finite-memory strategies is often based on the notion of *belief*. The belief of a player at some moment of the play is the set of states she thinks the game could possibly be in, according to the signals she received so far.

**Definition 5 (Belief).** *From an initial set of states  $L \subseteq K$ , the belief of player 1 after receiving signal  $c$  (hence playing action  $i(c)$ ), is the set of states  $k$  such that there exists a state  $l$  in  $L$  and a signal  $d \in D$  with  $p(k, c, d \mid l, i(c), j(d)) > 0$ . The belief of player 1 after receiving a sequence of signals  $c_1, \dots, c_n$  is defined inductively by:*

$$\mathcal{B}_1(L, c_1, \dots, c_n) = \mathcal{B}_1(\mathcal{B}_1(L, c_1, \dots, c_{n-1}), c_n).$$

*Beliefs of player 2 are defined similarly.*

Our second main result is that for the qualitatively determined games of Theorem 1, finite-memory strategies are sufficient for both players. The amount of memory needed by these finite-memory strategies is summarized in Table 1 and detailed in Theorem 2.

|                     | Almost-surely | Positively |
|---------------------|---------------|------------|
| <b>Reachability</b> | belief        | memoryless |
| <b>Safety</b>       | belief        | doubly-exp |
| <b>Büchi</b>        | belief        |            |
| <b>Co-Büchi</b>     |               | doubly-exp |

**Table 1. Memory required by strategies.**

**Theorem 2 (Finite-memory is sufficient).** *Every reachability game is either won positively by player 1 or won surely by player 2. In the first case playing randomly any action is a positively winning strategy for player 1 and in the second case player 2 has a surely winning strategy with finite-memory  $\mathcal{P}(K)$  and update function  $\mathcal{B}_2$ .*

*Every Büchi game is either won almost-surely by player 1 or won positively by player 2. In the first case player 1 has*

<sup>1</sup>cf. [3] for a precise definition.

*an almost-surely winning strategy with finite-memory  $\mathcal{P}(K)$  and update function  $\mathcal{B}_1$ . In the second case player 2 has a positively winning strategy with finite-memory  $\mathcal{P}(\mathcal{P}(K) \times K)$ .*

The situation where a player needs the least memory is when she wants to win positively a reachability game. To do so, she uses a memoryless strategy consisting in playing randomly any action.

To win almost-surely games with reachability, safety and Büchi objectives, it is sufficient for a player to remember her belief. A canonical almost-surely winning strategy consists in playing randomly any action which ensures the next belief to be almost-surely winning<sup>2</sup>. Similar strategies were used in [6]. These two results are not very surprising: although they were not stated before as such, they can be proved using techniques similar to those used in [17, 6].

The element of surprise is the amount of memory needed for winning positively co-Büchi and safety games. In these situations, it is still enough for player 1 to use a strategy with finite-memory but, surprisingly perhaps, an exponential size memory is not enough. Instead doubly-exponential memory is necessary as will be proved in the next subsection.

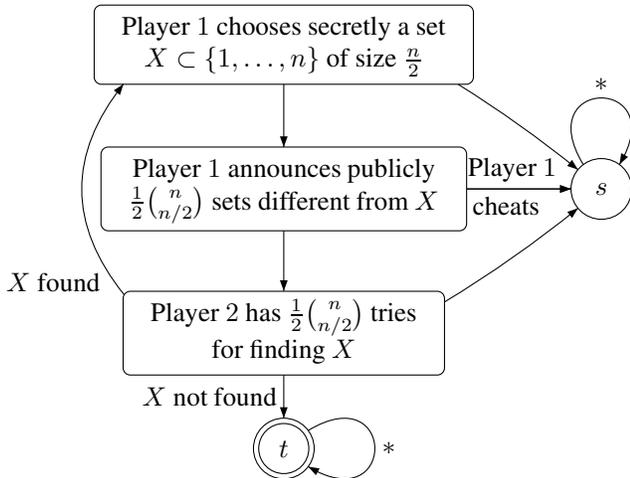
Doubly-exponential size memory is also sufficient. Actually for winning positively, it is enough for player 1 to make hypothesis about beliefs of player 2, and to store in her memory all pairs  $(k, L)$  of possible current state and belief of her opponent. The update operator of the corresponding finite-memory strategy uses numerous random choices so that the opponent is unable to predict future moves. More details are available in the proof of Theorem 4.

