OPERATIONS RESEARCH
Vol. 57, No. 4, July–August 2009, pp. 905–915
ISSN 0030-364X | EISSN 1526-5463 | 09 | 5704 | 0905

DOI 10.1287/opre.1080.0644
© 2009 INFORMS

Protocols with No Acknowledgment

Dinah Rosenberg
Laboratoire d'Analyse Géométrie et Applications, Institut Galilée, Université Paris Nord, 93430 Villetaneuse, France, and
Laboratoire d'Econométrie de l'Ecole Polytechnique, 75005 Paris, France, dinah@zeus.math.univ-paris13.fr

Eilon Solan
School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel, eilons@post.tau.ac.il

Nicolas Vieille
Département Finance et Economie, HEC, 78351 Jouy-en-Josas, France, vieille@hec.fr

We study a simple protocol for communication networks, in which users get no receipt acknowledgment of their requests. As a result, users hold partial and differential information over the state of the protocol. We characterize optimal behavior by viewing the protocol as a stochastic game with partial observation. We also study two classes of protocols that generalize this protocol.

Subject classifications: games/group decisions: stochastic.
Area of review: Optimization.
History: Received July 2006; revision received November 2006; accepted December 2007. Published online in *Articles in Advance* March 30, 2009.

1. Introduction

In many communication networks, such as the Internet, radio, cellular phone, and satellite networks, the communication medium is shared by multiple users. Often the problem of collision arises: if several users simultaneously attempt to use a channel, no information is transmitted. To minimize access collision, various protocols have been devised, such as the IEEE 802.11 (IEEE Standard 802.11a 1999), Aloha (Abramson 1970), and slotted Aloha (Roberts 1975). Using game-theoretic tools, the efficiency of these protocols has been studied, as well as the optimal strategies of the users (e.g., Altman et al. 2004a, b; Sagduyu and Ephremides 2003).

A key assumption that is made is that users know whether a collision occurs or not, and indeed, e.g., the IEEE 802.11 enables users to check whether the channel is busy. In this paper, we consider a situation in which users do not have this information.

For concreteness, consider the following stylized example. Two processors compete in sending packets over a single channel. The channel can transmit only a single packet at each time slot, and it is governed by a central protocol. The protocol requires sending a request before sending a packet. Thus, at every time slot each processor can either send a request, send a packet, or do nothing. If both processors send a request at the same time slot, a collision, which is not reported to the processors, occurs, and the protocol does not transmit a packet at the following time slot. If only one processor sends a request, and that processor sends a packet at the subsequent time slot, the protocol does transmit this packet. Otherwise, the request is offset, and the processor who made the request is penalized. The protocol then becomes free again, announces this fact to the processors, and waits for another request.[1]

Although the processors know when the protocol becomes free, its state becomes unknown after one time slot: if processor A sent a request, A does not know whether the protocol is waiting for its packet or whether B also sent a request; if A did nothing or sent a packet, the processor may either be free or waiting for B's packet.

As the users compete among themselves, the analysis requires the use of game-theoretic tools. The model that we use is that of recursive games. A *recursive game* is a stochastic game in which the payoff in nonabsorbing states is zero. Stochastic games were used by Sagduyu and Ephremides (2003) and Altman et al. (2004a) to model problems of access control. An overview of stochastic games, as well as some of their applications, can be found in Filar and Vrieze (1996) and Neyman and Sorin (2004).

In this paper, we analyze in detail the stylized protocol described above. We prove that the processors have a unique optimal strategy. An interesting consequence of our analysis is that this optimal strategy can be implemented by an automaton with three states.

We then generalize the example, and study two classes of recursive games that correspond to more complex protocols. In general recursive games with partial information the value need not exist; this happens when one player, by acting after his opponent, can guarantee more than he can when he acts first. We prove that in the two classes we study optimal strategies do exist, and we study the structure of the optimal strategies.

Our goal in this paper is not to develop a general theory of games with partial information. Rather, it is to show that situations with partial information, like ad hoc networks, can be modelled as recursive games and successfully analyzed using game-theoretic tools.

The relevant literature on stochastic games with partial information on the state is scarce. In search games, a searcher wishes to locate a target in minimal time, whereas the target tries to escape from the searcher; see Alpern and Gal (2002) and Gal and Howard (2005) for recent contributions that combine search and rendezvous aspects. In inspection games, an inspector verifies that an inspectee adheres to certain legal rules, whereas the inspectee has an interest in violating those rules; see Avenhaus et al. (2002) for a survey. The basic difficulty in such games is that the conditional distribution of the current state cannot serve as a state variable, as is the case for Markov decision processes with partial observation, see Arapostathis et al. (1993) or Monahan (1982). Indeed, in a game with partial information, the beliefs of the two players need not be commonly known. Moreover, the computation of this conditional distribution may simply be impossible without knowing the actual strategy of the other player. As a consequence, no dynamic programming principle holds.

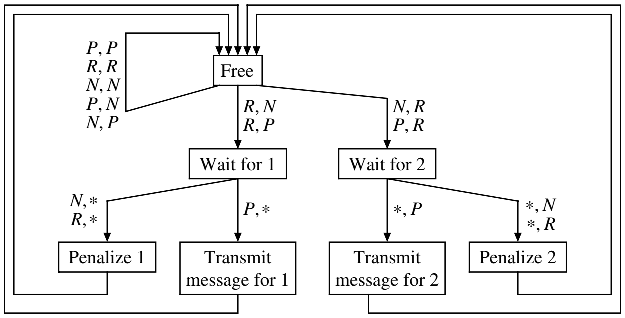
This paper is organized as follows. We analyze the protocol described above in §2. Section 3 lays out a general model of recursive games with no observation. It also contains the value existence results for two classes of protocols that generalize the protocol studied in §2, and includes a discussion of the structure of optimal strategies. Proofs appear in §4.

2. Analysis of the Protocol

We here analyze the protocol described in the introduction. We start by formally defining the game that corresponds to the protocol.

At every period each processor has three available actions—send a request, send a packet, and do nothing. For short, we denote these actions by R, P, and N, respectively. The protocol is shown in Figure 1.

Figure 1. The protocol.



The protocol remains free until exactly one of the processors sends a request. It then waits for a packet from the requesting processor. If a packet arrives, it is transmitted. Otherwise, the requesting processor must pay a penalty to the other processor for misusing the network. This penalty can be either monetary or nonmonetary, through, e.g., some priority given to the other processor in subsequent rounds. To analyze the situation as a game we need to attach a utility to each outcome. For simplicity, we assume that both the gain from sending a packet, and the loss due to the penalty, equal one.

We focus on competitive situations where each processor aims both at maximizing its long-run average payoff, and at minimizing that of the other processor. One simple modelling solution is to assume that a processor incurs a loss of one whenever the other processor successfully sends a packet.

Observe that the action N is dominated by P: when the protocol is free or busy waiting for the other processor, both actions have the same consequence; otherwise, it is preferable to use P than N. We can thus postulate that at every time slot the processors use one of the two actions P or R.

The situation can be modelled as a stochastic game as shown in Figure 2 (processor 1 is the row player and processor 2 is the column player).

The state s_0 corresponds to the protocol being free. For $i = 1, 2$, the state s_i corresponds to the protocol waiting for processor i. In each state, each processor has two actions, R and P. Transitions from s_0 are as depicted in Figure 2, whereas at state s_1 or s_2, after the processors choose their actions, a payoff is realized and the protocol moves back to s_0. Because the processors are informed when the protocol becomes free, the game effectively starts anew. We therefore study a single round of the game—until the first time the protocol restarts.

