Johannes Hörner
Department of Economics
Yale University
johannes.horner@yale.edu

Pauli Murto
Aalto University, visiting Yale
pauli.murto@aalto.fi

Econ 520a Advanced Microeconomic Theory, I

Fall 2015
Tu., Th., 2:30–3:50pm
HH28, Rm. 108

This course will be divided into two parts. The first is on the theory and methods of dynamic (including repeated) games; the second, on social learning and information aggregation.

Grading

There will be a written term paper to be turned in at the end of the semester. Students are **strongly** encouraged to come up with their own idea for a paper (obviously related to the course), and discuss it with either of us early enough in the semester. The purpose is to engage into some research, whether its goals are modest or not, rather than in a survey of the literature. Suggestions in terms of problems will be offered (but those won't be as easy as when you do your own model!).

Part I (J. Hörner)

We will provide an overview of dynamic games (repeated games and stochastic games, with or without private information). Within the context of discrete-time dynamic games with discounting (both qualifications to be understood throughout), we will survey all topics, with a focus on recent advances obtained in the last twenty years.

Topics will include, in the following order:

1. Repeated games with Imperfect Monitoring (RGIM)

   (a) Perfect Monitoring
   (b) Imperfect Public Monitoring
   (c) Imperfect Private Monitoring
2. Repeated Games with Incomplete Information (RGII)
   (a) Symmetric Learning
   (b) Private Information
      i. Strategic Types (Reputations)
      ii. General Payoff Types
3. Stochastic Games
4. Repeated Bayesian Games

The lectures will be based on lecture notes, which supplement readings of relevant papers. An extensive bibliography will be provided at the end of each set of lecture notes. The focus will be on recent results and open problems.

Nonetheless, there are two excellent textbooks one might like to consult for repeated games and related topics, namely:

Mailath, G., and L. Samuelson (2006). *Repeated Games and Reputations*, Oxford University Press, Oxford.

Mertens, J.-F., S. Sorin and S. Zamir (2015). *Repeated Games*, Cambridge University Press.

Part II (P. Murto)

We will focus on models of learning. In many applications, agents have private information about an unknown state variable that is common and payoff relevant to all agents. Firms have different opinions about market demand or the viability of new production technologies, buyers have different opinions about qualities of goods, traders have different opinions about asset values, etc. In such situations, agents may learn from each other through their interaction, and this learning feeds back into their behavior. We will study various models of such endogenous learning to figure out how learning influences agents' behavior and equilibrium outcomes. A central question is how dispersed information aggregates through agents' interaction.

We will first consider models where players learn by observing other players' behavior or outcome of their actions. Then we will move on to analyze markets and consider how prices aggregate dispersed information. The methodological starting point throughout is that the agents have a common prior and learning is through Bayesian updating.

Topics will include broadly:

1. Models of Observational Learning
2. Experimentation and Strategic Interactions
3. Information Aggregation in Large Auctions
4. Trading and Learning: Models of Market Microstructure

The lectures will be based on research articles. An extensive bibliography will be provided for each leacture. The focus will be on recent literature.

There are several textbooks that touch upon the topics of this course from somewhat different perspectives, e.g.:

Chamley, C. (2004). *Rational Herds. Economic Models of Social Learning*, Cambridge University Press.

Vives, X. (2010). *Information and Learning in Markets: The Impact of Market Microstructure*, Princeton University Press.

Repeated Games, Part I: Perfect Monitoring

Lecture Notes, Yale 2015, Johannes Hörner

September 3, 2015

This first set of notes introduces some background material that every student should be familiar with when the class starts.

I Introduction

Most interactions involve a dynamic element. Let us consider for now the prisoner's dilemma, for which the payoff matrix is given by

	L	R
T	$1,1$	$-L,1+G$
B	$1+G,-L$	$0,0$

Here, $G, L \in \mathbb{R}_+$ measure respectively the additional gain of one player when he defects while his opponent cooperates, and the loss of the latter player. It is also customary to assume that $G-L<1$, so that the sum of payoffs when both players cooperate (2) exceeds the sum when one defects $(1+G-L)$. The set of all payoff vectors that can be achieved in this game is represented in Figure 1.

The prisoner's dilemma is the normal form of a two-player, simultaneous-action game. What happens if this simultaneous-move game is repeated twice, or more generally, T times, and after each play, both players can observe what the opponent has done? How about if it is repeated infinitely often? How about if we consider other simultaneous-action games? This is the topic of **repeated games**, introduced in these notes.

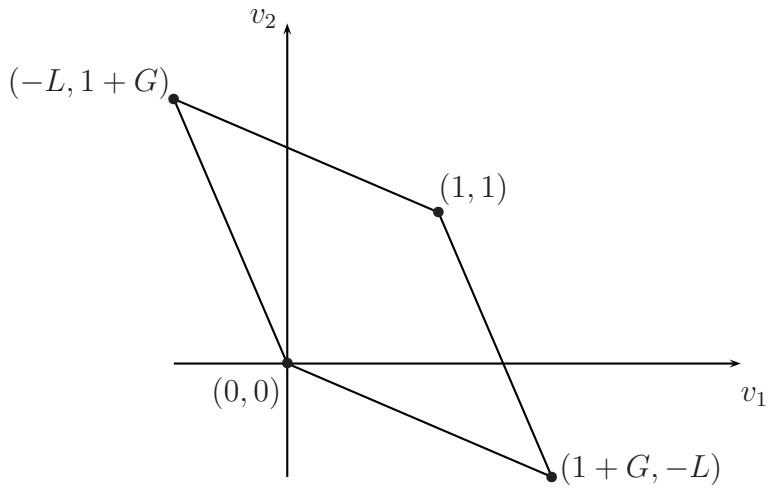


Figure 1: Feasible payoffs in the prisoner's dilemma

II Notation

The building block of the repeated game is the normal form game corresponding to each single interaction, which is referred to as the **stage game**. Here, we consider a finite stage game, and to distinguish strategies in the repeated game from those in the stage game, we shall refer to the choices in the stage game as **actions**. A stage game, therefore, is a triple (N, A, u), where $N = \{1, \ldots, n\}$ is the finite set of players, $A := \times_{i=1,\ldots,n} A_i$ is the Cartesian product of the finite action set of each player $i = 1, \ldots, n$, and $u : A \to \mathbb{R}^n$ is the utility vector $u(a) := (u_1(a), \ldots, u_n(a))$ that specifies the utility for each player for any given (pure) action profile $a \in A$. A mixed action for a player is an element of $\triangle A_i$, denoted α_i. We write $\alpha_i(a_i)$ for the probability assigned by the mixed action α_i to the pure action a_i. We shall consider the mixed extension of the function u. That is, we enlarge its domain to the set of mixed action profiles $\alpha \in \triangle A$ by setting, for each $i = 1, \ldots, n$,

$$u_i(\alpha) := \sum_{a \in A} \alpha(a) u_i(a),$$

where $\alpha(a)$ is the probability assigned to a by $\alpha \in \triangle A$. This (expected) utility is also referred to as the **reward**, instead of payoff, for reasons that will become clear. The set of feasible rewards is defined as

$$V := \text{co}\,\{u(a) : a \in A\}.$$

That is, it is defined as the set of rewards that can be achieved by convex combinations of payoffs from pure action profiles. Plainly, this is the same set as $\{u(\alpha) : \alpha \in \triangle A\}$. Note, however, that some payoffs in this set might require players to play correlated actions, because in general the set of independent mixed actions is a subset of the mixed action profiles, i.e. $\times_{i=1,\ldots,n} \triangle A_i \subsetneq \triangle A$. Therefore, it is customary in repeated games to assume that players have access to a public correlation device in every period, which allows them to replicate the play of correlated action profiles, without modeling it explicitly. We shall assume so whenever convenient. All results that are stated here can be proved without reference to a public randomization device, but the proofs become somewhat more complex.

The set V has been represented in Figure 1 above in the case of the prisoner's dilemma.

To define the repeated game, we must now specify the players' information. Periods are indexed by $t = 0, 1, \ldots, T$. The parameter T, called the **horizon** of the game, could be finite, in which case the game is said to be a **finitely repeated game**, or infinite, in which case this is an **infinitely repeated game**. Observe that, since we let time start at 0, if $T < \infty$, the game is actually repeated $T + 1$ times.

In this first set of notes, we assume that all players observe all realized actions at the end of the period (Note: this means that, if one player uses a mixed action α_i, his opponents will not observe the lottery itself, but only the realized action a_i). We write a^t for the action profile that is realized in period t. That is, $a^t = (a_1^t, \ldots, a_n^t)$ are the actions actually played in period t. A player's information set at the beginning of period t, therefore, is a vector $(a^0, a^1, \ldots, a^{t-1}) \in (A)^t$, for $t \geq 1$. We define the set of **histories of length** t as the set $H^t := (A)^t$, for $t \geq 1$, and denote its elements by h^t. This does not quite address the initial information set, and so, by convention, we set $H^0 := \{\varnothing\}$, and we interpret its single element h^0 as the initial information set. The set of **histories** is defined as $H := \cup_{t=0,\ldots,T} H^t$, with generic element $h \in H$.

A pure strategy for player i, then, is a map $s_i : H \to A_i$ that specifies for each history, what action to play. The set of pure strategies is denoted S_i. A behavior strategy is a function $\sigma_i : H \to \triangle A_i$, and the set of all behavior strategies is denoted Σ_i. We let, as usual, $S := \times_{i=1,\ldots,n} S_i$, $\Sigma := \times_{i=1,\ldots,n} \Sigma_i$, and write s, resp. σ, for a pure, resp. behavior, strategy profile. The statement of Kuhn's theorem, establishing the realization-equivalence between mixed and behavior strategies, applies to repeated games as well.

A strategy profile $\sigma \in \Sigma$ generates a distribution over terminal nodes, that is, over histories H^{T+1}. Again, it is clear what is meant when T is finite: if Z is finite, defining the probability space is simple. Let Z be the set of outcomes, and the set of events is the set $\mathcal{P}(Z)$ of all subsets

of Z. If $T = \infty$, the set of outcomes Z is the set of infinite sequences $(a^0, a^1, \ldots) \in A^{\mathbb{N}_0}$, and so it is infinite as well, and defining the set of events introduces technical details, which it is best to ignore.[1]

Note that every period of play begins a proper subgame, and since actions are simultaneous in the stage games, these are the only proper subgames, a fact that we must keep in mind when applying subgame-perfection.

When T is finite, we shall evaluate outcomes according to the average of the sum of rewards. That is, we define the function $v_i : H^{T+1} \to \mathbb{R}$ by setting, for all $h^{T+1} = (a^0, \ldots, a^T)$,

$$v_i(h^{T+1}) := (T+1)^{-1} \sum_{t=0}^{T} u_i(a^t).$$

Since a strategy profile generates a probability distribution, we can also extend the domain of the function v_i to all strategies $\sigma \in \Sigma$, by letting

$$v_i(\sigma) := (T+1)^{-1} \mathbb{E}_\sigma \left[\sum_{t=0}^{T} u_i(a^t) \right],$$

where the operator $\mathbb{E}_\sigma$ refers to the expectations under the probability distribution over terminal histories generated by σ. This is player i's **payoff** in the finitely repeated game.

There are several alternative ways of defining payoffs in the infinitely repeated game. We will focus on the case in which players discount future rewards using a common discount factor $\delta \in [0, 1)$. In this game, player i's payoff given some infinite history $h^\infty = (a^0, a^1, \ldots)$, is

$$v_i(h^\infty) := (1-\delta) \sum_{t=0}^{\infty} \delta^t u_i(a^t).$$

The normalization constant $(1-\delta)$ that appears in front is a way to make payoffs in the repeated game comparable to rewards in the stage game. Indeed, if a player receives a reward of 1 in every period, the unnormalized discounted sum is equal to $1 + \delta + \delta^2 + \cdots = 1/(1-\delta)$. Once it is normalized then, it is equal to 1 as well. Therefore, when considering the normalized discounted

[1] Formally, we consider the σ-algebra generated by the cylinders $\{h^t \times A^\infty : h^t \in H\}$ (where A^∞ are infinite sequences of action profiles), and use as probability distribution given σ the unique extension of the family of consistent probability distributions over finite histories that σ defines in the obvious way.

sum rather than the unnormalized one, the set of payoffs that are feasible in the repeated game becomes the same as the set of feasible rewards in the stage game, allowing for meaningful comparisons.

Since a strategy profile generates a probability distribution over infinite histories, we extend here as well the domain of the function v_i to all strategies $\sigma \in \Sigma$, by letting

$$v_i(\sigma) := (1-\delta)\mathbb{E}_\sigma \left[\sum_{t=0}^{\infty} \delta^t u_i(a^t) \right],$$

where the operator $\mathbb{E}_\sigma$ refers to the expectations under the probability distribution over infinite histories that is generated by σ. This is player i's **payoff** in the infinitely repeated game.

Given a stage game (A, u), define player i's **minmax payoff** as

$$\underline{v}_i := \min_{\alpha_{-i} \in \times_{j \neq i} \triangle A_j} \max_{a_i \in A_i} u_i(a_i, \alpha_{-i}).$$

This is the lowest reward player i's opponents can hold him down to by any independent choice of actions α_j, provided that player i correctly foresees α_{-i} and plays a best-reply to it. Observe that player i always has a best-reply in the set of pure strategies, and therefore restricting him to pure actions does not affect his minmax payoff. Let $\underline{\alpha}^i_{-i} \in \times_{j \neq i} \triangle A_j$ be a strategy for player i's opponents that attains the minimum in the definition. Such an action profile is called the **minmax profile**. We also write $\underline{\alpha}^i_i \in \triangle A_i$ for any action that is a best-reply for player i given $\underline{\alpha}^i_{-i}$ (if there are several such actions $\underline{\alpha}^i_{-i}$ or $\underline{\alpha}^i_i$, fix one of each in what follows).

If $v_i \geq \underline{v}_i$, the payoff $v_i \in \mathbb{R}$ is **individually rational** for player i, and if $v \geq \underline{v}_i$ for all $i = 1, \ldots, n$, the payoff vector v is **individually rational**. A payoff is strictly individual rational if all the inequalities hold strictly. The set of feasible and strictly individually rational payoffs is thus defined as

$$\underline{V} := \{v \in V | \forall i = 1, \ldots, n : v_i > \underline{v}_i\}.$$

The set $\underline{V}$ is represented below in the case of the prisoner's dilemma.

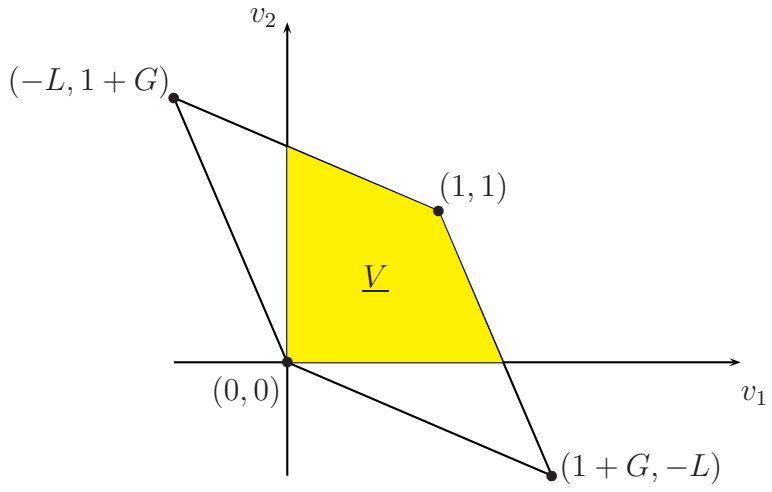


Figure 1: Feasible and strictly individual rational payoffs in the prisoner's dilemma

Given $T < \infty$, the finitely repeated game is denoted G^T, while the infinitely repeated game with discount factor δ is denoted G^δ. A **Nash equilibrium**, then, is a strategy profile $\sigma \in \Sigma$ such that, for all $i = 1, \ldots, n$, and all $\sigma_i' \in \Sigma$, $u_i(\sigma) \geq u_i(\sigma_i', \sigma_{-i})$. To define a **subgame-perfect Nash equilibrium** (SPNE), recall that, since any history $h \in H$ defines a subgame, $\sigma|_h$ denotes the restriction of $\sigma \in \Sigma$ to the subgame beginning at h. Therefore, a strategy profile $\sigma \in \Sigma$ is a subgame-perfect Nash equilibrium of the repeated game if for all histories $h \in H$, $\sigma|_h$ is a Nash equilibrium of the subgame. Observe that a subgame of a finitely repeated game is a finitely repeated game with a shorter horizon, while a subgame of the infinitely repeated game is an infinitely repeated game itself.

III Preliminary results

A **one-shot deviation** from a strategy $\sigma_i \in \Sigma_i$ by player i at $h \in H$ is a strategy σ_i' such that $\sigma_i'(h') = \sigma_i(h')$ for all $h' \neq h, h' \in H$, but $\sigma_i'(h) \neq \sigma_i(h)$. It is profitable, given σ_{-i}, if

$$v_i(\sigma_i'|_h, \sigma_{-i}|_h) > v_i(\sigma|_h).$$

The one-shot deviation principle, stated below, holds for all finite games (not just repeated ones), as well as infinitely repeated games, as long as payoffs are discounted. It is due to David Blackwell (1965).

Lemma 1 (One-shot deviation principle.) *A strategy profile $\sigma \in \Sigma$ is a subgame-perfect Nash equilibrium of G^δ if and only if no player has a profitable one-shot deviation.*

Proof: One direction is obvious: if σ is a SPNE, then there are no profitable deviations, whether one-shot or not. In the other direction, we must show that, whenever a profitable deviation exists, we can define a profitable one-shot deviation. Suppose then that σ is not a SPNE, that is, there exists $i = 1, \ldots, n$, $h^t \in H$, $\sigma_i' \in \Sigma_i$ such that

$$v_i(\sigma_i'|_{h^t}, \sigma_{-i}|_{h^t}) - v_i(\sigma|_{h^t}) > 0.$$

Because payoffs are discounted, it follows that there exists $T \in \mathbb{N}$ such that, defining σ_i'' by $\sigma_i''(h^\tau) = \sigma_i(h^\tau)$ for all $h^\tau \in H, \tau \geq t + T$, $\sigma_i''(h^\tau) = \sigma_i'(h^\tau)$ otherwise, and

$$v_i(\sigma_i''|_{h^t}, \sigma_{-i}|_{h^t}) - v_i(\sigma|_{h^t}) > 0.$$

That is, σ_i'' is a profitable deviation that only differs from σ_i at finitely many histories. We now proceed by induction. Consider all histories $h \in H^{t+T-1}$; either $v_i(\sigma_i''|_h, \sigma_{-i}|_h) > v_i(\sigma|_h)$ for some history h in this set, in which case we define σ_i''' as the one-shot deviation from σ_i at h, with $\sigma_i'''(h) = \sigma_i''(h)$, and this is a one-shot deviation; otherwise, we define σ_i''' by $\sigma_i'''(h^\tau) = \sigma_i(h^\tau)$ for all $h^\tau \in H, \tau \geq t + T - 1$, $\sigma_i'''(h^\tau) = \sigma_i'(h^\tau)$ otherwise, and again σ_i''' is a profitable deviation. Repeat. □

This is quite useful, because the set of strategies that a player could use from history h on, instead of $\sigma|_h$, is infinite, so it is convenient to have only a few alternatives to check.

Player i's minmax payoff is often referred to as player i's **reservation utility**. The reason for this name is the following result.

Lemma 2 *Player i's payoff is at least $\underline{v}_i$ in any Nash (and therefore, also, any subgame-perfect Nash) equilibrium of the (finitely or infinitely) repeated game, independently of $\delta \in [0, 1)$.*

Proof: Observe that player i can always use the following strategy $\sigma_i \in \Sigma_i$ in the repeated game:

$$\forall h \in H : \sigma_i(h) \in \text{argmax}_{a_i \in A_i} u_i(a_i, \sigma_{-i}(h)).$$

That is, the strategy of player i picks in every period some best-reply to the action profile played by players $-i$. This strategy may not be optimal, since it ignores the fact that the way other players play in the future possibly depends on how player i plays today. However, because all

players have the same information at the start of each period t, the probability distribution over the actions of players $-i$ in period t, conditional on player i's information, is the product of independent randomizations by player i's opponents. So in every period, player i guarantees at least the minimum over such mixed actions of the maximum over his best-replies, i.e. he guarantees at least $\underline{v}_i$ in every period, and so secures $(1-\delta)\sum_{t=0}^{\infty}\delta^t\underline{v}_i = \underline{v}_i$ in the repeated game G^δ, and also $(T+1)^{-1}\sum_{t=0}^{T}\underline{v}_i = \underline{v}_i$ in the finitely repeated game G^T. □

IV The finitely repeated game

To fix ideas, let us consider for now the prisoner's dilemma played twice. What are the subgame-perfect Nash equilibria of this game? Consider the second period ($t = 1$): independently of the history $h^1 \in H^1$, the strategy profile $\sigma|_{h^1}$ must be an equilibrium of the subgame, which is simply the game played once. Yet we know that, in the stage game, there is only one Nash equilibrium, in which both players defect. That is, in the second period, independently of the play in the first, in every subgame-perfect Nash equilibrium, players must play (D, D).

Consider now the first period. Players know that, in the second period, independently of their play in the first, they will play (D, D) in the second. Therefore, their action in the first period has no consequence on their reward in the second. Consequently, an optimal action for player i in the first period must be a best-reply in the stage game, given the action played by $-i$ in the first period. That is, players must play a Nash equilibrium of the stage game in the first period as well, which means that the unique SPNE in the twice repeated game is: "play D in both periods" (independently of the history).

Obviously, this reasoning did not rely on the prisoner's dilemma very much. What it relied on was that the stage game has a unique equilibrium. Nor did this reasoning rely on the fact that there were two periods only, rather than three, say. By induction, the proof can be extended to any finite number of periods. We have thus shown the following result.

Theorem 1 *Let G be a stage game that admits a unique Nash equilibrium α. Then, for any finite T, the unique subgame-perfect Nash equilibrium σ in G^T is such that:*

$$\forall i = 1, \ldots, n, \forall h \in H : \sigma_i(h) = \alpha_i.$$

Clearly also, what matters is not that the equilibrium of the stage game is unique, but that all equilibria of the stage game yield the same payoffs to all players. On the other hand, if a

stage game admits multiple Nash equilibria, with distinct payoffs, richer equilibria exist. Let us consider, for instance, the game of chicken:

	S	F
S	$2,2$	$0,3$
F	$3,0$	$-1,-1$

Both $(3,0)$ and $(0,3)$ are Nash equilibrium payoffs of this game. There are simple subgame-perfect Nash equilibria of the twice-repeated game in which, in the first period, the reward $(2,2)$ is received. For instance, consider the following strategy profile. In the first period, play S. In the second period, if $h^1 = (S,S)$ or $h^1 = (F,F)$, and assuming a public randomization device (where the device is flipping a coin), play (S,F) if Heads, and (F,S) if Tails. On the other hand, if $h^1 = (S,F)$, $\sigma_1(h^1) = F,\ \sigma_2(h^1) = S$, while if $h^1 = (F,S)$, $\sigma_1(h^1) = S,\ \sigma_2(h^1) = F$. In words, if a player unilaterally deviates in the first period, we play his least preferred equilibrium in the second period. To check that this is a SPNE, observe that in the second period, a Nash equilibrium of the stage game is always played. In the first period, if a player does not deviate, he receives $2 + (1/2) \times 3 + (1/2) \times 0 = 7/2$. On the other hand, by deviating, he receives $3 + 0 < 7/2$. (Observe that we are using the one-shot deviation principle here, since we assume that, if a player deviates in the first, he follows the equilibrium strategy in the second; observe also that a strategy profile must specify actions after every history, including the history (F,F) that is easy to overlook.)

The theorem stated above holds for subgame-perfect Nash equilibria, which is arguably the most natural solution concept in this environment. Since the proof relies on backward induction, it is easy to see that, in general, it is not valid for Nash equilibria. One important exception is the prisoner's dilemma (even then, the result only applies to the equilibrium outcome, not to the equilibrium strategies).

Lemma 3 *Let G be the two-player prisoner's dilemma. Then, for any finite T, the unique Nash equilibrium outcome of G^T has both players defect in all periods.*

Proof: Recall that, in the prisoner's dilemma, the unique Nash equilibrium payoff of the stage game is 0, and this is also player i's minmax payoff. For the sake of contradiction, let $(a^0, \ldots, a^T)$ be the outcome of a Nash equilibrium in pure strategies $s \in S$ of G^T such that a^t is not equal to (D,D) in some period t (the proof for the case of behavior strategies $\sigma \in \Sigma$ is analogous, but requires more notation). Let t' be the last such period, so that, for some $i = 1,2$, $a_i^{t'} = C$.

Consider the strategy s_i' defined by $s_i'(h^t) = s_i(h^t)$ for all $h^t \in H^t$, $t < t'$, $s_i'(h^t) = D$ for all $h^t \in H^t$, $t \geq t'$. The outcome is the same as that under s up to period $t' - 1$; it yields player i strictly more in period t' (since D is strictly dominant in the stage game), and it yields player i as much (0) in all remaining periods. Therefore, s_i' is a profitable deviation. □

The key of the proof is that, in the prisoner's dilemma, the unique Nash equilibrium payoff coincides with the minmax payoff.

V Infinitely repeated games

Returning to the prisoner's dilemma, what are SPNE of the infinitely-repeated game G^δ? Clearly, playing D in every period, independently of the history, is a SPNE: if this is how player $-i$ plays, player i can do no better than play the same way.

There are, however, other SPNE, provided δ is sufficiently close to one. For instance, consider the following strategy profile $\sigma \in \Sigma$, called **grim-trigger**. At the initial history h^0, $\sigma_i(h^0) = C$, for all $i = 1, \ldots, n$. For all $t \geq 1$, and all $h^t \in H^t$, $\sigma_i(h^t) = C$ if $h^t = ((C, C), \ldots, (C, C))$, that is, if both players have cooperated in all earlier periods. For all other histories h^t, $\sigma_i(h^t) = D$.

Let us verify that this is a SPNE for large enough δ. To do so, we shall extensively use the one-shot deviation principle. Consider first a history $h^t \neq ((C, C), \ldots, (C, C))$, $t \geq 1$; after such a history, both players are supposed to defect in every period. Clearly, this is a SPNE of the subgame (since this is, in fact, as already mentioned, a SPNE of the game itself).

Consider now a history $h^t = ((C, C), \ldots, (C, C))$, or $h^t = h^0$. Both players are supposed to cooperate. If player i cooperates, he receives

$$(1 - \delta)(1 + \delta \times 1 + \delta^2 \times 1 + \cdots) = 1.$$

By deviating once, and then reverting to the equilibrium strategy (which calls both of them to defect, since player i will have defected), he gets

$$(1 - \delta)(1 + G + \delta \times 0 + \delta^2 \times 0 + \cdots) = (1 - \delta)(1 + G).$$

So, if $1 \geq (1 - \delta)(1 + G)$, that is, if $\delta \geq 1 - (1 + G)^{-1}$, it is not a profitable deviation either.

The key was that, if players are sufficiently patient, the gains from a one-shot deviation can be made arbitrarily small relative to what a player might lose after such a deviation.

Can we construct other equilibrium payoffs? Can we generalize the argument to other games? The answer to both questions is yes, and is provided by the so-called **folk theorem**, stated and proved below. (Its statement and proof given below are due to Fudenberg and Maskin (1986)). Given some convex set $B \in \mathbb{R}^n$, the **dimension** of B, denoted $\dim B$, is the maximum number of linearly independent vectors in B. Topologically, the condition that $\dim B = n$ is equivalent to the condition that B has non-empty interior.

Theorem 2 (The Folk Theorem.) *Assume that* $\dim V = n$. *For all* $v \in V$, $v_i > \underline{v}_i$, *there exists* $\underline{\delta} < 1$ *such that, for all* $\delta \in (\underline{\delta}, 1)$, *there is a subgame-perfect Nash equilibrium of* G^δ *with payoff* v.

To put it differently, whenever V has non-empty interior, any of the feasible and strictly individually rational payoff vectors is a SPNE payoff for sufficiently high discount factors.

Proof: (i) For simplicity, suppose that there is a pure action profile $a \in A$ with $u(a) = v$. The proof for the general case follows essentially the same lines. Assume first that the minmax profile $\underline{\alpha}^i_{-i}$ against each player i is in pure strategies, so that the deviations from this profile are certain to be detected. At the end of this proof, case (ii) sketches how to modify the proof for the case of mixed minmax profiles.

Choose a vector $v' \in \operatorname{int} V$, the interior of V, and an $\varepsilon > 0$ such that, for each $i = 1, \ldots, n$,

$$\underline{v}_i < v'_i < v_i,$$

and the vector defined by

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

is in V. (The full-dimensionality assumption ensures that such $v'(i)$ exist for some $\varepsilon > 0$ and v'.)

Again, to avoid the details of public randomizations, assume that for each i there is a pure action profile $a(i)$ with $u(a(i)) = v'(i)$. Let $w^j_i := u_i(\underline{\alpha}^j)$ denote player i's payoff when minmaxing player j. Choose $N \in \mathbb{N}$ such that, for all i,

$$\max_a u_i(a) + N\underline{v}_i < \min_a u_i(a) + Nv'_i.$$

This is the punishment length such that, for discount factors close to 1, deviating once and then being minmaxed for N periods is worse than getting the lowest payoff once and then N periods

of v_i'.

Now consider the following strategy profile:

Play begins in **phase** I. In phase I, play action profile a, where $u(a) = v$. Play remains in phase I so long as in each period either the realized action is a or the realized action differs from a in two or more components. If a single player j deviates from a, then play moves to phase II_j.

Phase II_j: Play (the appropriate component of) $\underline{\alpha}^j$ each period. Continue in phase II_j for N periods so long as in each period either the realized action is $\underline{\alpha}^j$ or the realized action differs from $\underline{\alpha}^j$ in two or more components. Switch to phase III_j after N successive periods of phase II_j. If during phase II_j a single player i's action differs from $\underline{\alpha}_i^j$, begin phase II_i. (Note that this construction makes sense only if $\underline{\alpha}^j$ is a pure action profile; otherwise the"realized action" cannot be the same as $\underline{\alpha}^j$.)

Phase III_j: Play (the appropriate component of) $a(j)$, and continue to do so unless in some period a single player i fails to play $a_i(j)$. If a player i does deviate, begin phase II_i.

To show that these strategies are subgame perfect, it suffices to check that in every subgame no player can gain by deviating once and then conforming to the strategies thereafter.

In phase I, player i receives at least v_i from conforming, and he receives at most

$$(1-\delta)\max_a u_i(a) + \delta(1-\delta^N)\underline{v}_i + \delta^{N+1}v_i',$$

by deviating once. Since v_i' is less than v_i, the deviation will yield less than v_i for δ sufficiently large. Similarly, if player i conforms in phase III_j, $j \neq i$, then player i receives $v_i' + \varepsilon$. His payoff from deviating is at most

$$(1-\delta)\max_a u_i(a) + \delta(1-\delta^N)\underline{v}_i + \delta^{N+1}v_i',$$

which is less than $v_i' + \varepsilon$ when δ is sufficiently large.

In phase III_i, player i receives v_i' from conforming and at most

$$(1-\delta)\max_a u_i(a) + \delta(1-\delta^N)\underline{v}_i + \delta^{N+1}v_i',$$

from deviating once. The inequality $\max_a u_i(a) + N\underline{v}_i < \min_a u_i(a) + Nv_i'$ ensures that the deviation is unprofitable for δ sufficiently close to 1.

If player i conforms in phase II_j, $j \neq i$, when there are N' periods of phase II_j remaining

(including the current period), her payoff is

$$(1-\delta^{N'})w_i^j + \delta^{N'}(v_i' + \varepsilon).$$

If she deviates, she is minmaxed for the next N periods; the play in phase III_j will then give her v_i' instead of the $v_i' + \varepsilon$ she would get in phase III_j if she conformed now. Once again, the ε differential once phase III_j is reached outweighs any short-term gains when δ is close to 1. Finally, if player i conforms in phase II_i (i.e., when she is being punished) then when there are $N' \leq N$ periods of punishment remaining, player i's payoff is

$$q_i(N') := (1-\delta^{N'})\underline{v}_i + \delta^{N'} v_i' < v_i'.$$

If she deviates once and then conforms, she receives at most $\underline{v}_i$ in the period in which she deviates (because the opponents are playing $\underline{\alpha}^i_{-i}$) and her continuation payoff is then $q_i(N) \leq q_i(N'-1)$.

