# GAMES WITH A RECURSIVE STRUCTURE

JEAN-MICHEL COULOMB
*The University of York*
*York, UK*

**Abstract.** This chapter is based on a lecture of Jean-François Mertens. Two main topics are dealt with: ($i$) The reduction of a general (stochastic) game model to various combinatorial descriptions; ($ii$) the use of consistent probability distributions on the Universal Belief Space in order to exploit a recursive structure of zero-sum games.

These constructions lead to a conjecture that would guarantee the existence of the max min and its characterization whenever the information received by the maximizer is finer than that received by the minimizer.

## 1. Introduction

### 1.1. DESCRIPTION OF THE MODEL

Throughout the chapter, $S$ is a finite state space and $A$ (resp. $B$) is a finite set of actions available for player I (resp. player II) in any state $s \in S$. There is no loss of generality since by duplicating actions we can insure that the action sets are the same regardless of the state.

Any pair of payoffs (one for player I and one for player II) $r = (r^I, r^{II})$ remains in a compact set $R \subset \mathbb{R}^2$. The signals $m^I \in M^I$ (resp. $m^{II} \in M^{II}$) received by player I (resp. II) will be his only source of information. As above $M^I$ (resp. $M^{II}$) is finite. Let us denote by $M$ the product $M^I \times M^{II}$.

Our game model $\Gamma(I\!P)$ is described by:

(i) A probability measure $I\!P \in \Delta(S \times M)$.
(ii) A transition probability $Q: \ S \times A \times B \to \Delta(R \times S \times M)$.

Observe that $I\!P$ is seen as a variable whereas $Q$ is fixed. Later on we shall see why such a point of view is useful. The game $\Gamma(I\!P)$ unfolds as follows:

- An initial state $s_1 \in S$ and an initial pair of signals $(m_1^I, m_1^{II}) \in M$ are chosen according to $I\!P$. Then, player I (resp. II) receives the signal $m^I \in M^I$ (resp. $m^{II} \in M^{II}$).
- At stage $n$, assuming that the game is in state $s_n$, player I (resp. II) selects an action $a_n \in A$ (resp. $b_n \in B$) on the basis of his private information $m_1^I, \ldots, m_n^I$ (resp. $m_1^{II}, \ldots, m_n^{II}$). Subsequently, a new state $s_{n+1}$, a pair of signals $(m_n^I, m_n^{II})$ and a pair of payoffs $(r_n^I, r_n^{II})$ are chosen according to $Q(s_n, a_n, b_n)$.

Observe that the players are not told the current state unless it is included in their signal. For both players, any belief regarding the state or the previous actions of the opponent is based on past and current signals received so far. It is usual to assume that the signal of player I (resp. II) contains his previous action.

This model includes *information lags*. This means that the signals might not be communicated immediately to the players. The way information lags are modelled is as follows: each state $\tilde{s} \in \tilde{S}$ of a finite-state machine specifies which information should be given and retains any new information to be disclosed later on.

Instead of $I\!P$ (item $(i)$), our model is characterized by a probability measure $I\!P' \in \Delta(S \times \tilde{S} \times M)$ which selects the state $s \in S$, the state of the machine $\tilde{s} \in \tilde{S}$ and the pair of signals $m \in M$.

The transition probability $Q$ (item $(ii)$) is modified so as to include an effect of the current $\tilde{s}$ on the signals. Formally, this means that it is of the form $Q: \ \tilde{S} \times S \times A \times B \to \Delta(R \times S \times M)$. Moreover, the new state of the machine is chosen by the following transition probability $\tilde{Q}$ : $\tilde{S} \times A \times B \times R \times S \to \Delta(\tilde{S})$ involving the new state that has already been chosen using $Q$.

We obtain the exact formulation of the previous model when we consider a larger state space, $S' = S \times \tilde{S}$, with a corresponding transition probability $Q' = Q'(Q, \tilde{Q})$.

### 1.2. RECURSIVE STRUCTURE

In Section 2, we shall operate a few transformations of the initial model of Section 1.1. The final reduction of Section 2.1 is the one the reader should keep in mind since we will use it in Sections 5 and 6. In those sections, we shall make the assumption that the game $\Gamma(I\!P)$ is zero-sum ($r^I + r^{II} = 0$).

The *recursive structure* appears when one links the value of the (finite) $T$-stage game $\Gamma_T(I\!P)$ with the value of some $(T-1)$-stage games $\Gamma_{T-1}(I\!P')$ where $I\!P'$ is an "updated" initial probability measure (see formula (4) in Section 5).

The first idea is to reformulate our game on a different state space called the *Universal Belief Space* (Section 3). The *types* of the players play the role of the signals. We show that the recursive structure is transferred to this new game (Proposition 3) and that the value of finite or discounted games is preserved.

At first sight, it is not at all clear how this "artificial" game could be useful in studying uniform properties. However, let us assume that, when playing $\Gamma(I\!P)$, the signal of player II is included in the signal of player I. Then, it turns out that the game defined on the Universal Belief Space can be formulated as a stochastic game with the state space being the set of player II's types (Section 6).

The next idea is as follows: if this stochastic game has a value (in the spirit of [1]) or a max min, then the game $\Gamma(I\!P)$ has a max min (Proposition 4).

## 2. Reductions

### 2.1. MODELS WITH PARTITIONS

First, we shall refine the initial model so that payoffs and signals are related to particular partitions of the state space. For this purpose we have to define a set of more complex states. Somehow, such states include the previous ones, the payoffs and the signals.

Next, observe that the payoffs could be renormalized so that any pair of payoffs $r = (r_n^I, r_n^{II})$ is spanned by $X_1 = (1,0),\ X_2 = (0,1)$ and $X_3 = (0,0)$. Furthermore, since eventually we will be interested in dealing with the expected payoff, there is no loss of generality in assuming that $r$ is replaced by a probability measure on the set $X = \{X_1, X_2, X_3\}$.

The new state space is $\bar{S} = X \times S \times M$. The probability $I\!P$ on $S$ is rewritten as a probability $\bar{I\!P}$ on $\bar{S}$ (the pair of payoffs is $X$). Accordingly, we change the domain and the image set of the transition probability $Q$ (item $(ii)$). Let us denote the modified transition probability by $\bar{Q}$ with

$$\bar{Q}:\ \bar{S} \times A \times B \to \Delta(\bar{S}).$$

Now, our model is characterized by:

(i) A finite state space $\bar{S}$.