Figure 2. The corresponding stochastic game.

s_2

	R	P
R	1^{s_0}	-1^{s_0}
P	1^{s_0}	-1^{s_0}

s_0

	R	P
R	s_0	s_1
P	s_2	s_0

s_1

	R	P
R	-1^{s_0}	-1^{s_0}
P	1^{s_0}	1^{s_0}

A strategy of player 1 is a sequence $\sigma = (\sigma_n)_{n\geqslant 1}$, where $\sigma_n\colon \{R, P\}^{n-1} \to [0, 1]$, with the interpretation that $\sigma_n(a_1, \ldots, a_{n-1})$ is the probability assigned to the action R in stage n, after playing $a_1, \ldots, a_{n-1}$ in the first $n-1$ stages. We will sometimes drop the subscript n from σ_n and simply write $\sigma(\vec{a})$.

Because the game is symmetric, the value must be zero if it exists, and the optimal strategies of both players are identical.

Proposition 2.1. *The game has a value. The strategy* $\sigma^\star$ *that is defined by* $\sigma^\star_1 = 2/3$ *and, for* $n > 1$,

$$\sigma^\star_n(a_1, \ldots, a_{n-1}) = \begin{cases} \frac{2}{3} & \textit{if } a_{n-1} = P, \\ \frac{1}{2} & \textit{if } a_{n-2} = P \textit{ and } a_{n-1} = R, \\ 0 & \textit{if } a_{n-2} = a_{n-1} = R, \end{cases}$$

is the unique optimal strategy (*modulo events that occur with probability* 0).

The strategy $\sigma^\star$ can be implemented by the automaton in Figure 3.

Casual intuition suggests that an optimal strategy may exist that would depend only on the *last* move. Indeed, after processor 1 plays R, his state of ignorance is the same, regardless of his earlier moves: Either the protocol reinitializes itself and both processors know that the state is s_0, or the game is currently in state s_0 or s_1. For the same reason, after processor 1 plays P he should reason that either s_0 or s_2 is possible. As the theorem asserts, no such optimal strategy exists.

However, the two situations (the state of the game after player 1 plays R or P) differ in an important respect. In state s_2, player 1's decision is irrelevant. Therefore, after player 1 plays P, he may safely assume that the current state is s_0. This suggests that it might be optimal for player 1 to start the strategy anew, after he has played P. The strategy $\sigma^\star$ has exactly this property. By contrast, in state s_1, player 1's decision is payoff relevant. In loose terms, after player 1 plays R, it is important to assess the likelihood of both states s_0 and s_1; the whole sequence of past actions provides some information in this respect.

Proof. *Step* 1. The strategy $\sigma^\star$ guarantees zero.

We will prove that $\sigma^\star$ guarantees zero even if player 2 is told at any stage whether the entry (P, P) is played. A fortiori, this implies that $\sigma^\star$ guarantees zero in our game.

Figure 3. The strategy $\sigma^\star$.

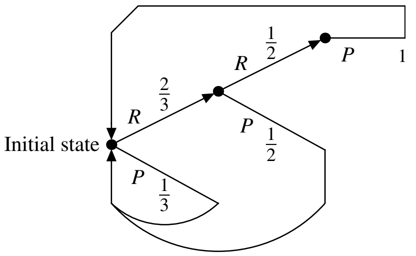


Figure 4. The auxiliary game.

s_2

	R	P
R	1^*	-1^*
P	1^*	-1^*

s_0

	R	P
R	s_0	s_1
P	s_2	0^*

s_1

	R	P
R	-1^*	-1^*
P	1^*	1^*

Whenever player 1 plays P, the automaton moves to its initial state, and its behavior restarts. Whenever player 2 is told that (P, P) has just been played, either the current round is over, or the protocol is in s_0, and player 1 starts anew.

Hence, we need only prove that for every strategy τ of player 2, the expected payoff $\tilde{\gamma}(\sigma^\star, \tau)$ until the action (P, P) is played, or until the protocol restarts, is nonnegative. Thus, to study optimal behavior we can study the auxiliary game in Figure 4.

An asterisked entry means a transition to an absorbing state with the corresponding payoff. It follows that whenever in the original game depicted in Figure 2 player 2 knows the state of the strategy of player 1 and the state of the protocol, a round ends, and this is captured by an absorbing state.

We will prove that in the auxiliary game the strategy $\sigma^\star$ guarantees zero. By Kuhn's theorem (Kuhn 1953), it is sufficient to show that $\sigma^\star$ guarantees zero against any *pure* strategy τ of player 2, that is, a deterministic sequence of actions. Observe that player 2's actions after he played P for the first time are not relevant: either player 1 played P, and the pair of actions played is (P, P), or player 1 plays R and the game moves to s_1. (The actions of player 2 at s_1 are irrelevant.)

Thus, we will consider sequences τ of the form $R^k P$, $k \in \mathbf{N} \cup \{0, \infty\}$, that play R for a certain number of stages and then P.

For $k = 0$, one has

$$\tilde{\gamma}(\sigma^\star, P) = \tfrac{2}{3}\big(\tfrac{1}{2} \times 1 + \tfrac{1}{2} \times (-1)\big) = 0. \tag{1}$$

Indeed, with probability $2/3$, player 1 plays R at the first stage, and then at the second stage he plays both actions with equal probability, and with probability $1/3$ he plays P at the first stage and the game restarts.

Similarly, for $k=1$,

$$\tilde{\gamma}(\sigma^\star, RP) = -\tfrac{1}{3} + \tfrac{2}{3} \times \tfrac{1}{2} = 0. \tag{2}$$

Indeed, with probability 2/3 player 1 plays R at the first stage and the game remains at s_0. Then, at the second stage he plays both actions with equal probability, so with probability 1/2 the game moves to s_1. At the third stage he plays P and receives 1. With probability 1/3 player 1 plays P at the first stage, the game moves to s_2 and the payoff is -1.

In the same way, we obtain

$$\tilde{\gamma}(\sigma^\star, RRP) = \tfrac{1}{3} - \tfrac{2}{3} \times \tfrac{1}{2} = 0 \tag{3}$$

and

$$\tilde{\gamma}(\sigma^\star, RRRP) = \tfrac{1}{3} + \tfrac{2}{3} \times \tfrac{1}{2} - \tfrac{2}{3} \times \tfrac{1}{2} = \tfrac{1}{3}. \tag{4}$$

If $k \geqslant 4$, then $\tilde{\gamma}(\sigma^\star, R^k P) = 1/3 + 2/3 \times 1/2 + 2/3 \times 1/2 = 1$ because the probability that player 1 plays P at least once in the first three stages is 1.

Step 2. $\sigma^\star$ is the unique optimal strategy that restarts whenever P is played.