(ii) The above construction assumes that player i would be detected if she failed to play $\underline{\alpha}^i_{-i}$ in phase II_j. This need not be the case if $\underline{\alpha}^i_{-i}$ is a mixed strategy. In order to be induced to use a mixed minmax action, player i must receive the same normalized payoff for each action in the action's support. Since these actions may yield different payoffs in the stage game, inducing player i to mix requires that her continuation payoff be lower after some of the pure actions in the support than after others. Now, in the strategies of part (i), the exact continuation payoffs for player i in phase III_j, $j \neq i$, were irrelevant (the essential requirement was that player i's payoff be higher in phase III_j than in phase III_i). Thus, players can be induced to use mixed actions as punishments by specifying that each player i's continuation payoff in phase III_j, $j \neq i$, vary with the actions player i chose in phase II_j in such a way that each action in the support of $\underline{\alpha}^i_{-i}$ gives player i the same overall payoff. □

VI On the Role of some of the Assumptions

A Short-run Players

Up to this stage, we have assumed that players were all **long-run**, that is, infinitely-lived. In many economic applications, it might make more sense to think of some of them as being **short-run**, and only concerned about the current period. By this, we do not necessarily mean

that they live for only one period, but that intertemporal incentives cannot be designed for them: the many small customers of a large firm might be hard to monitor, and so it is optimal for them to behave myopically. Similarly, when modeling the interactions between a government and taxpayers, a central bank and depositors, etc., it is implausible to assume that the "small" players' actions are not myopic best-replies.

This is modeled as follows. Suppose that players $i = 1, \ldots, L$, $L \leq I$, are long-run players, whose objective is to maximize the average discounted sum of rewards, with discount factor $\delta < 1$. Players $j \in SR := \{L+1, \ldots, I\}$ are short-run players, each representative of which plays only once. Let

$$B : \times_{i=1}^{L} \Delta(A_i) \to \times_{j=L+1}^{I} \Delta(A_j)$$

be the correspondence that maps any mixed action profile $(\alpha_1, \ldots, \alpha_L)$ for the long-run players to the corresponding static equilibria for the short-run players. That is, for each $\alpha \in \text{graph}B$, and each $j > L$, α_j maximizes $u_j(\cdot, \alpha_{-j})$.

Clearly then, we must redefine the set of feasible and individually rational payoffs to account for this constraint. We let

$$\underline{v}_i := \min_{\alpha \in \text{graph}B} \max_{a_i \in A_i} u_i(a_i, \alpha_{-i}),$$

where $i = 1, \ldots, L$. We let

$$V := \text{co}\{u(\alpha) : \alpha \in \text{graph}B\}$$

and

$$\underline{V} := \{v \in V : \forall i = 1, \ldots, L : v_i > \underline{v}_i\}.$$

One might conjecture that the folk theorem extends to this setting, with this revised notion of minmax.

But consider the following example, where Player 1 is long-run and Player 2 is short-run.

		Player 2		
		L	M	R
Player 1	U	$4, 0$	$0, 1$	$-1, -100$
	D	$2, 2$	$1, 1$	$0, 3$

Let p be the probability with which Player 1 plays D. Player 2's best-reply is M if $p \in [0, 1/2]$, L if $p \in [1/2, 100/101]$, and R if $p \geq 100/101$. There are three static Nash equilibria: the pure

(D, R), a second in which $p = 1/2$ and 2 mixes between M and L; and a third in which $p = 100/101$ and 2 mixes between L and R.

The highest payoff for Player 1 which lies in V is 3, achieved by $p = 1/2$ and L. But this payoff cannot be achieved in an equilibrium of the repeated game. To see this, let $\bar{v}$ denote the highest equilibrium for Player 1, given δ (more formally, consider the supremum). Suppose $\bar{v} > 2$. Because his continuation payoff after the first period cannot exceed $\bar{v}$, it holds that

$$\bar{v} \leq (1-\delta)u_1(\alpha^1) + \delta\bar{v},$$

where α^1 is the action profile played in the first period. It follows that $u_1(\alpha^1) \geq \bar{v} > 2$, and so Player 2 must play L with positive probability. But Player 2 is only willing to do so if Player 1 randomizes between both his actions. This means that Player 1 must be willing to play D, in which case she gets at most

$$(1-\delta)\cdot 2 + \delta\bar{v} < \bar{v},$$

a contradiction, since her payoff is supposed to be $\bar{v}$.

The point is clear: Player 1 cannot be compensated for playing a pure action within the support of his mixed action that would not maximize his reward; hence a long-run player's payoff is bounded above by

$$\bar{v}_i := \max_{\alpha \in \text{graph}B} \min_{a_i \in \text{support}(\alpha_i)} u_i(a_i, \alpha_{-i}).$$

We then define

$$V^* := \{v \in V : \forall i = 1, \ldots, L : v_i \leq \bar{v}_i\}.$$

A construction similar to the one of the folk theorem yields:

Theorem 3 *Assume that* $\dim \text{proj}_{i=1,\ldots,L} V = L$. *For all* $v \in V$, *there exists* $\underline{\delta} < 1$ *such that, for all* $\delta \in (\underline{\delta}, 1)$, *there is a subgame-perfect Nash equilibrium of* G^δ *with payoff* v.

B Dimension and Interiority

The following two examples (due to Fudenberg and Maskin, 1986; and Forges, Mertens and Neyman, 1986) illustrate some of the subtleties involved with dropping full-dimensionality of V. Consider first the following three-player game.

Player 2

Player 1		L	R	L	R
	U	$1,1,1$	$0,0,0$	$0,0,0$	$0,0,0$
	D	$0,0,0$	$0,0,0$	$0,0,0$	$1,1,1$
		A		B	

The minmax payoff is clearly 0. Yet it is easy to see that all subgame-perfect Nash equilibrium payoffs are at least $1/4$. Suppose not, and focus on the equilibrium that gives a player (and hence all players) his lowest equilibrium payoff, $\underline{v}$ (formally, consider the infimum). Let α_i denote the probability with which players use their first action (*i.e.*, U, L, A) in the first period. Note that each player i can secure as a reward at least

$$\max\{\alpha_j\alpha_k, (1-\alpha_j)(1-\alpha_k)\},$$

where i, j, k are distinct. Now, without loss, assume that $\alpha_1 \leq \alpha_2 \leq \alpha_3$. If $\alpha_2 \leq 1/2$, then $(1-\alpha_1)(1-\alpha_2) \geq 1/4$ and player 3 can secure a reward of $1/4$. If instead $\alpha_2 \geq 1/2$, then $\alpha_2\alpha_3 \geq 1/4$ and the same holds from player 1's point of view. That is, at least one player can secure $1/4$ as a reward. His continuation payoff being at least $\underline{v}$, it follows that

$$\underline{v} \geq (1-\delta)\frac{1}{4} + \delta\underline{v},$$

and so $\underline{v} \geq 1/4$.

Clearly, full dimensionality fails in this example: all three players have the same payoff function and so the same preferences over action profiles. To formalize this, we introduce the notion of **equivalent utilities**: player i and j have equivalent utilities if there exists two constants $b > 0$ and $c \in \mathbb{R}$ such that $u_i(a) = b \cdot u_j(a) + c$ for all a. This means that u_i and u_j are two utility representations of the same preference relation over actions. Let us now partition the set of players N into subsets $\{N_s\}_{s=1}^S$ with equivalent utilities. We define the **effective minmax payoff** of player $i \in N_s$ as

$$\underline{\underline{v}}_i = \min_a \max_{j \in N_s} \max_{a_j} u_i(a_j, a_{-j}).$$

Clearly, $\underline{\underline{v}}_i \geq \underline{v}_i$ and the two coincide when no other player has utility equivalent to i. We write $\underline{\underline{\alpha}}^i$ for the action profile achieving the effective minmax payoff.

The same reasoning as in the example implies that a player's lowest equilibrium payoff is no less than the effective minmax payoff (check!). It is not hard to prove that, if we use the effective

minmax payoff instead of the minmax payoff, we can drop the dimensionality assumption in the statement of the folk theorem. It is worth noting that the dimensionality assumption is actually not necessary with two players. To see this, suppose that $N = \{1, 2\}$ and that the players have equivalent utilities (there is nothing to show otherwise). Without loss, take $u_1 = u_2$. Then a feasible payoff vector (v, v) that is (strictly) individually rational in the usual sense must satisfy

$$v = v_1 = v_2 > \max\{u_1(\underline{\alpha}^1), u_2(\underline{\alpha}^2)\}.$$

Hence $\underline{\underline{v}}_1 \geq \max\{u_1(\underline{\alpha}^1), u_2(\underline{\alpha}^2)\}$. Yet as candidate for $\underline{\underline{\alpha}}^1$ we can take the mutual minmax vector $(\underline{\alpha}_1^2, \underline{\alpha}_2^1)$, which gives that $\underline{\underline{v}}_1 \leq \max\{u_1(\underline{\alpha}^1), u_2(\underline{\alpha}^2)\}$. Hence $\underline{\underline{v}}_1$ (and similarly $\underline{\underline{v}}_2$) are equal to $\max\{u_1(\underline{\alpha}^1), u_2(\underline{\alpha}^2)\}$, and so the set of strictly individually rational payoffs for the two notions of minmax payoffs coincide. See Wen (1994) for details.

Consider now the second example. Note that players 2 and 3 can each secure 0, and that the sum of their payoffs is also 0. Hence, the set of feasible and (weakly) individually rational payoffs is $\{(v, 0, 0) : v \in [0, 1]\}$.

Yet $(0, 0, 0)$ is the only Nash equilibrium payoff of the discounted game (no matter the value of $\delta < 1$). Otherwise, there is a first stage at which players play (U, L, A) or (D, R, B) with positive probability. Without loss, suppose it is (U, L, A). Then player 2 can secure a positive payoff by playing the equilibrium before and including this stage, and L afterwards, giving him an expected payoff that is strictly positive, a contradiction.

Note that this argument did not involve subgame perfection. And note also that players do not have equivalent utilities. But the problem is that $\underline{V}$ is empty in this example. Recall that the statement of the folk theorem requires that all players get a strictly individually rational payoffs: this is simply infeasible here.

Player 2

Player 1		L	R	L	R
	U	$1, 1, -1$	$0, 0, 0$	$0, 0, 0$	$0, 0, 0$
	D	$0, 0, 0$	$0, 0, 0$	$0, 0, 0$	$1, -1, 1$
			A		B

VII Finitely Repeated Games Revisited

There are three ways to induce a player to randomize. One is to adjust finely his continuation payoff by relying on future pure action choices that are finely calibrated. This is the approach

followed originally (see D. Fudenberg and E. Maskin (1986), *Journal of Economic Theory*, "On the Dispensability of Public Randomization in Discounted Repeated Games," **53**, 428–438). An alternative is to have players randomize, so that the opponent's exact probability of choosing actions (in the future) achieve precisely the right payoff. (This is the approach used under imperfect private monitoring, although it applies to perfect monitoring as well, see J. Hörner and W. Olszewski, "The Folk Theorem for Games with Private Almost-Perfect Monitoring," *Econometrica*, **2006**, 1499–1544). Both approaches rely on the horizon being infinite. The third approach does not, and hence is critical in implementing mixed minmaxing in finitely repeated games. Because the underlying idea extends beyond that particular application, we develop it here. It is due to O. Gossner (1995, "The Folk Theorem for Finitely Repeated Games with Mixed Strategies," *International Journal of Game Theory*, **24**, 95–107). It is based on approachability theory, see the end of these notes for a short survey, and is closely related to the idea of statistically testing whether a player is indeed taking actions with the right frequencies when he is supposed to randomize. However, testing the frequency is not enough. For instance, because of discounting, a player would "frontload" the more profitable action and postpone playing the less profitable ones, subject to the frequency requirement. Plainly, one should also test for these actions being "uniformly" distributed over time. Approachability sidesteps this problem. Instead of testing whether he truly randomizes, we check whether, *conditional* on each action profile by $-i$, player i gets the frequency right; because we control for the action profile of the others, we ensure in this fashion that their payoff is *as if* player i was randomizing.

To be completed.

VIII Literature

The folk theorem under perfect monitoring (with discounting) is established in Fudenberg and Maskin, 1984 ("The Folk Theorem for Repeated Games With Discounting or With Incomplete Information," *Econometrica*, **54**, 533–554.) A weaker version involving Nash reversion was known since the 70s (Friedman, J., 1971. "A noncooperative Equilibrium for Supergames," *Review of Economic Studies*, **38**, 1–12). The result presented has been improved over the years. Abreu, Dutta and Smith, 1994 ("The Folk Theorem for Repeated Games: A NEU condition," *Econometrica*, **62**, 939–948) weakens full-dimensionality, and Wen, 1994 ("The "Folk Theorem" for Repeated Games with Complete Information," *Econometrica*, **62**, 949–954) weakens it further to a necessary and sufficient condition. The folk theorem was long known for undiscounted

payoff criteria (for instance, Rubinstein, 1977, "Equilibrium in Supergames," CRIMEGT Rm 25).

Abreu 1988 ("On the Theory of Infinitely Repeated Games with Discounting" *Econometrica*, **56**, 383–396) does not focus on the folk theorem. Rather, he shows that, under perfect monitoring, attention can be restricted to strategies that are relatively simple: any equilibrium payoff that can be achieved can be done so by specifying $n+1$ outcome paths, one that must be followed as long as no player deviates, and one for each player in case of deviation by this player. Sorin 1986 ("On Repeated Games with Complete Information," *Mathematics of Operations Research*, **11**, 147–160) provides a wealth of results regarding the structure of equilibrium and feasible payoffs. In particular, he proves that a public randomization device can be dispensed with, a construction that was refined for the purpose of the perfect folk theorem by Fudenberg and Maskin, 1991 ("On the Dispensability of Public Randomization in Discounted Repeated Games," *Journal of Economic Theory*, **53**, 428–438)

As for finitely repeated games, the main result is due to Benoît and Krishna 1985 ("Finitely Repeated Games," *Econometrica*, **53**, 905–922). They prove a folk theorem assuming that players have distinct Nash payoffs. This condition has been weakened by Smith 1995 ("Necessary and Sufficient Conditions for the Perfect Finite Horizon Folk Theorem," *Econometrica*, **63**, 425–430). Both papers assume a public randomization device. Unlike in the case of infinitely repeated games, the public randomization plays a substantial role in their construction, and Gossner 1995 ("The Folk Theorem for Finitely Repeated Games with Mixed Strategies," *International Journal of Game Theory*, **24**, 95–107) provides a proof without it.

A Supplementary Material: Approachability

An important tool in repeated games is given by *approachability theory*, a tool introduced by Blackwell (1956, " An analog of the minimax theorem for vector payoffs," *Pacific Journal of Mathematics*, **6**, 1–8). This short survey follows closely Sorin (*A First Course on Zero-sum Repeated Games*, Springer, 2001).

Consider a two player *vector-valued* zero-sum game. A is an $I \times J$ matrix with coefficients in $\mathbb{R}^K$. At each stage n, Player 1 (resp., Player 2) chooses a move, i_n in I (resp., j_n in J). The corresponding vector payoff, $g_n = A_{i_n j_n}$ is then announced. Denote by h_n the sequence of payoffs before stage n (this is, at least, the information available to both players at stage n) and let $\bar{g}_n := \frac{1}{n}\sum_{m=1}^{n} g_m$ be the average payoff up to stage n. Let also $\|A\| = \max_{i\in I, j\in J, k\in K} |A_{ij}^k|$.

Definition 1 *A set C in $\mathbb{R}^K$ is **approachable** by Player 1 if for any $\epsilon > 0$ there exists a strategy*

σ and N such that, for any strategy τ of Player 2 and any $n \geq N$:

$$\mathbb{E}_{\sigma,\tau}(d_n) \leq \epsilon,$$

where d_n is the Euclidean distance $d(\bar{g}_n, C)$.

*A set C in $\mathbb{R}^K$ is **excludable** by Player 1 if for some $\delta > 0$, the set $C_\delta^c := \{z; d(z, C) \geq \delta\}$ is approachable by him.*

A dual definition holds for Player 2.

Notice that if Player 1 can approach C, Player 2 cannot exclude C, he could however approach C^c. The analog of the minmax theorem would be: if Player 1 cannot approach C then Player 2 can exclude it. The next analysis will prove this to be true within the class of convex set (counterexamples can be found for general sets C).

From the definitions it is enough to consider closed sets C and even their intersection with the closed ball of radius $\|A\|$.

Given s in $S = \Delta(I)$, define $sA = \text{co}\,\{\sum_i s_i A_{ij}; j \in J\}$, and similarly At, for t in $T = \Delta(J)$. If Player 1 uses s his expected payoff will be in sA.

The first result is a sufficient condition for approachability based on the following notion:

Definition 2 *A closed set C in $\mathbb{R}^K$ is a **B**-set for Player 1 if : for any $z \notin C$, there exists a closest point $y = y(z)$ in C to z and a mixed move $s = s(z)$ in S, such that the hyperplane through y orthogonal to the segment $[yz]$ separates z from sA.*

Note that for any $x \in sA$, any point on the line $[xz]$ closed to z will be closer to y than z itself. This geometric consideration is the basis of the next result.

Theorem 4 *Let C be a **B**-set for Player 1. Then C is approachable by that player. More precisely with a strategy satisfying $\sigma(h_{n+1}) = s(\bar{g}_n)$, whenever $\bar{g}_n \notin C$, one has:*

$$\mathbb{E}_{\sigma\tau}(d_n) \leq \frac{2\|A\|}{\sqrt{n}} \quad \forall \tau,$$

and d_n converges $P_{\sigma\tau}$-a.s. to 0.

Proof. Let Player 1 use a strategy σ as above. Denote $y_n = y(\bar{g}_n)$. Then one has:

$$\begin{aligned} d_{n+1}^2 &\leq \|\bar{g}_{n+1} - y_n\|^2 \\ &= \|\frac{1}{n+1}(g_{n+1} - y_n) + \frac{n}{n+1}(\bar{g}_n - y_n)\|^2 \\ &= \left(\frac{1}{n+1}\right)^2 \|(g_{n+1} - y_n)\|^2 + \left(\frac{n}{n+1}\right)^2 d_n^2 + \frac{2n}{(n+1)^2}\langle g_{n+1} - y_n, \bar{g}_n - y_n\rangle. \end{aligned}$$

The property of $s(\bar{g}_n)$ implies that:

$$\mathbb{E}(\langle g_{n+1} - y_n, \bar{g}_n - y_n\rangle \mid h_{n+1}) \leq 0,$$

since $\mathbb{E}(g_{n+1} \mid h_{n+1})$ belongs to $s(\bar{g}_n)A$.

We can assume C included in the ball of radius $\|A\|$, hence $\mathbb{E}(\|(g_{n+1} - y_n)\|^2 | h_{n+1}) \leq 4\|A\|^2$ and

$$\mathbb{E}(d_{n+1}^2 \mid h_{n+1}) \leq \frac{4\|A\|^2}{(n+1)^2} + \left(\frac{n}{n+1}\right)^2 d_n^2,$$

which implies by induction

$$\mathbb{E}(d_n^2) \leq \frac{4\|A\|^2}{n}.$$

To get almost sure convergence, let $e_n = d_n^2 + \sum_{m>n} \frac{4\|A\|^2}{m^2}$ so that, by $\mathbb{E}(d_{n+1}^2 \mid h_{n+1}) \leq \frac{4\|A\|^2}{(n+1)^2} + (\frac{n}{n+1})^2 d_n^2$: $\mathbb{E}(e_{n+1} \mid h_{n+1}) \leq e_n$. Thus e_n is a positive supermartingale that majorizes d_n^2 and satisfies $\mathbb{E}(e_n) \leq 4\|A\|^2(\frac{1}{n} + \sum_{m>n} \frac{1}{m^2})$. Hence it converges to 0 a.s. and d_n also. ■

Corollary 5 *For any s in S, sA is approachable by Player 1 with the constant strategy s*

It follows that a necessary condition for a set C to be approachable by Player 1 is that for any t in T, $At \cap C \neq \emptyset$, otherwise C would be excludable by Player 2. In fact this condition is also sufficient for convex sets.

Theorem 6 *Assume C closed and convex in $\mathbb{R}^K$. C is a **B**-set for Player 1 iff $At \cap C \neq \emptyset$, for all t in T.*

*In particular a set is approachable iff it is a **B**-set.*

Proof. By the previous Corollary, it is enough to show that if $At \cap C \neq \emptyset$, for all t, C is a **B**-set.

The idea is to reduce the problem to the one dimensional case and to use the minmax theorem.

In fact, let $z \notin C$, y be its projection on C and consider the game with real payoff $\langle y - z, A\rangle$. Since $At \cap C \neq \emptyset$ for all t, it implies that its value is at least $\langle y - z, y\rangle = \min_{c \in C}\langle y - z, c\rangle$. Hence

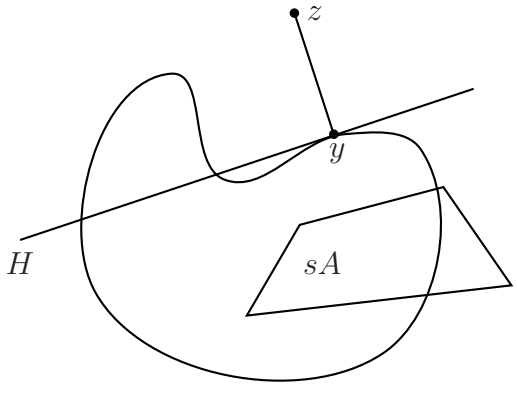


Figure 1: Approachability

there exists an optimal strategy s for Player 1 such that $\langle y - z, \sum_i s_i A_{ij} \rangle \geq \langle y - z, y \rangle$ for any j in J, which shows that sA is on the opposite side of the hyperplane and the result follows. ■

Repeated Games, Part II: Imperfect Public Monitoring

Lecture Notes, Yale 2015, Johannes Hörner

September 3, 2015

I An Example

This example follows Abreu, Milgrom and Pearce (1991). Consider the problem of a team of 5 workers, or players, who are working together on a project forever. We shall consider two variants of this infinitely repeated game, one in which their efforts reduce the probability of unfortunate events, and one in which it increases the probability of desirable ones. In both cases, denote each player's individual effort by $e_i \in [0,1]$, $i = 1, \ldots, 5$. Period length is $\Delta > 0$, and the cost of effort over some interval is $(e_i + e_i^2)\,\Delta/2$. At the beginning of each period, each player independently chooses his effort level, and his choice is not observed by the other players. Players all discount payoffs at rate $\delta := e^{-r\Delta}$, for some $r > 0$. For concreteness, let us pick $e^{-r} = 1/10$. Throughout, we assume $\Delta < 1/6$.

A First Variant: The "Bad News" Case

In the first variant, effort reduces the probability of an accident in each period. Each accident yields a (dis)utility of -1 whenever it occurs (the corresponding utility if there is no accident is normalized to 0), and its probability in a period is given by

$$\left(6 - \sum_j e_j\right)\Delta.$$

Accidents are the only events that are observed. Player i's reward in a period is then given by

$$-\left(6 - \sum_j e_j + \frac{e_i + e_i^2}{2}\right)\Delta.$$

In the one-shot game, each player maximizes this reward by choosing $e_i = 1/2$, which yields a reward of $-(31/8)\Delta$. Note that the social optimum involves all players putting in effort $e_i = 1$, for a reward of -2Δ.

What is the best equilibrium in the repeated game? Note that, because of the quadratic cost, players will never find it optimal to randomize.[1] Further, because the only way to punish players is by increasing the probability of accidents, which affects all players equally, we can expect equilibria to be symmetric, *i.e.*, $e_i = e$.[2]

So let us characterize the best symmetric equilibrium, under the assumption that players have access to a public randomization device. Let us denote by v^H the payoff of the best symmetric equilibrium, and let us also write v^L for the worst (symmetric) equilibrium payoff. We refer to H and L as states. Achieving v^H requires specifying an action e^H to be played in the initial period, as well as a continuation payoff from the second period onward according to whether there is an accident or no accident: w_A^H, w_N^H. Similarly, achieving v^L calls for an action e^L to be played, followed by continuation payoffs w_A^L, w_N^L. Because players could implement the continuation equilibrium immediately, it must be that

$$w_A^H, w_N^H, w_A^L, w_N^L \in \left[v^L, v^H\right]. \tag{1}$$

Furthermore, e^H, e^L must be Nash equilibria, with payoffs v^H, v^L, of the one-shot game with payoff functions

$$\begin{aligned}&\Delta\left(6-4e^k-e_i^k\right)\left[(1-\delta)\left(-1-\Delta\frac{e_i^k+(e_i^k)^2}{2}\right)+\delta w_A^k\right]\\&+\left(1-\Delta\left(6-4e^k-e_i^k\right)\right)\left[-(1-\delta)\Delta\frac{e_i^k+(e_i^k)^2}{2}+\delta w_N^k\right],\end{aligned} \tag{2}$$

with $k = H$ and L. Clearly, the payoff v^H (resp. v^L) is increasing in e^H (resp., e^L),[3] and so, to maximize the gap between v^H and v^L (so as to maximize incentives), we want to choose the largest (resp. smallest) effort consistent with (2), subject to (1). Note that, given our randomization device, we can achieve any payoff w in the range $\left[v^L, v^H\right]$ simply by using the randomization device to coordinate on v^L and v^H with probability $\left(v^H - w\right)/\left(v^H - v^L\right)$ and $\left(w - v^L\right)/\left(v^H - v^L\right)$, respectively. Therefore, players can restrict attention to equilibria in which, after any history,

[1] That is, if a player were indifferent between two distinct effort levels, he would strictly prefer the expected effort level, as its cost would be lower than the expected cost.

[2] Proving this is the case requires a little work, though.

[3] Recall that first-best effort is 1.

and as a function of the public randomization device, either e^H or e^L is played. So define

$$q_A^H := \frac{v^H - w_A^H}{v^H - v^L}, \quad q_N^H := \frac{v^H - w_N^H}{v^H - v^L},$$

as well as

$$q_A^L := \frac{w_A^L - v^L}{v^H - v^L}, \quad q_N^L := \frac{w_N^L - v^L}{v^H - v^L},$$

as the probabilities of switching states, as a function of the initial state (L or H) and the occurrence or not of an accident. It is not hard to see that, to maximize (resp., minimize) v^H (resp., v^L), we must set

$$q_N^H = 0, \quad q_N^L = 0;$$

that is, to encourage effort in state H, it is best to avoid all punishment if an accident is avoided. Similarly, to discourage effort in state L, we should only consider switching back to state H if an accident is observed. The switching probabilities can then be found by taking first-order conditions with respect to effort e_i^j in (2) and simplifying:[4]

$$(1-\delta)\left(e^H - \frac{1}{2}\right) = \delta q_A^H \left(v^H - v^L\right), \quad (1-\delta)\left(\frac{1}{2} - e^L\right) = \delta q_A^L \left(v^H - v^L\right).$$

Solving for the switching probabilities, and plugging back into (2) gives

$$v^k = \left((9/2)\left(e^k\right)^2 - 4e^k - 3\right)\Delta.$$

So it is best to set $e^H = 1$, $e^L = 4/9$, and these satisfy (1) if Δ is small and δ is high enough (more precisely, if $\Delta \leq .158$, given $e^{-r} = 1/10$). This gives us

$$v^H = -5\Delta/2, \quad v^L = -35\Delta/9$$

Note that v^H is independent of r. We picked $e^{-r} = 1/10$, but nothing changes if players are more patient. In particular, we do not get efficiency as $\delta \to 1$.

[4]These must hold even if effort is extreme, because otherwise one could decrease the probability of "punishment" (or increase the probability of "reward") and still get the same extreme effort, while increasing (or decreasing) v^H (resp. v^L).

B Second Variant: The "Good News" Case

In this variant, effort increases the probability of a desirable event —say, a sale, worth 1 to each player. Hence, let us assume that the probability of a sale is given by:[5]

$$\left(1+\sum_j e_j\right)\Delta.$$

The cost of effort is the same as before. We can analyze the game as before, solving for payoffs

$$w_S^H, w_N^H, w_S^L, w_N^L \in \left[v^L, v^H\right],$$

according to whether a sales occurs or not. Subject to this condition, the effort levels e^H, e^L must be Nash equilibria, with payoffs v^H, v^L, of the one–shot game with payoff functions

$$\begin{aligned}&\left(1+4e^k+e_i^k\right)\Delta\left[(1-\delta)\left(1-\Delta\frac{e_i^k+\left(e_i^k\right)^2}{2}\right)+\delta w_S^k\right]\\&+\left(1-\left(1+4e^k+e_i^k\right)\Delta\right)\left[-(1-\delta)\Delta\frac{e_i^k+\left(e_i^k\right)^2}{2}+\delta w_N^k\right],\end{aligned}$$

with $k=H$ and L. This can be analyzed as before. The main difference is that the probabilities of switching after sales should, quite intuitively, be set equal to zero. But the consequences are huge: as you can easily check, the unique solution is to set $e^L=e^H=1/2$, which is the Nash equilibrium of the one-shot game: no collusion can be sustained, independently of patience.

C Conclusions

We can draw several implications from this example:

1. Recursive methods appear to be applicable to games with imperfect monitoring. The next section will formalize the heuristic methods used here. This is the topic of Section II.

2. The folk theorem need not hold once monitoring is imperfect. In this example, in fact, the highest payoff is bounded away from the efficient payoff, no matter how patient players are. If there are sufficient conditions for the folk theorem, our example must fail those. It

[5] This keeps the range of probabilities the same as in the first variant.

is natural to wonder, then, what these sufficient conditions are, and this will be the topic of Section IV.

3. When the folk theorem fails, details matter a great deal. A characterization of the equilibrium payoff set for $\delta \to 1$ will be provided in Section III. In our example, the key distinction is whether the rare event is good news or bad news. What is the intuition behind the different results across variants? The key lies in the informativess of the bad news signal, which is the one that must trigger a punishment. In the first example, accidents are informative, as the likelihood ratio

$$\frac{\left(6-\sum_j e_j\right)\Delta}{\left(6-\sum_{j\neq i} e_j - e_i'\right)\Delta}$$

is independent of Δ and sensitive to e_i': accidents carry information about actions taken by the players. In the second example, in contrast, the likelihood ratio

$$\frac{1-\left(1+\sum_j e_j\right)\Delta}{1-\left(1+\sum_j e_{j\neq i}+e_i'\right)\Delta}$$

converges to 1, independently of e_i', as $\Delta \to 0$: bad news signal are no longer informative, and it is no longer possible to use them to collude effectively.

II Recursive Methods

A Notations

Attention is restricted throughout to infinitely repeated games. A repeated game with imperfect public monitoring specifies, in addition to the set of players $N = \{1, \ldots, n\}$, and action profiles A, a set of signals Y (finite), and, for each action profile $a \in A$, a distribution $\pi(\cdot \mid a)$ on Y, the **monitoring structure**. The interpretation is straightforward: as a function of the action profile a played in a given period, the signal $y \in Y$ is drawn according to the distribution $\pi(\cdot \mid a)$. Actions are not observed by other players; on the other hand, the signal y is publicly observed. We write $\pi(y \mid \alpha) = \sum_a \alpha(a)\,\pi(y \mid a)$ for the distribution of signals induced by a mixed action $\alpha \in \triangle A$.

For consistency, rewards are usually first defined as maps

$$g_i : A_i \times Y \to \mathbb{R},$$

so that a player's realized reward $g_i(a_i, y)$ carries no more information than what he already knows, or observes, namely a_i and y. Given some action profile a, player i's expected reward is then

$$u_i(a) := \sum_{y \in Y} g_i(a_i, y)\, \pi(y \mid a),$$

whose average discounted sum he seeks to maximize. It has been customary to use the function u as the primitive of the repeated game, rather than g. In this fashion, fixing u, we can examine how the quality of the monitoring affects the equilibrium payoff set without having to worry about how the change in monitoring affects the set of feasible payoffs, as it "mechanically" would if we were to take g as a primitive.

Therefore, we shall take u as a primitive, and ignore g from now on, but it is important to keep in mind that players cannot infer anything from $u_i(a)$ beyond what they already know, a_i and y.

A repeated game with imperfect public monitoring, then, is a collection $\left(N, A, Y, \pi(\cdot \mid a)_{a \in A}, u\right)$, along with a discount factor δ. Note that, up to period t, player i has observed an element of $H_i^t := (A_i \times Y)^t$, corresponding to the actions he has played, and the public signals he has observed.[6] This is the set of **private histories** h_i^t. Players share some information, namely the sequence of public signals, or **public history** $h^t \in H^t := Y^t$. Perfect monitoring is the special case in which $Y = A$, and $\pi(y \mid a) = 1$ iff $y = a$, so that action profiles are perfectly observed. Of course, our interest primarily lies in the case in which the monitoring is not perfect, though everything we shall prove applies to perfect monitoring as well.

B Definitions

What we are trying to achieve here is a "simple," "recursive" representation of the set of equilibrium payoffs. We shall cheat, and define jointly the solution concept, a refinement of sequential equilibrium, and the recursive representation, which led to our choice.

Definition 1 *The strategy σ_i is **public** if it is measurable with respect to the public history: for*

[6] If a public randomization device is assumed, it is understood that each player also has observed its realizations in all previous periods.

all t and sequences $(y^s, a_i^s, \hat{a}_i^s)_{s=0}^t$,

$$\sigma_i\left(a_i^1 y^1, a_i^2 y^2, \ldots, a_i^t y^t\right) = \sigma_i\left(\hat{a}_i^1 y^1, \hat{a}_i^2 y^2, \ldots, \hat{a}_i^t y^t\right).$$

If players $-i$ use public strategies, they disregard their private information going forward. Therefore, player i has nothing to gain from conditioning on his private information either, and he has a best-reply that is public as well. This implies that public strategies are closed under best-replies.