(ii) A partition $\Pi^I$ (resp. $\Pi^{II}$) of $\bar{S}$ corresponding to the information available to player I (resp. II).

(iii) A partition $\{W^I, W^{II}, W^0\}$ of $\bar{S}$ such that $W^I$ (resp. $W^{II}$) is the set of states with payoffs $X_1$ (resp. $X_2$). The former may be seen as the set of winning states of player I (resp. player II).

A play unfolds as follows:

- An initial state $\bar{s}_1$ is selected at stage 1 according to $\bar{I\!P} \in \Delta(\bar{S})$.
- At any stage $n \geq 1$, if the state is $\bar{s}_n$, then player I (resp. II) is informed of the element of the partition $\Pi^I$ (resp. $\Pi^{II}$) containing $\bar{s}_n$. Given that additional information, the players simultaneously select an action $a$ and $b$. The next state $\bar{s}_{n+1}$ is chosen according to the probability measure $Q(\bar{s}_{n+1}, a, b)$.

## 2.2. COMBINATORIAL MODEL

In this section we wish to give an insight about how to eliminate the stochastic features of our model. First we shall enlarge the state space. Let us introduce

$$\hat{S} = \bar{S} \cup (\bar{S} \times A) \cup (\bar{S} \times A \times B).$$

The transition probability $\bar{Q}$ is extended in a deterministic way. Given $\bar{s} \in \bar{S}$, if player I chooses the action $a \in A$, then the next state is $(\bar{s}, a)$. It is now player II's turn. If player II selects the action $b \in B$, then the next state is $(\bar{s}, a, b)$. Finally, a new state $\bar{s}' \in \bar{S}$ is chosen according to $\bar{Q}(\bar{s}, a, b)$, and so on.

Next, one wants to remove the stochastic nature of $\bar{Q}$. Notice that this last reduction step provides an equivalence with the previous formalizations in a limited sense only when the game is not zero-sum ([3],[4]). For simplicity, we shall restrict ourselves to the zero-sum case and in addition assume that the probability measures are rational and therefore have a smallest common denominator $m$. (It is unclear how limiting this assumption is but we guess that if the supports of the distributions are preserved, a small perturbation of the probability measures should not affect the model too much.)

With our rationality assumption, to any probability measure there corresponds a canonical partition of $Z_m$ (the usual group of relative integers modulo $m$), i.e., each element of the partition is related to a particular outcome of the lottery. Let us replace the initial action set $A$ (resp. $B$) by the following one $A' = A \times Z_m$ (resp. $B' = B \times Z_m$). If the current state is $\bar{s} \in \bar{S}$ and if player I chooses the action $a' = (a, z_I)$ then the subsequent state is $(\bar{s}, a')$. Next, if player II selects $b' = (b, z_{II})$, the following state is $(\bar{s}, a', b')$. To close the cycle, one looks at $z = z_I + z_{II}$ and the next state is the outcome corresponding to the element of the partition of $Z_m$, associated with $\bar{Q}(\bar{s}, a, b)$, that contains $z$.

Let us consider the average payoff of $T$ full cycles. In both models (with or without randomization), it is a finite game and therefore it has a value. Are both values equal and how are the optimal strategies related to each other?

The values are indeed the same and any optimal strategy in the model with randomization induces an optimal strategy in the one without randomization in a simple way. At stage $t$, in the latter model, player I (for instance) chooses his action $a_t$ according to an optimal strategy in the former model and $z_t$ according to the uniform probability on $Z_m$. The verification is left to the reader as an exercise. The argument for dealing with some version of the asymptotic payoff would be a bit more technical, but basically the same (use of the "Haar" measure on $Z_m \times Z_m \times \ldots$).

To summarize, here is the description of our combinatorial model:

- We have a partition $S = S^I \cup S^{II}$ of our state space. $S^I$ (resp. $S^{II}$) is the set of states such that it is player I's (resp. II's) turn to play. Each set $S^I$ and $S^{II}$ is partitioned by $\Pi^I$ and $\Pi^{II}$ so that it represents the information available to player I and II respectively.
- The set of actions $A$ (resp. $B$) of player I (resp. II) is a subset of the set of mappings $f:\ S^I \to S^{II}$ (resp. $S^{II} \to S^I$).

## 3. Universal Belief Space

To play the game $\Gamma(\mathbb{P})$, one first selects randomly an initial state in the finite state space $S$ and a pair of signals $m = (m^I, m^{II})$ in a finite product set $M = M^I \times M^{II}$. It is particularly interesting to isolate such a mechanism and to study it for its own sake. The rest of the chapter will show how useful it could be to analyze $\Gamma(\mathbb{P})$.

**Definition 1** *An Information Scheme $\mathcal{I}$ is a pair $(\Omega, \mathbb{P})$ where*

- *$\Omega = S \times E^I \times E^{II}$, where $E^I$ (resp. $E^{II}$) is a set of signals.*
- *$\mathbb{P} \in \Delta(\Omega)$, with finite support.*

*A triple $\omega = (s, e^I, e^{II})$ is chosen according to $\mathbb{P}$. Then, $e^I$ (resp. $e^{II}$) is communicated to player I (resp. II).*

Observe that the players know $\mathbb{P}$, they know that each knows and so on... . It is beyond the scope of the present chapter to elaborate formally on this, but any pair of signals given to the players generates an infinite sequence of hierarchical inferences for each player about the beliefs of his opponent [2]. Given $\omega \in \Omega$, player I (resp. II) has a conditional distribution on $S \times E^{II}$ (resp. $S \times E^I$) since he knows $e^I = e^I(\omega)$ (resp. $e^{II} = e^{II}(\omega)$). Therefore, player I (resp. II) has a marginal probability measure $t_1^I(\omega)$ (resp. $t_1^{II}(\omega)$) on $\Delta(S)$. Let us denote by $T_1^I$ (resp. $T_1^{II}$) the set of such probability measures. Consequently player I (resp. II) has a probability measure $t_2^I(\omega)$ (resp. $t_2^{II}(\omega)$) on $S \times T_1^{II}$ (resp. $S \times T_1^I$). The set of such probability measures is denoted by $T_2^I$ (resp. $T_2^{II}$). Thus, one obtains a sequence of sets $T_1^I, T_2^I, \ldots$ for player I and similarly for player II. With $T_0^I = \{1\}$ (resp. $T_0^{II} = \{1\}$), let us introduce the uniquely defined mapping $f_0^1 :\ T_1^I \to T_0^I$ (resp. $g_0^1 :$