Let σ be an optimal strategy of player 1 that restarts whenever P is played. Then, σ is described by a sequence $(\pi_k)_{k\geqslant 0}$, where π_k is the probability that R is played *at least* k times between two consecutive Ps. Note that $\pi_0 = 1$. With the above notations, any such optimal strategy must satisfy $\tilde{\gamma}(\sigma, R^k P) \geqslant 0$ for $k = 0, 1, 2$. After some algebraic manipulations, we obtain that these conditions amount to

$$2\pi_2 - \pi_1 \leqslant 0, \tag{5}$$

$$2\pi_3 - \pi_2 - \pi_1 + 1 \leqslant 0, \tag{6}$$

$$2\pi_4 - \pi_3 - \pi_2 + 2\pi_1 - 1 \leqslant 0. \tag{7}$$

These inequalities imply that

$$\pi_2 \leqslant \pi_{1/2}, \tag{8}$$

$$2\pi_3 \leqslant \pi_2 + \pi_1 - 1, \tag{9}$$

$$2\pi_4 \leqslant \pi_3 + \pi_2 - 2\pi_1 + 1 \leqslant \tfrac{1}{2} - \tfrac{3}{4}\pi_1, \tag{10}$$

where one uses (8)–(9) to derive (10). Because $\pi_4 \geqslant 0$, (10) implies that $\pi_1 \leqslant 2/3$. Because $\pi_3 \geqslant 0$, Equations (8) and (9) imply that $\pi_1 \geqslant 2/3$. This implies $\pi_1 = 2/3$, $\pi_2 = 1/3$, and $\pi_3 = 0$. Let x_n be the probability of playing R after n times R. We have

$$1 - \pi_1 = 1 - x_0 = \tfrac{1}{3}, \tag{11}$$

$$\pi_1 - \pi_2 = x_0(1 - x_1) = \tfrac{1}{3}, \tag{12}$$

$$\pi_2 - \pi_3 = x_0 x_1 (1 - x_2) = \tfrac{1}{3}, \tag{13}$$

so that $x_0 = 2/3$, $x_1 = 1/2$, $x_2 = 0$, and $\sigma = \sigma^\star$.

Step 3. $\sigma^\star$ is the unique optimal strategy.

Let $\sigma \neq \sigma^\star$ be arbitrary. Let t_k be the stage of the kth time in which player 1 plays P (by convention $t_0 = 0$). Because $\sigma \neq \sigma^\star$, there is k such that with positive probability σ differs from $\sigma^\star$ between stages $t_k + 1$ and t_{k+1}. By Step 2, conditional on the game being at s_0 at stage t_k, there is a pure reply of player 2, $b_1, b_2, \ldots,$ that ensures the expected absorbing payoff between stages $t_k + 1$ and t_{k+1} is negative.

Denote by $\tau^\star$ the strategy of player 2 that is identical to $\sigma^\star$. By the symmetry of the game, $\tau^\star$ guarantees that the expected payoff is nonpositive. Because $\sigma = \sigma^\star$ until stage t_k, it follows that under $(\sigma, \tau^\star)$, with positive probability the game is not absorbed before stage t_k. One can verify that the following strategy for player 2 guarantees that the expected payoff is negative, so in particular σ is not optimal: player 2 follows $\tau^\star$ until the kth time he plays P, then he follows the sequence $b_1, b_2, \ldots$ until the next time the action pair (P, P) is played, and from then he resumes following $\tau^\star$. □

3. The Model and Results

We here present the game-theoretic model that we use—recursive games in which the players observe nothing but their own actions. We then show by means of an example that without qualifications the value need not exist. We then provide two classes of games that generalize the protocol studied above, and prove that such games have a value. Finally, we discuss the structure of the optimal strategies.

3.1. The General Model

For any finite set X, we denote by $\Delta(X)$ the set of probability distributions over X.

A *recursive game* is defined by (i) a finite state space S that is partitioned into two subsets S^0 and S^*, and an initial state $s_1 \in S^0$, (ii) finite actions sets A and B, (iii) a payoff function $r\colon S^* \to \mathbf{R}$, and (iv) a transition rule $q(\cdot \mid s, a, b) \in \Delta(S)$ for each $s \in S^0$, $a \in A$, $b \in B$.

The game is played as follows. Denote by $s_n \in S$ the state of the game at stage $n \in \mathbf{N}$. At each stage n the two players choose actions a_n and b_n in A and B, respectively. If $s_n \in S^0$, then the next state s_{n+1} is drawn according to $q(\cdot \mid s_n, a_n, b_n)$; if $s_n \in S^*$, then $s_{n+1} = s_n$. States in S^* are called *absorbing* because once the game reaches such a state, it never leaves it.

We let $\theta := \inf\{n \geqslant 1\colon s_n \in S^*\}$ be the stage at which the game is absorbed (with $\inf \varnothing = +\infty$). The payoff from player 2 to player 1 is $r(s_\theta)1_{\theta<+\infty}$. Note that this payoff is zero if $\theta = \infty$.

A (behavior) strategy σ of player 1 is a sequence $(\sigma_n)_{n\in\mathbf{N}}$, where σ_n describes player 1's behavior at stage n. We study games with no observations. In these games each player observes only his own past actions, and no player observes the actions of his opponent, nor the state of the game. Accordingly, σ_n is a map from A^{n-1} to $\Delta(A)$. Strategies τ of player 2 are defined analogously. We denote by Σ (respectively, T) the space of strategies of player 1 (respectively, player 2).

Together with the initial state, a pair (σ, τ) of strategies induces a probability distribution $\mathbf{P}_{\sigma,\tau}$ over the set $H_\infty = (S \times A \times B)^{\mathbf{N}}$, endowed with the product σ-field (generated by cylinder sets). Expectation w.r.t. $\mathbf{P}_{\sigma,\tau}$ is denoted by $\mathbf{E}_{\sigma,\tau}$. The payoff induced by the strategy pair (σ, τ) is simply $\gamma(\sigma, \tau) := \mathbf{E}_{\sigma,\tau}[r(s_\theta)1_{\theta<+\infty}]$.

The game has a *value* v if

$$v = \sup_{\sigma\in\Sigma} \inf_{\tau\in T} \gamma(\sigma, \tau) = \inf_{\tau\in T} \sup_{\sigma\in\Sigma} \gamma(\sigma, \tau).$$

As the following example shows, in general the value does not exist.

3.2. The Value Need Not Exist

Consider the game with three nonabsorbing states in Figure 5.

Here, if the entry (B, L) is chosen, with probability 1/2 the game is absorbed, and the absorbing payoff is 1, and with probability 1/2 it moves to s_1.

We argue that this game has no value. A pure strategy of player 1 reduces to a choice θ_1 of when to play B for the first time because once he plays B, the game is either absorbed or moves to s_1, where player 1's actions are irrelevant. Similarly, a pure strategy of player 2 reduces to a choice θ_2 of when to play R for the first time. This choice is either an integer or $+\infty$.

Thus, (i) if the choices of the two players match, the payoff is zero; (ii) if the choices of the two players are finite and different, the player who chooses the larger number gains 1/2; (iii) if the choice of exactly one player is finite, that player gains 1/2.

Figure 5. The game with three nonabsorbing states.

s_1

	L	R
P	s_1	-2^*
B	s_1	-2^*

s_0

	L	R
P	s_0	s_2, 1/2 -1^*, 1/2
B	s_1, 1/2 1^*, 1/2	0^*

s_2

	L	R
P	s_2	s_2
B	2^*	2^*

We argue that this game has no value. To this end, we show that for every ε and every strategy τ of player 2, player 1 has a strategy σ such that $\gamma(\sigma, \tau) \geqslant 1/2 - \varepsilon$. This will imply $\inf_\tau \sup_\sigma \gamma(\sigma, \tau) \geqslant 1/2$ and, by symmetry, $\sup_\sigma \inf_\tau \gamma(\sigma, \tau) \leqslant -1/2$, so that the value does not exist.

Fix a strategy τ of player 2, and let q be the probability that player 2 plays R at least once. If $q = 1$, then player 1 can obtain a payoff arbitrarily close to 1/2 by playing P for many stages, then B forever. If instead $q < 1$, we may choose a stage N such that $\mathbf{P}_\tau(\theta_2 \geqslant N \mid \theta_2 < \infty) \leqslant \varepsilon$. Let σ be the strategy of player 1 that plays P up to stage N and B afterwards. The payoff is at least $(1/2)q + 1(1-q) - \varepsilon > 1/2 - \varepsilon$.