### 3.2 Doubly-exponential memory is necessary to win positively safety games.

We now show that a doubly-exponential memory is necessary to win positively safety (and hence co-Büchi) games. We construct, for each integer  $n$ , a reachability game, whose number of state is polynomial in  $n$  and such that player 2 has a positively winning strategy for her safety objective. This game, called `guess_my_setn`, is described on Fig. 4. The objective of player 2 is to stay away from  $t$ , while player 1 tries to reach  $t$ .

We prove that whenever player 2 uses a finite-memory strategy in the game `guess_my_setn` then the size of the memory has to be doubly-exponential in  $n$ , otherwise the safety objective of player 2 may not be achieved with positive probability. This is stated precisely later in Proposition 1. Prior to that, we briefly describe the game `guess_my_setn` for fixed  $n \in \mathbb{N}$ .

<sup>2</sup>for reachability and safety games, we suppose without loss of generality that target states are absorbing.



**Figure 4. A game where player 2 needs a lot of memory to stay away from target state  $t$ .**

**Idea of the game.** The game  $\text{guess\_my\_set}_n$  is divided into three parts. In the first part, player 1 generates a set  $X \subseteq \{1, \dots, n\}$  of size  $|X| = n/2$ . There are  $\binom{n}{n/2}$  possibilities of such sets  $X$ . Player 2 is blind in this part and has no action to play.

In the second part, player 1 announces by her actions  $\frac{1}{2} \binom{n}{n/2}$  (pairwise different) sets of size  $n/2$  which are different from  $X$ . Player 2 has no action to play in that part, but she observes the actions of player 1 (and hence the sets announced by player 1).

In the third part, player 2 can announce by her action up to  $\frac{1}{2} \binom{n}{n/2}$  sets of size  $n/2$ . Player 1 observes actions of player 2. If player 2 succeeds in finding the set  $X$ , the game restarts from scratch. Otherwise, the game goes to state  $t$  and player 1 wins.

It is worth noticing that in order to implement the game  $\text{guess\_my\_set}_n$  in a compact way, we allow player 1 to cheat, and rely on probabilities to always have a chance to catch player 1 cheating, in which case the game is sent to the sink state  $s$ , and player 1 loses. That is, player 1 has to play following the rules without cheating else she cannot win almost-surely her reachability objective. However we do not need to allow player 2 to cheat. Notice that player 1 is better informed than player 2 in this game.

**Concise encoding.** We now turn to a more formal description of the game  $\text{guess\_my\_set}_n$ , to prove that it can be encoded with a number of states polynomial in  $n$ . There are three problems to be solved, that we sketch here. First, remembering set  $X$  in the state of the game would ask for an exponential number of states. Instead, we use a fairly standard technique: recall at random a single element  $x \in X$ .

In order to check that a set  $Y$  of size  $n/2$  is different from the set  $X$  of size  $n/2$ , we challenge player 1 to point out some element  $y \in Y \setminus X$ . We ensure by construction that  $y \in Y$ , for instance by asking it when  $Y$  is given. This way, if player 1 cheats, then she will give  $y \in X$ , leaving a positive probability that  $y = x$ , in which case the game is sure that player 1 is cheating and punishes player 1 by sending her to state  $s$  where she loses.

The second problem is to make sure that player 1 generates an exponential number of pairwise different sets  $X_1, X_2, \dots, X_{\frac{1}{2} \binom{n}{n/2}}$ . Notice that the game cannot recall even one set. Instead, player 1 generates the sets in some total order, denoted  $<$ , and thus it suffices to check only one inequality each time a set  $X_{i+1}$  is given, namely  $X_i < X_{i+1}$ . It is done in a similar but more involved way as before, by remembering randomly two elements of  $X_i$  instead of one.

The last problem is to count up to  $\frac{1}{2} \cdot \binom{n}{n/2}$  with a logarithmic number of bits. Again, we ask player 1 to increment a counter, while remembering only one of the bits and punishing her if she increments the counter wrongly.