Definition 2 *The strategy profile σ is a **public perfect equilibrium**, or **PPE**, if, for all i, σ_i is public, and for all public histories h^t, $\sigma|_{h^t}$ is a Nash equilibrium of the repeated game.*

This solution concept is a natural extension of subgame-perfection to imperfect monitoring: if players' strategies only condition on public events, we require that they are Nash equilibria conditional on any such event. Clearly, PPE are sequential equilibria. However, there are sequential equilibria that are not PPE, and there are well-known examples of repeated games in which, as $\delta \to 1$, efficient payoffs can be approximated by sequential equilibria, but not by PPE: the power of statistical tests to detect deviations can be improved by using all information a player has available, which includes his own privately observed actions.[7]

We let E_δ denote the set of PPE payoffs, given the repeated game and the discount factor δ. The main benefit of this solution concept is that the set of PPE payoffs is independent of the public history: of course, which PPE is selected as a continuation strategy profile depends on the public history, in general, but the set of PPE to select from does not. This is not true for sequential equilibria, as private histories provide private correlation devices whose structure depends, among others, on the period considered.

We now turn to the central tools for studying E_δ, introduced by Shapley (1953), Mertens & Parthasarathy (1986) and Abreu, Pearce and Stacchetti (1990, hereafer referred to as APS). Recall that, if W and Y are sets, W^Y is the set of functions from Y to W.

Definition 3 *Given $W \subset \mathbb{R}^n$, $\alpha \in \times_i \triangle A_i$ is **enforceable** on W if there exists $w \in W^Y$ such that α is a Nash equilibrium of the game $\Gamma_\delta(w)$ with action sets A_i and payoff function*

$$(1-\delta)\, u(\cdot) + \delta \sum_y \pi(y \mid \cdot)\, w(y). \tag{3}$$

[7] Kreps and Wilson (1982) define sequential equilibrium for finite games only. Throughout these notes, the definition is extended to infinitely repeated games by equipping both the set of strategies and the set of systems of beliefs with the uniform topology of uniform convergence over information sets.

The function w **enforces** α. The equilibrium payoff vector

$$v = (1-\delta)\, u(\alpha) + \delta \sum_{y} \pi(y \mid \alpha)\, w(y)$$

is **decomposed** by (α, w) on W, and it is **decomposable** on W if there exists such a pair $(\alpha, w) \in \times_i \triangle A_i \times W^Y$.

Let $\mathcal{P}(\mathbb{R}^n)$ denote the set of all subsets of $\mathbb{R}^n$. We define the map

$$\begin{aligned} B \;:\; & \mathcal{P}(\mathbb{R}^n) \to \mathcal{P}(\mathbb{R}^n) \\ & W \longmapsto B(W) = \{v \in \mathbb{R}^n : v \text{ is decomposable on } W\}. \end{aligned}$$

These are all the payoffs that can be obtained by decomposition, by using "continuation" payoffs from W only. We shall also write B_δ rather than B whenever convenient. Here are a couple of properties of the operator B:

1. B is a monotone operator, as follows from the definition of enforceability:

$$W \subset W' \Longrightarrow B(W) \subset B(W').$$

2. B maps compact sets into compact sets. If $\left(\alpha^k, w^k\right)$ is a sequence that decomposes v^k, and $\lim_k \left(\alpha^k, w^k\right) = (\alpha, w)$, then (α, w) decomposes $v = \lim_k v^k$. Hence, if W is compact (so that sequences $\left(\alpha^k, w^k\right) \in \times_i \triangle A_i \times W^Y$ have convergent subsequences), $B(W)$ is closed, and clearly bounded.

Definition 4 *The set $W \subset \mathbb{R}^n$ is **self-generating** if*

$$W \subset B(W).$$

The interest in self-generating sets follows from the following:

Theorem 1 *If $W \subset \mathbb{R}^n$ is self-generating and bounded, $B(W) \subset E_\delta$.*

Proof: Because W is self-generating, each $v \in B(W)$ is decomposed by some pair $(\alpha_v, w_v) \in \times_i \triangle A_i \times W^Y$. For each $v \in B(W)$, define the strategy σ, parametrized by $v' \in B(W)$, that starts at the beginning of the game in state v, and after any history h^t, given the current state

$v' \in B(W)$, specifies $\sigma(h^t) = \alpha_{v'}$, and moves to state $v'' = w_{v'}(y)$ in the next period, as a function of the realized signal y. We first show that states truly correspond to payoffs, i.e. that the payoff from playing σ is indeed v. By definition of σ,

$$\begin{aligned} v &= (1-\delta)\, u(\sigma(\varnothing)) + \delta \sum_{y^0} \pi\left(y^0 \mid \sigma(\varnothing)\right) w_v\left(y^0\right) \\ &= (1-\delta)\, u(\sigma(\varnothing)) + \delta \sum_{y^0} \pi\left(y^0 \mid \sigma(\varnothing)\right) \left[(1-\delta)\, u\left(\sigma\left(y^0\right)\right) + \delta \sum_{y^1} \pi\left(y^1 \mid \sigma\left(y^0\right)\right) w_{w_v(y^0)}\left(y^1\right)\right] \\ &= \cdots \\ &= (1-\delta) \sum_{s=0}^{t-1} \delta^s \sum_{h^s} \mathbb{P}_\sigma\left[h^s\right] u\left(\sigma\left(h^s\right)\right) + \delta^t \sum_{h^t} \mathbb{P}_\sigma\left[h^t\right] w_v\left[h^t\right], \end{aligned}$$

where $\mathbb{P}_\sigma[h^t]$ is the probability of h^t under σ, and $w_v[h^t]$ is the state that is obtained after public history h^t, starting from state v, given the strategy σ. Because $w_v[h^t] \in W$ and W is bounded, it follows that

$$v = \lim_{t\to\infty} (1-\delta) \sum_{s=0}^{t-1} \delta^s \sum_{h^s} \mathbb{P}_\sigma\left[h^s\right] u\left(\sigma\left(h^s\right)\right) = (1-\delta)\mathbb{E}_\sigma\left[\sum_{t=0}^{\infty} \delta^t u_i(a^t)\right],$$

as was to be shown. Similarly, $w_v[h^t]$ is the continuation payoff under σ given public history h^t. Optimality of σ_i then follows from the one-shot deviation principle, given that $\alpha_{i,v'}$ is optimal against $\alpha_{-i,v'}$, given continuation payoffs $w_{v'}$. □

Continuation payoffs from a PPE are equilibrium payoffs themselves. Hence, all equilibrium payoffs must be decomposable on E_δ, and so $E_\delta \subset B(E_\delta)$. Therefore, E_δ is self-generating, and by the previous theorem, it follows that $B(E_\delta) \subset E_\delta$. Hence:

Corollary 2 *It holds that*

$$E_\delta = B(E_\delta).$$

Hence, E_δ is a fixed-point of B. Note that, from Theorem 1, every bounded fixed-point of B must be a subset of E_δ (because it must be self-generating), and so E_δ is actually the largest bounded fixed-point of E_δ.

Because V, the set of feasible payoffs, is compact, and decomposable payoffs with respect to V must be feasible,

$$V^1 = B(V) \subset V,$$

and V^1 is compact. More generally, by monotonicity of W, the sequence

$$V^{k+1} = B\left(V^k\right),$$

with $V^0 = V$, is (weakly) decreasing and compact. Furthermore, because

$$E_\delta \subset V \Longrightarrow E_\delta = B\left(E_\delta\right) \subset B\left(V\right) = V^1,$$

the set V^1 (and similarly V^k) is non-empty, containing E_δ. Let $V^\infty = \lim_k V^k$. Because V^k is a decreasing sequence of compact sets, V^∞ is compact.

Lemma 1 *The set V^∞ is self-generating.*

Proof: Fix $v \in V^\infty$. Then $v \in V^k$ for all k, and so there exists a sequence $\left(\alpha^k, w^k\right)$ that decomposes v with $w^k \in V^{k-1}$. We want to show that v is decomposable on V^∞, and the obvious candidate for decomposition is $(\alpha, w) = \lim_k \left(\alpha^k, w^k\right)$ (the sets are compact, so that we can assume that the sequence converges). We must show that $w \in V^\infty$. Suppose $w\left(y\right) \notin V^\infty$ for some y. Because V^∞ is closed, there exists a compact neighborhood $\mathcal{N}$ of $w\left(y\right)$ such that $\mathcal{N} \cap V^\infty = \varnothing$. Because $w^k \to w$, there exists $K \in \mathbb{N}$ such that $w^k\left(y\right) \in \mathcal{N}$ for all $k \geq K$, and so, for all $k > K$,

$$\mathcal{N} \cap \left(\cap_{k' \leq k} V^{k'}\right) \neq \varnothing.^8$$

The collection $\left\{\mathcal{N}, V^k : k \in \mathbb{N}\right\}$ has the finite intersection property, so $\mathcal{N} \cap V^\infty \neq \varnothing$, a contradiction. □

It follows immediately that:

Corollary 3 *It holds that*

$$E_\delta = V^\infty,$$

and E_δ is compact.

This suggests an immediate numerical algorithm for computing E_δ (see Judd, Yeltekin and Conklin, 2003). But comparative statics follow as well from this characterization, such as:

[8] Because $w^{k+1}(y) \in V^k = \cap_{k' \leq k} V^{k'}$.

Lemma 2 *If $W \subset \mathbb{R}^n$ is bounded, convex, and self-generating for δ, then it is also self-generating for $\delta' > \delta$.*

Proof: Suppose that $v \in W \subset B_\delta(W)$ is decomposed by (α, w) given δ, and define

$$w' := \frac{\delta' - \delta}{\delta'(1-\delta)} v + \frac{\delta(1-\delta')}{\delta'(1-\delta)} w.$$

Note that w' is in W, as $v, w \in W$ and W is convex. Then

$$\begin{aligned} & (1-\delta')\, u(\cdot) + \delta' \sum_y \pi(y \mid \cdot)\, w'(y) \\ = \; & \frac{\delta' - \delta}{1-\delta} v + \frac{1-\delta'}{1-\delta} \left[(1-\delta)\, u(\cdot) + \sum_y \delta \pi(y \mid \cdot)\, w(y) \right], \end{aligned}$$

and so α is enforced by w' given δ'. □

Therefore, if E_δ is convex, then it is contained in $E_{\delta'}$ for all $\delta' > \delta$. But there are well-known examples in which E_δ is not convex, no matter how large δ is (Yamamoto, 2010). However, this is trivially the case if we assume that players have access to a public randomization device. In fact, we can then assume that w only take as values the extreme points of E_δ, and by Carathéodory's theorem, we only need to randomize over $n+1$ extreme points. Alternatively, APS show that attention can be restricted to extreme points if the set of signals is "large:"suppose that signals are distributed absolutely continuously with respect to Lebesgue measure on a subset of $\mathbb{R}^k$, for some $k \geq 1$. Then if $W \subset \mathbb{R}^n$ is compact, convex and self-generating, we can choose the decomposition (α, w) such that w only takes values on the extreme points of W. While their proof relies on Lyapunov's theorem, a more transparent (but closely related) argument relies on Dubins-Spanier's "fair cake-cutting" theorem:

Theorem 4 *Let $\mu_1, \ldots, \mu_n$ be nonatomic probability measures on a measurable space (S, Σ). Given any $\beta_1, \ldots, \beta_m \geq 0$ with $\sum \beta = 1$, there is a partition $\{E_1, \ldots . E_m\}$ of S such that for all $i = 1, \ldots, n$, and all $j = 1, \ldots, m$, $\mu_i(E_j) = \beta_j$. In fact, there is a sub-σ-algebra $\hat{\Sigma}$ on which all the measures agree, which is rich in the sense that for every $r \in [0,1]$, there is $E \in \hat{\Sigma}$, $\mu_i(E) = r$.*

This means that we can define a random variable that is uniformly distributed, and whose distribution is independent of the action profile chosen (pick $\mu_j = \pi(\cdot|a)$, where j runs over the

set of action profiles). Hence, the result follows. A more surprising result of APS is that, under some additional conditions, the continuation payoff *must* take values in the extreme points of W.

III Characterization as $\delta \to 1$

We now strive to obtain a simpler characterization as $\delta \to 1$. By simpler, we mean: (i) a characterization that does not depend on δ (as we are taking limits, we may hope to be able to do so); (ii) a characterization that is not a fixed-point characterization.

We first define a local version of self-generation (first introduced by Fudenberg, Levine and Maskin, 1994) that simplify some of the analysis. We now index the operator B by the relevant discount factor.

Definition 5 *The set $W \subset \mathbb{R}^n$ is **locally self-generating** if $\forall v \in W, \exists \delta < 1$, and an open neighborhood $\mathcal{N}_v$ of v such that $\mathcal{N}_v \cap W \subset B_\delta(W)$.*

Note that, for compact sets, local self-generation suffices to establish self-generation for some high enough discount factor, as we can take a finite subcover of the open cover $\mathcal{N}_v$, and use as discount factor the highest one for this finite subcover.

To eliminate the discount factor, let us proceed heuristically for now. If α is a Nash equilibrium of the one-shot game $\Gamma_\delta(w)$ with payoff $v \in E_\delta$, then, subtracting δv on both sides of (3) and dividing through by $1-\delta$, we obtain

$$v = u(\alpha) + \sum_{y \in Y} \pi(y \mid \alpha) x(y),$$

where, for all y,

$$x(y) := \frac{\delta}{1-\delta}(w(y) - v), \text{ or } w(y) = v + \frac{1-\delta}{\delta} x(y).$$

Thus, provided that the equilibrium payoff set is convex, the payoff v is also in $E_{\tilde{\delta}}$ for all $\tilde{\delta} > \delta$, because we can use as continuation payoff vectors $\tilde{w}(y)$ the re-scaled vectors $w(y)$ (see Figure 1). Conversely, provided that the normal vector to the boundary of E_δ varies continuously with the boundary point, then any set of payoff vectors $w(y)$ that lie in one of the half-spaces defined by this normal vector (*i.e.*, such that $\lambda \cdot (w(y) - v) \leq 0$, or equivalently, $\lambda \cdot x(y) \leq 0$) must also lie in E_δ for discount factors close enough to one. In particular, if we seek to identify the payoff v

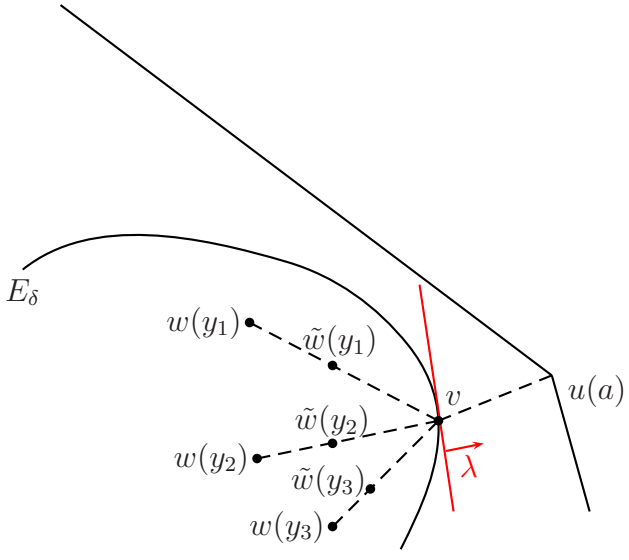


Figure 1: Continuation payoffs as a function of the discount factor

that maximizes $\lambda \cdot v$ on E_δ for δ close enough to 1, given $\lambda \in \mathbb{R}^n$, it suffices to compute the score

$$k(\lambda) := \sup_{x,v,\alpha} \lambda \cdot v,$$

such that α is a Nash equilibrium with payoff v of the game $\Gamma(x)$ whose action sets are A_i and payoff function is given by $u(a) + \sum_{y \in Y} \pi(y \mid a)x(y)$, and subject to the constraints $\lambda \cdot x(y) \leq 0$ for all y. Note that the discount factor no longer appears in this program. Furthermore, the unknown set E_δ no longer appears in the constraints.

We thus obtain a half-space $\mathcal{H}(\lambda) := \{v \in \mathbb{R}^n : \lambda \cdot v \leq k(\lambda)\}$ that contains E_δ for every δ. This must be true for all vectors $\lambda \in \mathbb{R}^n$. Let $\mathcal{H} := \bigcap_{\lambda \in \mathbb{R}^n} \mathcal{H}(\lambda)$. We thus have:

$$\overline{\lim_{\delta \to 1}} \; E_\delta \subset \mathcal{H}.$$

What is more remarkable is that the converse inclusion holds as well, if $\mathcal{H}$ has non-empty interior. That is:

Theorem 5 *If int* $\mathcal{H} \neq \varnothing$, *then*

$$\lim_{\delta \to 1} \; E_\delta = \mathcal{H}.$$

Proof: (Sketch of) The set $\mathcal{H}$ is compact and convex, and so it can be approximated by a

convex set $W \subset \text{int } \mathcal{H}$ whose normal vector to the boundary of W varies continuously with the boundary point. We show that W is locally self-generating. If $v \in \text{int } W$, then there exists an open neighborhood $\mathcal{N}$ of v, $\mathcal{N} \subset \text{int } W$, and $\delta < 1$ such that $\forall v' \in \mathcal{N}$, $v' = (1-\delta)\, u(\alpha) + \delta w$ with α a Nash equilibrium of the stage game and $w \in W$.

Assume now that v is a boundary point. Let $\lambda \in \mathbb{R}^n$ denote the normal vector, and let $k := \lambda \cdot v < k(\lambda)$ (recall that $W \subset \text{int } \mathcal{H}$). Fix (α, w^*) that decomposes some v^* such that $\lambda \cdot v^* \in (k, k(\lambda))$. Note that $\lambda \cdot x^*(y) \leq 0$ (where $x^* := \delta(w^* - v^*)/(1-\delta)$). To make those strict inequalities, pick v' such that $k < \lambda \cdot v' < \lambda \cdot v^*$ and note that (α, w') decomposes v', where

$$w' := v' + \frac{1-\delta}{\delta} x', \quad x' := x^* - v^* + v'.$$

Furthermore, $\lambda \cdot x' \leq \lambda \cdot (v' - v^*) =: -\kappa < 0$. We must find $\varepsilon > 0$, $\delta < 1$ such that, for all $v'' \in W$, $\|v'' - v\| < \varepsilon$, v'' is decomposed by (α, w''), where

$$w'' = v'' + \frac{1-\delta}{\delta}(x' - v' + v''),$$

with $w'' \in W$. The difficult constraint is $w'' \in W$. Note that $\max_y \|w''(y) - v''\| = O(1-\delta)$, while

$$\lambda \cdot w''(y) \leq \lambda \cdot v'' - \frac{1-\delta}{\delta}\kappa.^{9}$$

The existence of such δ, ε then follows from the smoothness of W at v.[10] See Figure 2. □

Note that this result does not only characterize the limit, it establishes the existence of this limit.

This characterization can be adapted to the case in which some of the players are "short-run," *i.e.*, myopic ($\delta = 0$). Suppose that players $i = 1, \ldots, L$, $L \leq n$, are long-run players, whose objective is to maximize the average discounted sum of rewards, with discount factor $\delta < 1$. Players $j \in SR := \{L+1, \ldots, n\}$ are short-run players, each representative of which plays only once. Let

$$B : \times_{i=1}^{n} \triangle A_i \to \times_{j=L+1}^{n} \triangle A_j$$

[9] To see this, note that, since $v'' \in W$, $\lambda \cdot v'' \leq k < \lambda v'$.

[10] All that is required is that the boundary of W has no "kink" at v, as what we need is that points whose scores are at least $\frac{1-\delta}{\delta}\kappa$ lower than the score of a point in W, yet within a distance of order $1-\delta$ of this point, be also in W.

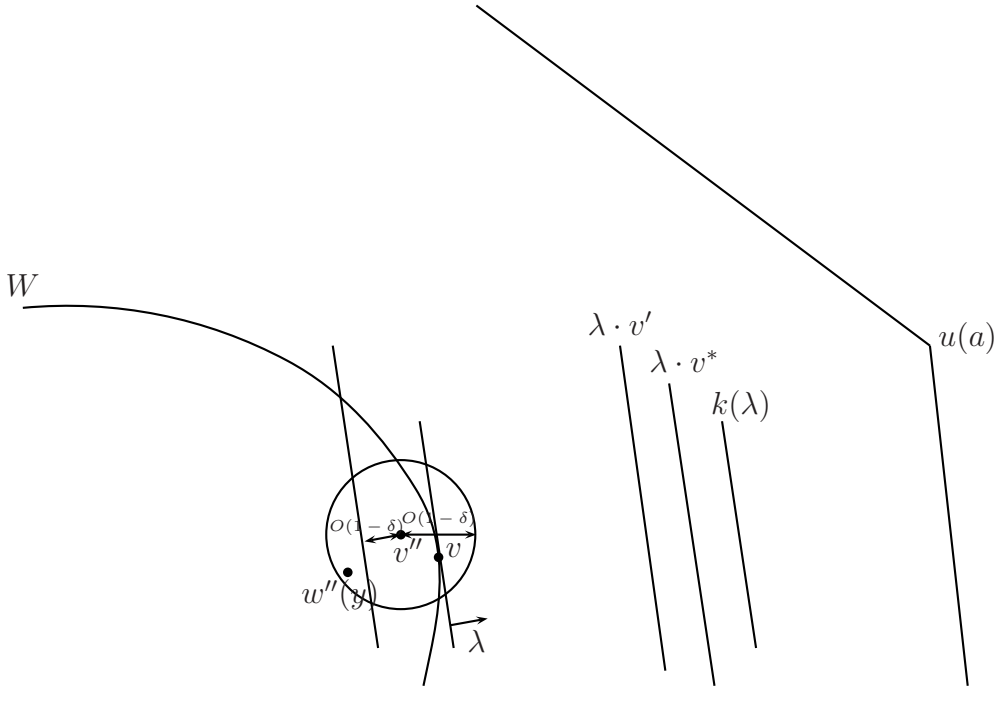


Figure 2: Construction in the sketch

be the correspondence that maps any mixed action profile $(\alpha^1, \ldots, \alpha^n)$ for the long-run players to the corresponding static equilibria for the short-run players in state. That is, for each $\alpha \in$ graph$B$, and each $j > L$, α^j maximizes $u_j(\cdot, \alpha_{-j})$. The characterization goes through if we "ignore" the short-run players and simply require that v be a Nash equilibrium payoff of the game $\Gamma(x)$ for the long-run players, achieved by some $\alpha \in$ graph B.

IV The Folk Theorem

Given our characterization of the limiting payoff set, we are left with the task of finding sufficient conditions under which $\mathcal{H}$ is equal to the feasible and individually rational payoff set. Recall that the main ingredient is the following maximization program $\mathcal{P}(\lambda)$, parameterized by $\lambda \in \mathbb{R}^n$:

$$\sup_{v,x,\alpha} \lambda \cdot v,$$

where the supremum is taken over all $v \in \mathbb{R}^I$, $x : Y \to \mathbb{R}^n$, and $\alpha \in \times_i \triangle A_i$ such that

(i) α is a Nash equilibrium with payoff v of the game $\Gamma(x)$;

(ii) For each $y \in Y$, $\lambda \cdot x(y) \leq 0$.

For a fixed α, the feasible set of $\mathcal{P}(\lambda)$ is non-empty if and only if α is **admissible**, in the sense that, for all i, if there exists $\nu \in \triangle A_i$ such that, for all y,

$$\sum_k \nu^k \pi(y \mid a_i^k, \alpha_{-i}) = \pi(y \mid \alpha),$$

(here, k runs over i's actions) then

$$\sum_k \nu^k u_i(a_i^k, \alpha_{-i}) \leq u_i(\alpha).$$

Indeed, it follows from Fan (1956) that there exists $x : Y \to \mathbb{R}^n$ such that α is a Nash equilibrium of the game $\Gamma(x)$ if and only if α is admissible. Considering any i for which $\lambda_i \neq 0$, we may add (or subtract if $\lambda_i < 0$) a constant to $x_i(y)$, independent of y, so that the constraint **(ii)** is satisfied.

It follows from duality that $\mathcal{P}(\lambda)$ is equivalent to (*i.e.*, gives the same value as) $\tilde{\mathcal{P}}(\lambda)$ given by

$$\sup_{\alpha \in \times_i \triangle A_i, \alpha \text{ admissible}} \min \sum_i \lambda_i u_i \left(\hat{\alpha}_i, \alpha_{-i}\right),$$

where the minimum is over $(\hat{\alpha}_i)_{\{i:\lambda_i \neq 0\}}$, $\sum_k \hat{\alpha}_i \left(a_i^k\right) = 1$, with $\lambda_i > 0 \Rightarrow (\alpha_i(a_i^k) = 0 \Rightarrow \hat{\alpha}_i(a_i^k) \leq 0, \alpha_i(a_i^k) = 1 \Rightarrow \hat{\alpha}_i(a_i^k) \geq 1)$, $\lambda_i < 0 \Rightarrow (\alpha_i(a_i^k) = 0 \Rightarrow \hat{\alpha}_i(a_i^k) \geq 0, \alpha_i(a_i^k) = 1 \Rightarrow \hat{\alpha}_i(a_i^k) \leq 1)$, and

$$\lambda_i \neq 0 \Rightarrow \hat{\pi}\left(y\right) := \pi\left(y \mid \hat{\alpha}_i, \alpha_{-i}\right) \geq 0.$$

Proof: (of the dual representation)
Fix throughout some strategy (α) such that α is admissible. We can rewrite $\mathcal{P}(\lambda)$ as

$$\max_{x,v} \lambda \cdot v$$

over x and v such that, for all i,

$$\sum_y \pi\left(y \mid \alpha\right) x_i\left(y\right) - v_i = -u_i\left(\alpha\right),$$

and, for all i, k,

$$\sum_y \left[\pi\left(y \mid a_i^k, \alpha_{-i}\right) - \pi\left(y \mid \alpha\right)\right] x_i\left(y\right) \leq u_i\left(\alpha\right) - u_i\left(a_i^k, \alpha_{-i}\right),$$

as well as, for all y,

$$\lambda \cdot x(y) \leq 0.$$

This is a linear program for (x, v). The first set of constraints ensure that α yields the payoff v, the second that playing α is a Nash equilibrium, and the third is the same constraint as **(ii)**. Because we assumed that α is admissible, the feasible set is non-empty, and because the value of this program is bounded above by $k(\lambda)$, it is finite. We shall consider the dual of this linear program. It is

$$\min -\sum_i \gamma_i u_i(\alpha) + \sum_{i,k} \nu_i^k \left(u_i(\alpha) - u_i\left(a_i^k, \alpha_{-i}\right)\right)$$

over $\gamma_i \in \mathbb{R}, \nu_i^k \geq 0, \eta_y \geq 0$, such that, for all i, y,

$$\pi(y \mid \alpha)\gamma_i - \sum_k \left[\pi(y \mid \alpha) - \pi\left(y \mid a_i^k, \alpha_{-i}\right)\right] \nu_i^k + \lambda_i \eta_y = 0,$$

where k runs through the actions of player i, and

$$-\gamma_i = \lambda_i.$$

Let $B := \{i : \lambda_i \neq 0, i \in I\}$ and $B' := \{i : \lambda_i = 0, i \in I\}$. Substituting $-\gamma_i = \lambda_i$ into the dual program, we get

$$\min \sum_{i \in B} \left[\lambda_i u_i(\alpha) + \sum_k \left(u_i(\alpha) - u_i\left(a_i^k, \alpha_{-i}\right)\right) \nu_i^k\right] + \sum_{i \in B'} \left[\sum_k \left(u_i(\alpha) - u_i\left(a_i^k, \alpha_{-i}\right)\right) \nu_i^k\right]$$

over $\nu_i^k \geq 0, \eta_y \geq 0$, such that, for all y and $i \in B$,

$$\pi(y \mid \alpha) + \sum_k \left[\pi(y \mid \alpha) - \pi\left(y \mid a_i^k, \alpha_{-i}\right)\right] \frac{\nu_i^k}{\lambda_i} = \eta_y,$$

for all y and $i \in B$,

$$\sum_k \left[\pi(y \mid \alpha) - \pi\left(y \mid a_i^k, \alpha_{-i}\right)\right] \nu_i^k = 0.$$

For $i \in B'$, $\sum_k \left[\pi(y \mid \alpha) - \pi\left(y \mid a_i^k, \alpha_{-i}\right)\right] \nu_i^k = 0$ can be satisfied by setting

$$\nu_i^k = \alpha_i(a_i^k) \geq 0.$$

Moreover, by admissibility,

$$\sum_k \left(u_i(\alpha) - u_i\left(a_i^k, \alpha_{-i}\right)\right) \nu_i^k \geq 0$$

whenever $\sum_k \left[\pi(y \mid \alpha) - \pi\left(y \mid a_i^k, \alpha_{-i}\right)\right] \nu_i^k = 0$. Hence, we can remove the constraints associated with $i \in B'$.

For $i \in B$, define $\xi_i^k := \nu_i^k / \lambda_i$. The dual program can be reduced to

$$\min \sum_{i \in B} \lambda_i \left[u_i(\alpha) + \sum_k \left(u_i(\alpha) - u_i\left(a_i^k, \alpha_{-i}\right)\right) \xi_i^k\right]$$

over $\nu_i^k \geq 0, \eta_y \geq 0$, such that, for all y and $i \in B$

$$\pi(y \mid \alpha) + \sum_k \left(\pi(y \mid \alpha) - \pi\left(y \mid a_i^k, \alpha_{-i}\right)\right) \xi_i^k = \eta_y.$$

Define $\hat{\alpha}_i \in \mathbb{R}^{|A_i|}$ by, for all a_i^k, $i \in B$,

$$\hat{\alpha}_i\left(a_i^k\right) = \alpha_i\left(a_i^k\right) + \alpha_i\left(a_i^k\right) \sum_{k'} \xi_i^{k'} - \xi_i^k.$$

It can be easily verified that

$$\pi(y \mid \hat{\alpha}_i, \alpha_{-i}) = \pi(y \mid \alpha) + \sum_k \left(\pi(y \mid \alpha) - \pi\left(y \mid a_i^k, \alpha_{-i}\right)\right) \xi_i^k.$$

Note that $\sum_k \hat{\alpha}_i\left(a_i^k\right) = 1$ for all $i \in B$. We can rewrite our problem as

$$\min \sum_{i \in B} \lambda_i u_i(\hat{\alpha}_i, \alpha_{-i}),$$

over $(\hat{\alpha}_i)_i$, $\sum_k \hat{\alpha}_i\left(a_i^k\right) = 1$, with $\lambda_i > 0 \Rightarrow (\alpha_i(a_i^k) = 0 \Rightarrow \hat{\alpha}_i(a_i^k) \leq 0, \alpha_i(a_i^k) = 1 \Rightarrow \hat{\alpha}_i(a_i^k) \geq 1)$, $\lambda_i < 0 \Rightarrow (\alpha_i(a_i^k) = 0 \Rightarrow \hat{\alpha}_i(a_i^k) \geq 0, \alpha_i(a_i^k) = 1 \Rightarrow \hat{\alpha}_i(a_i^k) \leq 1)$, as well as, and $\eta_y \geq 0$, such that, for all y and $i \in B$,

$$\pi(y \mid \hat{\alpha}_i, \alpha_{-i}) = \eta_y. \tag{4}$$

Since $\sum_{i \in B} \lambda_i u_i(\hat{\alpha}_i, \alpha_{-i}) = \sum_{i \in B} \lambda_i u_i(\hat{\alpha}_i, \alpha_{-i}) + \sum_{i \in B'} \lambda_i u_i(\hat{\alpha}_i, \alpha_{-i})$ no matter how we define $u_i(\hat{\alpha}_i, \alpha_{-i})$ for $i \in B$, adding back $0 = \sum_{i \in B'} \lambda_i u_i(\hat{\alpha}_i, \alpha_{-i})$ we can rewrite our problem without

using η_y as follows:

$$\min \sum_i \lambda_i u_i \left(\hat{\alpha}_i, \alpha_{-i}\right),$$

over $(\hat{\alpha}_i)_i$, $\sum_k \hat{\alpha}_i \left(a_i^k\right) = 1$, with $\lambda_i > 0 \Rightarrow (\alpha_i(a_i^k) = 0 \Rightarrow \hat{\alpha}_i(a_i^k) \leq 0, \alpha_i(a_i^k) = 1 \Rightarrow \hat{\alpha}_i(a_i^k) \geq 1)$, $\lambda_i < 0 \Rightarrow (\alpha_i(a_i^k) = 0 \Rightarrow \hat{\alpha}_i(a_i^k) \geq 0, \alpha_i(a_i^k) = 1 \Rightarrow \hat{\alpha}_i(a_i^k) \leq 1)$, and

$$\lambda_i \neq 0 \Rightarrow \hat{\pi}\left(y\right) := \pi\left(y \mid \hat{\alpha}_i, \alpha_{-i}\right) \geq 0$$

Taking the supremum over admissible $(\alpha_s)_s$, this gives us precisely $\tilde{\mathcal{P}}(\lambda)$. □

In a slight departure from the notations of the set of notes on perfect monitoring, let $\underline{V}$ denote the set of feasible and weakly individually rational payoff vectors. Denote by e_i the i-th basis vector in $\mathbb{R}^n$. The direction λ is a **coordinate direction** if $\lambda = \lambda_i e_i$ for some $\lambda_i \in \mathbb{R}$, $\lambda_i \neq 0$. It is a **non-coordinate direction** otherwise. We denote the set of non-coordinate directions by Λ^{nc}, and of coordinate directions by Λ^c. Denote such a direction λ^i. Finally, let $Ex(A)$ denote the set of (necessarily pure) action profiles achieving some extreme point of the feasible payoff set V.