$T_1^{II} \to T_0^{II}$) and starting from it let us define $f_n^{n+1}: \; T_{n+1}^I \to T_n^I$ in a straightforward way since $T_{n+1}^I = \Delta(S \times T_n^{II})$. What we have is a projective system, so let us define $T^I = \lim_{\leftarrow} T_n^I$ (resp. $T^{II} = \lim_{\leftarrow} T_n^{II}$). Any element of $T^I$ is called a type of player I and any element of $T^{II}$ is a type of player II. Observe that $T^I$ (resp. $T^{II}$) is a Hausdorff space homeomorphic to $\Delta(S \times T^{II})$ (resp. $\Delta(S \times T^I)$) (the set of regular probability measures endowed with the weak$^\star$-topology. [2], [4]).

**Definition 2** *The Universal Belief Space is defined as* $\tilde{\Omega} = S \times T^I \times T^{II}$.

To any $\omega = (s, e^I, e^{II})$ there corresponds canonically a type $\underline{t}^I(e^I)$ (resp. $\underline{t}^{II}(e^{II})$) of player I (resp. II). Therefore, there is a canonical mapping $\phi: \; \Omega \to \tilde{\Omega}$ defined by $\phi(\omega) = (s, \underline{t}^I(e^I), \underline{t}^{II}(e^{II}))$ with $\omega = (s, e^I, e^{II})$. Notice that $I\!P \circ \phi^{-1}$ is a probability measure $\tilde{I\!P}$ on $\tilde{\Omega}$ with finite support.

The probability measure $\tilde{I\!P}$ is consistent in the following sense:

$$\tilde{I\!P}[s, t^I | t^{II}] = t^{II}[s, t^I] \tag{1}$$

and similarly with the type of player I.

Notice that if $\underline{t}^{II}(e^{II}) = t^{II}$, then

$$I\!P[\{(s, e^I) | t^I = \underline{t}^I(e^I)\} | e^{II}] = t^{II}[s, t^I] \tag{2}$$

and similarly, if $\underline{t}^I(e^I) = t^I$, then

$$I\!P[\{(s, e^{II}) | t^{II} = \underline{t}^{II}(e^{II})\} | e^I] = t^I[s, t^{II}]. \tag{3}$$

Observe that $(\tilde{\Omega}, \tilde{I\!P})$ itself is a particular information scheme, denoted by $\tilde{\mathcal{I}}$. It has an interesting property: it does not need to be known to the players. It is sufficient for them to know that the types are chosen according to a consistent probability measure. Even if they know it, their type would coincide with the types that have been randomly chosen.

## 4. Finite Games and Information Schemes

Let $I$ and $J$ be two finite action sets and let $G = (G^s)_{s \in S}$ be a family of $I \times J$-matrix zero-sum games $[G^s_{i,j}]_{i,j}$. Given an information scheme $\mathcal{I}$, let us consider the following zero-sum game $G^{\mathcal{I}}$:

- Choose $\omega = (s, e^I, e^{II})$ and communicate the signals to the players according to $\mathcal{I}$.
- After receiving his signal, player I (resp. II) selects an action $i \in I$ (resp $j \in J$).
- If the actual state is $s$, then the payoff of player I is $G^s_{i,j}$.

Since $G^{\mathcal{I}}$ is a finite game, it has a value, denoted by $v(G^{\mathcal{I}})$. In parallel to $G^{\mathcal{I}}$, we will consider the game $G^{\tilde{\mathcal{I}}}$. Likewise, its value is denoted by $v(G^{\tilde{\mathcal{I}}})$.

One way to use his signal $e^I$ would be for player I to take into account his type $t^I = \underline{t}^I(e^I)$ to implement a strategy of $G^{\tilde{\mathcal{I}}}$. It would seem that in doing so, there is a loss of information. But there is none because we are dealing with a zero-sum game.

**Proposition 1** – *If $\sigma$ guarantees an amount $w$ in $G^{\tilde{\mathcal{I}}}$, so does its implementation in $G^{\mathcal{I}}$.*

- *In particular, the implementation of any optimal strategy from $G^{\tilde{\mathcal{I}}}$ is an optimal strategy in $G^{\mathcal{I}}$.*
- *Furthermore, $v(G^{\mathcal{I}}) = v(G^{\tilde{\mathcal{I}}})$.*

**Proof.** Let $\tilde{x} : T^I \to \Delta(I)$ be a strategy of player I that guarantees $w$ in $G^{\tilde{\mathcal{I}}}$ and let us denote by $\bar{x} : E^I \to \Delta(I)$ the strategy $\tilde{x} \circ \underline{t}^I$ of player I in $G^{\mathcal{I}}$. It is called the *implementation* of $\tilde{x}$ in $G^{\mathcal{I}}$.

Let $\bar{y} : E^{II} \to J$ be a strategy of player II in $G^{\mathcal{I}}$. Let us denote by $\tilde{y} : T^{II} \to \Delta(J)$ the strategy of player II in $G^{\tilde{\mathcal{I}}}$ defined as follows ($j \in J$):

$$\tilde{y}(t^{II})[j] = \frac{\mathbb{P}[\{e^{II} | \underline{t}^{II}(e^{II}) = t^{II},\ \bar{y}(e^{II}) = j\}]}{\tilde{\mathbb{P}}[t^{II}]}.$$

This technique is described as *averaging* $\bar{y}$ with respect to the type of player II.