In this game transitions are random. By duplicating each action, it is straightforward to obtain a game with deterministic transitions and no value.

3.3. Two Classes of Games

The depth of the simple protocol we have studied in §2 is 3—it remains in its initial state as long as the pair of actions chosen by the two processors is not one of a specific set of "desirable" action pairs. Once a desirable action pair was chosen, the protocol switches to a new state, and then it initializes itself. We call protocols with this structure "one-step protocols," and study the corresponding games in §4.1.

Another way to look at the simple protocol is as follows. As long as the processors choose the *same* action, the state of the protocol does not change. Once they choose different actions the protocol changes its state, and after one step it initializes itself. We define a class of "matching protocols," that have a similar structure, as follows. The set of possible states is divided into two subsets. The initial state is in the first subset, and as long as the processors choose the same action, the state of the protocol remains at the first subset. Once the players choose different actions, the new state of the protocol is in the second subset, and after one additional time slot the protocol initializes itself. We study the corresponding game in §4.2.

A state $s \in S^0$ is called *penultimate* if the subsequent state is absorbing, whatever the players play: $q(S^* \mid s, a, b) = 1$ for every $a \in A$ and $b \in B$. Denote by S_P the set of penultimate states. If the initial state is penultimate then the game is equivalent to a one-shot game, and in particular the value exists. A state is *standard* if it is neither absorbing nor penultimate. In Figure 4, states s_1 and s_2 are penultimate, and state s_0 is standard.

Definition 3.1. A recursive game is a *one-step game* if there is exactly one standard state, which is the initial state.

In a one-step game, once play leaves the initial state, the players make one last choice of action (in case the game reaches a penultimate state), and then the game is absorbed. However, the players may not know when the play actually leaves the initial state, nor to which penultimate state it moved.

Theorem 3.2. *One-step games have a value.*

Definition 3.3. A recursive game is called a *matching game* if (i) the action sets of the two players coincide, and (ii) for every standard state, all off-diagonal entries lead with probability 1 to penultimate states, or absorbing states. Formally, the game is a matching game if (i) $A = B$ and (ii) $q(S^* \cup S_P \mid s, a, a') = 1$ for every standard state s and every $a, a' \in A$ such that $a \neq a'$.

The game in Figure 4 is a matching game. In matching games each player knows that if it so happens that the game is still in a standard state, the past actions of the opponent *must have matched* his own past actions. In such a case, at every stage each player can calculate the probability of being in a given standard state $s \in S_P$, *conditional* on being in S_P.[2] If the game is indeed in a standard state, then so far the actions of the two players matched, and in particular the players calculate the same conditional probability.

This is the key observation that ensures that the value exists in matching games.

Theorem 3.4. *Matching games have a value.*

One property that is common to both one-step games and matching games is the following. Let p_n^i be the conditional distribution over standard states given player i's past actions and given the game is still in a standard state. In general, if the actions of the opponent are not known, p_n^i is not well defined. However, both in one-step games and in matching games this quantity is well defined. We conjecture that in every recursive game in which p_n^i is well defined for every $n \in \mathbf{N}$ and $i = 1, 2$, the value exists.

Comment 3.5. Our results go through if we replace "penultimate states" with "k-penultimate states"; for $k \in \mathbf{N}$, a state s is called k-penultimate if once s is visited, in at most k stages the game reaches an absorbing state, whatever the players play. The proofs are similar to the ones we provide here.

Comment 3.6. In some applications the situation is not completely competitive, and it would be interesting to study nonzero-sum recursive games with no observations. We leave this issue for future research.

3.4. Optimal Strategies

In this section, we discuss the structure of optimal strategies. We show that in general simple optimal strategies need not exist, and we point out why this happens. The examples we study exhibit the importance of beliefs on beliefs, as is customary in games with differential information.

In the example of §2, the optimal strategy restarts whenever P is played. There, one of the distinguishing features of the action P at s_0 is that, once played, player 1 can deduce from the structure of the game that either (i) the game is already over, or (ii) it is in state s_0, or (iii) it is in state s_2, where player 1's move is irrelevant. We say that such an action is *conservative*. Formally, an action a of player i is conservative if there is a standard state s such that after playing a, either the game is in s, or the actions of player i no longer affect the game.

Figure 6. The game in Example 2.

s_0

	L	R
P	2^*, 1/4 s_0, 1/4 s_1, 1/2	3^*, 1/4 s_0, 1/4 s_1, 1/2
B	1^*, 1/4 s_0, 1/4 s_1, 1/2	4^*, 1/4 s_0, 1/4 s_1, 1/2

s_1

	L	R
P	5^*	1^*
B	5^*	1^*

A strategy of player i has the *renewal property* if the way it plays after each time a conservative action is played depends only on the identity of that action, and not on the play in previous stages.

We here analyze the extent to which optimal strategies with such renewal properties do exist. Our results are quite negative.

Example 2. We consider the following one-step game, with two actions for both players. At the initial state s_0, the game moves with probability 1/2 to state s_1, and remains in s_0 with probability 1/4, regardless of the actions chosen. With probability 1/4, the game reaches some absorbing state, with payoff as indicated in Figure 6.

At state s_1, the action of player 1 is payoff irrelevant. Hence, both actions, P and B, are conservative. In the light of Example 1, the question arises whether player 1 has an optimal—or ε-optimal—strategy that restarts after either P or B is played. In other words, does there exist a time-independent optimal strategy? The answer is negative.

The probability of being at s_0 in stage n is $(1/4)^{n-1}$, whereas the probability of being at s_1 is $(1/2)(1/(4^{n-2}))$ for $n \geqslant 2$, and 0 for $n = 1$. In other words, for every $n \geqslant 2$, the conditional probability of being at s_0 in stage n, provided the game has not been absorbed yet, is 2/3.

Hence, effectively, at stage 1 the players play the matrix game

$$\begin{array}{c|c|c|} & L & R \\ \hline P & 2 & 3 \\ \hline B & 1 & 4 \\ \hline \end{array}$$

and at all subsequent stages they play the matrix game

$$\begin{array}{c|c|c|} & L & R \\ \hline P & 9/3 & 7/3 \\ \hline B & 7/3 & 9/3 \\ \hline \end{array}.$$

Figure 7. The game in Example 3.

s_0

	L	R
P	s_0	s_1
B	s_0	-1^*

s_1

	L	R
P	-2^*	-2^*
B	1^*	1^*

In particular, the unique optimal strategies of both players play a different action at stage 1 and at all subsequent stages.

The reason for the failure can be traced back. Although player 1 can always safely assume that the current state is the initial one, this does not end the story. Indeed, player 1 should take into account that player 2's state of uncertainty evolves through time, and player 1 may wish to exploit this fact.

Example 3. We now provide a more disturbing example in Figure 7.

In this game, as long as player 2 chooses the left column, the game remains in state s_0. As soon as player 2 plays R, the game either moves to s_1, or to an absorbing state with payoff -1. In state s_1, player 2's decision is payoff irrelevant.

Consider the action B at state s_0. After playing B, player 1 knows that either the game is by now over, or that it is still in the initial state. Thus, B is a conservative action. Moreover, he knows that if the game is not over, player 2 *also* knows that s_0 is the current state. However, as we show below, there is no optimal strategy of player 1 that restarts after B.