**Proposition 1.** *Player 2 has a finite-memory strategy with  $3 \times 2^{\frac{1}{2} \cdot \binom{n}{n/2}}$  different memory states to win positively  $\text{guess\_my\_set}_n$ .*

*No finite-memory strategy of player 2 with less than  $2^{\frac{1}{2} \cdot \binom{n}{n/2}}$  memory states wins positively  $\text{guess\_my\_set}_n$ .*

*Proof.* The first claim is quite straightforward. Player 2 remembers in which part she is (3 different possibilities). In part 2, player 2 remembers all the sets proposed by player 1 ( $2^{\frac{1}{2} \cdot \binom{n}{n/2}}$  possibilities). Between part 2 and part 3, player 2 inverts her memory to remember the sets player 1 did not propose (still  $2^{\frac{1}{2} \cdot \binom{n}{n/2}}$  possibilities). Then she proposes each of these sets, one by one, in part 3, deleting the set from her memory after she proposed it. Let us assume first that player 1 does not cheat and plays fair. Then all the sets of size  $n/2$  are proposed (since there are  $2 \cdot \frac{1}{2} \cdot \binom{n}{n/2}$  such sets), that is  $X$  has been found and the game starts another round without entering state  $t$ . Else, if player 1 cheats at some point, then the probability to reach the sink state  $s$  is non zero, and player 2 also wins *positively* her safety objective.

The second claim is not hard to show either. The strategy of player 1 is to never cheat, which prevents the game from entering the sink state. In part 2, player 1 proposes the sets  $X$  in a lexicographical way and uniformly at random. Assume by contradiction that player 2 has a counter strategy with strictly less than  $2^{\frac{1}{2} \cdot \binom{n}{n/2}}$  states of memory that wins positively the safety objective. Consider the end of part 2, when player 1 has proposed  $\frac{1}{2} \cdot \binom{n}{n/2}$  sets. If there are less than  $2^{\frac{1}{2} \cdot \binom{n}{n/2}}$  states the memory of player 2 can be in, then

there exists a memory state  $m_*$  of player 2 and at least two sets  $A, B$  among the  $\frac{1}{2} \cdot \binom{n}{n/2}$  sets proposed by player 1 such that the memory of player 2 after  $A$  is  $m_*$  with non zero probability and the memory of player 2 after  $B$  is  $m_*$  with non zero probability. Now,  $\overline{A} \cup \overline{B}$  has strictly more than  $\frac{1}{2} \cdot \binom{n}{n/2}$  sets of  $n/2$  elements. Hence, there is a set  $X \in \overline{A} \cup \overline{B}$  with a positive probability not to be proposed by player 2 after memory state  $m_*$ . Without loss of generality, we can assume that  $X \notin A$  (the other case  $X \notin B$  is symmetrical). Now, for each round of the game, there is a positive probability that  $X$  is the set in the memory of player 1, that player 1 proposed sets  $A$ , in which case player 2 has a (small) probability not to propose  $X$  and then the game goes to  $t$ , where player 1 wins. Player 1 will thus eventually reach the target state with probability 1, hence a contradiction. This achieves the proof that no finite-memory strategy of player 2 with less than  $2^{\frac{1}{2} \cdot \binom{n}{n/2}}$  states of memory is positively winning.  $\square$

## 4 Decidability.

We turn now to the algorithms which compute the set of supports that are almost-surely or positively winning for various objectives.

**Theorem 3** (Deciding positive winning in reachability games). *In a reachability game each initial distribution  $\delta$  is either positively winning for player 1 or surely winning for player 2, and this depends only on  $\text{supp}(\delta) \subseteq K$ . The corresponding partition of  $\mathcal{P}(K)$  is computable in time  $\mathcal{O}(G \cdot 2^K)$ , where  $G$  denotes the size of the description of the game. The algorithm computes at the same time the finite-memory strategies described in Theorem 2.*

As often in algorithmics of game theory, the computation is achieved by a fix-point algorithm.

*Sketch of proof.* The set of supports  $\mathcal{L} \subseteq \mathcal{P}(K)$  surely-winning for player 2 are characterized as the largest fix-point of some monotonic operator  $\Phi : \mathcal{P}(\mathcal{P}(K)) \rightarrow \mathcal{P}(\mathcal{P}(K))$ . The operator  $\Phi$  associates with  $\mathcal{L} \subseteq \mathcal{P}(K)$  the set of supports  $L \in \mathcal{L}$  that do not intersect target states and such that player 2 has an action which ensures that her next belief is in  $\mathcal{L}$  as well, whatever action is chosen by player 1 and whatever signal player 2 receives. For  $\mathcal{L} \subseteq \mathcal{P}(K)$ , the value of  $\Phi(\mathcal{L})$  is computable in time linear in  $\mathcal{L}$  and in the description of the game, yielding the exponential complexity bound.  $\square$

To decide whether player 1 wins almost-surely a Büchi game, we provide an algorithm which runs in doubly-exponential time and uses the algorithm of Theorem 3 as a sub-procedure.