Note that:

$$\begin{aligned}
\mathbb{E}_\omega[\sum_i \bar{x}(e^I)[i] G^s_{i,j(e^{II})}] &= \sum_{e^{II}} \mathbb{P}[e^{II}] \sum_{s,e^I,i} \tilde{x}(\underline{t}^I(e^I))[i] G^s_{i,j(e^{II})} \mathbb{P}[s, e^I | e^{II}] \\
&= \sum_{t^{II}} \tilde{\mathbb{P}}[t^{II}] \sum_{s,t^I,i,j} \tilde{x}(t^I)[i] y(t^{II})[j] G^s_{i,j} t^{II}[s, t^I]
\end{aligned}$$

by application of (2). By consistency (1) we obtain

$$\mathbb{E}_\omega[\sum_i \bar{x}(e^I)[i] G^s_{i,j(e^{II})}] = \tilde{\mathbb{E}}_{\tilde{\omega}}[\sum_{i,j} \tilde{x}(t^I)[i] y(t^{II})[j] G^s_{i,j}].$$

Remember that, in $G^{\tilde{\mathcal{I}}}$, $\tilde{x}$ guarantees $w$ to player I. This implies that

$$\mathbb{E}_\omega[\sum_i \bar{x}(e^I)[i] G^s_{i,j(e^{II})}] \geq w.$$

Hence in particular $v(G^{\mathcal{I}}) \geq v(G^{\tilde{\mathcal{I}}})$. A reverse inequality could be shown, thus establishing that the two values are the same and that the implementation of optimal strategies produces optimal strategies. ∎

Next, we state a property that will play an important role in Section 6.

**Proposition 2** *Seen as a function $v(\tilde{I\!P})$ of a consistent probability measure $\tilde{I\!P}$ with finite support, $v(G^{\tilde{\mathcal{I}}})$ is affine.*

**Proof.** Let $\tilde{I\!P}_1$ and $\tilde{I\!P}_2$ be two consistent probability measures with finite support. Let us denote by $\tilde{\mathcal{I}}_1$ (resp. $\tilde{\mathcal{I}}_2$) the corresponding information scheme. We can define a new information scheme $\tilde{\mathcal{I}}$ as follows:

- Choose $\tilde{I\!P}_1$ (resp. $\tilde{I\!P}_2$) with probability $\lambda$ (resp $1-\lambda$). The players are informed of the outcome of the lottery.
- Play $G^{\tilde{\mathcal{I}}_1}$ (resp. $G^{\tilde{\mathcal{I}}_2}$) if the lottery has drawn $\tilde{I\!P}_1$ (resp. $\tilde{I\!P}_2$).

Observe that, since we are dealing with consistent probability measures, the players do not need to know which one has been drawn by the lottery. Therefore, $\tilde{\mathcal{I}}$ is equivalent to choosing the type of the players with the consistent probability measure $\lambda\tilde{I\!P}_1 + (1-\lambda)\tilde{I\!P}_2$. ∎

## 5. Games with a Recursive Structure

From now on, we deal with the formalization of Section 2.1. Let us use our initial notations, namely $S$ for the state space and $I\!P$ for the initial lottery (instead of $\bar{S}$ and $\bar{I\!P}$) and assume that $\Gamma(I\!P)$ is zero-sum. The set of winning states $W^I$ for player I is simply denoted by $W$.

As the game unfolds, a family of past and present signals (i.e., elements of the partition where the actual state $s$ falls) is available to player I (resp. player II) at any stage. The current payoff depends only on the current state and any assessment about the current state is based solely on those signals. How should the players treat that information in order to construct an optimal or an $\epsilon$-optimal strategy? Remember that for both players, his current signal contains his previous action.

At any stage, it seems that we have a particular information scheme with very large sets of signals, except that the players ignore which strategy has been chosen by his opponent. Therefore, the players don't know what theoretical probability measure induced those signals. As we said when defining information schemes, the knowledge of this probability measure is crucial.

Nevertheless, we are going to show that given any pair of strategies, the corresponding consistent probability measure with finite support on the Universal Belief Space $\tilde{\Omega}$ is useful. We will define an auxiliary game which will have the same value as the initial one whenever the minmax theorem applies to the latter. In this section we show recursive formulae satisfied by finitely repeated or discounted games. In the next section we will address the question of infinitely repeated games.

Let us consider the $T$-stage game $\Gamma_T(I\!P)$. Note that it is a finite game and it begins, at *stage 0*, by an initial lottery that amounts to an information scheme (because this time $I\!P$ is known by the players). Let us denote by $\tilde{I\!P}$

the corresponding consistent probability measure on $\tilde{\Omega}$. As we have shown in Proposition 1, the value of $\Gamma_T(I\!\!P)$ is also the value of $\Gamma_T(\tilde{I\!\!P})$, hence a function $v_T(\tilde{I\!\!P})$ of $\tilde{I\!\!P}$. It is possible to implement a pair of optimal strategies which depend on the initial type ($t_1^I$ or $t_1^{II}$) instead of the first signal ($m_1^I$ or $m_1^{II}$).

Now, let $\sigma$ be an optimal strategy of player I such that $\sigma = \sigma_1, \bar{\sigma}$ where $\sigma_1 : T^I \to \Delta(A)$ and $\bar{\sigma}$ is a strategy in the $T-1$-stage game, starting at stage 2, that does not involve the first signal $m_1^I$ received but the initial type $t_1^I$. Similarly, let us denote by $\tau = \tau_1, \bar{\tau}$ an optimal strategy of player II such that $\tau_1 : T^{II} \to \Delta(B)$ and $\bar{\tau}$ is a strategy in the $T-1$-stage game, starting right after stage 1, that does not involve the first signal $m_1^{II}$ received but the initial type $t_1^{II}$.

To play at stage 2, the information used by player I (resp. II) is his initial type $t_1^I$ (resp. $t_1^{II}$) and his current signal $m_2^I$ (resp. $m_2^{II}$). The corresponding distribution $I\!\!P^{\sigma_1,\tau_1}$ on $S \times T^I \times T^{II}$ can be calculated using $(\sigma_1, \tau_1)$ and $\tilde{I\!\!P}$. Therefore, assuming that the players have used the pair $(\sigma_1, \tau_1)$ at the first stage, the game that the players face after that is nothing other than $\Gamma_{T-1}(I\!\!P^{\sigma_1,\tau_1})$ where the previous types are added to the new signals (given by the partitions $\Pi^I$ and $\Pi^{II}$ as usual). Let us denote by $\tilde{I\!\!P}^{\sigma_1,\tau_1}$ the corresponding consistent probability measure. The value of $\Gamma_{T-1}(I\!\!P^{\sigma_1,\tau_1})$ is a function $v_{T-1}(\tilde{I\!\!P}^{\sigma_1,\tau_1})$ of $\tilde{I\!\!P}^{\sigma_1,\tau_1}$.