Under what conditions does $\lim_{\delta \to 1} E_\delta = \underline{V}$? Assuming throughout that $\underline{V}$ has non-empty interior, it reduces to finding conditions under which, in every direction $\lambda \in \mathbb{R}^n$,

$$k(\lambda) = \max_{v \in \underline{V}} \lambda \cdot v. \tag{5}$$

Depending on the direction λ, the maximum on the right-hand side is achieved either by some $a \in Ex(A)$, or some minmax action profile $\underline{\alpha}^i$ (for coordinate directions $\lambda^i = \lambda_i e_i$, $\lambda_i < 0$).

Therefore, among the set of sufficient conditions for a folk theorem, we start with:

Assumption A1: For every i, some $\underline{\alpha}^i$, is admissible.

Let us define the matrix $\Pi_i\left(\alpha_{-i}\right)$ as the $|A_i| \times |Y|$-matrix whose (a_i, y)-th entry is $\pi\left(y \mid a_i, \alpha_{-i}\right)$. Further, given α, the matrix $\Pi_{ij}\left(\alpha\right)$ is defined as

$$\Pi_{ij}\left(\alpha\right) = \begin{pmatrix} \Pi_i\left(\alpha_{-i}\right) \\ \Pi_j\left(\alpha_{-j}\right) \end{pmatrix}$$

Note that this matrix has maximal rank $|A_i| + |A_j| - 1$, because $\alpha_i \Pi_i\left(\alpha_{-i}\right) = \alpha_j \Pi_j\left(\alpha_{-j}\right)$.

Admissibility for some $\alpha \in \times_i \triangle A_i$ is automatically satisfied if the matrix $\Pi_i\left(\alpha_{-i}\right)$ has full row rank for every i: this is what Fudenberg, Levine and Maskin define as **individual full rank**

for α. But admissibility is clearly a weaker requirement.

If λ is a coordinate direction $\lambda^i = \lambda_i e_i$, $\lambda_i < 0$, no further assumptions are necessary for (5), since player i is the only one whose action the minimum is taken with respect to in the dual, but then again, $\underline{\alpha}^i$ dictates that he takes a best-reply (which is what the minimum calls for, given that $\lambda_i < 0$).

How about other directions? We need to make sure that the only $\hat{\alpha}_i$ allowed by the constraints in the dual is actually α_i (we are interested, of course, in $\alpha = a \in Ex(A)$).

Let $Q_i(a) := \{\pi(\cdot \mid a_{-i}, a_i^k) \mid a_i^k \in A_i \setminus \{a_i\}\}$ be the set of distributions over signals as player i's action varies over all his actions but a_i. Let $C_i(a)$ denote the convex cone with vertex 0 spanned by $Q_i(a) - \pi(\cdot \mid a)$. Note now that the restriction on $\hat{\alpha}$, when $\alpha = a$ is pure, is that $\pi(\cdot \mid \hat{\alpha}_i, a_{-i}) - \pi(\cdot \mid a) \in -C_i(a)$ whenever $\lambda_i > 0$, and $\pi(\cdot \mid \hat{\alpha}_i, a_{-i}) - \pi(\cdot \mid a) \in C_i(a)$ whenever $\lambda_i < 0$. Finally, we need admissibility, for which it suffices that 0 is not a conical combination of $Q_i(a) - \pi(\cdot \mid a)$, *i.e.*, there exists no positive weights on actions of i different that a_i such that the resulting average distribution over signals coincide with the one induced by a_i, given a_{-i}. We thus impose:

Assumption A2: For every $a \in Ex(A)$, every pair of players i, j, $C_i(a) \cap -C_j(a) = C_i(a) \cap C_j(a) = \{0\}$, and 0 is not a conical combination of $Q_i(a) - \pi(\cdot \mid a)$.

The next theorem is now an immediate corollary.

Theorem 6 *(The folk theorem) If (i) $\underline{V}$ has non-empty interior, and (ii) Assumptions **A1** and **A2** hold, then*

$$\lim_{\delta \to 1} E_\delta = \underline{V}.$$

This theorem was proved under slightly stronger assumptions by Fudenberg, Levine and Maskin (1994), and under the currently stated ones by Kandori and Matsushima (1998). The dual representation suggests further possible weakenings, but we shall not do so. On the contrary, we note that, from the dual representation, a stronger assumption can be made that implies **A2**, namely, we may assume that the profile a has **pairwise full rank**: that is, for i and j the matrix $\Pi_{ij}(a)$ has rank $|A_i| + |A_j| - 1$ for all pairs i, j. This is the assumption originally made by Fudenberg, Levine and Maskin (1994).

We may now finally come back to our original example. How come the folk theorem failed? Leaving aside the fact that the characterization of the asymptotic payoff set does not apply in this case ($\mathcal{H}$ has empty interior, as the set of equilibrium payoffs is an interval) the problem

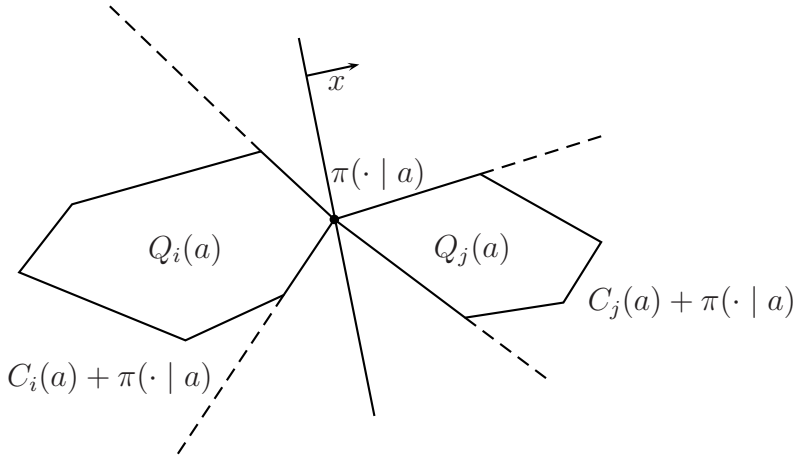


Figure 3: On the role of $C_i(a) \cap C_j(a) = \{0\}$

is that one of the rank conditions failed. In fact, both rank conditions failed: there are only two signals, but a continuum of actions, so clearly players cannot be made indifferent across all actions. But individual full rank is not the main problem, as deviations from full effort can still be statistically detected. The main problem is that players cannot identify deviators (which pairwise full rank requires), and therefore, cannot provide incentives without punishing everyone; the key insight behind FLM is that providing incentives does not require losing any efficiency, if as a function of the signal, the "aggregate payoff" is re-distributed across players so as to keep the sum fixed. This was, as mentioned, the key idea behind pairwise directions and pairwise full rank.

Figures 3 and 4 illustrate the role of the assumptions. These figures represent probability distributions and sets of such distributions. As is clear from Figure 3, if $C_i(a) \cap C_j(a) = \{0\}$, one can find a direction x such that, by punishing one player in that direction while simultaneously rewarding his opponent, incentives are aligned: to avoid the punishment (or reap the reward) both players have an incentive to select a in terms of the actions available to them, in terms of the signal distributions they can generate via unilateral deviations. Similarly, Figure 4 illustrates that, if $C_i(a) \cap -C_j(a) = \{0\}$, one can find a direction x such that incentives are aligned when both players' payments go in the same direction, so that they are both simultaneously punished or rewarded.

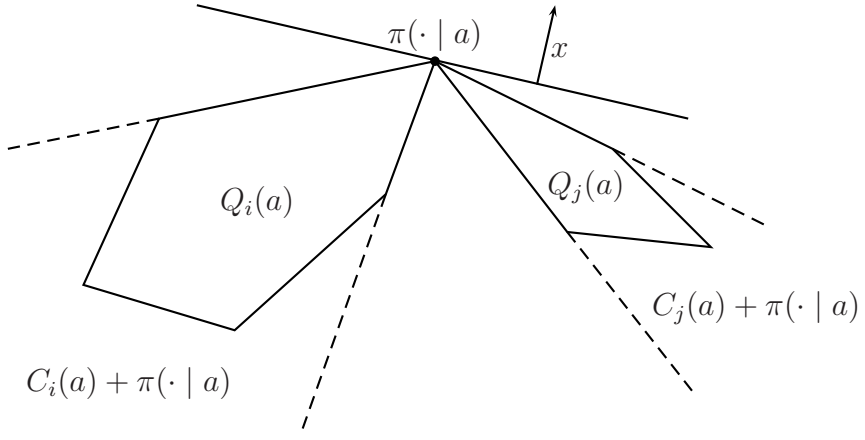


Figure 4: On the role of $C_i(a) \cap -C_j(a) = \{0\}$

V Some Qualifications and Extensions

A What if $\operatorname{int} \mathcal{H} = \varnothing$?

This may happen. Fortunately, the scoring algorithm by Fudenberg and Levine (1994) has been refined by Fudenberg, Levine and Takahashi (2007) to cover this case. The extension is sufficiently intuitive that no proof will be given. Given a set $X \subseteq \mathbb{R}^n$, let $\operatorname{aff} X := \{\sum_{k=1}^K \lambda_k x_k : K \in \mathbb{N}, x_k \in X, \sum_{k=1}^K \lambda_k = 1\}$ denote the affine hull of X. (Note that λ_k can take negative values.) That is, an affine hull is simply the sum of some vector $y \in \mathbb{R}^n$ and a linear subspace $U(X) \subseteq \mathbb{R}^n$, namely $\operatorname{aff} X = y + U(X)$. The dimension of $\operatorname{aff} X$ is defined to be the dimension of $U(X)$. Given some affine hull X, let $\mathcal{P}(\lambda, X)$ be the program

$$k(\lambda, X) := \sup_{x,v,\alpha} \lambda \cdot v,$$

such that α is a Nash equilibrium with payoff v of the game $\Gamma(x)$ whose action sets are A_i and payoff function is given by $u(a) + \sum_{y \in Y} \pi(y \mid a) x(y)$, and subject to the constraints $\lambda \cdot x(y) \leq 0$ for all y, as well as $x(y) \in X$ for all y.

This is a simple modification of the program that we have defined in Section III, with the added constraint that x takes values in X. Let $\mathcal{H}(\lambda, X) := \{v \in \mathbb{R}^n : \lambda \cdot v \leq k(\lambda)\}$ and $\mathcal{H}(X) := \bigcap_{\lambda \in \mathbb{R}^n, \lambda \in U(X)} \mathcal{H}(\lambda, X) \cap X$.

We now define the finite sequence $\{X^k\}_{k=1}^K$ as follows: $X^1 = \mathbb{R}^n$; for $k \geq 1$, if $\dim \mathcal{H}(X^k) =$

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

Figure 5: Prisoner's dilemma

$\dim X^k$, set $X^{k+1} = \mathcal{H}(X^k)$ and $K = k$ (so, stop the sequence); if $\dim \mathcal{H}(X^k) < \dim X^k$, set $X^{k+1} = \text{aff}\mathcal{H}(X^k)$ and continue.

Because at each step of the sequence (as long as it does not terminate) the dimension of X^k diminishes by at least one, the sequence cannot be of length larger than n. It turns out that

$$\lim_{\delta \to 1} E_\delta = \mathcal{H}(X^K).$$

We note that Theorem 5 is the special case in which the procedure ends in one step, because $\text{int}\mathcal{H} = \text{int}\mathcal{H}(\mathbb{R}^n) \neq \varnothing$. As a corollary of this generalization, it follows that $\lim_{\delta \to 1} E_\delta$ always exists.

B Are the Full Rank Assumptions Necessary?

The example of the first section and the ensuing analysis strongly suggests that some conditions such as individual full rank and pairwise full rank might not just be sufficient, but also necessary to support cooperation when incentives in the stage game are misaligned. Yet the restriction to public strategies raises the possibility that these assumptions are driven by the solution concept (PPE) rather than the "fundamentals," and that perhaps these assumptions could be dispensed with a weaker solution concept.

By now, there are some nice examples of games in which private strategies improve on public ones (See, for instance, Kandori and Obara, 2006). Indeed, it is "easy" to construct examples in which cooperation can be supported in the two-player prisoner's dilemma example despite lack of pairwise full rank.[11] Consider the game in Figure 5. Suppose that there are two signals, $y = \underline{y}, \bar{y}$, with the probability of the signal $\bar{y}$ being p if both agents cooperate, q if only one does, and r when neither cooperates. Assume that $1 > p > q > r > 0$, so that the likelihood of $\bar{y}$ increases with the number of cooperators. We assume sufficient noise; specifically, suppose that $2(p-q) < 1-p$ and $r > \max\{2q-p, (3p^2+4q^2-6pq)/p\}$.

[11] This possibility contradicts the famous example of Radner, Myerson and Maskin (1986), but their analysis is restricted to PPE, seemingly unwittingly.

Solving for the limit set (as discounting vanishes) of public perfect equilibrium payoffs – equilibria in public strategies– has become a straightforward exercise, thanks to the characterization of Fudenberg and Levine (1994) and its refinement by Fudenberg, Levine and Takahashi (FLT, 2004). Consider the score in direction $(-1,1)$, namely the maximum of

$$u^2(\alpha) - u^1(\alpha),$$

where $\alpha = (\alpha^1, \alpha^2)$ is the strategy profile played (*i.e.*, α^i is the probability that player i plays C). It is then a simple matter of algebra to check that, given our restrictions on the parameters, maximizing this difference over action profiles α, subject to incentive compatibility and budget-balance yields a score of 0, achieved by setting $\alpha = x = 0$. Fudenberg and Levine's result implies that all payoff vectors (v^1, v^2) must then satisfy $v^2 - v^1 \leq 0$. By considering the direction $(-1,1)$, we then get that it must hold that $v^1 = v^2$: all equilibrium payoffs must be symmetric payoffs. This implies that int $\mathcal{H} = \varnothing$, and Theorem 5 does not apply, although inspection of the proof shows that one direction of the theorem does, namely, $\limsup_{\delta \to 1} E_\delta \subset \mathcal{H}$. This is where the refinement of FLT applies. Loosely speaking, FLT's theorem states that, whenever int $\mathcal{H} = \varnothing$, one should recompute the scores $\mathcal{P}(\lambda)$ by adding one more constraint, namely, the vector x lies in the smallest affine subspace of $\mathbb{R}^n$ that contains $\mathcal{H}$. Taking the intersection over resulting half spaces, we then obtain a "new" set $\mathcal{H}'$ contained in this affine subspace, and provided this one has non-empty interior, *relative* to the affine subspace, we have identified the limit equilibrium payoff set.[12]

Applying this refinement, we may thus re-visit the efficient direction $(1,1)$, and maximize

$$u^1(\alpha) + u^2(\alpha),$$

subject to budget-balance, $x^1(y) + x^2(y) \leq 0$, and symmetry, $x^1(y) = x^2(y)$. Again, simple algebra show that, given the requirement of symmetry, we can do no better than set $\alpha = 0$ and $x = 0$: all equilibrium payoffs must satisfy $v^1 + v^2 \leq 0$, and we are done, as this implies that the unique PPE payoff is $(0,0)$, independently of the discount factor. Yet at the cost of a more complicated construction, one can prove that cooperation can be approached, by relying on ideas inspired by private monitoring.

Yet *some* minimum assumptions are required. Even some version of pairwise full rank is

[12] If $\mathcal{H}'$ does not have non-empty relative interior, this procedure must be iterated, requiring now the constraint that x must lie in the smallest affine subspace of $\mathbb{R}^n$ that contains $\mathcal{H}'$. In this fashion, the limit set of PPE can always be solved for.

	W	S
W	$w-c, w-c, B$	$w-c, w, 0$
S	$w, w-c, 0$	$w, w, 0$

Table 1: Open

	W	S
W	$b, 0, 0$	$b, 0, 0$
S	$b, 0, 0$	$b, 0, 0$

Table 2: Delegate to A_1

	W	S
W	$0, b, 0$	$0, b, 0$
S	$0, b, 0$	$0, b, 0$

Table 3: Delegate to A_2

required, at least if there are three players or more. The following is an example due to Tomala (unpublished). Consider the following matrix game.

The interpretation is as follows. It is a game between a principal (P) and two agents (A_1, A_2). The principal is the owner of a firm and the agents are employees. Each day, the principal decides whether to open the firm or not. If he does, agents may work or shirk, these choices being unobservable. The firm produces an output only when both agents work. An agent who shirks gets his daily wage without paying the effort cost. If the principal decides not to open the firm, he has only two options, delegating the firm to agent A_1 or to agent A_2.

There, B is the expected profit of the principal when he opens the firm and both player work, w is the wage paid to each agent and c is the cost of effort. We assume $w - c > 0$. Then, b is the expected profit of either agent when he operates the firm by himself. We assume that $(Open, W, W)$ is the only surplus-efficient outcome, that is, $B + 2w - 2c > 2w$ (or $B > 2c$) and $B > b$. Two important properties of this one-shot game are:

1. The minmax is 0 for each player, so that all feasible payoffs are individually rational.

2. The set of feasible payoffs has non-empty interior.

The question now is whether the efficient outcome can be (approximately) implemented by an equilibrium of the repeated game when players are patient. Our main point is that, if the principal cannot tell the difference between agent A_1 shirking or agent A_2 shirking, then no matter how he tries to punish, he rewards one of the agents.

Specifically, assume that the action of the principal and the output (*i.e.*, the principal's payoff) are publicly observed. The public signal is given by the following table. We see that, for any (possibly mixed) action profile (a_1, a_2, a_P), the public signals are the same under (S, a_2, a_P) and under (a_1, S, a_P). It is neither essential that signals be public, nor be a deterministic function of actions.

Under this structure, any (not necessarily public perfect) equilibrium payoff $v = (v_1, v_2, v_P)$ satisfies $v_1 + v_2 \geq \min\{2w, b\}$. Let σ be an equilibrium of the repeated game. For $i = 1, 2$, let τ_i be the deviation of agent A_i that plays S always. Under the signaling structure, the distribution

	W	S
W	(O, B)	$(O, 0)$
S	$(O, 0)$	$(O, 0)$

Table 4: Open

	W	S
W	$(D_1, 0)$	$(D_1, 0)$
S	$(D_1, 0)$	$(D_1, 0)$

Table 5: Delegate to A_1

	W	S
W	$(D_2, 0)$	$(D_2, 0)$
S	$(D_2, 0)$	$(D_2, 0)$

Table 6: Delegate to A_2

of signals of the principal is the same under (τ_1, σ_{-1}) as under (τ_2, σ_{-2}). Thus, the actions of the principal also have the same distributions under these two strategy profiles. For each action of the principal $a \in \{O, D_1, D_2\}$, let

$$f_\sigma(a) = \mathbb{E}_\sigma \left[\sum_{t \geq 1} (1-\delta)\delta^{t-1} \mathbf{1}_{\{a_t = a\}} \right]$$

be the expected discounted average number of times where the action a is played. Since the deviations τ_1, τ_2 are not identifiable, we have

$$\forall a \in \{O, D_1, D_2\}, f_{(\tau_1, \sigma_{-1})}(a) = f_{(\tau_2, \sigma_{-2})}(a) =: f(a).$$

Now, $U_i(\tau_i, \sigma_{-i}) = wf(O) + bf(D_i)$, and from the equilibrium condition $v_i \geq U_i(\tau_i, \sigma_{-i})$. It follows that

$$v_1 + v_2 \geq wf(O) + bf(D_1) + wf(O) + bf(D_2) \geq \min\{2w, b\}(f(O) + f(D_1) + f(D_2)) = \min\{2w, b\}.$$

As an immediate corollary, if $2w - 2c < b$, then the efficient outcome $(w - c, w - c, B)$ is not an equilibrium outcome of the repeated game, and equilibrium payoffs are bounded away from efficiency.

C Rates of Convergence

One might conclude from the folk theorems under imperfect public monitoring that, under appropriate rank conditions, it is irrelevant whether monitoring is imperfect or not. While this is certainly a valid viewpoint in the limit as $\delta \to 1$, it must be qualified for a fixed discount factor, and in terms of convergence rates.

First, for a fixed discount factor, Kandori (1992) has shown that as monitoring improves in the sense of Blackwell, the set of PPE weakly expands: hence, better monitoring cannot hurt. This improvement is quantified in Hörner and Takahashi (2015) in terms of convergence rate. They

show that the set of PPE payoffs E_δ approaches the set of individually rational payoff vectors $\underline{V}$ at rate (at least) $(1-\delta)^{1/2}$ under perfect monitoring, and that this rate is tight (namely, examples can be found where this is precisely the rate of convergence). On the other hand, under public monitoring, this rate dips to $(1-\delta)^{1/4}$: under standard individual and pairwise full rank assumptions, $E_\delta \to \underline{V}$ at least as fast as rate $(1-\delta)^{1/4}$,[13] and examples can be found where convergence occurs precisely at this rate. Hence, imperfect monitoring comes at a cost.

VI Literature

The literature on repeated games with imperfect public monitoring was motivated, among others, by Green and Porter 1984 ("Noncooperative Collusion under Imperfect Price Information," *Econometrica*, **52**, 87–100). The initial example, which illustrates the difference between good news and bad news, is due to Abreu, Milgrom and Pearce 1991 ("Information and Timing in Repeated Partnerships," *Econometrica*, **59**, 1713–1733).

Abreu, Pearce and Stacchetti developed the main ideas behind self-generation (Abreu, D., D. Pearce, and E. Stacchetti 1990, "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, **58**, 1041–1063), though similar ideas were introduced in Mertens, J.-F. and T. Parthasarathy 1987 ("Equilibria for Discounted Stochastic Games," C.O.R.E. Discussion Paper 8750). As mentioned, the operator was introduced by Shapley in 1953 ("Stochastic Games," *Proceedings of National Academy of Science*, **39**, 1095–1100).

The topological structure of E_δ is not well-studied. As mentioned in the text, E_δ need not be increasing in δ when one does not assume a public randomization device, no matter whether one restricts attention to arbitrarily high discount factors or not, and a counter-example under perfect monitoring can be found in Yamamoto (2010, "The Use of Public Randomization in Discounted Repeated Games," *International Journal of Game Theory*, **39**, 431–443).

The numerical algorithm mentioned in the text is developed in Judd, Yeltekin and Conklin, 2003 ("Computing Supergame Equilibria," *Econometrica*, **71**, 1239–1254). Finally, the "fair cake-cutting" algorithm that is mentioned in Section II is due to Dubins and Spanier 1961 ("How to cut a cake fairly," *American Mathematical Monthly*, **68**, 1–17). Dubins and Spanier prove the first part of Theorem 4, involving a finite partition. The second part of Theorem 4, involving an arbitrary number $r \in [0,1]$, is due to Border, Ghirardato and Segal 2008 ("Unanimous Subjective Probabilities," *Economic Theory*, **34**, 383–387).

[13] This means that $d(E_\delta, \underline{V}) \leq M(1-\delta)^{1/4}$, for some constant M and all $\delta < 1$.

The limit characterization of equilibrium payoffs based on scores is due to Fudenberg and Levine 1994 ("Efficiency and Observability with Long-Run and Short-Run Players," *Journal of Economic Theory*, **62**, 103–135), who went on to use it, as we have done in these notes, to prove the folk theorem in Fudenberg, Levine and Maskin 1994 ("The Folk Theorem with Imperfect Public Information," *Econometrica*, **62**, 997–1040). The slightly weaker, but also easier to interpret conditions for the folk theorem given here were introduced by Kandori and Matsushima 1998 ("Private Observation, Communication and Collusion," *Econometrica*, **66**, 627–652) in an environment with private monitoring. Kandori 2003 ("Randomization, Communication, and Efficiency in Repeated Games with Imperfect Public Monitoring," *Econometrica*, **71**, 345–353) shows that, if players could in addition communicate, the folk theorem would hold under the same full-dimensionality assumptions as under perfect monitoring. The "technical" result due to Fan used in Section IV on linear inequalities is stated in "Systems of Linear Inequalities," in *Linear Inequalities and Related Systems*, H. W. Kuhn and A. W. Tucker, Ed., Paper 5. Annals of Mathematics Studies, Vol. 38, Princeton: Princeton Univ. Press.

Finally, Fudenberg, Levine, and Takahashi 2007 ("Perfect Public Equilibrium when Players are Patient," *Games and Economic Behavior*, **61**, 27–49) show how the scoring algorithm must be adjusted to account for the possibility that int $\mathcal{H} = \varnothing$, and give a general characterization that applies to that case as well. The dual characterization can be found in Hörner, Takahashi and Vieille 2014 ("On the Limit Perfect Public Equilibrium Payoff Set in Repeated and Stochastic Games," *Games and Economic Behavior*, **85**, 70–83). The famous example showing how PPE can be restrictive, mentioned in Section V, is due to Radner, Myerson and Maskin, 1986 ("An Example of a Repeated Partnership Game with Discounting and with Uniformly Inefficient Equilibria," *Review of Economic Studies*, **53**, 59–69).

The impact of imperfect monitoring is studied in Kandori 1992 ("The use of information in repeated games with imperfect monitoring," *Review of Economic Studies*, **59**, 581–594). The rates of convergence are derived in Hörner and Takahashi (2015, "How fast do equilibrium payoff sets converge in repeated games?" working paper, Yale University).

Repeated Games, Part III: Imperfect Private Monitoring

Lecture Notes, Yale 2015, Johannes Hörner

September 3, 2015

I Motivation

There are at least two possible approaches to oligopoly:

- Green and Porter (1984), and before Bresnahan (1982), model oligopoly as follows (roughly): firms choose quantities, which is private information, and they observe the market price, which is a stochastic function of the aggregate output.

- On the other hand, following Stigler (1964), there might be cases that are better described as follows: firms choose prices, which are private information, and they observe their sales, privately as well, which are a function of the vector of prices. Think for instance, of a few, large customers buying from a few firms. The specific deals are likely not to be publicly disclosed.

These two models have a fundamental difference: in the first case, there is some common, public information, namely the history of market prices. These prices are somewhat random, but they allow nevertheless players to coordinate on them. In the second case, there is no public information whatsoever.

The first example illustrates public monitoring that we have studied before, at least in the case in which players (firms) only condition their actions on the public history. It is worth pointing out here that players could possibly do better if their continuation strategy also depended on their private actions, because it would allow players to make a better inference about the likelihood that a deviation has occurred, given any public signal, the market price. Of course, if the strategy profile is pure, the difference does not really matter, as what actions players take in equilibrium is

known, so that such **private strategies** add nothing. If players randomize after some histories, however, there might be a substantial difference. In fact, the benefit of such private strategies might be good enough a reason for players to randomize in the first place. Such examples are known, but this fascinating topic remains largely unknown.

In this set of notes we shall model the second example, namely when all signals are private.[1] This represents an area in which there are only few results. All methods that were introduced so far turn out to be of no use as far as we know, and we shall have to start from scratch again.

II Notation

We stick to the same notation as before, whenever possible. Players are denoted $i \in N := \{1, \ldots, n\}$. Each of them picks an action $a_i \in A_i$, a finite set. Each player then receives a private signal $y_i \in Y_i$, a finite set as well. Note the important difference with public monitoring: each signal is now indexed by the player who receives it. The private signal of player i contains all the information that player i has, in addition to his knowledge of his own actions played.

We denote action profile by $a = (a_1, \ldots, a_n)$ and signal profile by $y = (y_1, \ldots, y_n)$. Signals are determined by the action profile that is realized. Namely, for each action profile a, $\pi(y \mid a)$ represent the probability of signal profile y. We shall denote by $\pi_i(\cdot \mid a)$ the marginal distribution of π on Y_i, *i.e.* $\pi_i(y_i \mid a) = \sum_{y_{-i}} \pi(y_i, y_{-i} \mid a)$. Given some mixed action profile $\alpha \in \triangle A$, we also define $\pi(\cdot \mid \alpha) := \sum_{a \in A} \pi(\cdot \mid a)\alpha(a)$, where $\alpha(a)$ denotes the probability that α attaches to a.

A monitoring structure is thus a collection $\pi = \{\pi(\cdot \mid a) : a \in A\}$. Some special cases and definitions are introduced below:

- Public monitoring: $Y_i = Y$, and $\pi(y \mid a) = 0$ for all a if $y_i \neq y_j$ for some i, j.

- Perfect monitoring: $Y_i = A$ and $\pi(y \mid a) = 0$ if $y_i \neq a$ for some i.[2]

- ε**-perfect monitoring**: $Y_i = A_{-i}$ and $\pi_i(a_{-i} \mid a) > 1 - \varepsilon$ for all a.

- **Conditional Independence**: $\pi(y \mid a) = \times_i \pi_i(y_i \mid a)$ for all a.

- **Full Support**: $\pi(y \mid a) > 0$ for all y, a.

[1] To the extent that strategies are necessarily private in such an environment, any progress in this area also furthers our understanding of the role of private strategies under public monitoring.

[2] There is of course some flexibility. We could have alternatively defined perfect monitoring as: $Y_i = A_{-i}$ and $\pi(y \mid a) = 0$ if $y_i \neq a_{-i}$ for some i. Viewed this way, perfect monitoring is equivalent to 0-perfect monitoring.

We can start by assuming that payoffs depend on both a player's action and his signal: $g_i(y_i, a_i)$. Of course, in some cases it is natural to assume that his payoff depends directly on the other players' actions, but recall that we assumed that y_i is all the information that a player gets, so if his payoff contained additional information, we should just redefine Y_i appropriately to include it.

We define the expected reward as $u_i(a) = \Sigma_y \pi(y \mid a) g_i(y_i, a_i)$, and we shall fix u_i rather than g_i, even when we vary the monitoring structure π. As in the case of imperfect monitoring, our purpose is to understand how imperfect information affects the ability to achieve cooperative outcomes, not how it affects expected values, given an action profile. So, to keep things conceptually separate, we shall assume that u_i is fixed throughout, and extends its domain to mixed actions as usual.

A repeated game with imperfect private monitoring, then, specifies a collection

$$(N, A, (Y_i)_i, (\pi(\cdot \mid a)_{a \in A}, u),$$

along with a common discount factor $\delta < 1$. As usual, payoffs refer to the average discounted expected sum of rewards.

Notations for histories and strategies must be amended, of course. A private history for player i is an element $h_i^t = \left(y_i^0, a_i^0, \ldots, y_i^{t-1}, a_i^{t-1}\right) \in H_i^t$ (set $H_i^0 := \{\varnothing\}$). Set $H_i := \cup_t H_i^t$. A (behavior) strategy for player i is a collection of maps $\sigma_i = (\sigma_i^t)_{t=0}^{\infty}$, with $\sigma_i^t : H_i^t \to \triangle A_i$. Given some strategy σ_i and private history h_i^t, we shall write $\sigma_i|_{h_i^t}$ for the continuation strategy induced by σ_i, given that history; *i.e.*, $\sigma_i|_{h_i^t}(h_i^\tau) = \sigma_i(h_i^t, h_i^\tau)$. Player i's set of strategies is denoted Σ_i, with $\Sigma := \times_i \Sigma_i$. A strategy profile induces a distribution over infinite profiles of private histories in the natural way.

We shall use sequential equilibrium as a solution concept, but it matters little: if full support is assumed (and we shall typically do so, to avoid technicalities), any Nash equilibrium has a sequential equilibrium with the same equilibrium path, because any player's information set off the equilibrium path must follow the player's own deviation.

Lemma 1 *Under the full support assumption, if $\sigma \in \Sigma$ is a Nash equilibrium of the game with discount factor δ, there exists a sequential equilibrium which has the same equilibrium path.*

Proof: For each $i \in N$, define the pure strategy $\sigma_i' : H_i \to A_i$ as follows: given h_i^t, σ_{-i}, compute i's belief about $-i$'s continuation strategy via Bayes' rule. Because payoffs are linear in strategies,

there exists a pure-strategy best-reply for player i, and define $\sigma_i'(h_i^t)$ as (any of) the pure action(s) that such a pure-strategy best-reply might specify at h_i^t.

Let $H_i' \subseteq H_i$ be the set of histories that are assigned positive probability under σ, and let $\hat{\sigma} \in \Sigma$ be defined as: (i) $\hat{\sigma}_i(h_i^t) = \sigma_i(h_i^t)$ if $h_i^t \in H_i'$, and (ii) $\hat{\sigma}_i(h_i^t) = \sigma_i'(h_i^t)$ otherwise. Because σ_i and $\hat{\sigma}_i$ specify the same actions on path, and full support is assumed, they induce the same beliefs after any history. By construction then, $\hat{\sigma}$ is now sequentially rational on and off-path, and the system of beliefs is clearly consistent. □

III Examples

A A negative result (Matsushima, 1991)

We shall start our analysis with a stark result: strict subgame-perfect equilibria, such as grim-trigger, might fail to be even Nash equilibria for private monitoring structures that are arbitrarily close to perfect.