Lemma 3.7. *The value of the game is zero.*

Proof. We exhibit an optimal strategy for both players. Plainly, the strategy of player 2 that plays L at every stage yields a payoff zero. On the other hand, assume that at the outset of the game player 1 flips a fair coin and then follows one of the sequences $PBPB\cdots$ or $BPBP\cdots$ of actions, depending on the outcome.[3] Against such a strategy, assume that player 2 plays R for the first time at stage n. With probability 1/2, player 1 plays B at that stage, and the final payoff is -1; with probability 1/2, he plays P at stage n, and then B at stage $n+1$, with a final payoff of $+1$. Thus, the expected payoff is zero, regardless of the strategy of player 2. □

Denote by σ^* the above strategy of player 1. One can actually prove the following.

Lemma 3.8. *The strategy σ^* is the unique optimal strategy of player* 1.

The proof is relegated to §4.

When one relaxes the optimality condition and requires only approximate (ε-) optimality, it is sometimes the case that this allows for simpler strategies, see, e.g., Flesch et al. (1998). Because in our model all that a player observes is his own past actions, a class of simple strategies is the class of move-independent strategies.

Definition 3.9. A strategy $\sigma=(\sigma_n)$ is *move independent* if σ_n is a constant function for every $n\in\mathbf{N}$.

It is not difficult to show that in the game of Figure 4 there is no move-independent ε-optimal strategy for $\varepsilon>0$ sufficiently small. The calculations are not enlightening, and therefore omitted.

4. Proofs

4.1. Proof of Theorem 3.2

Recall that the inequality

$$\sup_{\sigma\in\Sigma}\inf_{\tau\in T}\gamma(\sigma,\tau)\leqslant\inf_{\tau\in T}\sup_{\sigma\in\Sigma}\gamma(\sigma,\tau) \tag{14}$$

always holds.

If $\sup_{\sigma\in\Sigma}\inf_{\tau\in T}\gamma(\sigma,\tau)=\inf_{\tau\in T}\sup_{\sigma\in\Sigma}\gamma(\sigma,\tau)=0$, the result holds and the value is zero. Thus, by (14) we may assume that either

$$\sup_{\sigma\in\Sigma}\inf_{\tau\in T}\gamma(\sigma,\tau)<0 \quad\text{or}\quad \inf_{\tau\in T}\sup_{\sigma\in\Sigma}\gamma(\sigma,\tau)>0.$$

W.l.o.g. we assume the former, and we multiply all payoffs so that $v=\sup_{\sigma\in\Sigma}\inf_{\tau\in T}\gamma(\sigma,\tau)=-1$. We denote by $M\geqslant 1$ a uniform bound on the payoffs in the game.

Because γ is not a continuous function over the product set of strategy profiles, we cannot apply a standard minmax theorem to prove the result. Instead, we will consider a restricted strategy set for player 2, and apply a minmax theorem over the corresponding constrained game.

For every $\varepsilon>0$, denote by T_ε the set of strategies τ such that after any sequence of actions the probability to play each action is at least ε. Formally, denote

$$\Delta_\varepsilon(B)=\{\beta\in\Delta(B)\colon\, \beta_b\geqslant\varepsilon,\ \forall b\in B\}.$$

Then, $\tau\in T_\varepsilon$ if and only if $\tau_n(\vec{b})\in\Delta_\varepsilon(B)$ for every finite sequence of actions $\vec{b}$. We will prove two lemmas.

Lemma 4.1.

$$\sup_{\sigma\in\Sigma}\inf_{\tau\in T_\varepsilon}\gamma(\sigma,\tau)=\inf_{\tau\in T_\varepsilon}\sup_{\sigma\in\Sigma}\gamma(\sigma,\tau).$$

Denote $v_\varepsilon=\sup_{\sigma\in\Sigma}\inf_{\tau\in T_\varepsilon}$. Because $T_{\varepsilon_1}\subseteq T_{\varepsilon_2}$ whenever $\varepsilon_1\geqslant\varepsilon_2$, the function $\varepsilon\mapsto v_\varepsilon$ is monotonic nondecreasing, and therefore the limit $\lim_{\varepsilon\to 0}v_\varepsilon$ exists.

Lemma 4.2.

$$\lim_{\varepsilon\to 0}v_\varepsilon\leqslant v=-1.$$

Before proving the two lemmas, let us see why they imply Theorem 3.2. By the definition of v_ε, Lemma 4.2, Equation (14), because $T_\varepsilon \subset T$, and by Lemma 4.1,

$$\begin{aligned}\lim_{\varepsilon\to 0}\sup_{\sigma\in\Sigma}\inf_{\tau\in T_\varepsilon}\gamma(\sigma,\tau) &= \lim_{\varepsilon\to 0} v_\varepsilon \leqslant -1\\ &= \sup_{\sigma\in\Sigma}\inf_{\tau\in T}\gamma(\sigma,\tau)\\ &\leqslant \inf_{\tau\in T}\sup_{\sigma\in\Sigma}\gamma(\sigma,\tau)\\ &\leqslant \lim_{\varepsilon\to 0}\inf_{\tau\in T_\varepsilon}\sup_{\sigma\in\Sigma}\gamma(\sigma,\tau)\\ &= \lim_{\varepsilon\to 0}\sup_{\sigma\in\Sigma}\inf_{\tau\in T_\varepsilon}\gamma(\sigma,\tau).\end{aligned}$$

Therefore, $\sup_{\sigma\in\Sigma}\inf_{\tau\in T}\gamma(\sigma,\tau) = \inf_{\tau\in T}\sup_{\sigma\in\Sigma}\gamma(\sigma,\tau)$, and the value exists.

Proof of Lemma 4.1. The set Σ of strategies is $\prod_{n\geqslant 1}\Delta(A)^{A^{N-1}}$, which is convex. When endowed with the product topology, it is a compact metric space. Similarly, T_ε is convex and compact. Moreover, the payoff function is bilinear over the set of mixed strategies, which, by Kuhn's Theorem (Kuhn 1953), are equivalent to behavior strategies. We now argue that the payoff function $\gamma(\sigma,\tau)$ is continuous over $\Sigma\times T_\varepsilon$.

To see this, observe that the assumptions imply that the per-stage probability of absorption is strictly positive. That is, there is $\rho>0$ such that for every $\alpha\in\Delta(A)$ and $\beta\in\Delta_\varepsilon(B)$, we have $q(s_1\mid s_1,\alpha,\beta)<1-\rho$. Indeed, because the function $q(s_1\mid s_1,\cdot,\cdot)$ is continuous over the compact set $\Delta(A)\times\Delta_\varepsilon(B)$, if this is not the case, then there are $\alpha\in\Delta(A)$ and $\beta\in\Delta_\varepsilon(B)$ satisfying $q(s_1\mid s_1,\alpha,\beta)=1$. Because β gives positive weight to each action, this implies that $q(s_1\mid s_1,\alpha,\beta')=1$ for every $\beta'\in\Delta_\varepsilon(B)$. However, this implies that by playing the mixed action α player 1 guarantees a payoff zero, which contradicts the assumption.

By Fan's (1953) fixed-point theorem, the game has a value. □

Proof of Lemma 4.2. *Step* 1. Structure of the proof.

Fix $\delta\in(0,1)$. We first prove in Step 2 that for every σ there is $\varepsilon_\sigma>0$ and a strategy $\tau\in T_{\varepsilon_\sigma}$ of player 1 such that $\gamma(\sigma,\tau)\leqslant v+\delta$. We then use in Step 3 a compactness argument to show that ε_σ is bounded away from zero. This implies that there is $\varepsilon>0$ such that for every $\sigma\in S$,

$$\inf_{\tau\in T_\varepsilon}\gamma(\sigma,\tau)\leqslant v+\delta,$$

so that $v_\varepsilon\leqslant v+\delta$. Because δ is arbitrary, the result follows.