Obviously the pair $(\bar{\sigma}, \bar{\tau})$ is optimal in $\Gamma_{T-1}(I\!\!P^{\sigma_1,\tau_1})$. Once again, one may assume that instead of depending on the initial type and the first signal, it depends on the second type ($t_2^I$ or $t_2^{II}$), calculated with $I\!\!P^{\sigma_1,\tau_1}$.

Hence, $(\bar{\sigma}, \bar{\tau})$ is optimal in $\Gamma_{T-1}(\tilde{I\!\!P}^{\sigma_1,\tau_1})$ where the initial lottery is a consistent probability measure which draws types. Therefore the pair remains optimal in the game $\Gamma_{T-1}(\tilde{I\!\!P}^{\sigma_1',\tau_1'})$ for any consistent probability measure $\tilde{I\!\!P}^{\sigma_1',\tau_1'}$ associated with a different first mixed action $\sigma_1' : T^I \to \Delta(A)$ (resp. $\tau_1' : T^{II} \to \Delta(B)$) of player I (resp. II).

Let us denote by $\tau'$ the strategy $\tau_1', \bar{\tau}$. Since $\sigma$ is optimal and since by the previous remark $(\bar{\sigma}, \bar{\tau})$ achieves $v_{T-1}(\tilde{I\!\!P}^{\sigma_1,\tau_1'})$, we have

$$\frac{1}{T}\tilde{I\!\!P}[W] + \frac{T-1}{T} v_{T-1}(\tilde{I\!\!P}^{\sigma_1,\tau_1'}) \geq v_T(\tilde{I\!\!P}).$$

Hence, by taking the infimum with respect to the arbitrary $\tau_1'$, we obtain

$$\frac{1}{T}\tilde{I\!\!P}[W] + \frac{T-1}{T} \inf_{\tau_1} v_{T-1}(\tilde{I\!\!P}^{\sigma_1,\tau_1}) \geq v_T(\tilde{I\!\!P}).$$

Similarly, we could show that

$$\frac{1}{T}\tilde{I\!\!P}[W] + \frac{T-1}{T} \sup_{\sigma_1} v_{T-1}(\tilde{I\!\!P}^{\sigma_1,\tau_1}) \leq v_T(\tilde{I\!\!P}).$$

This is enough to prove that $(\sigma_1,\tau_1)$ is a saddle point of $(\sigma_1',\tau_1')\mapsto v_{T-1}(\tilde{I\!P}^{\sigma_1',\tau_1'})$ and that

$$\begin{aligned} v_T(\tilde{I\!P}) &= \frac{1}{T}\tilde{I\!P}[W]+\frac{T-1}{T}\min_{\tau_1}\max_{\sigma_1} v_{T-1}(\tilde{I\!P}^{\sigma_1,\tau_1}) \\ &= \frac{1}{T}\tilde{I\!P}[W]+\frac{T-1}{T}\max_{\sigma_1}\min_{\tau_1} v_{T-1}(\tilde{I\!P}^{\sigma_1,\tau_1}). \end{aligned} \tag{4}$$

Instead of uniformly averaging the payoff of the first $T$ stages as we previously did, we could truncate the $(1-\lambda)$-discounted average (with $\lambda\in(0,1)$). It is still a finite game; therefore it has a value $v_{T,\lambda}(\tilde{I\!P})$ and a recursive equation similar to (4):

$$v_{T,\lambda}(\tilde{I\!P})=\lambda\tilde{I\!P}[W]+(1-\lambda)\min_{\tau_1}\max_{\sigma_1} v_{T-1,\lambda}(\tilde{I\!P}^{\sigma_1,\tau_1}).$$

Taking the limit when $T\to\infty$, the value $v_\lambda(\tilde{I\!P})$, we obtain

$$v_\lambda(\tilde{I\!P})=\lambda\tilde{I\!P}[W]+(1-\lambda)\min_{\tau_1}\max_{\sigma_1} v_\lambda(\tilde{I\!P}^{\sigma_1,\tau_1}). \tag{5}$$

We have exploited the fact that the payoff of a discounted game can be uniformly approximated by the payoff of a finite game.

Now, our goal is to define a new game $\tilde\Gamma(\tilde{I\!P})$ which could play the role of a "formal" representation of $\Gamma$. We proceed as follows: its state space is the set of consistent probability measures $\tilde{I\!P}$ with finite support. An action of player I (resp. player II) is a mapping $\sigma:\ T^I\to\Delta(A)$ (resp. $\tau:\ T^{II}\to\Delta(B)$). The current payoff is $\tilde{I\!P}[W]$ and given a pair $(\sigma,\tau)$, the next state is $\tilde{I\!P}^{\sigma,\tau}$ (deterministic transition). Obviously, the definition of a game implies the full specification of the information available to the players at each stage. For the moment let us assume that his current type is the only information available to a player. Our goal is to show the existence of a "value" which satisfies a recursive equation. In the discounted game case, we obtain a fixed-point equation analogous to (5). Later on, we will enrich the structure of $\tilde\Gamma(\tilde{I\!P})$ (Section 6) without affecting the fixed-point equation.

**Proposition 3** *The game $\tilde\Gamma_T(\tilde{I\!P})$ (resp. $\tilde\Gamma_\lambda(\tilde{I\!P})$), with initial state $\tilde{I\!P}$, has a value $\tilde v_T(\tilde{I\!P})$ (resp. $\tilde v_\lambda(\tilde{I\!P})$) that is the same as $v_T(\tilde{I\!P})$ (resp. $v_\lambda(\tilde{I\!P})$), hence as $v_T(I\!P)$ (resp. $v_\lambda(I\!P)$) as well.*

**Proof.** When $T=0$, the proposition is clearly true. Suppose that the proposition is true for $T-1$, $T>1$. One derives from (4) that:

$$\begin{aligned} \frac{1}{T}\tilde{I\!P}[W]+\frac{T-1}{T}\min_{\tau_1}\max_{\sigma_1}\tilde v_{T-1}(\tilde{I\!P}^{\sigma_1,\tau_1}) &= \frac{1}{T}\tilde{I\!P}[W] \\ &+ \frac{T-1}{T}\max_{\sigma_1}\min_{\tau_1}\tilde v_{T-1}(\tilde{I\!P}^{\sigma_1,\tau_1}). \end{aligned}$$

This is enough to guarantee that the value $\tilde{v}_T(\tilde{I\!\!P})$ exists and satisfies the same recursive equation as $v_T(\tilde{I\!\!P})$. Therefore, we have $\tilde{v}_T(\tilde{I\!\!P}) = v_T(\tilde{I\!\!P})$.