Consider the standard, two-player prisoner's dilemma (but the result holds for all n-player games with a unique Nash equilibrium).

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} 1,1 & -L, 1+G \\ 1+G, -L & 0,0 \end{array} \right) \end{array}$$

Assume that monitoring is conditionally independent and has full support.

Observe that any strategy profile induces a belief about the opponent's histories that we can compute (it is quite complicated in general). Let us denote by $\mathbb{P}[h_{-i}^t \mid h_i^t]$ the probability that player i assigns to $-i$ having observed h_{-i}^t given that he has observed h_i^t himself. It might be desirable to restrict attention to equilibria that satisfy two intuitive properties:

1. pure strategies (*i.e.*, no randomization after any history);

2. independence of irrelevant information (III): given any two histories $h_i^t, h_i'^t \in H_i^t$, if for all $h_{-i}^t \in H_{-i}^t$,
$$\mathbb{P}[h_{-i}^t \mid h_i^t] = \mathbb{P}[h_{-i}^t \mid h_i'^t],$$

then the continuation strategies after h_i^t and $h_i'^t$ are identical: $\sigma_i|_{h_i^t} = \sigma_i|_{h_i'^t}$.

The second condition requires that a player should play the same way after two histories that lead to the same beliefs about the other players' private histories. Note that these properties are satisfied by grim-trigger under perfect monitoring: it is a pure strategy, and a player's continuation strategy is grim-trigger itself whenever he observes that all players cooperated so far, and defection otherwise. Yet:

Theorem 1 *The only equilibrium σ satisfying the two properties is the repetition of the Nash equilibrium of the stage game after all histories: $\sigma_i(h_i^t) = D$ for all $h_i^t \in H_i$.*

Proof: Suppose that $\sigma \in \Sigma$ satisfies the two properties. Observe that, for all y_{-i}, y_i, and (i) for all $a_{-i}^0 \neq \sigma_{-i}(\varnothing)$,

$$\mathbb{P}[a_{-i}^0, y_{-i} \mid \sigma_i(\varnothing), y_i] = 0,$$

which is independent of y_i, and (ii) for $a_{-i}^0 = \sigma_{-i}(\varnothing)$,

$$\mathbb{P}[\sigma_{-i}(\varnothing), y_{-i} \mid \sigma_i(\varnothing), y_i] = \frac{\pi(y_i, y_{-i} \mid \sigma(\varnothing))}{\pi_i(y_i \mid \sigma(\varnothing))} = \times_{j \neq i} \pi_j(y_j \mid \sigma(\varnothing)),$$

which is independent of y_i as well.[3] So $\sigma_i|_{\sigma_i(\varnothing), y_i}$ is independent of y_i (by *III*): the initial observation that i gets does not affect his beliefs about h_{-i}^t. Similarly, $\sigma_{-i}|_{\sigma_{-i}(\varnothing), y_{-i}}$ is independent of y_{-i}. This means that $\sigma_i(\varnothing)$ must be a best-reply to $\sigma_{-i}(\varnothing)$. So $\sigma(\varnothing)$ is a Nash equilibrium of the stage game. The result follows by induction. □

This result implies that any positive result must relax one of two assumptions: pure strategies, or independence of irrelevant information. We shall see that either relaxation allows for positive results. We shall illustrate this with two examples.

[3]The assumption of pure strategies is used at the step where the ratio is simplified: as you can check, it is typically not true that $\pi(y \mid \alpha) = \times_i \pi_i(y_i \mid \alpha)$ for $\alpha \in \times_i \triangle A_i$ non-degenerate, even under conditional independence.

B Relaxing Pure Strategies (Bhaskar and van Damme, 2002)

We consider the following two-period game, with *no discounting*. In the first period, the game is

$$\begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} 2,2 & -1,3 \\ 3,-1 & 0,0 \end{pmatrix} \end{array}$$

In the second period, the game is given by

$$\begin{array}{c} \\ G \\ B \end{array} \begin{array}{c} \begin{array}{cc} G & B \end{array} \\ \begin{pmatrix} 3,3 & 0,0 \\ 0,0 & 1,1 \end{pmatrix} \end{array}$$

Signal sets are $Y_i = \{c, d\}$, where $\pi_i(c \mid a_i, C) = 1-\varepsilon$, $\pi_i(d \mid a_i, D) = 1-\varepsilon$, for some $\varepsilon \in (0, 1/2)$, and all i, a_i. Further, let us assume conditional independence and full support.

Observe first that with pure strategies, we cannot get (C, C) followed by (G, G), by the reasoning of the previous theorem (without requiring independence of irrelevant information): because the two pure-strategy equilibria of the second period game are strict, and because signals are conditionally independent, player i's action in the second period cannot depend on the signal that he receives if player i uses a pure strategy in the first round. Incentives cannot be provided for players to play C in the first, and so the highest payoff that we can achieve in a pure-strategy equilibrium is $0 + 3$.

Consider now the following mixed-strategy equilibrium: Player i plays C with probability $\mu \in (0, 1)$, to be determined, followed by B if and only if $h_i = (D, d)$.

Let us examine if and when this is an equilibrium. Note that the probability that a player observes Dd given that that the other player observes Dd is given by

$$\mathbb{P}[Dd \mid Dd] = \frac{(1-\mu)^2 (1-\varepsilon)^2}{(1-\mu)(\mu\varepsilon + (1-\mu)(1-\varepsilon))} \to 1 \text{ as } \varepsilon \to 0.$$

So, if ε is small enough, the probability of $-i$ playing B when player i observes (D, d) is close enough to one for B to be the best-reply. Also,

$$\mathbb{P}[Dd \mid Cd] = \frac{(1-\mu)\mu(1-\varepsilon)\varepsilon}{\mu(\mu\varepsilon + (1-\mu)(1-\varepsilon))} \to 0 \text{ as } \varepsilon \to 0.$$

Similarly,

$$\mathbb{P}[Dd \mid Cc] \to 0 \text{ as } \varepsilon \to 0, \mathbb{P}[Dd \mid Dc] \to 0 \text{ as } \varepsilon \to 0,$$

so that, if ε is small enough, the probability of $-i$ playing G is close enough to one conditional on any $h_i \neq (D, d)$. Consequently, playing G is the unique best-reply.

However, we must also make sure that players are indifferent between both actions, for them to be willing to randomize. The payoffs are respectively:

- Payoff from C : $\mu(2+3) + (1-\mu)(-1 + (1-\varepsilon)3 + \varepsilon \cdot 0)$;
- Payoff from D : $\mu(3 + 3(1-\varepsilon) + \varepsilon \cdot 0) + (1-\mu)\left(0 + (1-\varepsilon)^2 \cdot 1 + 3\varepsilon^2\right)$;

These expressions are linear in μ, so we can solve. Observe that, for $\varepsilon = 0$, we get:

$$5\mu + 2(1-\mu) = 6\mu + 1 - \mu, \text{ or } \mu = 1/2.$$

This means that the expected payoff is $(2+3)/2 + (-1+3)/2 = 7/2 > 3$: we do better than without mixing, but not as well as we could hope for (namely $2+3$). Clearly, we can get a solution achieving a nearby payoff for $\varepsilon > 0$ small enough as well.

The good news is that mixing can help. The trick is that by having players randomize in the first period, players entertain uncertainty about the other player's past action. The players' information is conditionally independent given pure action profiles, not mixed ones: after all, player $-i$'s information contains in particular the action that he played, and player i's information contains the private signal that he receives, and as long as signals are informative and both actions are assigned positive probability, this generates correlation between the players' information. Therefore, players are able to coordinate in the second, and this generates incentives.

The bad news is that we are still getting a payoff that is lower than what we could have hoped for. What is the problem? The only punishment we have is severe: namely, $3 - 1 = 2$. If cheating in the first period triggers a punishment –as incentives require that it does– players must be given incentives to cheat: after all, cheating only yields one extra dollar! So, we must lower the probability that the punishment occurs (to lower its expected value), and we do so by only playing the bad equilibrium when both players actually play D.

To get further, assume now that players have access to a randomization device: namely, players commonly observe a public randomization device at the end of the first period. Namely, a draw x from the uniform distribution on $[0, 1]$ is commonly observed.

Now we can scale punishments. Consider the following strategies: Play C in the first period with probability $\mu \in (0,1)$, followed by B if and only if either ($y_i = d$ or $a_i = D$) *and* $x \leq \lambda$ for some λ to be determined. Let us consider the case $\varepsilon \to 0$. The payoffs are:

- Payoff from C : $\mu(2+3) + (1-\mu)(-1+(1-\lambda)3+\lambda\cdot 1)$;

- Payoff from D : $\mu(3+(1-\lambda)3+\lambda\cdot 1) + (1-\mu)(0+(1-\lambda)3+\lambda\cdot 1)$;

Solving, we get:

$$\mu = \frac{1}{2\lambda},$$

so by taking $\lambda \sim 1/2$, we get $\mu \sim 1$, and so the payoff is approximately 5. The point is, in a repeated game, we shall get more freedom to scale punishments, and therefore, we will not need a public randomization any longer.

C Relaxing Independence of Irrelevant Information (Kandori, 1991)

We consider now the following two-period game, with *no discounting*. In the first period, the two players play

$$\begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} 2,2 & -1,3 \\ 3,-1 & 0,0 \end{pmatrix} \end{array}$$

In the second period instead, the game is given by

$$\begin{array}{c} \\ G \\ B \end{array} \begin{array}{c} \begin{array}{cc} G & B \end{array} \\ \begin{pmatrix} 3,3 & 0,2 \\ 4,-2 & -1,-1 \end{pmatrix} \end{array}$$

The second period game admits a unique equilibrium. It is in mixed strategies, with probability $1/2$ assigned to each action, yielding a payoff $(3/2, 1/2)$.

Observe that under perfect monitoring, we can only get $(0,0) + (3/2, 1/2)$ in a subgame-perfect Nash equilibrium: since in the second period, there is a unique equilibrium, it follows by backward induction that players will play the unique Nash equilibrium in the first period as well.

Consider the same monitoring structure as before: $Y_i = \{c, d\}$, where $\pi_i(c \mid a_i, C) = 1-\varepsilon$, $\pi_i(d \mid a_i, D) = 1-\varepsilon$, some $\varepsilon \in (0, 1/2)$, all i, a_i, and assume conditional independence and full support.

Consider the following strategy profile: Play C in the first period; in the second period, play G with probability μ_i (to be determined) in the second period if $y_i = c$, and play B with probability 1 if $y_i = d$. Observe that this violates *III*, as signals are independent, and players know their opponent's past action.

If player i plays C, he expects G to be played with probability

$$(1-\varepsilon)\,\mu_{-i},$$

independently of the signal he receives; so we set $\mu_i := 1/(2(1-\varepsilon))$, which implies that, after C, player 1 is indeed indifferent between both actions. This gives an expected payoff of $2+3/2 = 7/2$.

If player 1 plays D, for small enough ε, he assigns high probability to player 2 playing B in the second period, and his best-reply is then G. So 1's payoff is then

$$3 + 3\varepsilon\mu_2 = 3 + 3\frac{\varepsilon}{2(1-\varepsilon)} \leq \frac{7}{2} \text{ if } \varepsilon \leq 1/4.$$

The same argument applies to player 2.

Hence, we have constructed an equilibrium in which players play C in the first and get $7/2$: note that this is better than what can be achieved under perfect monitoring! The key here is not to create correlation about past information. On the contrary, strategies are "uncoupled:" we use the indifference to specify actions in the second period that react to private signals so as to provide incentives in the first. Player i's strategy in the second period only serves the purpose of "controlling" $-i$'s payoff so as to punish or reward him as a function of the private signal. Players no longer coordinate.

To conclude, the ideas underlying these two examples are very different.

- In the first one, randomization (and almost-perfect monitoring) generates the correlation across signals (that is otherwise absent because of conditional independence) that allows successful coordination. In the second period, on the equilibrium path, players must perform Bayesian inference as a function of their signal to compute their best-reply. Because best-replies depend on those beliefs, this approach has been called the *belief-based* approach in the literature. Successful generalizations to the infinitely repeated prisoner's dilemma include Sekiguchi (1997) and Bhaskar and Obara (2002).

- In the second one, which also uses randomization, but requires moreover that players violate *III* on the equilibrium path, players are actually indifferent between different actions, and

play differently according to their signal nevertheless. Because the set of best-replies in the second period does not depend on the signal received (at least on the equilibrium path), this approach has been referred to as the *belief-free approach*. It has been applied to the prisoner's dilemma by Piccione (2002) and Ely and Välimäki (2002).

The belief-free approach has proved more tractable so far. The next section investigates this approach more systematically.

IV The Belief-free Approach

A The Prisoner's Dilemma

Consider again the two-player prisoner's dilemma in its general form:

$$\begin{array}{cc} & \begin{array}{cc} C & \quad D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} 1,1 & -L,1+G \\ 1+G,-L & 0,0 \end{array} \right) \end{array}$$

where $G - L < 1$, so that (C, C) is the action profile that maximizes the players' sum of payoffs. This game is infinitely repeated with discount factor δ close enough to one (the lower bound will be specified as we proceed). Assume for now that **monitoring is perfect**.

We describe a strategy profile by a finite automaton. After any history, player i can be in one of two states. In the Good state, he plays C and transits in the following period to the other state, the Bad state, if and only if he observes the (perfectly informative) signal d; even then, he does so only with probability p, to be determined. In the Bad state, he plays D and transits in the following period to the Good state if and only if he observes the signal c, and even then, does so only with probability r, to be determined. Think of the initial state as the Good state for now, although we shall re-visit this issue later. See Figure 1.

We claim that we can pick r and p so that each player is indifferent between both actions, independently of his opponent's state. To see this, observe that we can solve the system of equations

$$\begin{aligned} V^G &= (1-\delta)\cdot 1 + \delta V^G = (1-\delta)(1+G) + \delta\left(pV^B + (1-p)V^G\right), \\ V^B &= (1-\delta)\cdot 0 + \delta V^B = (1-\delta)(-L) + \delta\left(rV^G + (1-r)V^B\right) \end{aligned}$$

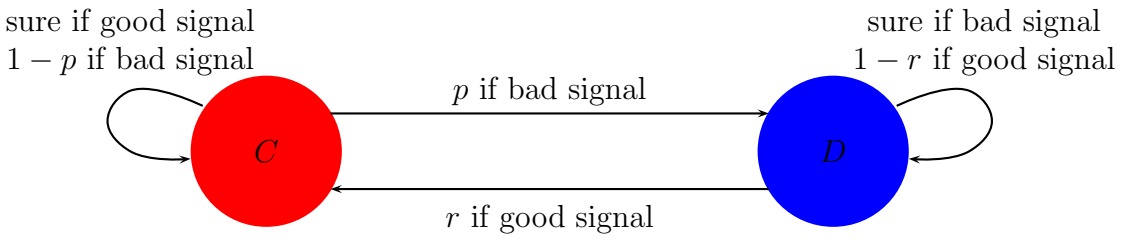


Figure 1: A two-state Automaton

for V^B, V^G, r and p. These are the equations giving the payoff from cooperating and defecting in the Good and the Bad state, and that require these payoffs to be equal (across actions). The solution is actually:

$$V^G = 1, V^B = 0, p = \frac{1-\delta}{\delta}G,\ r = \frac{1-\delta}{\delta}L,$$

which is an admissible solution provided δ is large enough (*i.e.*, $\delta \geq \max\{\frac{G}{1+G}, \frac{L}{1+L}\}$). If player $-i$ starts out in the Good state, player i's payoff in the repeated game is 1. If $-i$ starts out in the Bad state, it is 0. Because players are indifferent between both actions after all histories, they are indifferent between starting in either state. So we can achieve any payoff in $[0,1]^2$ by assuming that player $-i$ appropriately randomizes between the two possible states. See Figure 2.

Now, consider the same system of equations with almost-perfect monitoring. It is easy to see that one can still solve the system of equations (after all, it is linear in the probabilities), and that, furthermore, V^G and V^B continuously tend to 1 and 0 as monitoring becomes perfect.

This is an example of a **belief-free equilibrium**, a sequential equilibrium (leaving aside beliefs) that has the additional feature that each player plays a best-reply independently of the private history observed by his opponents. Formally,

Definition 1 *A strategy profile $\sigma \in \Sigma$ is a belief-free equilibrium if for all h_i^t, h_{-i}^t, $\sigma_i|_{h_i^t}$ is a best-reply to $\sigma_{-i}|_{h_{-i}^t}$.*

While this definition is restrictive, such equilibria always exist: the repetition of any stage-game Nash equilibrium, for instance, is a belief-free equilibrium. Belief-free equilibria have remarkable properties. The definition implies, in particular, that all actions in $\cup_{h_i^t}$supp $s_i(h_i^t)$ (*i.e.*, all actions that are in the support of $s_i(h_i^t)$ for some h_i^t) are best-replies to $\sigma_{-i}|_{h_{-i}^t}$. In the equilibrium of

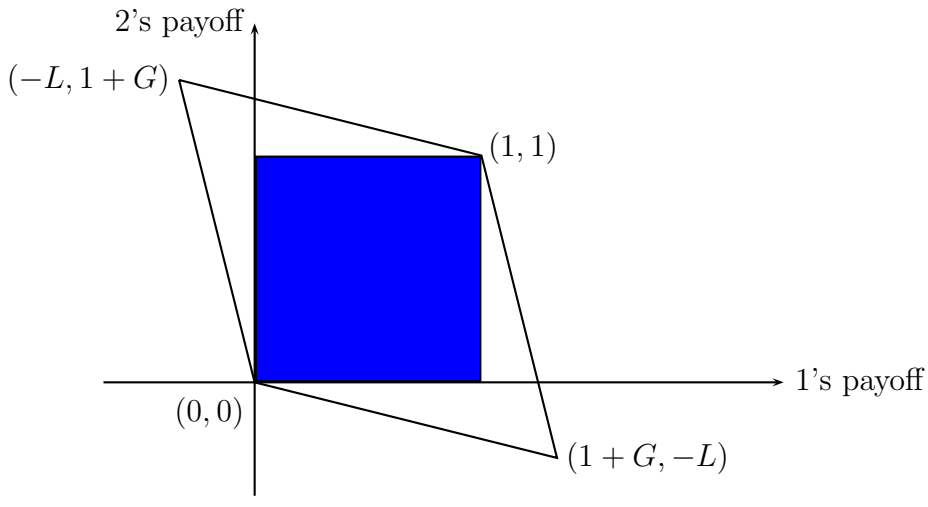


Figure 2: Equilibrium payoffs spanned by the two-state automaton

our example, this set is $\{C, D\}$, the set of all actions, but there is no reason that we need *all* actions to be optimal, or that it must be independent of calendar time. What matters is that this set of optimal actions $\mathcal{A}_i^t \subseteq A_i$ be independent of i's private history. For instance, we could construct a belief-free equilibrium such that, in every prime period, the set of players' optimal actions, independently of the history, is $\{D\}$, and $\{C, D\}$ in all the others.

Nevertheless, it is clear that, in order to construct nontrivial equilibria, we need players to be at least occasionally indifferent over multiple actions, for otherwise there would be no possibility of adjusting continuation strategies to the signals received, thereby conveying incentives.

To guess how to determine *all* belief-free equilibrium payoffs, let us try to construct other non-trivial belief-free equilibria in the prisoner's dilemma, achieving asymmetric payoffs. As usual, it is more convenient to describe such equilibria with the help of a public randomization device (for patient enough players, one would use calendar time as a convexification device). The randomization device allows us to describe the strategies in a stationary fashion. We stick with perfect monitoring, but as in the previous example, the strategies that we will describe are robust to imperfect private monitoring, in the sense that all indifference or strict preference conditions can be satisfied for small enough noise by varying continuously the probabilities of transitions.

Let us call the set of optimal actions $\mathcal{A}^t = \times_i \mathcal{A}_i^t$ the **regime** that applies in period t. In the prisoner's dilemma, there are only nine possible regimes, which makes the analysis manageable. The public randomization device determines a probability distribution over the regimes.

Let us try an equilibrium in which there are the following three regimes:

1. player i has strict incentives to play C, while his opponent has strict incentives to play D a fraction μ of the time (*i.e.*, $\mathcal{A}^t = \{C\} \times \{D\}$);

2. player i has strict incentives to play C, while player $-i$ is indifferent over all actions a fraction λ of the time (*i.e.*, $\mathcal{A}^t = \{C\} \times \{C, D\}$);

3. both players are indifferent over both actions a fraction $1 - \mu - \lambda$ of the time (*i.e.*, $\mathcal{A}^t = \{C, D\} \times \{C, D\}$).

These regimes will be referred to as the first, second and third regime, respectively. Given this specification, which defines what actions players are willing to take, what is the worst that can happen to player i? The worst occurs when player $-i$ plays D whenever the regime allows him to do so: for player 1, he can play D if he wishes to in the third regime, while player 2 can do so both in the second and the third regime. Similarly, the best that can happen to a player is that his opponent plays C whenever this is an action available to him given the regime. As in the previous example, let us refer to these two behaviors as the Bad and Good state, respectively. Each player is then rewarded or punished by his opponent according to which one of these behaviors the other player adopts, and how he transits from one behavior to the other, as a function of the signals. We must ensure that we have enough leeway to provide the appropriate incentives to each player: we must check that he is indifferent across all actions in his regime, and weakly prefers them to all others.

Let us denote player i's payoff in the Good and Bad states V_i^G and V_i^B, respectively. These are the values before the realization of the public randomization device. Let us also write V_{ik}^G, V_{ik}^B for these payoffs given that the regime is $k = 1, 2, 3$.

$$V_1^G = \mu V_{11}^G + \lambda V_{12}^G + (1 - \mu - \lambda) V_{13}^G, V_1^B = \mu V_{11}^B + \lambda V_{12}^B + (1 - \mu - \lambda) V_{13}^B,$$

and similarly for player 2. What is important to note is that this is player 1's payoff *independently* of the specific strategy that he chooses, as long as he keeps picking actions within the regime that the public randomization device determines. Of course, we must verify that he is indeed

indifferent. Considering for instance, the third regime, we may write

$$\begin{aligned} V_{13}^G &= (1-\delta)\cdot 1 + \delta \sum_{y_2} \pi_2(y_2 \mid CC)[p_3(y_2)V_1^B + (1-p_3(y_2))V_1^G] \\ &= (1-\delta)\cdot(1+G) + \delta \sum_{y_2} \pi_2(y_2 \mid DC)[p_3(y_2)V_1^B + (1-p_3(y_2))V_1^G] \end{aligned}$$

where $p_k(y_i)$ refers to the probability with which player i switches to the Bad state, if he starts in the Good state, the regime is k and the signal he receives is y_i (as mentioned, we assume perfect monitoring, but we might as well write the general expression.)

Further, define M_{ik}, m_{ik}, $i = 1, 2$, $k = 1, 2$, by

$$V_{ik}^G = (1-\delta)M_{ik} + \delta V_i^G, V_{ik}^B = (1-\delta)m_{ik} + \delta V_i^B.$$

Substituting, we get

$$\begin{aligned} M_{13} &= 1 - \frac{\delta}{1-\delta} \sum_{y_2} \pi_2(y_2 \mid CC) p_3(y_2)(V_1^G - V_1^B) \\ &= (1+G) - \frac{\delta}{1-\delta} \sum_{y_2} \pi_2(y_2 \mid DC) p_3(y_2)(V_1^G - V_1^B), \end{aligned}$$

or

$$M_{13} = 1 - \sum_{y_2} \pi_2(y_2 \mid CC) x_3^G(y_2) = 1 + G - \sum_{y_2} \pi_2(y_2 \mid DC) x_3^G(y_2),$$

where $x_3^G(y_2) := \delta p_3(y_2)(V_1^G - V_1^B)/(1-\delta)$. Similarly, for the Bad state,

$$m_{13} = \sum_{y_2} \pi_2(y_2 \mid DD) x_3^B(y_2) = -L + \sum_{y_2} \pi_2(y_2 \mid CD) x_3^B(y_2),$$

where $x_3^G(y_2) := \delta r_3(y_2)(V_1^G - V_1^B)/(1-\delta)$, and $r_k(y_i)$ is the probability that player i switches to the Good state, given that he is in the Bad state, that the current regime is k, and that his signal is y_i.

Note that we must have $x_k^G(y_i) \geq 0, x_k^B(y_i) \geq 0$. Ignoring any further constraints on x_k^G, x_k^B, how high (resp. low) can we drive M_{13} (resp. m_{13})? Clearly, $M_{13} \leq 1$, and $m_{13} \geq 0$. Furthermore, under perfect monitoring, we can get both as equalities (for instance, set $x_3^G(y_2) = 0$ for the signal that arises under CC, and set $x_3^G(y_2) = G$ for the signal arising under DC).

To bolster our intuition, let us carry out the same exercise for the first regime. By the same

steps, we get

$$\begin{aligned} M_{11} &= -L - \sum_{y_2} \pi_2(y_2 \mid CD) x_1^G(y_2) \geq 0 - \sum_{y_2} \pi_2(y_2 \mid DD) x_1^G(y_2), \\ m_{11} &= -L + \sum_{y_2} \pi_2(y_2 \mid CD) x_1^B(y_2) \geq 0 + \sum_{y_2} \pi_2(y_2 \mid DD) x_1^B(y_2). \end{aligned}$$

Note that we only need weak preferences here, not equalities. It follows that, with perfect monitoring, $M_{11} = -L, m_{11} = 0$.

A moment's reflection leads to the conclusion that, quite generally, under perfect monitoring,

$$M_i^{\mathcal{A}} = \max_{\alpha_{-i} \in \triangle \mathcal{A}_{-i}} \min_{a_i \in \mathcal{A}_i} u_i(a_i, \alpha_{-i}), m_i^{\mathcal{A}} = \min_{\alpha_{-i} \in \triangle \mathcal{A}_{-i}} \max_{a_i \in A_i} u_i(a_i, \alpha_{-i}),$$

where $M_i^{\mathcal{A}}, m_i^{\mathcal{A}}$ are the relevant bounds given regime $\mathcal{A}$. The intuition is simple: the upper bound is the lowest payoff that i can get, given that he must be willing to take the "worst" action for him in $\mathcal{A}_i$, maximized over the other player's action in $\mathcal{A}_{-i}$; on the other hand, while considering the lower bound, we must ensure that player i does not gain from any action available to him, not only those in $\mathcal{A}_i$.

$$M_{12} = 1, m_{12} = 0,$$

and similarly for player 2,

$$M_{21} = 1 + G, M_{22} = 1, M_{23} = 1, m_{21} = 1 + G, m_{22} = 1 + G, m_{23} = 0.$$

Of course, we need to check that $V_i^G > V_i^B$, or equivalently, that $M_i > m_i$, $i = 1, 2$. This puts bounds on λ and μ, namely

$$\begin{aligned} -\mu L + (1 - \mu) \cdot 1 &> 0, \text{ and } \lambda + (1 - \lambda - \mu) > \lambda (1 + G), \\ \text{or } 1 - \lambda (1 + G) &> \mu, \text{ and } \mu < \frac{1}{1 + L}. \end{aligned}$$

As long as these inequalities are strict, we can then pick δ large enough so that the necessary values of x_k^G, x_k^B are feasible (note that, from their definition, we can achieve a fixed value of x_k^G, x_k^B with an arbitrarily small difference $V_i^G - V_i^B$ provided δ is high enough.)

By choosing μ close to $(1 + L)^{-1}$, we get a payoff vector that tends to $(0, 1 + G/(1 + L))$, which is one of the two asymmetric vertices of the feasible and individually rational payoff set of the prisoner's dilemma.

By varying the values of μ and λ, we can then obtain all missing payoffs from the feasible and individually rational payoff set. We then get a folk theorem for the two-player prisoner's dilemma that is robust to slight perturbations in the monitoring structure.

We have established the following result.

Theorem 2 *Consider the prisoner's dilemma. Fix any v in int $\underline{V}$. There exists $\underline{\delta} < 1$, $\bar{\varepsilon} > 0$, $\forall \delta \in (\underline{\delta}, 1)$, and all ε-perfect monitoring structures, $\varepsilon < \bar{\varepsilon}$, there exists an equilibrium of the repeated game given δ with payoff v.*

B A General Characterization

Fix an arbitrary two-player repeated game with imperfect private monitoring. Let $BF(\delta)$ denote the set of belief-free equilibrium payoffs given δ. Given subsets $\mathcal{A}_i \subseteq A_i$, define

$$M_i^{\mathcal{A}_1 \times \mathcal{A}_2} = \sup v_i$$

such that there exists $\alpha_{-i} \in \triangle \mathcal{A}_{-i}$, $x_i : \mathcal{A}_{-i} \times Y_{-i} \to \mathbb{R}_+$, with

$$v_i \geq u_i(a_i, \alpha_{-i}) - \sum_{a_{-i}} \sum_{y_{-i}} \pi_{-i}(y_{-i} \mid a_i, a_{-i}) \alpha_{-i}(a_{-i}) x_i(a_{-i}, y_{-i}),$$

with equality for all $a_i \in \mathcal{A}_i$. Similarly, define

$$m_i^{\mathcal{A}_1 \times \mathcal{A}_2} = \inf v_i$$

such that there exists $\alpha_{-i} \in \triangle \mathcal{A}_{-i}$, $x_i : \mathcal{A}_{-i} \times Y_{-i} \to \mathbb{R}_+$, with

$$v_i \geq u_i(a_i, \alpha_{-i}) + \sum_{a_{-i}} \sum_{y_{-i}} \pi_{-i}(y_{-i} \mid a_i, a_{-i}) \alpha_{-i}(a_{-i}) x_i(a_{-i}, y_{-i}),$$

Suppose that there exists a distribution p over the sets of possible regimes $\mathcal{A}$ such that

$$\sum_{\mathcal{A}} p(\mathcal{A})(M_i^{\mathcal{A}} - m_i^{\mathcal{A}}) > 0$$

for both i. This is the **positive** case. Let $\mathcal{P}$ denote the set of distributions satisfying this inequality weakly. Then it can be shown that:

$$\lim_{\delta\to 1} BF(\delta) = \bigcup_{p\in\mathcal{P}} \times_{i=1}^{2} \left[\sum_{\mathcal{A}} p(\mathcal{A}) m_i^{\mathcal{A}}, \sum_{\mathcal{A}} p(\mathcal{A}) M_i^{\mathcal{A}}\right]. \tag{1}$$

If instead $\mathcal{P}$ is empty, the limit of $BF(\delta)$ is the convex hull of the static Nash equilibrium payoffs. (Note that the two cases are not exhaustive, so there is a third case that we do not treat here.) In the case in which monitoring is perfect, those results generalize to N players, but those product sets might be outside of the feasible payoff set

$$V(p) := \text{co}\ \left\{\sum_{\mathcal{A}} p(\mathcal{A}) u(a) \mid a \in \mathcal{A}, \mathcal{A}_i \subseteq A_i\right\},$$

and so the statement becomes

$$\lim_{\delta\to 1} BF(\delta) = \bigcup_{p\in\mathcal{P}} V(p) \cap \times_i \left[\sum_{\mathcal{A}} p(\mathcal{A}) m_i^{\mathcal{A}}, \sum_{\mathcal{A}} p(\mathcal{A}) M_i^{\mathcal{A}}\right].$$

in the positive case, which requires $V(p) \cap \times_i [\sum_{\mathcal{A}} p(\mathcal{A}) m_i^{\mathcal{A}}, \sum_{\mathcal{A}} p(\mathcal{A}) M_i^{\mathcal{A}}]$ to be N-dimensional for some p.

There are plenty of games that, even under perfect monitoring, fall into the negative case. Besides, even in the positive case, it is not generally true that the set defined by (1) coincides with the set of feasible and individually rational payoffs. Therefore, belief-free equilibria, as tractable as they might be, do not suffice to establish a folk theorem in general, even under perfect monitoring (except in the prisoner's dilemma!). This leaves open the question of robustness of the folk theorem. Furthermore, even in the special case of the prisoner's dilemma, belief-free equilibria do not achieve efficiency unless monitoring is arbitrarily precise. Nevertheless, we shall see that they are important building blocks.

V Generalizations

A Non-negligible Noise

Our first negative result pertained to monitoring structures that satisfy conditional independence. Our first positive result for non-negligible noise applies precisely to this case.

Assume that the game being played is a two-player prisoner's dilemma (though we know now that this is a very special game). We assume that signals are informative: for all a_{-i}, there exists $y_{-i}^C, y_{-i}^D \in Y_{-i}$ such that

$$\pi_{-i}(y_{-i}^C \mid C, a_{-i}) > \pi_{-i}(y_{-i}^C \mid D, a_{-i}), \text{ and } \pi_{-i}(y_{-i}^D \mid D, a_{-i}) > \pi_{-i}(y_{-i}^D \mid C, a_{-i}). \tag{2}$$

Note that these signals could depend on a_{-i}. Given that signals are noisy, the trick will be to apply periodically some statistical test to make better decisions as to when to switch from the Good state to the Bad state. This requires introducing some notation.