**Theorem 4** (Deciding almost-sure winning in Büchi games). *In a Büchi game each initial distribution  $\delta$  is either almost-surely winning for player 1 or positively winning for player 2, and this depends only on  $\text{supp}(\delta) \subseteq K$ . The corresponding partition of  $\mathcal{P}(K)$  is computable in time  $\mathcal{O}(2^{2^G})$ , where  $G$  denotes the size of the description of the game. The algorithm computes at the same time the finite-memory strategies described in Theorem 2.*

*Sketch of proof.* The proof of Theorem 4 is based on the following ideas.

First, suppose that from every initial support player 1 can win the *reachability objective* with positive probability. Since this positive probability can be bounded from below, repeating the same strategy can ensure that Player 1 wins the Büchi condition with probability 1. According to Theorem 3, in the remaining case there exists a support  $L$  surely winning for player 2 for her co-Büchi objective.

We prove that in case player 2 can *force the belief* of player 1 to be  $L$  someday with positive probability from another support  $L'$ , then  $L'$  is positively winning as well for player 2. This is not completely obvious because in general player 2 cannot know exactly *when* the belief of player 1 is  $L$ . For winning positively from  $L'$ , player 2 plays totally randomly until she guesses randomly that the belief of player 1 is  $L$ , at that moment she switches to a strategy surely winning from  $L$ . Such a strategy is far from being optimal, because player 2 plays randomly and in most cases she makes a wrong guess about the belief of player 1. However player 2 wins positively because there is a chance she is lucky and guesses correctly at the right moment the belief of player 1.

Player 1 should surely avoid her belief to be  $L$  or  $L'$  if she wants to win almost-surely. However, doing so player 1 may prevent the play from reaching target states, which may create another positively winning support for player 2, and so on...

Using these ideas, we prove that the set  $\mathcal{L}_\infty \subseteq \mathcal{P}(K)$  of supports almost-surely winning for player 1 for the Büchi objective is the largest set of initial supports from where

- (†) player 1 has a strategy for winning positively the reachability game while ensuring at the same time her belief to stay in  $\mathcal{L}_\infty$ .

Property (†) can be reformulated as a reachability condition in a new game whose states are states of the original game augmented with beliefs of player 1, kept hidden to player 2.

The fix-point characterization suggests the following algorithm for computing the set of supports positively winning for player 2:  $\mathcal{P}(K) \setminus \mathcal{L}_\infty$  is the limit of the sequence  $\emptyset = \mathcal{L}'_0 \subsetneq \mathcal{L}'_0 \cup \mathcal{L}'_1 \subsetneq \mathcal{L}'_0 \cup \mathcal{L}'_1 \subsetneq \mathcal{L}'_0 \cup \mathcal{L}'_1 \cup \mathcal{L}''_2 \subsetneq \dots \subsetneq \mathcal{L}'_0 \cup \dots \cup \mathcal{L}'_m = \mathcal{P}(K) \setminus \mathcal{L}_\infty$ , where

- (a) from supports in  $\mathcal{L}''_{i+1}$  player 2 can surely guarantee the safety objective, under the hypothesis that player 1 beliefs stay outside  $\mathcal{L}'_i$ ,
- (b) from supports in  $\mathcal{L}'_{i+1}$  player 2 can ensure with positive probability the belief of player 1 to be in  $\mathcal{L}''_{i+1}$  someday, under the same hypothesis.

The overall strategy of player 2 positively winning for the co-Büchi objective consists in playing randomly for some time until she decides to pick up randomly a belief  $L$  of player 1 in some  $\mathcal{L}'_i$ . She forgets the signals she has received up to that moment and switches definitively to a strategy which guarantees (a). With positive probability, player 2 is lucky enough to guess correctly the belief of player 1 at the right moment, and future beliefs of player 1 will stay in  $\mathcal{L}'_i$ , in which case the co-Büchi condition holds and player 2 wins.  $\square$

Property  $\dagger$  can be formulated by mean of a fix-point according to Theorem 3, hence the set of supports positively winning for player 2 can be expressed using two nested fix-points. This should be useful for actually implementing the algorithm and for computing symbolic representations of winning sets.

## 5 Complexity and special cases.

In this section we show that our algorithms are optimal regarding complexity. Furthermore, we show that these algorithms enjoy better complexity in restricted cases, generalizing some known algorithms [17, 6] to more general subcases, while keeping the same complexity.