To prove the statement about the discounted value, introduce as above the truncated discounted game. ∎

## 6. Player I Has More Information Than Player II

In this section we restrict ourselves to the case where player I always has more information than player II. In other words, the signal received by player I includes the signal received by player II. This will have among other consequences the fact that one can use the auxiliary game on the Universal Belief Space to study uniform properties.

Clearly, by (1), any consistent probability measure $\tilde{I\!\!P}$ is spanned by finitely many types $t^{II}$. Notice that we restrict ourselves to considering types $t^{II}$ of player II inducing a marginal probability measure with finite support on $T^I$ such that any type $t^I$ in the support includes $t^{II}$. In other words, the marginal probability measure induced by $t^I$ on $T^{II}$ is $\delta_{t^{II}}$ ("Dirac" measure).

What can we say about the probability measure induced by such a type $t^{II}$ on $\tilde{\Omega}$ ?

**Lemma 1** *The probability measure induced on $\tilde{\Omega}$ by $t^{II}$ is consistent.*

**Proof.** Observe that

$$\begin{aligned}
\tilde{I\!\!P}[s, t^I | t^{II}] &= \frac{\tilde{I\!\!P}[s, t^I, t^{II}]}{\tilde{I\!\!P}[t^{II}]} \\
&= \frac{\tilde{I\!\!P}[s, t^I]}{\tilde{I\!\!P}[t^{II}]} \\
&= \frac{\tilde{I\!\!P}[s|t^I] \times \tilde{I\!\!P}[t^I]}{\tilde{I\!\!P}[t^{II}]} \\
&= \tilde{I\!\!P}[s|t^I] \times \frac{\tilde{I\!\!P}[t^I, t^{II}]}{\tilde{I\!\!P}[t^{II}]} \\
&= \tilde{I\!\!P}[s|t^I] \times \tilde{I\!\!P}[t^I|t^{II}].
\end{aligned}$$

By application of (1), we obtain:

$$t^{II}[s, t^I] = t^{II}[t^I] \times t^I[s] = t^{II}[t^I] \times t^I[s, t^{II}].$$

This means that $t^{II}$ induces a consistent probability measure on $\tilde{\Omega}$. ∎

We shall slightly modify the structure of $\tilde{\Gamma}(\tilde{I\!\!P})$. Following Lemma 1 and Proposition 2 the value $\tilde{v}_\lambda(\tilde{I\!\!P})$ of $\tilde{\Gamma}_\lambda(\tilde{I\!\!P})$ can be decomposed as $\sum_{t^{II}} \tilde{I\!\!P}[t^{II}] \times \tilde{v}_\lambda(t^{II})$ where we identify $t^{II}$ with the corresponding consistent probability measure induced on $\tilde{\Omega}$. We thus choose as new state space the set of types $t^{II}$ of player II. Let us reformulate the recursive equation, derived from (5). If the initial type of player II is $t^{II}$, if player I uses a tuple $\alpha$ of mixed action $\alpha(t^I) \in \Delta(A)$ for any $t^I$ in the support of $t^{II}$ and if player II selects the pure action $b \in B$, then with probability, say $p(t^{II}, \alpha, b)[t'^{II}]$, the next type of player II is $t'^{II}$. We obtain:

$$\tilde{v}_\lambda(t^{II}) = \lambda t^{II}[W] + (1-\lambda) \min_{\beta \in \Delta(B)} \max_{\alpha} \sum_{t'^{II}} \beta[b] \times p(t^{II}, \alpha, b)[t'^{II}] \times \tilde{v}_\lambda(t'^{II}). \tag{6}$$

We define a new game $\hat{\Gamma}(\tilde{I\!P})$ that unfolds as follows: an initial state $t^{II}$ is chosen according to $\tilde{I\!P}$ and after that a play proceeds as in a stochastic game (the current state being known to both players). For simplicity, the type of player II (state) at stage $t$ is denoted by $\mu^t$. Obviously, one has $\tilde{v}_\lambda(\mu^1) = \hat{v}_\lambda(\mu^1)$. In the theory of stochastic games [1], a fixed-point equation identical to (6) is one of the sufficient ingredients to prove the existence of the uniform value. In addition to (6), we shall request that, for any initial state $\mu^1$, $\lim_{\lambda\to 1} \hat{v}_\lambda(\mu^1) = \hat{v}(\mu^1)$ exists. Observe that, if such is the case, $\hat{v}(\tilde{I\!P}) = \sum_\mu \tilde{I\!P}[\mu] \times \hat{v}(\mu)$ is the natural candidate for the value of $\hat{\Gamma}(\tilde{I\!P})$ (considering the asymptotic average payoff). However, we should be careful to specify the information available to the players in addition to the current state $\mu^t$ at stage $t$. For an argument similar to that of [1] to hold, one should assume that each player is informed of the actions of his opponent.

A crucial idea of the proof to come (Proposition 4) is to implement $\epsilon$-optimal strategies for $\hat{\Gamma}(\tilde{I\!P})$ into $\Gamma(I\!P)$. Remember that in $\Gamma(I\!P)$, the signal of player I contains the signal of player II. This implies that the previous action of player II is known to player I (since the signal of player II contains his previous action). Therefore, in $\hat{\Gamma}(\tilde{I\!P})$, we assume that at stage $t+1$ player I is informed of the action $b_t \in B$ chosen by player II at stage $t$. The information received by player II in $\hat{\Gamma}(\tilde{I\!P})$ is not as straightforward. Nothing tells us that player II knows the previous action of player I in $\Gamma(I\!P)$. Looking carefully at what we want to prove (Proposition 4), the existence of the max min of $\Gamma(I\!P)$, observe that it is as if player II knew the strategy of player I. This has the following consequence for the information provided to player II in $\hat{\Gamma}(\tilde{I\!P})$.

A pure action of player I in any state $\mu^t$ is a tuple of actions in $A$. Only actions corresponding to types $t^I$ of the support of $\mu^t$ need to be specified. There is no loss of generality in considering mixed actions that are a tuple of mixed $\alpha \in \Delta(A)$. Based on what we said above, it is convenient to assume that, at any stage, player II is informed of the previous mixed action chosen by player I.