Given a_{-i}, let $f_i^C(r, T, \tau)$ denote the probability that player $-i$ receives exactly r signals equal to y_{-i}^C out of T periods, when $-i$ plays a_{-i} in all T periods and i plays C in exactly τ periods out of the T periods. Given a_{-i}, define $f_i^D(r, T, \tau)$ the same way, replacing C with D, and y_{-i}^C with y_{-i}^D. Let $F_i^C(r, T, \tau) := \sum_{s=0}^{r} f_i^C(s, T, \tau)$, and define similarly F_i^D. A function h on $\tau = 1, \ldots, T$, is **single-peaked** if

$$h(\tau) \geq h(\tau + 1) \Rightarrow h(\tau + 1) \geq h(\tau + 2).$$

Here are some basic facts whose proofs are omitted.

1. $f_i^D(r, T-1, \tau)$ is single-peaked in τ;
2. for all $c > 0$, there exists $\{r_i^D(T)\}_{T=1}^{\infty}$ such that:
   (a) $\lim_{T\to\infty} F_i^D(r_i^D(T), T, 0) = 1$;
   (b) $\lim_{T\to\infty} F_i^D(r_i^D(T), T, T) = 0$;
   (c) $\liminf_{T\to\infty} T \cdot f_i^D(r_i^D(T), T-1, 0) \geq c$.

Next, we define $x_i^C \leq 0$ by

$$u_i(DC) - u_i(CC) = G = x_i^C \cdot \left(F_i^D(r_i^D(T), T, T) - F_i^D(r_i^D(T), T, 0)\right),$$

as well as $z_i^C := 1 + (1 - F_i^D(r_i^D(T), T, 0))x_i^C$. Finally, define $x_i^D \geq 0$ by

$$u_i(DD) - u_i(CD) = L = x_i^D \cdot (\pi_{-i}(y_{-i}^C \mid CD)^T - \pi_{-i}(y_{-i}^C \mid DD)^T),$$

and $z_i^D := \pi_{-i}(y_{-i}^C \mid DD)^T x_i^D$.

It is time to describe the structure of the equilibrium strategies. The horizon will be divided into *review phases* of length T. During those T periods, each player will stick with one action,

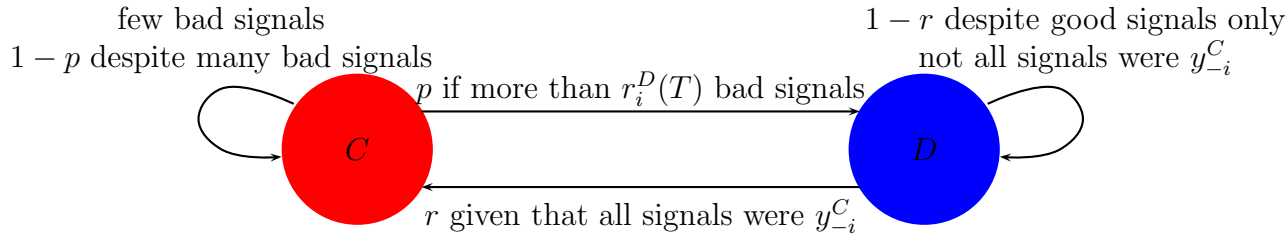


Figure 3: Review Strategies

either C or D. At the end of the phase, as a function of the action that he was playing, and the number of signals (y_i^C, or y_i^D, depending on the case), he will either stick with the same play for the next round, or switch to the other phase.

This is an immediate and natural generalization of the example in the previous section, and so this strategy can loosely be summarized by Figure 3.

In fact, it is enough to assume that transitions to the Good state from the Bad state can only occur if *all* signals were good. On the other hand, transitions to the Bad state occur, with positive probability, if and only if there are too many bad signals.

More formally, in the Good state, player i plays C all the time. A sequence $\{y_i^1, \ldots, y_i^T\}$ of signals received in a round is either an element of $\Omega_i^{Pass} := \{\#y_i^D \leq r_i^D(T)\}$, or of Ω_i^{Fail} otherwise. If the sequence of signals that i receives about $-i$ is in Ω_i^{Pass}, he transits to the Bad state with a probability such that $-i$'s continuation payoff is z_{-i}^C; if it is in Ω_i^{Fail}, this continuation payoff is equal to $z_{-i}^C + \frac{1-\delta^T}{\delta^T} x_{-i}^C$. In the Bad state, player i plays D in all periods and player $-i$ passes the test if and only if all signals that i receives are equal to y_i^C, in which case i transits to the Good state with a probability that gives $-i$ a payoff of $z_{-i}^D + \frac{1-\delta^T}{\delta^T} x_{-i}^D$; otherwise, the test is failed and the continuation payoff is z_{-i}^D.

We claim that (i) player i's payoff if $-i$ is in the Good state is z_i^C, and it is z_i^D in the Bad state; (ii) conditional on either state of $-i$, player i is indifferent between playing C in all periods, or D in all periods, and prefers either to any other strategy.

As for (i), note that, if i plays C in all periods and $-i$ is in the Good state, player i's payoff is

$$(1-\delta^T)\cdot 1 + \delta^T \left(z_i^C + (1 - F_i^D(r_i^D(T), T, 0)) \frac{1-\delta^T}{\delta^T} x_i^C \right),$$

while by playing D all the time he gets

$$(1-\delta^T)(1+G)+\delta^T\left(z_i^C+(1-F_i^D(r_i^D(T),T,T))\frac{1-\delta^T}{\delta^T}x_i^C\right)=z_i^C,$$

and the indifference follows from the definition of x_i^C. Furthermore, note that, by property (a), $\lim_{T\to\infty} z_i^C = 1$. Similarly, if player $-i$ is in the Bad state,

$$z_i^D=\delta^T\left(z_i^D+\frac{1-\delta^T}{\delta^T}\pi_{-i}(y_{-i}^C\mid DD)^T x_i^D\right)=-(1-\delta^T)L+\delta^T\left(z_i^D+\frac{1-\delta^T}{\delta^T}\pi_{-i}(y_{-i}^C\mid CD)^T x_i^D\right),$$

by definition of x_i^D and z_i^D. Note that $\lim_{T\to\infty} z_i^D = 0$.

We now show that no other strategy does better. This is where conditional independence turns out to play a key role. Note that, conditional on either state of the opponent, and hence, conditional on his constant action, signals received by i carry no information about those that player $-i$ has received about i. Therefore, player i has nothing to gain by conditioning his strategy within a round on the sequence of signals that he receives, so that it suffices to consider strategies of i in which his action is a deterministic function of calendar time. Furthermore, given that the tests do not take into account the timing of the signals received, it is always better for player i to play D before C within a round, given discounting, if indeed he plans to play both actions at some point.

We are left with considering strategies that can be described by a sequence $DDD\cdots CCC$, in which player i plays D for τ periods and C for the remaining $T-\tau$ periods. Denote by $V(\tau)$ the payoff from following such a strategy, fixing the state of player $-i$. We already know that $V(0)=V(T)$, and we wish to show that this common value exceeds $V(\tau)$ for all $\tau=1,\ldots,T-1$.

Consider first the case in which player $-i$ is in the Good state (so that he plays C throughout the phase). Then

$$\begin{aligned} V(\tau)-V(0) &= (1-\delta^\tau)G+\delta^T\left(F_i^D(r_i^D(T),T,0)-F_i^D(r_i^D(T),T,\tau)\right)\frac{1-\delta^T}{\delta^T}x_i^C \\ &= G(1-\delta^T)\left(\frac{1-\delta^\tau}{1-\delta^T}-g(\tau)\right), \end{aligned}$$

using the definition of x_i^C, where

$$g(\tau):=\frac{F_i^D(r_i^D(T),T,0)-F_i^D(r_i^D(T),T,\tau)}{F_i^D(r_i^D(T),T,0)-F_i^D(r_i^D(T),T,T)}.$$

We now need the following lemma.

Lemma 2 *The function h defined by $h(\tau) := g(\tau) - \tau/T$, $\tau = 1, \ldots, T-1$, is single-peaked.*

Proof. First, we show that g is single-peaked. Indeed, for $\tau \leq T-1$,

$$\begin{aligned}\Delta g(\tau) := g(\tau+1) - g(\tau) &= \frac{F_i^D(r_i^D(T), T, \tau) - F_i^D(r_i^D(T), T, \tau+1)}{F_i^D(r_i^D(T), T, 0) - F_i^D(r_i^D(T), T, T)} \\ &= \frac{\pi_{-i}(y_{-i}^D \mid D, a_{-i}) - \pi_{-i}(y_{-i}^D \mid C, a_{-i})}{F_i^D(r_i^D(T), T, 0) - F_i^D(r_i^D(T), T, T)} F_i^D(r_i^D(T), T-1, \tau),\end{aligned}$$

and note that the first term is independent of τ, while the second is single-peaked, by the first statistical fact.

Next, observe that $\Delta g(0) \geq f_i^D(r_i^D(T), T-1, 0) \times \text{Constant} \geq \text{Constant}/T$, for T large enough, by the third statistical fact. Because $g(0) = 0$, $g(1) > 1/T$.

Finally, assume that $h(\tau) \geq h(\tau+1)$. Then

$$\Delta g(\tau) < \frac{\tau + 1 - \tau}{T} = \frac{1}{T} \leq \Delta g(0),$$

and so $\Delta g(\tau+1) < \frac{1}{T} = \frac{(\tau+2)-(\tau+1)}{T}$, or $h(\tau+1) > h(\tau+2)$. Therefore, h is single-peaked. ■

Because $V(T) = V(0)$, it follows from Lemma 2 that $V(\tau) - V(0) \leq 0$ for all $\tau = 1, \ldots, T-1$, for δ high enough (note that $(1-\delta^\tau)/(1-\delta^T) \to t/T$).

Second, consider the case in which player $-i$ is in the Bad state (so that he plays D throughout the phase). Then

$$\begin{aligned}V(\tau) - V(0) &= (1-\delta^\tau)L + \delta^T \pi_{-i}(y_{-i}^C \mid CD)^{T-\tau} \left(\pi_{-i}(y_{-i}^C \mid DD)^\tau - \pi_{-i}(y_{-i}^C \mid CD)^\tau\right) \frac{1-\delta^T}{\delta^T} x_i^D \\ &= L(1-\delta^T)\left(\frac{1-\delta^\tau}{1-\delta^T} - g(\tau)\right),\end{aligned}$$

using the definition of x_i^D, where

$$g(\tau) := \pi_{-i}(y_{-i}^C \mid CD)^{T-\tau} \frac{\pi_{-i}(y_{-i}^C \mid DD)^\tau - \pi_{-i}(y_{-i}^C \mid CD)^\tau}{\pi_{-i}(y_{-i}^C \mid DD)^T - \pi_{-i}(y_{-i}^C \mid CD)^T}.$$

To show that $V(\tau) - V(0) \leq 0$ for high enough δ, it will suffice here as well to prove that, for all

$\tau = 1, \ldots, T-1$, $\tau/T - g(\tau) < 0$. Simplifying,

$$g(\tau) = \frac{1-K^\tau}{1-K^T}, \text{ where } K := \frac{\pi_{-i}(y_{-i}^C \mid DD)}{\pi_{-i}(y_{-i}^C \mid CD)} < 1.$$

Note that g is increasing and concave in τ. Hence, because $V(0) = V(T)$, it follows that $\tau/T - g(\tau) < 0$ for all $\tau = 1, \ldots, T-1$, and so $V(\tau) - V(0) \leq 0$ for all $\tau = 1, \ldots, T-1$, for δ high enough.

Putting all together, we have shown that

Theorem 3 *Consider the prisoner's dilemma. Assume that monitoring is conditionally independent, and that signals are informative, i.e.* (2) *is satisfied. Fix any $v \in (0,1)^2$. There exists $\underline{\delta} < 1$ such that, for all $\delta \in (\underline{\delta}, 1)$, there exists an equilibrium of the repeated game given δ with payoff v.*

The result can be generalized to asymmetric payoffs as well, and to other stage games. Quite generally, with such a construction, any payoff that can be obtained as a belief-free equilibrium payoff under perfect monitoring for high enough discount factors, can also be obtained as an equilibrium payoff under private, but conditionally independent monitoring, provided that signals are informative, for high enough discount factors.

This, however, is not satisfactory on two accounts. As we have mentioned, the prisoner's dilemma is a very special stage game, and belief-free equilibria do not suffice for the folk theorem under perfect monitoring in general. Second, conditional independence is a very special monitoring structure.

B Is the Folk Theorem Robust?

An issue that was repeatedly mentioned is whether the folk theorem under perfect monitoring is robust to slight, private, perturbations in the monitoring structure. By now, it is known that the following theorem holds:

Theorem 4 *Consider any stage-game (possibly with more than 2 players). Fix any v in int $\underline{V}$. There exists $\underline{\delta} < 1$, $\bar{\varepsilon} > 0$, $\forall \delta \in (\underline{\delta}, 1)$, and all ε-perfect monitoring structures, $\varepsilon < \bar{\varepsilon}$, there exists an equilibrium of the repeated game given δ with payoff v.*

The proof of this result is too lengthy for this set of notes. Nevertheless, this result is based on belief-free equilibria too (though obviously it must involve equilibria that are not belief-free).

It is roughly based on the following insight: replace the stage game by the normal form of the finitely repeated game, for some length to be carefully chosen. Then, given a payoff $v \in \text{int } \underline{V}$ to be achieved, there exists $\left\{T, \mathcal{S}_i, s_i^G, s_i^B\right\}_{i=1,2}$, with $T \in \mathbb{N}$, $\mathcal{S}_i \subset \Sigma_i^T$ (the set of strategies in the finitely repeated game), and s_i^G, $s_i^B \in \mathcal{S}_i$ such that, for all δ close enough to one,

$$\min_{\mathcal{S}_i} v_i\left(s_i, s_{-i}^G\right) > v_i > \max_{\Sigma_i^T} v_i\left(s_i, s_{-i}^B\right),$$

where v_i is the payoff in the T-period finitely repeated game.

Strategy s_{-i}^G is the "Good" strategy that secures player i at least v_i on average over T periods, provided only player i uses some strategy within $\mathcal{S}_i$. Strategy s_{-i}^B is the "Bad" strategy that keeps player i's average payoff below v_i, independently of i's strategy $s_i \in S_i^T$. In each *block* of the supergame, player $-i$ uses either s_{-i}^G or s_{-i}^B. By suitably choosing the probability with which player $-i$ sticks to or changes his finitely-repeated game strategy from one block to the next, as a function of his observations in the last block only, he ensures that player i is indifferent across all the elements in $\mathcal{S}_i$ at the beginning of each block, independently of his private history, provided only that noise and discounting are low enough. In turn, because player i is indifferent across these elements, it is optimal for him to condition his choice of s_i^G or s_i^B within each block on his observations in the last block. The payoff v_i is then exactly achieved by specifying appropriately the probability that player $-i$ plays s_{-i}^G in the initial block.

Thus, the time horizon is divided into T-period blocks. These are not review phases, as their purpose is not to improve on the quality of the signal (indeed, the theorem assumes that signals are arbitrarily precise). Rather, they are chosen so that the normal-form of the finitely repeated games has properties similar to the one-shot prisoner's dilemma.

In equilibrium, any strategy of player i that adheres within each future block to an element of $\mathcal{S}_i$ is optimal, independently of player $-i$'s history. More precisely, let $s_i'^n \mid h_i^{nT}$ denote the restriction of $s_i' \mid h_i^{nT}$ to the $(n+1)$-st block. Given s_{-i}, any strategy s_i' such that

$$\forall m \geq n, \forall h_i^{mT} \quad s_i'^m \mid h_i^{mT} \in \mathcal{S}_i,$$

for all histories h_i^{mT} following history h_i^{nT}, yields an optimal continuation strategy $s_i \mid h_i^{nT}$, independently of h_{-i}^{nT}.

In the prisoner's dilemma, it is enough to pick $T = 1$: this is the construction that we have see in Section A. In general, T depends both on the stage game and the payoff vector v. When $T > 1$, a *block equilibrium* need not be belief-free, as a player's set of optimal actions within

a block may depend on his private history. However, this dependence is limited to the *recent history* –the finite, terminal segment of the player's private history of those actions taken and signals observed within the current block.

Because block equilibria need not be belief-free, sequential rationality within each block raises difficulties under imperfect private monitoring, affecting the way $\mathcal{S}_i$, s_i^G, s_i^B and T are defined. Because of these difficulties, these strategies are actually not explicitly specified, but their existence follows from the application of a fixed-point theorem.

This result actually establishes that the limit set of equilibrium payoffs (as monitoring noise vanishes and $\delta \to 1$) includes $\underline{V}$. The converse inclusion is not obvious, as all sequential equilibria are considered. In fact, it is known that the converse inclusion is not generally true: one can construct equilibria achieving payoffs that drive a player's payoff *below* his minmax payoff. Necessary and sufficient conditions on the monitoring structure for this not to happen are given in Gossner and Hörner (2010), and are satisfied for the case of vanishing noise, as long as one defines almost-perfect monitoring within structures for which $Y_i = A_{-i}$, as we have (see Notation, Section II).

Similarly, one might wonder whether the folk theorem under imperfect public monitoring is robust. The answer is also known to be a qualified yes, provided that the rank assumptions are strengthened, and, in the case of two players, one replaces the minmax payoff by some static Nash equilibrium payoff as a lower bound on the set of payoffs to be achieved.

C Open Questions

There are many. Of course, one would like to find natural assumptions under which the folk theorem holds under private monitoring. There are a few results that improve on those mentioned in these notes, but they either make some assumption on the correlation of signals, or impose a lower bound on the number of signals (larger than what is required under public monitoring). While these results confirm that the folk theorem extends beyond the case of conditional independence, the set of assumptions that are made remain unsatisfactory, as they are tied to a very specific proof strategy, and do not reduce to the standard assumptions under public monitoring.

It seems unlikely that such assumptions will be found without getting a handle on the structure of equilibrium payoffs first. As in the case of public monitoring, finding a recursive structure requires defining the solution concept simultaneously (perfect public equilibrium, in the case of

public monitoring). While it can be shown that belief-free equilibria satisfy properties that parallel those of perfect public equilibria, they are too restrictive.

This is the main challenge that private monitoring raises: what "state" variables can we use to write equilibrium payoffs recursively? Players' continuation strategies depend on their private histories, and must be sensitive to those histories if incentives are to be provided. But these strategies must be best-replies to the opponents' strategies, and so it seems that we must determine the player's belief about these continuation strategies, and hence about the private histories that the other players have observed, conditional on one's own private history. Belief-free strategies eschewed the problem by requiring strategies to be optimal independently of those. This is too strong. But these beliefs cannot be directly used as state variables: first, the support of these beliefs are private histories, which are elements of a set that varies over time, as histories become longer and longer. Second, given these beliefs, strategies form a correlated equilibrium of the game, not a Nash equilibrium, and worse, this is only true on the equilibrium path: as soon as a player deviates, the distribution over players' beliefs is no longer common knowledge.

There have been recent attempts to weaken the notion of belief-free equilibria, and to find computational methods to deal with belief-based equilibria. All these attempts, however, only provide ways of *checking* whether a given strategy profile is an equilibrium or not. Given that we are unable to describe the equilibrium strategies that achieve the folk theorem even in the case of public monitoring, such results, as remarkable as they are, remain far from what is needed.

VI Literature

The motivating papers from industrial organization that were mentioned in the introduction are Green, Edward J. and Porter, Robert H., 1984 ("Noncooperative Collusion under Imperfect Price Information," *Econometrica*, **52**, 87–100), T. Bresnahan, 1982 ("The oligopoly solution concept is identified," *Economics Letters*, **10**, 87–92), and Stigler, G., 1964 ("A Theory of Oligopoly," *Journal of Political Economy*, **72**, 44–61.) A recent application of private monitoring to international trade is Park, J.-H., 2014. "Enforcing International Trade Agreements with Imperfect Private Monitoring," *Review of Economic Studies*, **81**, 473–500.

The negative result mentioned first is due to Matsushima, H., 1991 ("On the theory of repeated games with private information : Part I: anti-folk theorem without communication," *Economics Letters*, **35**, 253–256.). Half a decade passed before someone constructed actually a non-trivial equilibrium for the prisoner's dilemma. This feat was accomplished (under some

restrictions) by Sekiguchi, T., 1997 ("Efficiency in Repeated Prisoner's Dilemma with Private Monitoring," *Journal of Economic Theory*, **76**, 345–361), a PhD student at the time. Before that, a two-period example had been constructed by M. Kandori, 1991 ("Cooperation in finitely repeated games with imperfect private information," mimeo), which is the one presented in Section III.C. The example in Section III.B. is due to V. Bhaskar and E. van Damme, 2002 ("Moral hazard and private monitoring," *Journal of Economic Theory*, **102**, 16–39).

The first example of a belief-free equilibrium in an infinitely repeated game is due to Piccione, M., 2002 ("The Repeated Prisoner's Dilemma with Imperfect Private Monitoring," *Journal of Economic Theory*, **102**, 70–83). This example was drastically simplified by Ely, J. and J. Välimäki, 2002 ("A Robust Folk Theorem for the Prisoner's Dilemma," *Journal of Economic Theory*, **102**, 84–105). Bhaskar, V., and I. Obara (2002, "Belief-Based Equilibria in the Repeated Prisoners' Dilemma with Private Monitoring," *Journal of Economic Theory*, **102**, 40–69) also proved the folk theorem for the prisoner's dilemma under almost-perfect monitoring by using a construction that is not belief-free and generalizes Sekiguchi's construction. A nice survey of the results up to 2002 can be found in Kandori, M. (2002), "Introduction to Repeated Games with Private Monitoring, *Journal of Economic Theory*, **102**, 1–15.

The systematic analysis of belief-free equilibria is due to Ely, J., J. Hörner and W. Olszewski, 2005 ("Belief-free Equilibria in Repeated Games," *Econometrica*, **73**, 377–415) for two players. The generalization of the characterization to more players, for the case of almost-perfect monitoring, is due to Yamamoto, Y., 2007 ("A Limit Characterization of Belief-Free Equilibrium Payoffs in Repeated Games," *Journal of Economic Theory*, **144**, 802–824.) Yamamoto also generalized the folk theorem under almost perfect monitoring to the case of the n-players prisoner's dilemma ("Efficiency Results in N Player Games with Imperfect Private Monitoring," (2007), *Journal of Economic Theory*, **138**, 382–413.)

Review strategies were introduced by Radner, R., 1986 ("Repeated Partnership Games with Imperfect Monitoring and No Discounting," *Review of Economic Studies*, **53**, 43–58), and Matsushima, H., 2004 ("Repeated Games with Private Monitoring: Two Players," *Econometrica*, **72**, 823–852) is the one who thought of combining them with the construction of Ely and Välimäki in order to extend results from the case of almost-perfect monitoring to the case of conditionally independent private, but not almost-perfect monitoring. This construction was extended somewhat by Ely, Hörner and Olszewski, 2005 (Section V. A. is based on it) and more substantially, recently by Yamamoto, 2011 ("Repeated Games with Private and Independent Monitoring," mimeo, Harvard) to more general games –but still under the restriction to conditionally inde-

pendent monitoring.

The robustness of the folk theorem under perfect monitoring to almost-perfect monitoring was established in Hörner, J. and W. Olszewski, 2006 ("The Folk Theorem for Games with Private Almost-Perfect Monitoring," *Econometrica*, **74**, 1499–1544), which, as mentioned, does not use belief-free equilibria *per se*, but involves a construction that bears a strong similarity with such strategies. This paper did not show that the set of equilibrium payoffs is not larger than $\underline{V}$, the set of feasible and individually rational payoffs, and surprisingly, it might actually be. However, if the definition of almost-perfect monitoring assumes $Y = A_i$ (as was assumed in Section II of these notes, but almost-perfect monitoring can be defined more generally), this follows from Gossner, O. and J. Hörner, 2010 ("When is the lowest equilibrium payoff in a repeated game equal to the minmax payoff?" *Journal of Economic Theory*, **145**, 63–84.)

Building on Mailath, G. J., and S. Morris, 2002 ("Repeated Games with Almost-Public Monitoring," *Journal of Economic Theory*, **102**, 189–228), Hörner, J. and W. Olszewski, 2009 ("How Robust is the Folk Theorem with Imperfect Public Monitoring?" *Quarterly Journal of Economics*, **124**, 1773–1814) show also the folk theorem under imperfect public monitoring is robust to small perturbations.

Recently, some papers have tried to extend the folk theorem to private monitoring without requiring conditional independence. Building among others on insights of Obara, I., 2009 ("Folk Theorem with Communication," *Journal of Economic Theory*, **144**, 120–134), Sugaya, T., 2010 ("Belief-Free Review-Strategy Equilibrium without Conditional Independence," mimeo, Princeton University) shows that it holds generically when sufficiently many signals are available –but the bound is larger than what we know is appropriate under imperfect public monitoring, so we have not yet found the adequate generalization of FLM to private monitoring. Indeed, Fong, Gosnner, Hörner and Sannikov 2011 ("Efficiency in the Repeated Prisoner's Dilemma with Imperfect Private Monitoring," mimeo, Yale University) show that it is possible to get efficiency in the prisoner's dilemma without conditional independence with fewer signals. Unfortunately, their construction requires a lower bound on the informativeness of the signals. Sugaya 2013 ("A general folk theorem with private monitoring," mimeo, Stanford University) shows that, provided there are sufficiently many signals, this difficulty can be overcome.[4]

General recursive methods are still missing. Belief-free equilibria were generalized somewhat by Kandori, M., 2011 ("Weakly Belief-Free Equilibria in Repeated Games With Private Monitor-

[4]His paper attempts to prove a general folk theorem under private monitoring. Unfortunately, several of the steps in the proof are difficult to understand. If correct, it is a very important paper.

ing," *Econometrica*, **79**, 877–89), but unlike for belief-free equilibria, there is no characterization of the equilibrium payoff set for such weakly belief-free equilibria. More generally, some methods for *checking* whether some given strategy profiles are equilibria under private monitoring for some private monitoring structure are known. See Kandori, M. and I. Obara, 2010 ("Towards a Belief-Based Theory of Repeated Games with Private Monitoring: An Application of POMDP," mimeo, Tokyo University) and Phelan, C. and A. Skrzypacz, 2012 ("Beliefs and Private Monitoring," *Review of Economic Studies*, **79**, 1637–1660). Unfortunately, one must guess the strategy profile (as represented by a finite automaton) first, and as is known, a folk theorem will require to consider strategy profiles whose representation through finite automata requires arbitrary many states, making such methods only a very first step in a general characterization.

Results for the undiscounted case are much cleaner. See Lehrer, E., 1990 ("Nash Equilibria of n-player Repeated Games with Semi-Standard Information," *International Journal of Game Theory*, **19**, 191–217).

Repeated Games, Part IV: Stochastic Games

Lecture Notes, Yale 2015, Johannes Hörner

September 3, 2015

Stochastic games generalize repeated games, by allowing the stage game to depend on some "state" variable. Despite their name (coined by Shapley, who introduced these games in 1953), the evolution of this variable need not be random, although stochastic transitions are allowed. For instance, alternating move games are stochastic games.

Importantly, transitions may depend on actions –capital stock depends on prior investment, unemployment on previous macroeconomic shocks, the party in power on the performance of the previous incumbent, etc.

We will restrict ourselves to finite stochastic games, in which states, actions and public signals are drawn from finite sets. For now also, we assume that the state that prevails is common knowledge.

Existence of equilibrium under discounting can then be established by focusing on the counterpart of the repetition of static Nash equilibrium in repeated games (the easiest way to prove equilibrium existence in repeated games as well), namely, so-called *Markov* equilibria, in which actions only depend on the current state, and ignore all other aspects of the history (see Fink, 1964, and Takahashi, 1962). Solving for such equilibria is important, as it helps us understand the basic structure of the game we are dealing with –just as it is important to know what the stage game Nash equilibria in a repeated game are. But focusing on those (as is often done in applications) would be equivalent to focusing on perpetual defection in the repeated prisoner's dilemma, and we are led to investigate subgame-perfect Nash equilibria, perfect public equilibria, etc., of a stochastic game just as we have done in repeated games.

	$\overline{0}$	$\overline{1}$
$\overline{0}$	1	0
$\overline{1}$	0^*	1^*

Figure 1: Big Match

I Two Famous Examples of Non-Irreducible Games

A The Big Match

This is an example of Gillette (1957) whose solution is due to Blackwell and Ferguson (1968). This is a zero-sum game. The game matrix is as follows. The superscript "$*$" refers to the fact that, for those such designated action profiles, play remains "frozen" in that action profile thereafter. That, is there are three "states," two of which are absorbing. As soon as player 1 plays $\overline{1}$, one of the two absorbing states is reached, according to player 2's action, with payoffs no longer evolving from that point on. The game starts in the "non-absorbing" state corresponding to the payoff matrix.

As a motivation, think about player 1 trying to guess player 2's action. Whenever his guess is correct, he gains \$1. However, if he ever predicts $\overline{1}$, then the game "ends," and the ensuing rewards are equal to the one obtained in that round.

Blackwell and Ferguson (1968) were not concerned with discounted payoffs, unlike us, but rather with the limit of means. Nonetheless, we will show that in their game, as far as the value is concerned, it makes no difference, but it makes an important difference in terms of strategies.

We start with the "easy" criterion: discounting. In that case, it is not entirely obvious but nonetheless intuitive that players have *stationary* strategies, as long as the game is not observed, and we will assume as much. That is, given δ, there exists a value, v, and actions $\alpha_1 \in \Delta(A_1)$, $\alpha_2 \in \Delta(A_2)$, or equivalently probabilities $p, q \in [0,1]$ of playing $\overline{0}$, such that

$$\begin{aligned} v &= \max_p \{(1-\delta)pq + \delta pv + \delta(1-p)(1-q)\}, \\ \text{and } v &= \min_q \{(1-\delta)pq + \delta pv + \delta(1-p)(1-q)\}. \end{aligned}$$

It is fairly obvious that neither plays wants to play a pure strategy, and hence must be indifferent between their two actions, so that

$$v = (1-\delta)q + \delta v = (1-q) = (1-\delta)p + \delta pv = \delta pv + (1-p),$$

or

$$v = q = 1/2, \text{ and } p = (2-\delta)^{-1}.$$

We note that, as $\delta \to 1$, $v, q \to 1/2$, and $p \to 1$: in the limit, player 1 plays T for sure.

Our focus in these lectures is on the discounted criterion, but it is instructive to take a short detour here. Consider now the undiscounted game, with payoffs defined as (for player 1)

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} u(a_t).$$

Unlike in the discounted case, we argue that it is not sufficient to consider stationary strategies, namely, denoting the corresponding expected limit payoff $V(p,q)$, it holds that

$$\max_p \min_q V(p,q) = 0 < \frac{1}{2} = \min_q \max_p V(p,q).$$

To see the left-hand side equality: if $p < 1$, then by choosing $q = 1$, 2 guarantees 0; if $p = 1$, then $q = 0$ gives 0 as well. For the right-hand side equality: if $q = 1/2$, then $V(p, 1/2) = 1/2$ for all p, and if $q \neq 1/2$, then $\max\{V(1,q), V(0,q)\} > 1/2$.

Hence, either the value of this game does not exist, or it involves non-stationary strategies. Blackwell and Ferguson showed that player 1 can guarantee a limiting average reward as close to $1/2$ as he likes, by carefully taking into account the opponent's behavior, *i.e.*, his past actions, in the process of choosing his own actions. (But there is no way that player 1 can guarantee *exactly* $\frac{1}{2}$, as is not hard to see.)

Indeed, fix $K \in \mathbb{N}$, and let 0_T (resp, 1_T) denote the number of rounds up to but not including T in which player 2 's action is $\overline{0}$ (resp., $\overline{1}$). Further, set $k_T = 0_T - 1_T$, so that

$$k_T = 0_T - 1_T = (T - 1 - 1_T) - 1_T = T - 1 - 2 \cdot 1_T.$$

We define a class of strategies for player 1 as follows, in period T, provided the game is not yet absorbed, σ_K plays $\overline{1}$ with probability

$$\frac{1}{(k_T + K + 1)^2}.$$

Note that if $k_T = -K$, then the game ends with probability 1. We will show that, for every

sequence of actions by player 2, $a_2^T := (a_{2,1}, \ldots, a_{2,T})$, it holds that

$$\mathbb{E}_{\sigma_K, a_2^T} \frac{1}{T} \sum_{t=1}^{T} u(a_t) \geq \frac{K}{2(K+1)} - \frac{K+1}{2T},$$

so that whichever (pure, and hence also, by linearity of payoffs, mixed) strategy player 2 uses, the same inequality holds, and so, for every $\varepsilon > 0$, player 1 can guarantee $\frac{K}{2(K+1)} - \varepsilon$ over a long enough horizon.