The special cases that we consider regard inclusion between knowledges of players. To this end, we define the following notion. If at each moment of the game the belief of player  $x$  is included in the one of player  $y$ , then player  $x$  is said to have more information (or to be better informed) than player  $y$ . It is in particular the case when for every transition, the signal of player 1 contains the signal of player 2.

### 5.1 Lower bound.

We prove here that the problem of knowing whether the initial support of a reachability game is almost-surely winning for player 1 is 2EXPTIME-complete. The lower bound even holds when player 1 is more informed than player 2.

**Theorem 5.** *In a reachability game, deciding whether player 1 has an almost-surely winning strategy is 2EXPTIME-hard, even if player 1 is more informed than player 2.*

*Sketch of proof.* We do a reduction from the membership problem for EXPSPACE alternating Turing machines. Let

$\mathcal{M}$  be an EXPSPACE alternating Turing machine, and  $w$  be an input word of length  $n$ . From  $\mathcal{M}$  and  $w$  we build a stochastic game with partial observation such that player 1 can achieve almost-surely a reachability objective if and only if  $w$  is accepted by  $\mathcal{M}$ . The idea of the game is that player 2 describes an execution of  $\mathcal{M}$  on  $w$ , that is, she enumerates the tape contents of successive configurations. Moreover she chooses the rule to apply when the state of  $\mathcal{M}$  is universal, whereas player 1 is responsible for choosing the rule in existential states. When the Turing machine reaches its final state, the play is won by player 1. In this game, if player 2 really implements some execution of  $\mathcal{M}$  on  $w$ , player 1 has a surely winning strategy if and only if  $w$  is accepted by  $\mathcal{M}$ .

This reasoning holds under the assumption that player 2 effectively describes the execution of  $\mathcal{M}$  on  $w$  consistent with the rules chosen by both players. However, player 2 could cheat when enumerating successive configurations of the execution. To prevent player 2 from cheating, it would be convenient for the game to remember the tape contents, and check that in the next configuration, player 2 indeed applied the chosen rule. However, the game can remember only a logarithmic number of bits, while the configurations have a number of bits exponential in  $n$ . Instead, we ask player 1 to pick any position  $k$  of the tape, and to announce it to the game (player 2 does not know  $k$ ), which is described by a linear number of bits. The game keeps the letter at this position together with the previous and next letter on the tape. This allows the game to compute the letter  $a$  at position  $k$  of the *next* configuration. As player 2 describes the next configuration, player 1 will announce to the game that position  $k$  has been reached again. The game will thus check that the letter player 2 gives is indeed  $a$ . This way, the game has a positive probability to detect that player 2 is cheating. If so, the game goes to a sink state which is winning for player 1. To increase the probability for player 1 of observing player 2 cheating, player 1 has the possibility to restart the whole execution from the beginning whenever she wants. If player 2 cheats infinitely often, player 1 will detect it with probability one, and will win the game almost-surely.

We now have to take into account that player 1 could cheat: she could point a certain position of the tape contents at a given step, and point somewhere else in the next step. To avoid this kind of behaviour, a small piece of information about the position pointed by player 1 is kept secret in the state of the game. If player 1 is caught cheating, the game goes to a sink state losing for player 1.

This construction ensures that player 1 has an almost sure winning strategy if and only if  $w$  is accepted by the alternating Turing machine  $\mathcal{M}$ . Note that in the game described above player 1 does not have full information but has more information than player 2.  $\square$

## 5.2 Special cases.

A first straightforward result is that in a safety game where player 1 has full information, deciding whether she has an almost-surely winning strategy is in PTIME.

Now, consider a Büchi game. In general, as shown in the previous section, deciding whether the game is almost-surely winning for player 1 is 2EXPTIME-complete. However, it is already known that when player 2 has a full observation of the game the problem is EXPTIME-complete only [6]. We show that our algorithm keeps the same EXPTIME upper-bound even in the more general case where player 2 is more informed than player 1, as well as in the case where player 1 fully observes the state of the game.

**Proposition 2.** *In a Büchi game where either player 2 has more information than player 1 or player 1 has complete observation, deciding whether player 1 has an almost-surely winning strategy or not (in which case player 2 has a positively winning strategy) can be done in exponential time.*

*Sketch of proof.* In both cases, player 2 needs only exponential memory because if player 2 has more information, there is always a *unique* belief of player 1 compatible with her signals, and in case player 1 has complete observation her belief is always a singleton set.  $\square$

Note that the latter proposition does not hold when player 1 has more information than player 2. Indeed in the game from the proof of Theorem 5, player 1 does have more information than player 2 (but she does not have full information).