To recapitulate, in $\hat{\Gamma}(\tilde{I\!P})$, a strategy $\hat{\sigma}$ of player I is a family of $\hat{\sigma}_t$ $(t \geq 1)$ specifying a mixed action $\alpha_{t^I}(\mu^1, \ldots, \mu^{t-1}, b_1, \ldots, b_{t-1},) \in \Delta(A)$ for any $t^I$ in the support of $\mu^t$. The previous actions $\alpha_1, \ldots \alpha_{t-1}$ of player I do not appear explicitly, because if they did, then they could be replaced by what they are, a function of past states and past actions of player II. A strategy $\hat{\tau}$ of player II is a family of $\hat{\tau}_t$ which specifies

$$\beta_{\mu^t}(\mu^1, \ldots, \mu^{t-1}, \alpha_1, \ldots, \alpha_{t-1}, b_1, \ldots, b_{t-1}) \in \Delta(B).$$

From now on, $\gamma_T(\sigma, \tau)$ (resp. $\hat{\gamma}_T(\hat{\sigma}, \hat{\tau})$) denotes the expected average payoff associated with the pair $(\sigma, \tau)$ (resp. $(\hat{\sigma}, \hat{\tau})$) in $\Gamma_T(I\!P)$ (resp. $\hat{\Gamma}_T(\tilde{I\!P})$).

We assume that we can prove the existence of the maxmin $\hat{w}(\tilde{I\!P})$ of $\hat{\Gamma}(\tilde{I\!P})$. Formally, this is expressed in the form of the next statement.

**Assumption 1** *For any $\epsilon > 0$, player I has a strategy $\hat{\sigma}^\epsilon$ in $\hat{\Gamma}$ that guarantees $\hat{w}(\tilde{I\!P})$ up to $\epsilon$. In other words, there exists $T_0 > 0$ such that for any $T \geq T_0$ and any strategy $\hat{\tau}$ of player II*

$$\hat{\gamma}_T(\hat{\sigma}^\epsilon, \hat{\tau}) \geq \hat{w}(\tilde{I\!P}) - \epsilon. \tag{7}$$

*On the other hand, for any strategy $\hat{\sigma}$ of player I and any $\epsilon > 0$, there exists an $\epsilon$-best reply $\hat{\tau}(\hat{\sigma})$ for player II. This means that*

$$\limsup_{T\to\infty} \hat{\gamma}_T(\hat{\sigma}, \hat{\tau}(\hat{\sigma})) \leq \hat{w}(\tilde{I\!P}) + \epsilon. \tag{8}$$

A stronger requirement (satisfied in particular if $\hat{\Gamma}(\tilde{I\!P})$ has a uniform value) would be to assume in addition to (7) a stronger condition of uniformity with respect to the strategies of player I:

**Assumption 2** *For any $\varepsilon > 0$, there exists $T_0$ such that, given any strategy $\hat{\sigma}$ of player I, player II has an $\epsilon$-best reply $\hat{\tau}(\hat{\sigma})$ satisfying*

$$\hat{\gamma}_T(\hat{\sigma}, \hat{\tau}(\hat{\sigma})) \leq \hat{w}(\tilde{I\!P}) + \epsilon \tag{9}$$

*for any $T \geq T_0$.*

In the final part of the present chapter we shall deal with the proof of the following proposition.

**Proposition 4** *If assumption 1 is satisfied, then $\hat{w}(\tilde{I\!P})$ is the* max min *of $\Gamma(I\!P)$. Let us denote it by $w(I\!P)$.*
*Under assumption 2 one has in addition:*

$$\begin{aligned}\hat{w}(\tilde{I\!P}) &= \lim_{\lambda\to 1} \tilde{v}_\lambda(\tilde{I\!P}) \\ &= \lim_{\lambda\to 1} v_\lambda(I\!P).\end{aligned}$$

**Proof.** The definition of the max min of $\Gamma(I\!\!P)$ involves as above two conditions. For any $\epsilon > 0$:

*(i)* There exists a strategy $\sigma^\epsilon$ of player I and $T_0$ such that for any $T > T_0$ and any strategy $\tau$ of player II, we have

$$\gamma_T(\sigma^\epsilon, \tau) \geq v(I\!\!P) - \epsilon. \tag{10}$$

*(ii)* Against any strategy $\sigma$ of player I, there exists a strategy $\tau^\epsilon(\sigma)$ such that

$$\limsup_{T\to\infty} \ \gamma_T(\sigma, \tau^\epsilon(\sigma)) \leq v(I\!\!P) + \epsilon. \tag{11}$$

In order to establish $(i)$ and $(ii)$, some of our arguments will be reminiscent of those used for proving Proposition 1.

$(i)$ To start with, let us take any strategy $\hat{\sigma} = \hat{\sigma}_1, \hat{\sigma}_2, \dots$ of player I in $\hat{\Gamma}(\tilde{I\!\!P})$. Let us denote by $\sigma = \sigma_1, \sigma_2, \dots$ a strategy of player I in $\Gamma(I\!\!P)$ obtained as follows:

- After stage 0 (choice of the current state) in $\Gamma(I\!\!P)$, player I can calculate his type $t_1^I$ as well as player II's type $t_1^{II} = \mu^1$. Next, he can implement the appropriate component of $\hat{\sigma}_1$.
- Assume that player I has calculated a family of past and current types $t_1^I, \dots, t_\ell^I$ for himself. Included in those types, he obtains a family $\mu^1, \mu^2, \dots, \mu^\ell$ ($\ell \geq 1$) of past and current types for player II. Since player I knows the past actions $b_1, \dots, b_{\ell-1}$ of player II, player I implements $\hat{\sigma}_\ell$ in order to play in $\Gamma(I\!\!P)$ at stage $\ell$, thus obtaining $\sigma_\ell$.
  The next type $t_{\ell+1}^I$ is obtained as follows: $\mu^\ell$ induces a probability measure on $S$ (which gives the "law" of the current state $s_\ell$) as well as a probability measure with finite support on $T^I$. Given the current action $b_\ell$, player I can calculate the "law" of the next state $s_{\ell+1}$. From there, he can calculate his next type $t_{\ell+1}^I$ depending on his new signal.