Fix a_2^T and note that

1. if $a_{2,1} = \overline{1}$, then $\sigma_K|_{\sigma_K(\varnothing), a_{2,1}} = \sigma_{K-1}$;
2. if $a_{2,1} = \overline{0}$, then $\sigma_K|_{\sigma_K(\varnothing), a_{2,1}} = \sigma_{K+1}$.

Let $t_* := \inf\{t \in \mathbb{N} : a_{1,t} = \overline{1}\}$ be the first time the game is absorbed, and set

$$X_T = \begin{cases} 1/2 & \text{if } t_* > T \\ 1 & \text{if } t_* \leq T, \quad a_{2,t_*} = \overline{1} \\ 0 & \text{if } t_* \leq T, \quad a_{2,t_*} = \overline{0}. \end{cases}$$

We may think of X_T as player 1's payoff in the game modified such that he gets $1/2$ if absorption has not taken place by time T. By induction on T, we show that

$$\mathbb{E}_{\sigma^K, a_2^T} X_T \geq \frac{K}{2(K+1)} \forall K.$$

For $T = 1$, then

- If $a_{2,1} = \overline{1}$, then

$$\mathbb{E}_{\sigma^K, a_2^T} X_1 = \left(1 - \frac{1}{(K+1)^2}\right) \frac{1}{2} + \frac{1}{(K+1)^2} > \frac{1}{2} > \frac{K}{2(K+1)}.$$

- If $a_{2,1} = \overline{0}$, then

$$\mathbb{E}_{\sigma^K, a_2^T} X_1 = \left(1 - \frac{1}{(K+1)^2}\right) \frac{1}{2} = \frac{K(K+2)}{2(K+1)^2} > \frac{K}{2(K+1)}.$$

Suppose now that this is satisfied for $T = t_0$, and let us consider $T = t_0 + 1$. Then

- If $a_{2,1} = \overline{1}$, then

$$\begin{aligned}\mathbb{E}_{\sigma^K, a_2^T} X_{t_0+1} &= \left(1 - \frac{1}{(K+1)^2}\right) \mathbb{E}_{\sigma^{K-1}, a_2^T} X_{t_0} + \frac{1}{(K+1)^2} \\ &> \left(1 - \frac{1}{(K+1)^2}\right) \frac{K-1}{2K} + \frac{1}{(K+1)^2} = \frac{K}{2(K+1)}.\end{aligned}$$

- If $a_{2,1} = \overline{0}$, then

$$\mathbb{E}_{\sigma^K, a_2^T} X_{t_0+1} = \left(1 - \frac{1}{(K+1)^2}\right) \mathbb{E}_{\sigma^{K+1}, a_2^T} X_{t_0} > \left(1 - \frac{1}{(K+1)^2}\right) \frac{K+1}{2(K+2)} = \frac{K}{2(K+1)}.$$

Let $t^* := \min\{t_*, T\}$, and note that $k_{t^*} \geq -K$, so $0_{t^*} \geq (t^* - K - 1)/2$, so that

$$\begin{aligned}\mathbb{E}_{\sigma_K, a_2^T} \frac{1}{T} \sum_{t=1}^{T} u(a_t) &= \mathbb{E}_{\sigma_K, a_2^T} \frac{0_{t^*} + (T - t^*)\mathbf{1}_{\{a_{2,t^*} = \overline{1}\}}}{T} \\ &\geq \mathbb{E}_{\sigma_K, a_2^T} \frac{\frac{t^*-K-1}{2} + (T - t^*)\mathbf{1}_{\{a_{2,t^*} = \overline{1}\}}}{T} \\ &= \mathbb{E}_{\sigma_K, a_2^T} \frac{\frac{t^*}{2} + (T - t^*)\mathbf{1}_{\{a_{2,t^*} = \overline{1}\}}}{T} - \frac{K+1}{2T} \\ &= \mathbb{E}_{\sigma_K, a_2^T} \sum_{t=1}^{T} X_t - \frac{K+1}{2T} \\ &\geq \frac{K}{2(K+1)} - \frac{K+1}{2T},\end{aligned}$$

as was to be shown.

B Paris Match

We now consider a non-zero sum variant of the first game, which is called Paris Match (Sorin, 1986), in tribute to Sylvain Sorin's favorite tabloid.

This example is famous because, unlike in the previous one, it shows a disconnect between limit discounted and undiscounted equilibrium *payoff sets*. For us, however, it will be interesting because it will show that the equilibrium payoff set might be bounded *from above*, unlike what usually occurs in repeated games, in which individual rationality provided a lower bound to equilibrium payoffs.

Figure 2 shows the payoff matrix. As is clear, it is a variation on the Big Match. In particular,

	L	R
T	$(1,0)^*$	$(0,2)^*$
B	$(0,1)$	$(1,0)$

Figure 2: Paris Match

the minmax payoff of player 1 is precisely the value of Big Match, and so equal to 1/2, while the minmax payoff of player 2 is a simple variant of the Big Match, whose value is easily seen to be 2/3.

Figure 3 illustrates the feasible payoff set F, the minmax payoff vector $\underline{v}$ which is also the unique equilibrium payoff in the discounted game, and the set of undiscounted payoffs G –the boundary points of the feasible payoff set that dominate the minmax payoff. The reader interested in why this is the undiscounted equilibrium payoff set is referred to Sorin's (1986) paper.

Given δ, let w denote the maximum Nash equilibrium payoff of player 2, achieved by some strategy σ. Let w_{BL} (resp., w_{BR}) be the continuation payoff in this equilibrium, given that the first action profile is BL (resp., BR). Let p (resp., q) denote the probability with which player 1 (resp., 2) plays T (resp., L). Note that $p < 1$, for otherwise 2 would set $q = 0$, and the resulting payoff would violate individual rationality, and similarly $q < 1$ (otherwise player 1 would choose $p = 1$). Next, if $p = 0$, then because $q < 1$, we would have that w_{BR} is also a Nash payoff, and $w = \delta w_{BR} < w$ yields a contradiction. So $p > 0$. Similarly, $q = 0$ would imply $p = 0$, and so also $q > 0$. It follows that both p, q are in $(0, 1)$, both w_{BR}, w_{BL} are Nash payoffs, and players must be indifferent between both actions. That is,

$$w = (1-p)(1-\delta+\delta w_{BL}) = 2p + (1-p)\delta w_{BR}.$$

Hence,

$$w \leq (1-p)(1-\delta+\delta w), \text{ and } (1-p)(2-\delta w) \leq 2-w,$$

so that

$$(2-w)(1-\delta+\delta w) \geq w(2-\delta w),$$

and hence $2 \geq 3w$. Since $w \geq \underline{v}_2 = 2/3$, it follows that $w = \underline{v}_2$.

A similar reasoning applies to player 1. Let u be his maximum equilibrium payoff, with continuation payoffs u_{BR}, u_{BL}, which are Nash payoffs as well. We get

$$u = q = q\delta u_{BL} + (1-q)(1-\delta+\delta u_{BR}),$$

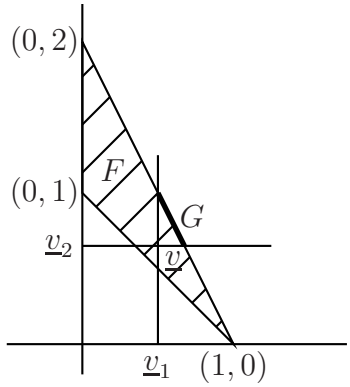


Figure 3: G and NE_δ in the Paris-Match Example

and so

$$u \leq q\delta u + (1-q)(1-\delta+\delta u),$$

and so, given $u = q$,

$$u \leq \delta u^2 + (1-u)(1-\delta+\delta u), \text{or } u \leq 1/2.$$

Hence $u = \underline{v}_1$, and $NE_\delta = \{\underline{v}\}$ for every $\delta \in [0, 1)$.

We skip the formal definition of Nash equilibrium payoffs for the undiscounted game, but it can be shown that the equilibrium payoff set is equal to $G = \{(x, 2(1-x)) : \underline{v}_1 \leq x \leq \underline{v}_2\}$. See Figure 3.

C Irreducible vs. Non-irreducible Games

The previous two examples illustrate that stochastic games can have some remarkable properties: discontinuities as $\delta \to 1$, and failure of the folk theorem, despite perfect monitoring. Not all games are as "badly" behaved as these two examples, however. There is an important class of games, studied in the next section, whose properties are similar to those from repeated games: so-called *irreducible* games. These are games in which no individual state, or more generally no proper subset of states is "absorbing." Before defining them more formally, it is worth motivating them by a simpler class of problems where they already play an important role, namely one-player problems, or *Markov decision processes*. A (finite) Markov decision process (or MDP) $\mathcal{M}$ is defined by a state space S, an action set is A, both finite, a reward function is $r : S \times A \to \mathbf{R}$, and a transition function is $p(\cdot \mid s, a)$. We let Σ denote the set of strategies in $\mathcal{M}$, which are maps from histories (past actions and states, including today's state) into (possibly mixed) actions

$\Delta(A)$. A (deterministic) Markov strategy σ^M is a pure strategy that only depends on today's state, that is, a map $S \to A$. Write Σ^M for the set of all Markov strategies.

For $\delta < 1$ and $N \in \mathbf{N}$, we let

$$v_\delta(s) := \max_{\sigma \in \Sigma} \mathbb{E}_{s,\sigma} \left[(1-\delta) \sum_{n=1}^{\infty} \delta^{n-1} r(s_n, a_n) \right],$$

and

$$v_N(s) := \max_{\sigma \in \Sigma} \mathbb{E}_{s,\sigma} \left[\frac{1}{N} \sum_{n=1}^{N} r(s_n, a_n) \right],$$

denote the values of the discounted and finite horizon versions of $\mathcal{M}$, as a function of the initial state s.

A MDP $\mathcal{M}$ together with a Markov strategy σ^M defines a Markov chain, denoted $\mathcal{M}(\sigma^M)$. An MDP $\mathcal{M}$ is *irreducible* if $\mathcal{M}(\sigma^M)$ is irreducible for all $\sigma^M \in \Sigma^M$.[1] It is *unichain* if for each σ^M, there is at most one ergodic class (and so a possibly empty transient class). (If the transient class is empty, it is then an irreducible MDP.) It is *multichain* if for some σ^M, the Markov chain has at least two ergodic classes.

We note that our two examples, ignoring their competitive nature, and viewing the two players as one, choosing $a \in A$ (to maximize, for instance, the sum of players' payoffs) are multichain MDP, as there are multiple absorbing states (each being an ergodic class). *Even* in the world of MDP, there is a significant difference between multichain and unichain MDP. When the MDP is irreducible (or more generally, unichain), the following result is known as the *Average Cost Optimality Equation* (ACOE), which ties together the discounted and limit of means payoff criteria, and provides a way to solve for the optimal strategy.

Proposition (ACOE). *There is a unique $v \in \mathbf{R}$ and a unique (up to an additive constant) map $\theta : S \to \mathbf{R}$ such that*

$$v + \theta(s) = \max_{a \in A} \left\{ r(s,a) + \mathbb{E}_{p(\cdot|s,a)} \theta(\cdot) \right\}, \text{ for all } s \in S. \tag{1}$$

In addition, $v = \lim_{\delta \to 1} v_\delta(s) = \lim_{N \to +\infty} v_N(s)$ for all $s \in S$.

[1] A Markov chain is irreducible if all states are communicating, that is, for all states $s, t \in S$, there exists an integer n such that the probability of $s_n = t$ is positive, given $s_0 = s$. An ergodic class $T \subseteq S$ is a set of states such that any two of them are communicating, and none of them communicates with a state not in T. A state s is transient if there is a positive probability that $s_n \neq s$ for all $n \geq 1$, given $s_0 = s$. A transient class is a maximum set of states that are transient.

Proof. We first prove the existence of a solution to (1). For $\delta < 1$ the dynamic programming principle writes

$$v_\delta(s) = \max_{a \in A} \left\{ (1-\delta) r(s,a) + \delta \mathbb{E}_{p(\cdot|s,a)} v_\delta(\cdot) \right\}, \text{ for all } s \in S. \tag{2}$$

Let $a^*(s)$ achieve the maximum in (2), so that $v_\delta(s) = (1-\delta) r(s, a^*(s)) + \delta \mathbb{E}_{p(\cdot|s,a^*_s)} v_\delta(\cdot)$ for each s. This implies that $\delta \mapsto v_\delta(s)$ is a bounded and rational function on $[0,1)$. In particular, both $v(s) := \lim_{\delta \to 1} v_\delta(s)$ and $\theta(s) := \lim_{\delta \to 1} \dfrac{v_\delta(s) - v(s)}{1-\delta}$ exist. Irreducibility readily implies that $v(s)$ is independent of s.

Equation (2) then rewrites as

$$v + (v_\delta(s) - v) = \max_{a \in A} \left\{ (1-\delta) r(s,a) + \delta \mathbb{E}_{p(\cdot|s,a)} \left[v_\delta(t) - v \right] + \delta v \right\}.$$

Equation (1) follows when dividing by $1-\delta$ and letting $\delta \to 1$.

We next prove uniqueness, and start with v. Let (v, θ) be a solution to (1), so that

$$\theta(s) = \max_{a \in A} \left\{ r(s,a) + \mathbb{E}_{p(\cdot|s,a)} \theta(\cdot) \right\} - v. \tag{3}$$

Substituting (3) into the right-hand side of (1) yields first

$$2v + \theta(s) = \max_{\sigma} \mathbb{E}_{s,\sigma} \left[r(s_1, a_1) + r(s_2, a_2) + \theta(s_3) \right],$$

and, by induction,

$$v + \frac{\theta(s)}{N} = \max_{\sigma} \mathbb{E}_{s,\sigma} \left[\frac{1}{N} \sum_{n=1}^{N} r(s_n, a_n) + \frac{\theta(s_{N+1})}{N} \right],$$

for each N. This implies that $\lim_{N \to \infty} v_N(s)$ exists and is equal to v.

We conclude with the uniqueness of θ. Let (v, θ) and (v, ψ) be two solutions to (1). This implies

$$\theta(s) - \psi(s) \leq \max_{a \in A} \mathbb{E}_{p(\cdot|s,a)} \left(\theta(\cdot) - \psi(\cdot) \right)$$

for each s. By irreducibility, it follows that $\theta(\cdot) - \psi(\cdot)$ is constant. □

We emphasize one of the major ingredients of the proof, namely, a power series expansion of

	L	R
T	0	1^*
B	1^*	0^*

Figure 4: A game where the value is not rational in $1-\delta$.

$v_\delta(s)$ in terms of δ, that is,

$$v_\delta(s) = v + (1-\delta)\theta(s) + o(1-\delta).$$

This property still holds for multichain MDPs. In fact, it holds that

Proposition. *There exists $\bar{\delta} < 1$ and a strategy $\sigma^M \in \Sigma^M$ such that σ^M is optimal for all $\delta \in [\bar{\delta}, 1)$. Furthermore, v_δ has an expansion in power series in $1-\delta$.*

We omit the proof. But this property, which holds for irreducible (zero-sum) irreducible stochastic games fails in the multichain case. As an example, consider the game given in Figure 4. Let p denote the probability of T, and q the probability of L. The value v_δ must satisfy

$$v_\delta = \max_p \min_q \{pq\delta v_\delta + p(1-q) + q(1-p)\} = \min_q \max_p \{pq\delta v_\delta + p(1-q) + q(1-p)\}.$$

As in the Big Match, it is readily verified that players cannot play pure, and so must be indifferent between their two actions, so that

$$v_\delta = p\delta v_\delta + 1 - p = p = q,$$

and so

$$v_\delta = p = q = \frac{1-\sqrt{1-\delta}}{\delta}.$$

More complicated examples exist, in which v_δ has an expansion whose first term is an arbitrary rational function of $1-\delta$.

Returning to MDPs, we know a more general version of the ACOE, that applies to the multichain case, and stated below for completeness.

Proposition (ACOE, multichain). *Consider the system in $v, \theta : S \to \mathbb{R}$:*

$$\begin{cases} v(s) & = \max_{a \in A} \mathbb{E}_{p(\cdot|s,a)} v(\cdot), \\ v(s) + \theta(s) & = \max_{a \in A(s)} \{r(s,a) + \mathbb{E}_{p(\cdot|s,a)} \theta(\cdot)\}, \end{cases}$$

	L	R
T	$(0,1)$	$(0,-1)^*$
B	$(2,2)$	$(-1,0)$

State 1

	L	R
T	$(2,2)$	$(0,1)^*$
B	$(-1,0)$	$(0,-1)$

State 2

Figure 5: Example 5, where some IR payoffs are not equilibrium payoffs

where $A(s) := \{a \in A : v(s) = \mathbb{E}_{p(\cdot|s,a)} v(\cdot)\}$. *It admits a unique solution* v, *and* $v = \lim_{\delta \to 1} v_\delta(s) = \lim_{N \to +\infty} v_N(s)$ *for all* $s \in S$.

Any strategy that achieves the maximum is optimal both for the average criterion and for the discounted case provided the discount factor is high enough.

We know of no such characterization for (finite) zero-sum stochastic games that are not irreducible, except in one special case. In general, it is known that the value satisfies

$$\lim_{\delta \to 1} v_\delta(s) = \lim_{N \to +\infty} v_N(s),$$

and that (as was clear in the example) players have optimal Markov strategies for fixed δ (although not in the limit, as the Big Match illustrates). The one exception for which the value is explicitly known covers a class of (non-irreducible) zero-sum games to which all our examples belonged, namely, games in which all states but one are absorbing (so called absorbing games). The formula for the value can be found in Laraki (2010).

Non-Irreducible Examples: Plainly, absorbing games are not the only problematic ones. Here are two further examples, one of which is not absorbing, due to Dutta (1995) that illustrate the subtleties with stochastic games. In those two examples, a superscript "$*$" means that, if the corresponding action profile is played in that state, the state switches with probability 1. In Example 5, because players can always elect to transit from one state to the other, the feasible set of payoffs $F(s)$, as a function of the state, is independent of s: $F(1) = F(2) =: F$ (viewed as an MDP, it is unichain). However, player 2's (asymptotic) minmax payoff $\underline{v}_2(s)$, namely,

$$\lim_{\delta \to 1} \min_{\sigma_1} \max_{\sigma_2} \mathbb{E}_\sigma \left[\sum_{n=0}^{\infty} (1-\delta)\delta^n u_2(s_n, a_n) \mid s_0 = s \right],$$

(assuming this limit exists in general, which can be shown), depends on the state: plainly, $\underline{v}_2(0) = 1 \neq 0 = \underline{v}_2(2)$. Consider now a pair $v(1), v(2) \in F$ such that $v_2(1) \in (1,2)$ and

	L	R
T	$(0,2)$	$(3,3)$
B	$(-1,0)$	$(-1,0)$

State 1

	L	R
T	$(3,0)$	$(0,-1)$
B	$(1,0)^*$	$(0,1)$

State 2

Figure 6: Example 6, where some IR payoffs are not equilibrium payoffs

$v_2(2) \in (0,1)$. [This is possible to do, by for instance publicly randomizing between (T,L) and (B,L).] These are strictly individually rational payoffs, yet they cannot be equilibrium payoffs. Indeed, consider the game starting in $s_0 = 2$. To give player 2 a strictly positive payoff, player 1 must (occasionally, with positive probability) play T. If he does so, player 2 can guarantee himself a long-run (or discounted, for high enough discount factors) average payoff of 1 by playing R in state 2 so as to take the game in state 1 with probability 1 eventually.

In Example 6, it is easily checked that $\underline{v_i}(s) = 0$, for $i = 1,2$, $s = 1,2$. Also

$$F(1) = \text{co}\{(0,2),(3,3),(-1,0)\}, \quad F(2) = \text{co}(F(1) \cup G),$$

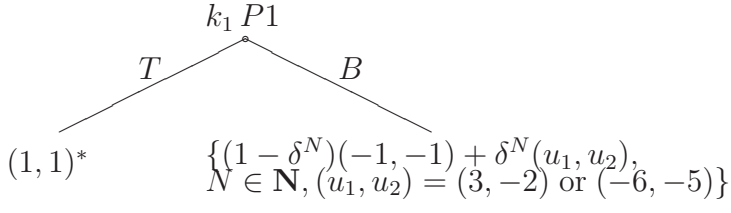
where $G = \text{co}\{(3,0),(0,-1),(0,1)\}$. So $F(1) \neq F(2)$. Consider a pair $v(1), v(2)$, with $v(2) \in G \backslash F(1)$, and $v_2(2) < 3/4$. In other words, this is a payoff generated by some public randomization over (T,L), (T,R) and (B,R). These are individually rational payoffs, yet to achieve those action B must be played. But then player 2 can deviate and play L, thereby taking the game to state 1, in which his payoff cannot be lower than $3/4$, in equilibrium, for otherwise player 1 would get less than his minmax payoff of 0. Hence, despite being individually rational, these cannot be equilibrium payoffs. The argument also applies to high enough discount factors.

An example of non-convergence. The following is from Renault and Ziliotto (2015, unpublished). The game starts with player 1 choosing between L, which leads to an absorbing state k_1 with payoff $(1,1)$, or R, which leads to state k_2. (This initial action entails no reward, or if you prefer, a 0 reward). The game with initial state k_2 is as follows.

$$\begin{array}{c} \\ L \\ R \end{array} \begin{array}{c} \begin{array}{ccc} L & M & R \end{array} \\ \begin{pmatrix} (-1,-1) \circlearrowleft & (-12,-11)^* & (-4,-7)^* \\ (-22,-12)^* & (3,-2)^* & (-9,-4)^* \end{pmatrix} \end{array}$$

That is, the subgame is absorbed as soon as the realized action profile isn't (T,L). We note that this subgame has two (absorbing) Nash equilibria, namely (B,M) and Player 1 randomizing

$(1/4, 3/4)$ between T and B while 2 randomizes $(1/4, 3/4)$ between M and R, yielding a payoff vector $(-5, -6)$. Finally, it is also optimal for players to play (T, L), but this does not lead to absorption. So any subgame-perfect Nash equilibrium of the subgame consists of a string of (T, L) followed (if ever) by one of the two absorbing Nash equilibria, leading to a reduced form game summarized in the following figure.



Let $\Delta = \{\delta \in [0,1), \exists N \geq 1, (1-\delta^N)(-1) + 3\delta^N = 1\} = \{(\frac{1}{2})^{1/N}, N \geq 1\}$. It is then clear that, for $\delta \notin \Delta$, $E_\delta \cap \{(u_1, u_2), u_2 \in (-1, 1)\} = \emptyset$, whereas, for $\delta \in \Delta_1$, $\{1\} \times (-1, 1) \subset E_\delta$. It follows that the set of equilibrium payoffs as $\delta \to 1$ might converge to different limits according to the subsequence of discount factors that are considered.

It is worth pointing out, however, that all convergent subsequences of equilibrium payoff sets have elements in common: in particular, it follows (almost immediately) from Shapley (1953) that the (convergent) payoff vector resulting from any Nash equilibrium in stationary strategies is an element that belongs to all those sets.

This property (of non-empty intersection) does not extend in general to the case in which states are not observed.

II Irreducible Games

Given that our interest lies in non-zero sum games, there is little hope to go beyond the irreducible case, given our lack of understanding of zero-sum games that are not irreducible. We will focus on those in what follows.

A Irreducible Games: An Example

Let us attempt to adapt FL's arguments to a specific example.

There are two states, $i = 1, 2$, and two players. Each player only takes an action in his own state: player i chooses L or R in state i. Actions are not observable, but affect transitions, so that players learn about their opponents' actions via the evolution of the state. If action L (R)

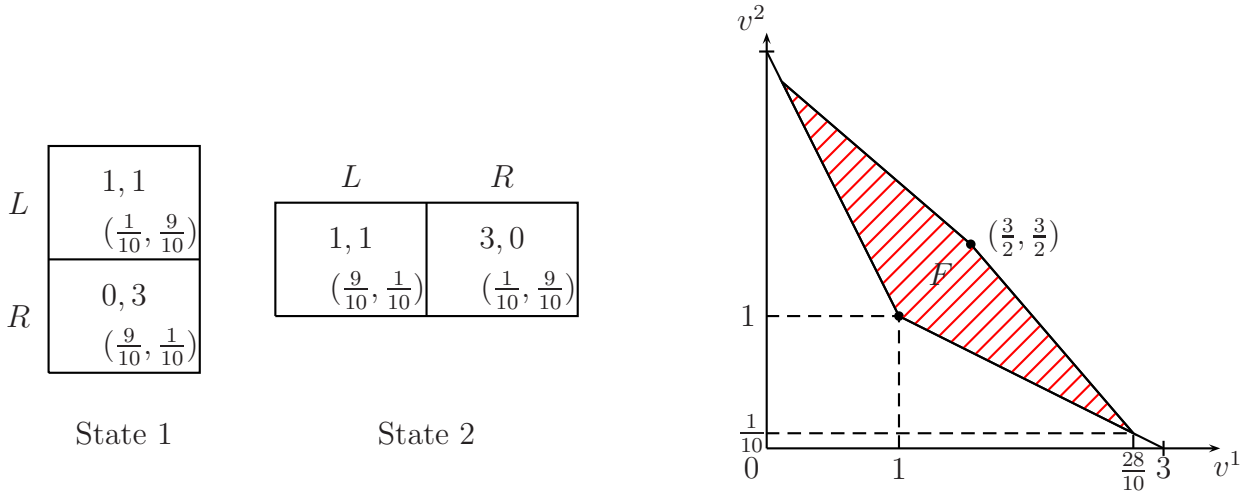


Figure 7: Rewards and Transitions in Example 1

is taken in state i, then the next state is again i with probability p^L (p^R). Let us pick here $p^L = 1 - p^R = 1/10$. Rewards are given in Figure 7 (transition probabilities to states 1 and 2, respectively, are given in parenthesis). Throughout, we refer to this game as Example 1.

Player i has a higher reward in state $j \neq i$, independently of the action. Moreover, by playing L, which yields him the higher reward in his own state, he maximizes the probability to switch states. Thus, playing the efficient action R requires intertemporal incentives, which are hard to provide absent public signals. Constructing an equilibrium in which L is not always played appears challenging, but not impossible: playing R in state i if and only if the state was $j \neq i$ in the previous two periods (or since the beginning of the game if fewer periods have elapsed) is an equilibrium for some high discount factor ($\delta \approx .823$). So there exist equilibrium payoffs above 1.

In analogy with what we have done for repeated games with public monitoring, we may now decompose the payoff vector in state $s = 1, 2$ as

$$v_s = (1-\delta) r(s, \alpha_s) + \delta \sum_t p(t \mid s, \alpha_s) w_t(s), \tag{4}$$

where t is the next state, $w_t(s)$ is the continuation payoff then, and $p(t \mid s, \alpha_s)$ is the probability of transiting from s to t given action α_s at state s. Fix $\lambda \in \mathbb{R}^I$. If v_s maximizes the score $\lambda \cdot v_s$ in all states $s = 1, 2$, then the continuation payoff in state t gives a lower score than v_t, independently of the initial state: for all s, t,

$$\lambda \cdot (w_t(s) - v_t) \leq 0. \tag{5}$$

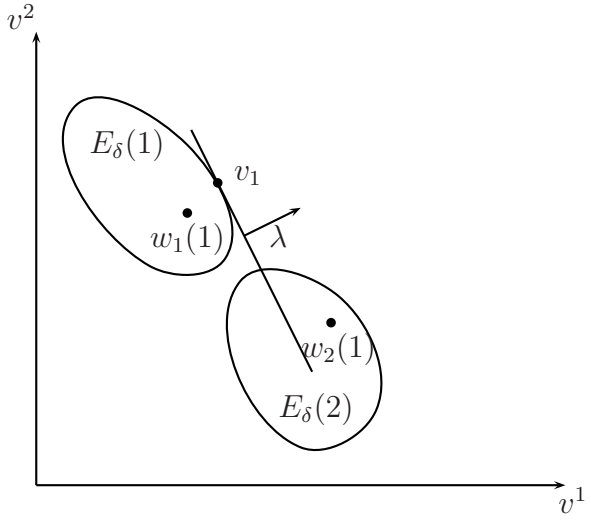
Our goal is to eliminate the discount factor. Note, however, that if we subtract δv_s on both sides of (4), and divide by $1-\delta$, we obtain

$$v_s = r(s,\alpha_s) + \sum_t p(t \mid s,\alpha_s)\frac{\delta}{1-\delta}(w_t(s) - v_s), \tag{6}$$

and there is no reason to expect $\lambda \cdot (w_t(s) - v_s)$ to be negative, unless $s = t$ (compare with public monitoring). Unlike the limiting set of feasible payoffs as $\delta \to 1$, the set of feasible rewards does depend on the state (in state 1, it is the segment $[(1,1),(0,3)]$; in state 2, the segment $[(1,1),(3,0)]$; see the right panel of Figure 7), and so the score $\lambda \cdot w_t(s)$ in state t might exceed the maximum score achieved by v_s in state s. Thus, defining x by

$$x_t(s) := \frac{\delta}{1-\delta}(w_t(s) - v_s),$$

we know that $\lambda \cdot x_s(s) \leq 0$, for all s, but not the sign of $\lambda \cdot x_t(s)$, $t \neq s$. On the one hand, we cannot restrict it to be negative: if $\lambda \cdot x_2(1) \leq 0$, then, because also $\lambda \cdot x_1(1) \leq 0$, by considering $\lambda = (1,0)$, player 1's payoff starting from state 1 cannot exceed his highest reward in that state (*i.e.*, 1). Yet we know that some equilibria yield strictly higher payoffs.



On the other hand, if we impose no restrictions on $x_t(s)$, $s \neq t$, then we can set v_s as high as we wish in (6) by picking $x_t(s)$ large enough. The value of the program to be defined would be unbounded. What is the missing constraint?

We do know that (5) holds for all pairs (s,t). By adding up these inequalities for $(s,t) = (1,2)$ and $(2,1)$, we obtain

$$\lambda \cdot (w_1(2) + w_2(1) - v_1 - v_2) \leq 0, \text{ or, rearranging, } \lambda \cdot (x_1(2) + x_2(1)) \leq 0. \tag{7}$$

Equation (7) has a natural interpretation in terms of *s-blocks*, as defined in the literature on Markov chains (see, for instance, Nummelin, 1984). When the Markov chain (induced by the players' strategies) is communicating, as it is in our example, we might divide the game into the subpaths of the chain between consecutive visits to a given state s. The score achieved by the continuation payoff once state s is re-visited on the subpath $(s_1, \ldots, s_k)$ (where $s_1 = s_k = s$) cannot exceed the score achieved by v_s, and so the difference in these scores, as measured by the sum $\lambda \cdot \sum_{j=1}^{k-1} x_{s_{j+1}}(s_j)$, must be negative. Note that the irreducibility assumption also guarantees that the limit set of feasible payoffs F (as $\delta \to 1$) is independent of the initial state, as shown in the right panel of Figure 7. To conclude, we obtain the program

$$\sup_{v,x,\alpha} \lambda \cdot v,$$

over $v \in \mathbb{R}^2$, $\{x_t(s) \in \mathbb{R}^2 : s,t = 1,2\}$, and $\alpha = (\alpha_s)_{s=1,2}$ such that, in each state s, α_s is a Nash equilibrium with payoff v of the game whose payoff function is given by $r(s,a_s) + \sum_t p(t \mid s,a_s)x_t(s)$, and such that $\lambda \cdot x_1(1) \leq 0$, $\lambda \cdot x_2(2) \leq 0$, and $\lambda \cdot (x_1(2) + x_2(1)) \leq 0$. Note that this program already factors in our assumption that equilibrium payoffs can be taken to be independent of the state.

It will follow from the main theorem of the next section that this is the right program. Perhaps it is a little surprising that the constraints involve unweighted sums of vectors $x_t(s)$, rather than, say, sums that are weighted by the invariant measure under the equilibrium strategy.

B Notation

We introduce stochastic games with public signals. At each stage, the game is in one state, and players simultaneously choose actions. Nature then determines the current reward (or flow payoff) profile, the next state and a public signal, as a function of the current state and the action profile. The sets S of possible states, I of players, A^i of actions available to player i, and Y of public signals are assumed finite. (Because states will often appear as subscripts, players are now promoted to superscripts.)

Given an action profile $a \in A := \times_i A^i$ and a state $s \in S$, we denote by $r(s,a) \in \mathbb{R}^I$ the reward profile in state s given a, and by $p(t, y \mid s, a)$ the joint probability of moving to state $t \in S$ and of receiving the public signal $y \in Y$. (As usual, we can think of $r^i(s,a)$ as the expectation given a of some realized reward that is a function of a private outcome of player i and the public signal only.)

We assume that at the end of each period, the only information publicly available to all players consists of nature's choices: the next state together with the public signal. When properly interpreting Y, this includes the case of perfect monitoring and the case of publicly observed rewards. Note however that this fails to include the case where actions are perfectly monitored, yet states are not disclosed. In such a case, the natural "state" variable is the (common) posterior belief of the players on the underlying state.