Step 2. Let σ be an arbitrary strategy of player 1, set $\delta_1=\min\{\delta/12M, 1/2\}$, and let $\tau_0\in T$ be a strategy such that

$$\gamma(\sigma,\tau_0)\leqslant\inf_{\tau\in\mathcal{T}}\gamma(\sigma,\tau)+\delta_1.$$

We first prove that $\pi:=\mathbf{P}_{\sigma,\tau_0}(\theta=+\infty)<3\delta_1$. Set $\eta=\delta_1/(2(M+1))$. Let $N_1\in\mathbf{N}$ be sufficiently large such that $\mathbf{P}_{\sigma,\tau_0}(N_1\leqslant\theta<+\infty)\leqslant\eta$. In particular, under (σ,τ_0), the probability is at most η that the game is in some penultimate state at stage N_1+1.

Let τ_1 be the strategy that (i) follows τ_0 up to stage N_1+1, and next (ii) plays a δ_1-best reply against the strategy induced by σ in the continuation game, given that absorption has not occurred prior to stage N_1+1.[4]

By the choice of τ_1, one has

$$\begin{aligned}&\gamma(\sigma,\tau_0)+\eta M+(1-\pi-\eta)(v+\delta_1)\\ &\qquad\geqslant\gamma(\sigma,\tau_1)\geqslant\gamma(\sigma,\tau_0)-\delta_1.\end{aligned}$$

Therefore, $(1-\pi)(-1+\delta_1)\geqslant-\delta_1-\eta M+\eta(-1+\delta_1)$, so that

$$1-\pi\leqslant\frac{2\delta_1}{1-\delta_1}\leqslant 3\delta_1,$$

as desired.

We now construct the strategy τ as a perturbed version of τ_0. Let $N_\star$ be sufficiently large such that $\mathbf{P}_{\sigma,\tau_0}(\theta\geqslant N_\star)<3\delta_1$. Set $\eta_1=\delta/6MN_*$. Let τ be the strategy that at every stage follows τ_0 with probability $1-\eta_1$ and with probability η_1 plays a random action that is chosen uniformly from B.

Then,

$$\gamma(\sigma,\tau)\leqslant\gamma(\sigma,\tau_0)+2\eta_1MN_\star+2M\cdot 3\delta_1\leqslant v+\tfrac{2}{3}\delta.\quad(15)$$

Observe that $\tau\in T_{\varepsilon_\sigma}$, with $\varepsilon_\sigma=\eta_1/|B|$.

Step 3. We now prove that ε_σ can be uniformly bounded away from zero. We argue by contradiction, and assume that there is a sequence (ε_n,σ^n) in $(0,1)\times\Sigma$, with $\lim_{n\to\infty}\varepsilon_n=0$ and

$$\inf_{\tau\in\mathcal{T}_{\varepsilon_n}}\gamma(\sigma^n,\tau)\geqslant v+\delta\quad\text{for each } n\in\mathbf{N}.\quad(16)$$

Because Σ is compact, up to a subsequence, we may thus assume that the sequence (σ^n) converges to some strategy σ. By Step 1 (see Equation (15)) there is $\varepsilon_\sigma>0$ and $\tau\in\mathcal{T}_{\varepsilon_\sigma}$, with $\gamma(\sigma,\tau)\leqslant\sup_{\sigma\in\Sigma}\inf_{\tau\in T}\gamma(\sigma,\tau)+2\delta/3$. However, because $\tau\in T_{\varepsilon_\sigma}$, $\gamma(\sigma,\tau)=\lim_{n\to+\infty}\gamma(\sigma_n,\tau)$, which contradicts (16). □

4.2. Proof of Theorem 3.4

Step 1. Structure of the proof.

The proof uses a variant of the vanishing discounting method. For every $\lambda\in[0,1)$, we denote by $\gamma_\lambda(\sigma,\tau)$ the λ-discounted expected payoff under the pair of strategies (σ,τ):

$$\gamma_\lambda(\sigma,\tau)=\mathbf{E}_{\sigma,\tau}[\lambda^{\theta-1}r(s_\theta)].$$

The function γ_λ is linear in both σ and τ (when viewed as probability distributions over $A^{\mathbf{N}}$ and $B^{\mathbf{N}}$), and jointly continuous. By Fan's (1953) fixed-point theorem, the λ-discounted game has a value v_λ:

$$v_\lambda := \max_{\sigma\in\Sigma}\min_{\tau\in\mathcal{T}}\gamma_\lambda(\sigma,\tau)=\min_{\tau\in\mathcal{T}}\max_{\sigma\in\Sigma}\gamma_\lambda(\sigma,\tau). \tag{17}$$

Set $v=\limsup_{\lambda\to 1}v_\lambda$. We will show that the value of the matching game is v. To this end, it is sufficient to construct for every $\varepsilon>0$ a strategy for player 1 that guarantees $\limsup_{\lambda\to 1}v_\lambda-\varepsilon$. Indeed, using a symmetric argument, player 2 would be able to guarantee $\liminf_{\lambda\to 1}v_\lambda+\varepsilon$. We fix throughout $\varepsilon>0$.

Step 2. Definition of the strategy.

Denote by σ_λ a λ-discounted optimal strategy of player 1, that is, a strategy that achieves the maximum in the second quantity in (17).

Recall that a pure strategy σ is an element of $\bigcup_{n\in\mathbf{N}}A^{A^{n-1}}$, so that a sequence of strategies converges if and only if every coordinate converges to a limit. Assume w.l.o.g. that the limit $\sigma_0:=\lim_{\lambda\to 0}\sigma_\lambda$ exists.

Because $\sigma_0=\lim_{\lambda\to 0}\sigma_\lambda$, and because $1=\lim_{\lambda\to 1}\lambda^t$ for every fixed t, we have for every $N\in\mathbf{N}$,

$$\mathbf{E}_{\sigma_0,\tau}[r(\theta)\mathbf{1}_{\theta\leqslant N}]=\lim_{\lambda\to 1}\mathbf{E}_{\sigma_\lambda,\tau}[\lambda^{\theta-1}r(\theta)\mathbf{1}_{\theta\leqslant N}].$$

Therefore, if τ is a strategy of player 2 that satisfies $\mathbf{P}_{\sigma_0,\tau}(\theta<+\infty)=1$, we have

$$\begin{aligned}\gamma(\sigma_0,\tau)&=\lim_{N\to\infty}\mathbf{E}_{\sigma_0,\tau}[r(\theta)\mathbf{1}_{\theta\leqslant N}]\\&=\lim_{N\to\infty}\lim_{\lambda\to 1}\mathbf{E}_{\sigma_\lambda,\tau}[\lambda^{\theta-1}r(\theta)\mathbf{1}_{\theta\leqslant N}].\end{aligned}$$

Because the last limit is uniform, this quantity is equal to

$$\begin{aligned}\lim_{\lambda\to 1}\lim_{N\to\infty}\mathbf{E}_{\sigma_\lambda,\tau}[\lambda^{\theta-1}r(\theta)\mathbf{1}_{\theta\leqslant N}]&=\lim_{\lambda\to 1}\gamma_\lambda(\sigma_\lambda,\tau)\\&\geqslant\limsup_{\lambda\to 1}v_\lambda.\end{aligned}$$

Hence, the only "problematic" strategies of player 2 are those that are not absorbing against σ_0. Because the game is a matching game, if the game is not absorbed it means that the actions of player 2 must have matched those of player 1, so that even though player 1 is not told the actions of player 2, he can deduce them from the assumption that the game is not absorbed. Observe that if this assumption is incorrect, and the game was already absorbed, then the actions chosen at this stage do not matter, so making an incorrect assumption cannot hurt player 1.