## 6 Conclusion.

We considered stochastic games with signals and established two determinacy results. First, a reachability game is either almost-surely winning for player 1, surely winning for player 2 or positively winning for both players. Second, a Büchi game is either almost-surely winning for player 1 or positively winning for player 2. We gave algorithms for deciding in doubly-exponential time which case holds and for computing winning strategies with finite memory.

The question ‘does player 1 have a strategy for winning positively a Büchi game?’ is undecidable [2], even when player 1 is blind and alone. An interesting research direction is to design subclasses of stochastic games with signals for which the problem is decidable.

## References

[1] R. J. Aumann. *Repeated Games with Incomplete Information*. MIT Press, 1995.

- [2] C. Baier, N. Bertrand, and M. Größer. On decision problems for probabilistic Büchi automata. In *Proc. of FOSSACS'08*, vol. 4972 of LNCS, pp. 287–301. Springer, 2008.
- [3] N. Bertrand, B. Genest, and H. Gimbert. Qualitative determinacy and decidability of stochastic games with signals. Technical Report hal-00356566, HAL archives ouvertes, January 2009. Previous version hal-00341288, November 2008.
- [4] D. Berwanger, K. Chatterjee, L. Doyen, T. A. Henzinger, and S. Raje. Strategy construction for parity games with imperfect information. In *Proc. of CONCUR'08*, vol. 5201 of LNCS, pp. 325–339. Springer, 2008.
- [5] K. Chatterjee, L. de Alfaro, and T. A. Henzinger. The complexity of stochastic Rabin and Streett games. In *Proc. of ICALP'05*, vol. 3580 of LNCS, pp. 878–890. Springer, 2005.
- [6] K. Chatterjee, L. Doyen, T. A. Henzinger, and J.-F. Raskin. Algorithms for omega-regular games of incomplete information. *Logical Methods in Computer Science*, 3(3), 2007.
- [7] K. Chatterjee, M. Jurdzinski, and T. A. Henzinger. Quantitative stochastic parity games. In *Proc. of SODA'04*, pp. 121–130. SIAM, 2004.
- [8] A. Condon. The complexity of stochastic games. *Information and Computation*, 96:203–224, 1992.
- [9] L. de Alfaro and T. A. Henzinger. Concurrent omega-regular games. In *Proc. of LICS'00*, pp. 141–154. IEEE, 2000.
- [10] L. de Alfaro, T. A. Henzinger, and O. Kupferman. Concurrent reachability games. *Theoretical Computer Science*, 386(3):188–217, 2007.
- [11] L. de Alfaro and R. Majumdar. Quantitative solution of omega-regular games. In *Proc. of STOC'01*, pp. 675–683. ACM, 2001.
- [12] H. Gimbert and F. Horn. Simple stochastic games with few random vertices are easy to solve. In *Proc. of FOSSACS'08*, vol. 4972 of LNCS, pp. 5–19. Springer, 2008.
- [13] E. Grädel, W. Thomas, and T. Wilke. *Automata, Logics and Infinite Games*, vol. 2500 of LNCS. Springer, 2002.
- [14] F. Horn. *Random Games*. PhD thesis, Université Denis-Diderot, 2008.
- [15] J.-F. Mertens and A. Neyman. Stochastic games have a value. In *Proc. of the National Academy of Sciences USA*, vol. 79, pp. 2145–2146, 1982.
- [16] A. Paz. *Introduction to probabilistic automata*. Academic Press, 1971.
- [17] J. H. Reif. Universal games of incomplete information. In *Proc. of STOC'79*, pp. 288–308. ACM, 1979.
- [18] J. Renault. The value of repeated games with an informed controller. Technical report, CEREMADE, Paris, Jan. 2007.
- [19] J. Renault and S. Sorin. Personal Communications, 2008.
- [20] D. Rosenberg, E. Solan, and N. Vieille. Stochastic games with imperfect monitoring. Technical Report 1376, Northwestern University, July 2003.
- [21] O. Serre and V. Gripon. Qualitative concurrent games with imperfect information. Technical Report hal-00360857, HAL archives ouvertes, February 2009.
- [22] L. S. Shapley. Stochastic games. In *Proc. of the National Academy of Sciences USA*, vol. 39, pp. 1095–1100, 1953.
- [23] S. Sorin. *A first course on zero-sum repeated games*. Springer, 2002.
- [24] J. von Neumann and O. Morgenstern. *Theory of games and economic behavior*. Princeton University Press, 1944.