For any $T > 0$, we shall construct a pure strategy $\hat{\tau} = \hat{\tau}(T)$ for player II in $\hat{\Gamma}(\tilde{I\!\!P})$ such that, at stage $t$, $\hat{\tau}_t$ depends on the past and current states $\mu^1, \dots, \mu^t$. Incorporating the knowledge of $\sigma$ (since it is fixed) player II can implement $\hat{\tau}$ in $\Gamma(I\!\!P)$. We shall show that $\tau = \tau(T)$ is a best reply against $\sigma$ in $\Gamma_T(I\!\!P)$. Observe that the following holds:

$$\gamma_T(\sigma, \tau) = \hat{\gamma}_T(\hat{\sigma}, \hat{\tau}). \tag{12}$$

For this construction, let us proceed recursively with respect to $T$ and for any initial probability distribution $I\!\!P$. When $T = 0$ it is obvious.

At stage 0 the state is chosen according to $I\!\!P$. Both players are informed of the element of their respective partition where the state falls. Since player I uses $\sigma$, his first mixed action $\sigma_1$ depends on his type.

Right after stage 1, if the type of player II is $\mu^1$ and if his action is $b_1$, then the players face $\Gamma_{T-1}(I\!P_{\mu^1,b_1})$ where the initial probability distribution $I\!P_{\mu^1,b_1}$ can be calculated using $\sigma_1$ and $b_1$. The truncated strategy $\sigma(\mu^1, b_1)$ of player I starting from stage 1 can be seen as the implementation of some strategy $\hat{\sigma}(\mu^1, b_1)$ of $\hat{\Gamma}(\tilde{I\!P}_{\mu^1,b_1})$. By application of the recursive hypothesis to the game $\Gamma_{T-1}(I\!P_{\mu^1,b_1})$, we obtain a pure best reply $\tau(\mu^1, b_1)$ that is the implementation of a strategy $\hat{\tau}(\mu^1, b_1)$ of the form we want. Clearly, one can obtain a pure best reply $\tau$ against $\sigma$ in $\Gamma_T(I\!P)$ by implementation of some $\hat{\tau}$ with the desired form, by selecting an optimal $b_1 : T^{II} \to B$ which player II selects at the first stage.

In particular, if applied to the strategy $\hat{\sigma}^\epsilon$ of player I in $\hat{\Gamma}$ introduced in assumption 1, (12) and (7) will imply that for $T \geq T_0$

$$\gamma_T(\sigma^\epsilon, \tau) \geq w(I\!P) - \epsilon,$$

that is (10).

*(ii)* The arguments that we are going to use here are quite similar to those of part *(i)*.

Let $\hat{\tau}$ be a strategy of player II in $\hat{\Gamma}(\tilde{I\!P})$. It may be impossible to directly implement $\hat{\tau}$ in $\Gamma(I\!P)$ because it normally involves the past mixed actions of player I which are unknown to player II. However, we deal with a reply and therefore assume that player II knows $\sigma$.

So let $\sigma$ be any strategy of player I in $\Gamma(I\!P)$. We aim at proving that there exists a strategy $\hat{\sigma} = \hat{\sigma}(\sigma)$ of player I in $\hat{\Gamma}(\tilde{I\!P})$ such that given $\hat{\tau}$ in $\hat{\Gamma}(\tilde{I\!P})$ one can "mimic" $(\hat{\sigma}, \hat{\tau})$ in $\Gamma(I\!P)$ in the sense that there exists $\tau = \tau(\sigma, \hat{\tau})$ which satisfies

$$\gamma_T(\sigma, \tau(\sigma, \hat{\tau})) = \hat{\gamma}_T(\hat{\sigma}(\sigma), \hat{\tau}). \tag{13}$$

The construction of $\hat{\sigma}$ is done stage by stage. Recall that $\sigma$ is a sequence of mappings $\sigma_t$, $t = 1, \ldots, \infty$, prescribing a mixed action depending on the past and present signals. At stage one, average $\sigma_1$ with respect to the type of player I to obtain $\hat{\sigma}_1$.

Let us assume that $\hat{\sigma}_1, \ldots, \hat{\sigma}_{t-1}$ corresponding to the first $t-1$ stages have been constructed. Once again, let us average $\sigma_t$ with respect to the current type of player I, but conditionally to the fact that the previous types of player II were $\mu^1, \ldots, \mu^{t-1}$ and his previous actions were $b_1, b_2, \ldots, b_{t-1}$. The crucial point is that this information is independent of the strategy of player II. Let us denote by $\hat{\sigma}_t$ the corresponding mixed action for stage $t$ in $\hat{\Gamma}(\tilde{I\!P})$.

As for player II at stage one he can implement $\hat{\tau}_1$ in $\Gamma(I\!P)$, thus obtaining $\tau_1$, since it depends only on the current state $\mu^1$. Knowing $\sigma$, he can compute

$\hat{\sigma}$, hence the sequence of types $\mu^1, \ldots, \mu^t$ as well. Using $\hat{\tau}$ this defines a strategy $\tau$ in $\Gamma(\mathbb{P})$.

To finish the proof, i.e., to construct an $\epsilon$-best reply against $\sigma$, let us apply the previous construction to an $\epsilon$-best reply $\hat{\tau}(\hat{\sigma}(\sigma))$ of player II in $\hat{\Gamma}(\tilde{\mathbb{P}})$ satisfying (8), hence (11) by (13).

Clearly under assumption 6.2 $w(\mathbb{P})$ is the limit of $v_T(\mathbb{P})$ when $T \to \infty$ (hence $w(\mathbb{P})$ is also the limit of $v_\lambda(\mathbb{P})$ when $\lambda \to 1$). ∎

**ACKNOWLEDGEMENTS.** The writing of this chapter has been supported by a grant from the Australian Research Council no. A69703141. The author wishes also to thank Jean-François Mertens and Sylvain Sorin for their comments.

## References

1. Mertens, J.-F. and Neyman, A. (1981) Stochastic games, *International Journal of Game Theory* **10**, 53–56.
2. Mertens, J.-F. and Zamir, S. (1985) Formulation of Bayesian analysis for games with incomplete information, *International Journal of Game Theory* **14**, 1–29.
3. Mertens, J.-F. (1990) Repeated games, in T. Ichiishi, A. Neyman and Y. Tauman (eds.), *Game Theory and Applications*, Academic Press, San Diego, CA, pp. 77–130.
4. Mertens, J.-F., Sorin, S. and Zamir, S. (1994) Repeated games, CORE Discussion Papers 9420, 9421, 9422, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.