The optimal strategy of player 1 that we construct will be a perturbation of σ_0. It is sufficiently close to σ_0 to ensure that the payoff is high when $\mathbf{P}_{\sigma_0,\tau}(\theta<+\infty)=1$.

Let $\eta>0$ be given. We first introduce a (stopping) time t. Informally, t is the first time starting from which player 1's behavior is almost pure (nonrandom). Formally, letting $\vec{a}_n$ denote the sequence of moves played by player 1 in the first $n-1$ stages, we set

$$\begin{aligned}t:=\inf\{n\colon \mathbf{P}_{\sigma_0}(a_{n+k}=a_k^*,\ \forall k\geqslant 1\mid a_1,\ldots,a_n)\geqslant 1-\eta,\\ \text{for some }\vec{a}^*=(a_1^*,a_2^*,\ldots)\in A^{\mathbf{N}}\}.\end{aligned}$$

If player 1 happens to play $\vec{a}_n$ in the first $n-1$ stages, then with very high probability, he will play the sequence $\vec{a}^*$ in the future.

If $t=+\infty$, after each stage n player 1 mixes between several pure strategies after stage n. It follows that on the event $\{t=+\infty\}$ absorption occurs with probability 1:

$$\mathbf{P}_{\sigma_0,\tau}(\theta<+\infty\mid t=+\infty)=1. \tag{18}$$

For every $m\geqslant t$, let σ^m be a strategy of player 1 with the following properties:

(P.1) It coincides with σ_0 up to stage $M:=\max\{m,t\}$.

(P.2) Among all strategies that satisfy (P.1) it maximizes the payoff of player 1 (up to ε), assuming player 2 follows $\vec{a}^*$ after stage t.

As m increases, the constraints imposed on player 1's strategy gets sharper, so that the corresponding payoff $\gamma(\sigma^m,\vec{a}^*)$ is nonincreasing in m.

The sequence σ^m may or may not differ from $\vec{a}^*$. If $\sigma^m=\vec{a}^*$, then σ^{m+1} may also be taken to be equal to $\vec{a}^*$. Accordingly, we denote $Q:=\sup\{m\colon \sigma^m\neq\vec{a}^*\}$ (Q can be finite or infinite), and we let μ be a probability distribution over $\{1,2,\ldots,Q\}\cup\{\infty\}$ that assigns a probability at most η to each integer $m\leqslant Q$.

The strategy σ is defined as follows. Choose $m\in\{1,2,\ldots,Q\}\cup\{\infty\}$ according to μ. Play σ_0 until stage t, then switch to the sequence σ^m (with $\sigma^\infty=\vec{a}^*$). In effect, once at stage t, player 1 chooses at random how long he will comply with $\vec{a}^*$, and then plays optimally, assuming that player 2 does follow $\vec{a}^*$.

Step 3. σ is good against any pure strategy of player 2.

Fix a pure strategy of player 2. Such a pure strategy is given by a sequence $\vec{b}\in B^{\mathbf{N}}$ of actions. We will compare $\gamma(\sigma,\vec{b})$ with $\limsup_{\lambda\to 1}\gamma_\lambda(\sigma_\lambda,\vec{b})$, and prove that $\gamma(\sigma,\vec{b})$ is higher than v, up to some small error term.

By (18), and because σ coincides with σ_0 up to stage t,

$$\mathbf{E}_{\sigma,\vec{b}}[r(s_\theta)1_{\theta\leqslant t}]=\lim_{\lambda\to 1}\mathbf{E}_{\sigma_\lambda,\vec{b}}[\lambda^{\theta-1}r(s_\theta)1_{\theta\leqslant t}].$$

At stage t, there is high probability that player 1 will play according to $\vec{a}^*$, which is close to σ_0, so that, if $\theta=t+1$, the two payoffs are close.

We now focus on the payoff that is obtained if $\theta>t+1$—that is, on the case where player 1's moves match $\vec{b}$ until stage t. Below, we condition on that event. It is therefore convenient to relabel stages starting from stage t or, alternatively, to assume that $t=1$. We separate the proof into two cases.

Case 1. $\vec{b}=\vec{a}^*$. Let $m\geqslant 1$ be given. Denote σ_λ^m the strategy that plays $\vec{a}^*$ during $m-1$ stages, and then follows σ_λ. Plainly,

$$\gamma_\lambda(\sigma_\lambda^m,\vec{a}^*)\leqslant\sup\gamma_\lambda(\cdot\,,\vec{a}^*), \tag{19}$$

where the supremum is taken over all sequences that coincide with $\vec{a}^*$ during $m-1$ stages. The right-hand side is nonnegative because $\gamma_\lambda(\vec{a}^*, \vec{a}^*) = 0$. It is positive if there exists a stage beyond stage m at which player 1 may mismatch $\vec{a}^*$ and reach a positive payoff. The corresponding undiscounted payoff is therefore higher than the discounted one. That is,

$$\gamma_\lambda(\sigma^m_\lambda, \vec{a}^*) \leqslant \sup \gamma(\,\cdot\,, \vec{a}^*) \leqslant \gamma(\sigma^m, \vec{a}^*) + \varepsilon.$$

The second inequality holds because σ^m is a strategy that attains the supremum up to ε.

Because $\sigma_0 = \lim_{\lambda\to 1} \sigma_\lambda$ mostly plays $\vec{a}^*$, under $(\sigma_\lambda, \vec{b})$ the game will be absorbed with high probability; hence, $\limsup_{\lambda\to 1} \|\gamma_\lambda(\sigma_\lambda, \vec{a}^*) - \gamma_\lambda(\sigma^m_\lambda, \vec{a}^*)\| \leqslant \varepsilon$. It follows that

$$\gamma(\sigma^m, \vec{a}^*) \geqslant v - 2\varepsilon.$$

Because $\gamma(\sigma, \vec{a}^*)$ is a convex combination of $\gamma(\sigma^m, \vec{a}^*)$, $m \geqslant 1$, this implies

$$\gamma(\sigma, \vec{a}^*) \geqslant v - 2\varepsilon.$$

Case 2. $\vec{b} \neq \vec{a}^*$. Because σ_0 mostly plays like $\vec{a}^*$, one has $\|\limsup_{\lambda\to 1} \gamma_\lambda(\sigma_\lambda, \vec{b}) - \gamma(\vec{a}^*, \vec{b})\| \leqslant \eta$ because $\theta < +\infty$ if the players follow $(\vec{a}^*, \vec{b})$.

Let n be the stage of first mismatch between $\vec{a}^*$ and $\vec{b}$. If play proceeds up to stage n, that is, if $\theta > n+1$, the probability is at least $1-\varepsilon$ that player 1 will play according to $\vec{a}^*$ in stages n and $n+1$, and the payoff will be $\gamma(\vec{a}^*, \vec{b})$. On the other hand, if play does not proceed up to stage n, it must be that player 1 was playing according to σ^m for some m. Because $\vec{b}$ coincides with $\vec{a}^*$ up to stage n, the payoff is that induced by $(\sigma^m, \vec{b})$, $\gamma(\sigma^m, \vec{a}^*)$. According to the analysis in Case 1, we thus proved that

$$\gamma(\sigma, \vec{b}) \geqslant \min\left\{\limsup_{\lambda\to 1} \gamma_\lambda(\sigma_\lambda, \vec{b}), \limsup_{\lambda\to 1} \gamma_\lambda(\sigma_\lambda, \vec{a}^*)\right\} - \varepsilon$$
$$\geqslant v - \varepsilon.$$

Step 4. σ is optimal.

We showed in Step 3 that

$$\gamma(\sigma, \vec{b}) \geqslant \limsup_{\lambda\to 1} \gamma_\lambda(\sigma_\lambda, \vec{b}) - \varepsilon$$

for every sequence of actions of player 2 $\vec{b}$. By Kuhn's theorem (Kuhn 1953), every strategy is a probability distribution over pure strategies. Because for every sequence of uniformly bounded r.v.s (X_n) one has $\mathbf{E}[\limsup_{n\to\infty} X_n] \geqslant \limsup_{n\to\infty} \mathbf{E}[X_n]$, we deduce that for every strategy τ of player 2,

$$\begin{aligned}\gamma(\sigma, \tau) &= \mathbf{E}_\tau[\gamma(\sigma, \vec{b})]\\ &\geqslant \mathbf{E}_\tau[\limsup_{\lambda\to 1} \gamma_\lambda(\sigma_\lambda, \vec{b})]\\ &\geqslant \limsup_{\lambda\to 1} \mathbf{E}_\tau[\gamma_\lambda(\sigma_\lambda, \vec{b})]\\ &= \limsup_{\lambda\to 1} \gamma_\lambda(\sigma_\lambda, \tau).\end{aligned}$$

Proof of Lemma 3.8. Let σ be an optimal strategy of player 1. In particular, it guarantees zero against every strategy of player 2. Given a stage $n \in \mathbf{N}$, we let B_n (respectively, P_n) denote the event: player 1 chooses B (respectively, P) at stage n.

Step 1. For each $n \geqslant 1$, $\mathbf{P}_\sigma(B_n) \leqslant 1/2$.

Let $n \geqslant 1$ be arbitrary. The strategy of player 2 that plays R for the first time in stage n yields -1 in the event B_n, and at most 1 otherwise. If $\mathbf{P}_\sigma(B_n) > 1/2$, this strategy yields negative expected payoff against σ, which contradicts the fact that σ guarantees zero.

Step 2. σ coincides with σ^* after any sequence of positive probability.

If player 2 plays R for the first time at stage $n \geqslant 1$, the payoff is 1 on the event $P_n B_{n+1}$, and at most -1 otherwise. Thus, $\mathbf{P}_\sigma(P_n B_{n+1}) \geqslant 1/2$. By Step 1, this implies $\mathbf{P}_\sigma(B_{n+1}) = 1/2$ for each $n \geqslant 1$. Next, let $x = \mathbf{P}_\sigma(B_1) \leqslant 1/2$, so that $\mathbf{P}(P_1 B_2) \geqslant 1 - x - 1/2$. If player 2 plays R at stage 1, the expected payoff is at most $x \times (-1) + 1/2 \times 1 + (1 - x - 1/2) \times (-2) = x - 1/2$. Because the expected payoff is nonnegative, this yields $x = 1/2$, so that $\mathbf{P}_\sigma(B_n) = 1/2$ for each $n \geqslant 1$. To conclude the proof, it is sufficient to prove that the probability is zero that player 1 plays B twice in a row.

For each $n \geqslant 1$,

$$\mathbf{P}_\sigma(B_{n+1}) = \tfrac{1}{2} = \mathbf{P}_\sigma(P_n B_{n+1}) + \mathbf{P}_\sigma(B_n B_{n+1}).$$

Because $\mathbf{P}_\sigma(P_n B_{n+1}) = 1/2$, this yields $\mathbf{P}_\sigma(B_n B_{n+1}) = 0$, and the result follows. □

Endnotes

[1] It is natural to assume that the processor who made the successful request be informed that either its packet was successfully transmitted (so that the processor keeps track of which packets were transmitted), or that it was penalized. For simplicity, we assume that this information is available to the two processors. It would be interesting to study the situation when only the processor who made the request has this information.

[2] But they need not be able to compute the probability of still being in a standard state.

[3] The behavior version of this strategy is the following. At stage $n \geqslant 1$, he chooses P and B with probability 1/2 each. From then on, he plays the action that he did not play in the previous stage.

[4] By Kuhn's theorem (Kuhn 1953), a strategy can be identified with a probability distribution over infinite sequences of actions. Given a measurable set $\mathcal{A} \subset A^{\mathbf{N}}$ of sequences, and a history $\vec{b} \in B^{N_1}$ of past moves, the probability assigned by the continuation strategy to $\mathcal{A}$ is the probability (computed with σ) that the sequence of actions played from stage $N_1 + 1$ on is in $\mathcal{A}$ conditional on the (unobservable to the players) event $\theta > N_1$ and on player 2 having played $\vec{b}$.

Acknowledgments

The authors thank Eran Shmaya for providing the example presented in §3.2, and two anonymous referees for their helpful comments. This research was supported by a grant from the Ministry of Science and Technology, Israel, and the Ministry of Research, France. The research of the second author was supported by the Israel Science Foundation, grant number 69/01.

References

Abramson, N. 1970. The Aloha system: Another alternative for computer communications. *AFIPS Conf. Proc.* **36** 295–298.

Alpern, S., S. Gal. 2002. Searching for an agent who may or may not want to be found. *Oper. Res.* **50** 311–323.

Altman, E., R. El Azouzi, T. Jiménez. 2004a. Slotted Aloha as a stochastic game with partial information. *Comput. Networks* **45** 701–713.

Altman, E., D. Barman, R. El Azouzi, T. Jiménez. 2004b. A game theoretic approach for delay minimization in slotted Aloha. *Proc. IEEE ICC*, Paris, 3999–4003.

Arapostathis, A., V. S. Borkar, E. Fernandez-Gaucherand, M. K. Gosh, S. I. Marcus. 1993. Discrete-time controlled Markov processes with average cost criterion: A survey. *SIAM J. Control Optim.* **31** 282–344.

Avenhaus, R., B. von Stengel, S. Zamir. 2002. Inspection games. R. J. Aumann, S. Hart, eds. *Handbook of Game Theory with Economic Applications*, Vol. III, Chapter 51. Elsevier.

Everett, H. 1957. Recursive games. *Contributions to the Theory of Games*, Vol. 3. *Annals of Mathematical Studies*, Vol. 39. Princeton University Press, Princeton, NJ, 47–78.

Fan, K. 1953. Minimax theorems. *Proc. National Acad. Sci. U.S.A.* **39** 42–47.

Filar, J. A., K. Vrieze. 1996. *Competitive Markov Decision Processes*. Springer.

Flesch, J., F. Thuijsman, O. J. Vrieze. 1998. Simplifying optimal strategies in stochastic games. *SIAM J. Control Optim.* **36** 1331–1347.

Gal, S., J. V. Howard. 2005. Rendezvous-evasion search in two boxes. *Oper. Res.* **53**(4) 689–697.

IEEE Standard 802.11a. 1999. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. IEEE, Piscataway, NJ.

Kuhn, H. W. 1953. Extensive games and the problem of information. H. W. Kuhn, A. W. Tucker, eds. *Contributions to the Theory of Games*, Vol. 2. *Annals of Mathematical Studies*, Vol. 28. Princeton University Press, Princeton, NJ.

Monahan, G. E. 1982. A survey of partially observable Markov decision processes: Theory, models and algorithms. *Management Sci.* **28** 1–16.

Neyman, A., S. Sorin. 2004. *Stochastic Games and Applications*. Springer.

Roberts, L. G. 1975. ALOHA packet system with and without slots and capture. *Comput. Comm. Rev.* **5** 28–42.

Sagduyu, Y. E., A. Ephremides. 2003. Power control and rate adaptation as stochastic games for random access. *Proc. 42nd IEEE Conf. Decision and Control*, Hawaii. IEEE, Piscataway, NJ, 4202–4207.