Thus, in the stochastic game, in each period $n = 1, 2, \ldots$, the state is observed, the stage game is played, and the corresponding public signal is then revealed. (Here, stages at denoted by n rather by t, as the latter variable will be used for states, alongside s). The stochastic game is parameterized by the initial state s_1, and it will be useful to consider all potential initial states simultaneously. The public history at the beginning of period n is then $h_n = (s_1, y_1, \ldots, s_{n-1}, y_{n-1}, s_n)$. We set $H_1 := S$, the set of initial states. The set of public histories at the beginning of period n is therefore $H_n := (S \times Y)^{n-1} \times S$. We let $H := \bigcup_{n \geq 1} H_n$ denote the set of all public histories. The private history for player i at the beginning of period n is a sequence $h_n^i = (s_1, a_1^i, y_1, \ldots, s_{n-1}, a_{n-1}^i, y_{n-1}, s_n)$, and we similarly define $H_1^i := S$, $H_n^i := (S \times A^i \times Y)^{n-1} \times S$ and $H^i := \bigcup_{n \geq 1} H_n^i$. Given a stage $n \geq 1$, we denote by s_n the state, a_n the realized action profile, and y_n the public signal in period n. We will often use the same notation to denote both these realizations and the corresponding random variables.

A (behavior) strategy for player $i \in I$ is a map $\sigma^i : H^i \to \triangle A^i$. Every pair of initial state s_1 and strategy profile σ generates a probability distribution over histories in the obvious way and thus also generates a distribution over sequences of the players' rewards. Players seek to maximize their payoffs, that is, average discounted sums of their rewards, using a common discount factor $\delta < 1$. Thus, the payoff of player $i \in I$ if the initial state is s_1 and the players follow the strategy profile σ is defined as

$$(1-\delta) \sum_{n=1}^{+\infty} \delta^{n-1} \mathbb{E}_{s_1,\sigma}[r^i(s_n, a_n)].$$

Public strategies are defined as usual. A strategy σ^i is public if it depends on the public history

only, and not on the private information. That is, a public strategy is a mapping $\sigma^i : H \to \triangle A^i$. A *perfect public* equilibrium (hereafter, PPE) is a profile of public strategies such that, given any period n and public history h_n, the strategy profile is a Nash equilibrium from that period on. Note that this class of equilibria includes Markov equilibria, in which strategies only depend on the current state and period. In what follows though, a *Markov* strategy for player i will be a public strategy that is a function of states only, *i.e.*, a function $S \to \triangle A^i$.[2] For each Markov strategy profile $\alpha = (\alpha_s)_{s \in S} \in (\times_{i \in I} \triangle A^i))^S$, we denote by $q_\alpha(t \mid s) := p(t \times Y \mid s, \alpha_s)$ the transition probabilities of the Markov chain over S induced by α.

We denote by $E_\delta(s) \subset \mathbb{R}^I$ the (compact) set of PPE payoffs of the game with initial state $s \in S$ and discount factor $\delta < 1$. All statements about convergence of, or equality between sets are understood in the sense of the Hausdorff distance $d(A, B)$ between sets A, B.

The next assumption is critical.

Assumption A: The limit set of PPE payoffs is independent of the initial state: for all $s, t \in S$,

$$\lim_{\delta \to 1} d(E_\delta(s), E_\delta(t)) = 0.$$

This is an assumption on endogenous variables. A stronger assumption on exogenous variables that implies Assumption **A** is *irreducibility*: For any (pure) Markov strategy profile $a = (a_s) \in A^S$, the induced Markov chain over S with transition function q_a is irreducible. Actually, it is not necessary that every Markov strategy gives rise to an irreducible Markov chain. It is clearly sufficient if there is some state that is accessible from every other state regardless of the Markov strategy.

Note also that, by redefining the state space to be $S \times Y$, one may further assume that only states are disclosed. That is, the class of stochastic games with public signals is no more general than the class of stochastic games in which only the current state is publicly observed. However, the Markov chain over $S \times Y$ with transition function $\tilde{q}_a(t, z \mid s, y) := p(t, z \mid s, a_s)$ need not be irreducible even if $q_a(t \mid s)$ is.

[2] In the literature on stochastic games, such strategies are often referred to as stationary strategies.

C The Characterization

Given a state $s \in S$ and a map $x : S \times Y \to \mathbb{R}^{S \times I}$, we denote by $\Gamma(s, x)$ the one-shot game with action sets A^i and payoff function

$$r(s, a_s) + \sum_{t \in S} \sum_{y \in Y} p(t, y \mid s, a_s) x_t(s, y),$$

where $x_t(s, y) \in \mathbb{R}^I$ is the t-th component of $x(s, y)$.

Given $\lambda \in \mathbb{R}^I$, we denote by $\mathcal{P}(\lambda)$ the maximization program

$$\sup_{v,x,\alpha} \lambda \cdot v,$$

where the supremum is taken over all $v \in \mathbb{R}^I$, $x : S \times Y \to \mathbb{R}^{S \times I}$, and $\alpha = (\alpha_s) \in (\times_{i \in I} \triangle A^i)^S$ such that

(i) For each s, α_s is a Nash equilibrium with payoff v of the game $\Gamma(s, x)$;

(ii) For each $T \subseteq S$, for each permutation $\phi : T \to T$ and each map $\psi : T \to Y$, one has $\lambda \cdot \sum_{s \in T} x_{\phi(s)}(s, \psi(s)) \leq 0$.

Denote by $k(\lambda) \in [-\infty, +\infty]$ the value of $\mathcal{P}(\lambda)$. We will prove that the feasible set of $\mathcal{P}(\lambda)$ is non-empty, so that $k(\lambda) > -\infty$ (Proposition 4), and that the value of $\mathcal{P}(\lambda)$ is finite, so that $k(\lambda) < +\infty$.

We define $\mathcal{H}(\lambda) := \{v \in \mathbb{R}^I : \lambda \cdot v \leq k(\lambda)\}$, and set $\mathcal{H} := \bigcap_{\lambda \in \mathbb{R}^I} \mathcal{H}(\lambda)$. Note that $\mathcal{H}$ is convex. Let S^1 denote the set of $\lambda \in \mathbb{R}^I$ of norm 1. Observe that $\mathcal{H}(0) = \mathbb{R}^I$, and that $\mathcal{H}(\lambda) = \mathcal{H}(c\lambda)$ for every $\lambda \in \mathbb{R}^I$ and $c > 0$. Hence $\mathcal{H}$ is also equal to $\bigcap_{\lambda \in S^1} \mathcal{H}(\lambda)$.

Our main result is a generalization of FL's algorithm to compute the limit set of payoffs as $\delta \to 1$.

Theorem 1 (Main Theorem)**.** *Assume that $\mathcal{H}$ has non-empty interior. Under Assumption **A**, $E_\delta(s)$ converges to $\mathcal{H}$ as $\delta \to 1$, for any $s \in S$.*

Note that, with one state only, our optimization program reduces to the algorithm of FL. Note that these propositions do not rely on Assumption **A**.

Proposition 1. *For every $\delta < 1$, we have the following.*

1. *$k(\lambda) \geq \min_{s \in S} \max_{w \in E_\delta(s)} \lambda \cdot w$ for every $\lambda \in S^1$.*

2. $\mathcal{H} \supseteq \bigcap_{s\in S} E_\delta(s)$.

We note that it need not be the case that $\mathcal{H} \supseteq E_\delta(s)$ for each $s \in S$.

Proposition 2. *Assume that $\mathcal{H}$ has non-empty interior, and let Z be any compact set contained in the interior of $\mathcal{H}$. Then $Z \subseteq E_\delta(s)$, for every $s \in S$ and δ large enough.*

The logic of the proof of Proposition 5 is inspired by FL and FLM, but differs in important respects. We here give a short and overly simplified account of the proof, that nevertheless contains some basic insights.

Let a payoff vector $v \in Z$ and a direction $\lambda \in S^1$ be given. Since v is interior to $\mathcal{H}$, one has $\lambda \cdot v < k(\lambda)$, and there thus exists $x = (x_t(s, y))$ such that v is a Nash equilibrium payoff of the one-shot game $\Gamma(s, x)$, and all inequality constraints on x hold with strict inequalities.

For high δ, we use x to construct equilibrium continuation payoffs w adapted to v in the discounted game, with the interpretation that $x_t(s, y)$ is the normalized (continuation) *payoff increment*, should (t, y) occur. Since we have no control over the sign of $\lambda \cdot x_t(s, y)$, the one-period argument that is familiar from repeated games does not extend to stochastic games. To overcome this issue, we will instead rely on large blocks of stages of fixed size. Over such a block, and thanks to the inequalities **(ii)** satisfied by x, we will prove that the sum of payoff increments is negative. This in turn will ensure that the continuation payoff at the end of the block is below v in the direction λ.

Since $\mathcal{H}$ is convex, it follows from these two propositions that

$$\mathcal{H} = \lim_{\delta\to 1} \bigcap_{s\in S} E_\delta(s).$$

This statement applies to all finite stochastic games with observable states and full-dimensional $\mathcal{H}$, whether they satisfy Assumption **A** or not. Theorem 1 then follows, given Assumption **A**.

D Connection with the ACOE

At first sight, the characterization above looks very different from the one in the one-player case, namely the ACOE. In fact, with one player, it turns out to be the same.

Corollary 2. *In the one-player case with irreducible transition probabilities, the set $\mathcal{H}$ is a singleton $\{v^*\}$, with $v^* = \lim_{\delta\to 1} v_\delta(s)$ for each $s \in S$. Moreover, there is a vector $x^* \in \mathbf{R}^S$ such*

that

$$v^* + x_s^* = \max_{a_s \in A} \left(r(s, a_s) + \sum_{t \in S} p(t \times Y | s, a_s) x_t^* \right) \tag{8}$$

holds for each $s \in S$*, and* $v = v^*$ *is the unique value solving (8) for some* $x \in \mathbf{R}^S$*.*

To get some intuition for this corollary, note first that, with one player, signals become irrelevant, and we might ignore them. Consider then the direction $\lambda = 1$. To maximize the player's payoff, we should increase the values of $x_t(s)$ as much as possible. So conjecture for a moment that all the constraints **(ii)** bind: for all $T \subseteq S$ and permutations $\phi : T \to T$, $\sum_{s \in T} x_{\phi(s)}(s) = 0$. Let us then set $x_t^* := x_t(\bar{s})$, for some fixed state $\bar{s} \in S$. Note that, for all $s, t \in S$,

$$x_t(s) = -x_{\bar{s}}(t) - x_s(\bar{s}) = x_t(\bar{s}) - x_s(\bar{s}) = x_t^* - x_s^*,$$

where the first two equalities use the binding constraints. Because a_s is a Nash equilibrium of the game $\Gamma(s, x)$ with payoff v^*, we have

$$v^* = \max_{a_s \in A} \left(r(s, a_s) + \sum_{t \in S} p(t \times Y | s, a_s) x_t(s) \right).$$

Using $x_t(s) = x_t^* - x_s^*$ gives the desired result. See Hörner, Sugaya, Takahashi and Vieille (2011) for details.

III Literature

An excellent introduction to zero-sum stochastic games can be found in Sorin (2002), *A First Course on Zero Sum Repeated Games*, Springer.

The model of a stochastic game has been introduced by Shapley, L.S. (1953), "Stochastic Games," *Proceedings of the National Academy of Sciences of the U.S.A.*, **39**, 1095–1100. Existence of (Markov) equilibria in stochastic games is due to Fink, A. M. (1964), "Equilibrium in a stochastic n-person game," *Journal of Science of the Hiroshima University, Series A-I*, **28**, 89–93, and Takahashi, M. (1962), "Stochastic Games with Infinitely Many Strategies," *Journal of Science of the Hiroshima University, Series A-I*, **26**, 123–124. The Big Match example was solved by Blackwell and Ferguson (1968, "The "Big Match"," *Annals of Mathematical Statistics*, **39**, 159–163), but it was suggested earlier (see Gillette, D., 1957, "Stochastic Games with Zero Stop Probabilities," in *Contributions to the Theory of Games, III*, M. Dresher, A.W. Tucker,

and P. Wolfe (eds.), Annals of Mathematical Studies, **39**, Princeton University Press, 179–187). Paris Match is defined and solved in Sorin, S. (1986, "Asymptotic properties of a non-zero-sum stochastic game," *International Journal of Game Theory*, **15**, 101–107). The value for zero-sum absorbing games can be found in Laraki, R. (2010), "Explicit formulas for repeated games with absorbing states," *International Journal of Game Theory*, **39**, 53–69. The game exhibited in Section I.C. that illustrates the non-convergence of E_δ is due to J. Renault and B. Ziliotto (2015, working paper, Toulouse).

Results for Markov Decision Processes can be found in many excellent textbooks, including Puterman, M. (2008, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley). A good source for Markov chains (referred to in Section II) is Nummelin, E. (1984, *General irreducible Markov chains and non-negative operators*, Cambridge Tracts in Mathematics 83, Cambridge University Press).

The recursive methods for analyzing PPE payoffs in stochastic games are due to Hörner, J., T. Sugaya, S. Takahashi and N. Vieille (2011), "Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \to 1$ and a Folk Theorem," *Econometrica*, **79**, 1277–1318. They derive a folk theorem generalizing FLM. An independent proof under slightly stronger assumptions is in Fudenberg, D. and Y. Yamamoto (2012), "The Folk Theorem for Irreducible Stochastic Games with Imperfect Public Monitoring," *Journal of Economic Theory*, **146**, 1664–1683. The first folk theorem for stochastic games, following Fudenberg and Maskin's folk theorem in its constructive approach, is due to Dutta, P. (1995), "A Folk Theorem for Stochastic Games," *Journal of Economic Theory*, **66**, 1–32.

IV Incomplete Information

We now turn to states that are *privately* observed. We start with the simplest case, in which there is only player who observes the state, and his opponent does not. In addition, we return to perfect monitoring for the time being.

A Sender-Receiver

Furthermore, assume that the informed player, player 1 (or P1 for short) does not take any payoff-relevant action (he is the *Sender*). His action $a \in A$ is a message (or *report*, or *announcement*) about the state of the world. The message is public. Upon observing the message, player 2 (P2 for short, or the *Receiver*) takes an action $b \in B$. Hence, formally, even

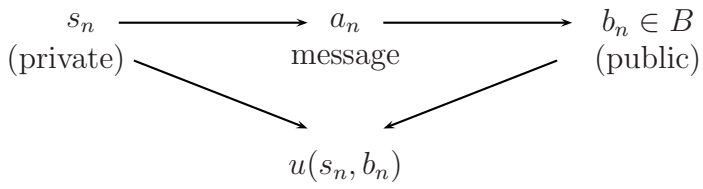


Figure 8: Structure of Sender-Receiver Games

leaving aside the incomplete information, this is not a standard repeated game, as in each stage an extensive-form game is being played.

P1's message is about the state of the world $s \in S$ which he privately observes at the start of each stage. Both S and B are finite.

Rewards are denoted $u(s_n, b_n) = (u_1(s_n, b_n), u_2(s_n, b_n))$. As mentioned, the action of P1 does not appear; but P2's actions matters for both players. Players discount future rewards at rate $\delta \in (0, 1)$.

Because there is no commitment by the players (they play in each stage the action they prefer), the well-known *revelation principle* does not apply here. It is **not** without loss to assume that there are as many messages in A as states. (Counterexamples are known for other games.) Yet let us assume as much: $|A| = |S|$. The structure is summarized in the figure below.

The state (s_n) follows a Markov chain over S, with transitions $p(\cdot \mid \cdot)$. It is assumed to be irreducible and aperiodic,[3] so that it admits a unique invariant distribution $\mu \in \triangle S$. Without loss, $\mu(s) > 0$ for all $s \in S$. For simplicity, assume that the initial state s_0 is drawn according to μ.

A strategy for P1 is a map $\sigma_1 : \cup_{n \geq 0}(S \times A \times B)^n \times S \to \triangle A$, while a strategy for P2 is a map $\sigma_2 : \cup_{n \geq 0}(A \times B)^n \to \triangle B$. A stationary strategy for P2 is simply a map $\beta : A \to \triangle B$, which summarizes P2's actions when told $a \in A$.

Consider the following simple example, with two states $S = \{L, R\}$, and two actions by the receiver, $B = \{\ell, r\}$. States are i.i.d. over time, with both states being equally likely. Here, $c \in (1, 2)$. We identify messages with states (so the two messages are R and L, but of course P1 can lie).

In the one-shot game, there exists a unique equilibrium, in which P2 plays r with probability 1. To see this, note that P1 strictly prefers r over ℓ, independently of the state. So it must be that all messages that are sent with positive probability induce the same action (possibly mixed)

[3]That is, there exists N, for every $s, t \in S$, and for all $n > N$, $p(s_n = s \mid s_0 = t) > 0$.

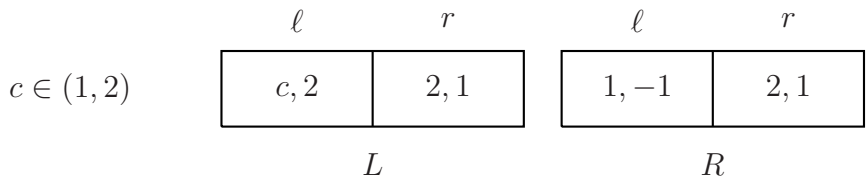


Figure 9: Rewards and Transitions in the Example

by P2. Because beliefs are a martingale, there is some message that induces a belief of P2 that assigns probability at least $1/2$ to state R; after this message, it is strictly optimal to play r; hence this is the only action that P2 ever takes with positive probability in equilibrium. The corresponding equilibrium payoff vector is $(2, 1)$.

Note that this equilibrium (in which, say, P1 sends both messages with equal probability, independently of the state, and P2 plays r no matter the message that he hears) is also an equilibrium of the repeated game, and it gives P2 its lowest individually rational payoff, because he always have the option of playing r no matter P1's strategy.

Remarkably, at least when players are patient enough, the dynamic game also admits equilibria that yield much higher payoffs to P2. More specifically, let us construct an equilibrium with payoff arbitrarily close to $\left(\frac{2+c}{2}, \frac{3}{2}\right)$ –the best possible payoff for P2. Note that P1 gets a payoff below the payoff he receives if he could commit to "shut up"!

We actually will construct two equilibria: one explicitly, to illustrate the possibility of getting something else than babbling; the second one is not fully explicit, but achieves the desired payoff asymptotically. First, we construct an equilibrium in which equilibrium payoffs are **not** $(2, 1)$. Consider the following strategy profile:

- If n is odd, P1 reports s_n truthfully, and P1 plays ℓ if told L, r if R.
- If n is even, P1 randomizes between both messages (independently of the state) and P2 plays the action he has **not** played in the previous period.
- If P2 deviates, both players switch to the repetition of the one-shot equilibrium (with payoff $(2, 1)$.

Under this strategy profile, payoffs are

$$\left(\frac{1}{1+\delta}\left(\frac{2+c}{2} + \delta\frac{5+c}{4}\right), \frac{1}{1+\delta}\left(\frac{3}{2} + \delta\frac{3}{4}\right)\right).$$

Because P2 gets 1 after deviating, and yet the conditional expected payoff if he does not is at least 1 in each stage, P2 does not deviate. As for P1, note that it suffices to consider a deviation on a pair of consecutive stages. He can deviate in two ways: announce L when the state is R; and announce R when state is L.

The first deviation gives 1 and then 2 (instead of 2 and then $(1+c)/2$). The second deviation gives 2 and then $(1+c)/2$ (instead of c and then 2). Picking $\delta \geq \frac{4-2c}{3-c}$ makes both deviations unprofitable.

The trick is that P1's choice is no longer whether to induce P2 to play ℓ rather than r: both actions are taken equally often. P1's choice is whether he prefers these actions to match the state or not.

We now turn to the second equilibrium: proving the existence of an equilibrium whose payoff approximates $\left(\frac{2+c}{2}, \frac{3}{2}\right)$. Fix an integer N "large enough." We divide play into "blocks" of length $2N$. Consider the following strategy σ_2 for P2 over a block: play ℓ if told L, play r if told R, until the first time one of the two reports has been sent N times or more. Once this occurs, P2 chooses in all remaining periods of the block the action that he has played **least** often up to that date.

It is not clear what P1's best-reply to this strategy is. But we claim that **any** best-reply must achieve a payoff of at least $\frac{2+c}{2} - \varepsilon$, where ε can be chosen to be as small as we wish but an appropriate choice of N and δ. This is because both actions will be played an equal number of times by P2.

Let σ_1 denote any (pure) best-reply by P1. By definition it is optimal given σ_2. And P2 can be deterred from deviating by threatening to revert to the repetition of the one-shot Nash equilibrium.

To conclude, in the repeated game, we can construct equilibria in which the sender is forced to speak, whether he likes it or not; the receiver, on the other hand, is never forced to listen, and so is guaranteed to get at least the payoff from the babbling equilibrium.

We now turn to a more general analysis. Let $\mathcal{M} \in \triangle(S \times A)$ to be the set of **copulas** based on μ. That is, $m \in \mathcal{M}$ if and only if the marginals of m on S and A are both equal to μ. This set is defined by linear inequalities, so it is a compact convex polyhedron. Let $m_0 \in \mathcal{M}$ denote the particular copula such that $m_0(s,s) = \mu(s)$ and $m_0(s,a) = 0$ if $s \neq a$. Hence, m_0 is the joint long-run average distribution of the sequence (s_n, a_n) when P1 reports truthfully. Any strategy by P1 with the property that the long-run average frequency of the reports matches the invariant distribution defines a copula; conversely, each copula can be obtained by some strategy of P1

with this property, and so we might think of these strategies in terms of copulas.

Given $m \in \mathcal{M}$ and a stationary strategy $\beta : A \to \triangle B$, let

$$v(m, \beta) := \sum_{(s,a)} m(s, a) u(s, \beta(\cdot \mid a)) \in \mathbb{R}^2$$

denote the expected payoff vector when reports are drawn according to m and P2 uses $\beta(\cdot \mid a)$. Let

$$\underline{v}^2 := \max_{b \in B} \sum_s \mu(s) u_2(s, b)$$

denote the payoff of P2 under the one-shot equilibrium in which reports are uninformative.

Definition: Let $V(\mathcal{M})$ denote the set of vectors $v(m_0, \beta)$ such that

$$v^1(m_0, \beta) \geq v^1(m, \beta) \quad \forall m \in \mathcal{M},$$

and

$$v^2(m_0, \beta) \geq \underline{v}^2.$$

Let $E(\delta)$ denote the set of sequential equilibrium payoffs in the game with discount factor δ. (Assume a public randomization device.) We have that:

Theorem: If int $V^s(\mathcal{M}) \neq \varnothing$, then

$$V(\mathcal{M}) \subseteq \liminf_{\delta \to 1} E(\delta).$$

The converse requires an assumption.

Theorem: Suppose that there exists non-negative numbers $(\alpha_s)_{s \in S}$ such that, for every $s' \in S$, $\sum_{s \in S \setminus \{s'\}} \alpha_s \leq 1$, and $p(s' \mid s) = \alpha_{s'}$ whenever $s' \neq s$. Then for every $\delta < 1$,

$$E(\delta) \subseteq V(\mathcal{M}).$$

The assumption of the theorem is strong, but it is always satisfied with two states. In general, it is "almost" equivalent to assuming that states follow a renewal process. The remarkable conclusion is that, under this assumption, *at least in terms of payoffs*, checking whether the average empirical frequency with which P1 takes each action matches the invariant distribution is the only relevant statistical test that needs to be run. Yet clearly, there are other statistical

tests that one could use. For instance, one could check whether the persistence in the reports (as measured, say, by the average run length) matches the persistence of the Markov chain.

V More General Results

A Examples

The previous analysis suggests that the empirical frequency of observed reports can be used as a way of disciplining players. Furthermore, under the assumption on transitions in the last example, it suggests that *nothing* more can be used as statistically useful information. This is not true in general, as the next example suggests.

Consider two firms that compete through prices. Their unit cost is private information and change over time: $s_1 \in \{L, H\}$, $s_2 \in \{M, V\}$, with $L < M < H < V < 1$. Each firm's cost is the same from one period to the next with probability $p \in (1/2, 1)$, and the draws are independent across firms. In each period, a consumer arrives who is willing to pay up to \$1 for one unit of the (indivisible) unit.

Suppose that firms could commit to a mechanism which, as a function of the reports they make, leads to an agreed-upon firm to sell at a price of 1. Suppose that this is done over a large enough horizon T (which may be thought as a block, as in the previous example, or the actual horizon length). Ideally, the low-cost firm should be the one getting the sale: this is efficient way of splitting the market.

If firms report truthfully, firm 2 makes a sale approximately a quarter of the time. As a result, profits are (approximately) $v_1 = \frac{1-L}{2} + \frac{1-H}{4}$ for firm 1 and $v_2 = \frac{1-M}{4}$ for firm 2. We claim that firm 2 has an incentive to lie *based on* firm 1's previous reports. Specifically, suppose that firm 2 deviates from truth-telling by reporting M if and only if firm 1 has reported H in the previous period. The deviation leads firm 2 to make sales in approximately $p\frac{T}{2}$ periods, earning a profit $p\frac{1-M}{4} + p\frac{1-V}{4}$, which exceeds $\frac{1-M}{4}$ if p is close enough to one.

Of course, such systematic lies will be detected. But what is the best test to use, and will it suffice to induce truth-telling?

The second example shows that, in general, we cannot expect to induce truth-telling always. This is a zero-sum two-player game in which player 1 has two private states, s^1 and $\hat{s}^1$, and player 2 has a single state, omitted. Player 1 has actions $A^1 = \{T, B\}$ and player 2 has actions $A^2 = \{L, R\}$. Player 1's reward is given by Figure 4. Recall that rewards are not observed. States s^1 and $\hat{s}^1$ are equally likely in the initial round, and transitions are action-independent,

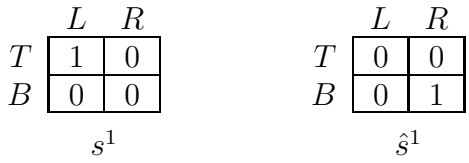
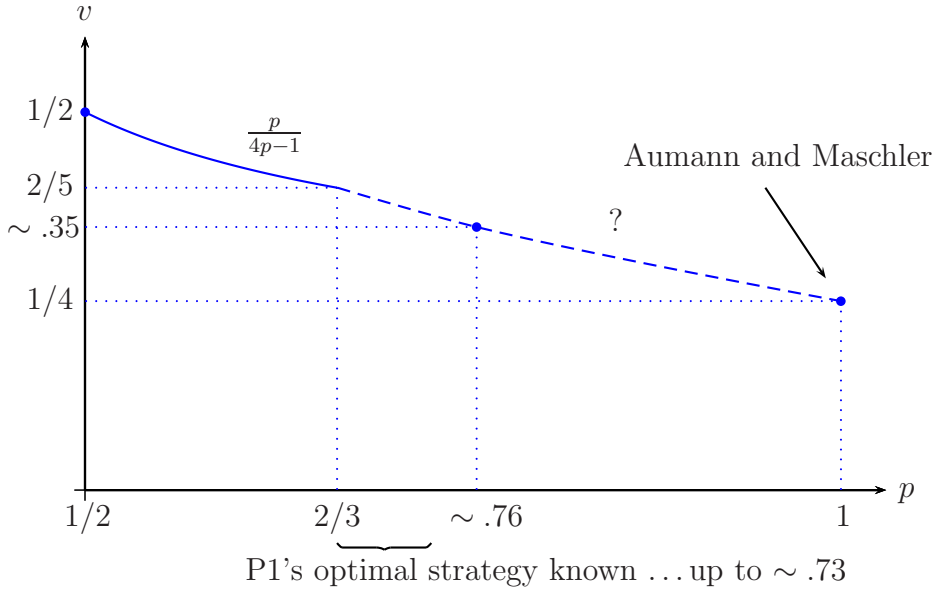


Figure 10: Player 1's reward in the Example

with $p \in [1/2, 1)$ denoting the probability that the state remains unchanged from one round to the next.

Set $M^1 := \{s^1, \hat{s}^1\}$, so that player 1 can disclose his state if he wishes to. Will he? By revealing the state, player 2 can secure a payoff of 0 by playing R or L depending on player 1's report. Yet player 1 can secure a payoff of 1/4 by choosing reports and actions at random. In fact, this is the (uniform) value of this game for $p = 1$ (Aumann and Maschler, 1995). When $p < 1$, player 1 can actually get more than this by trading off the higher expected reward from a given action with the information that it gives away. He has no interest in giving this information away for free through informative reports. Silence is called for.

Just because we may focus on the silent game does not mean that it is easy to solve. Its (limit) value for arbitrary $p > .719$ is still unknown (It is $p/(4p-1)$ for $p \in [1/2, .719]$). See the figure below. Because the optimal strategies depend on player 2's belief about player 1's state, the problem of solving for them is infinite-dimensional, and all that can be done is to characterize its solution via some functional equation.



What this example illustrates is that small message spaces are just as difficult to deal with as larger ones. When players hide their information, their behavior reflects their private beliefs, which calls for a state space as large as it gets.

The surprise, then, is that the literature on Markovian games manages to get positive results at all: in most games, efficiency requires coordination, and thus disclosure of (some) private information. As is turns out, existence is much easier to obtain in the independent private values environment, the focus of most of these papers. The zero-sum example involved both interdependent values and independent types, an ominous combination in mechanism design: with interdependent values, the uninformed player's payoff depends on the informed player's type, so that he cannot resist adjusting his action to the message he receives. This might hurt the informed player, who cannot be statistically disciplined into truth-telling, given independent types.

In our dynamic environment as well, positive results will obtain as soon as we impose private values or relax independent types.

Player 1 has two private states, s^1 and $\hat{s}^1$, and player 2 has a single state, omitted. Player 1 has actions $A^1 = \{T, B\}$ and player 2 has actions $A^2 = \{L, R\}$. Rewards are given by Figure 2 (values are private). The two types s^1 and $\hat{s}^1$ are i.i.d. over time and equally likely. Monitoring

	L	R
T	$1, 1$	$1, -1$
B	$0, -1$	$0, 1$

s^1

	L	R
T	$0, 1$	$0, -1$
B	$1, -1$	$1, 1$

$\hat{s}^1$

Figure 11: A two-player game in which the mixed minmax payoff cannot be achieved.

is perfect. To minmax player 2, player 1 must randomize uniformly, independently of his type. But clearly player 1 has a strictly dominant strategy in the repeated game, playing T in state s^1 and B in state $\hat{s}^1$. Even if player 1's continuation utility were to be chosen freely, it would not be possible to get player 1 to randomize in both states: to play B when his type is s^1, or T when his type is $\hat{s}^1$, he must be compensated by \$1 in continuation utility. But then he has an incentive to report his type incorrectly, to pocket this promised utility while playing his favorite action.

This example illustrates that fine-tuning continuation payoffs to make a player indifferent between several actions in several private states simultaneously is generally impossible to achieve with independent types. This still leaves open the possibility of a player randomizing for *one* of his types. This is especially useful when each player has only one type, like in a standard repeated game, as it then delivers the usual mixed minmax payoff. Indeed, the characterization below yields a minmax payoff somewhere in between the mixed and the pure minmax payoff, depending on the particular game considered. This example also shows that truth-telling is restrictive even with independent private values: in the silent game, player 1's unique equilibrium strategy minmaxes player 2, as he is left guessing player 1's action. Leaving a player in the dark about one's state can serve as a substitute for mixing at the action step. To achieve lower equilibrium payoffs, truth-telling must be abandoned, at least during punishments. As follows from Theorem 4 below, it is indeed possible to drive player 2's payoff down to his minmax payoff of 0 in equilibrium, as $\delta \to 1$.

Our final example has again two players. Player 1 has $K+1$ types, $S^1 = \{0, 1, \ldots, K\}$, while player 2 has only two types, $S^2 = \{0, 1\}$. Transitions do not depend on actions (omitted), and are as follows. If $s_n^1 = k > 0$, then $s_n^2 = 0$ and $s_{n+1}^1 = s_n^1 - 1$. If $s_n^1 = 0$, then $s_n^2 = 1$ and s_{n+1}^1 is drawn randomly (and uniformly) from S^1. In words, s_n^1 stands for the number of rounds until

the next occurrence of $s^2 = 1$. By waiting no more than K rounds, all reports by player 1 can be verified.

This example makes two closely related points. First, in order for player $-i$ to statistically discriminate between player i's states, it is not necessary that his set of signals (here, players $-i$'s states) be as rich as player i's, unlike in static mechanism design with correlated types. Two states for one player can be enough to cross-check the reports of an opponent with many more states, provided that states in later rounds are informative enough.

Second, the long-term dependence of the stochastic process implies that one player's report should not always be evaluated on the fly. It is better to hold off until more evidence is collected. Note that this is not the same kind of delay as the one that makes review strategies effective, taking advantage of the central limit theorem to devise powerful tests even when signals are independently distributed over time. It is precisely because of the dependence that waiting is useful here.

This raises an interesting statistical question: does the tail of the sequence of private states of player $-i$ contain indispensable information in evaluating the truthfulness of player i's report in a given round, or is the distribution of this infinite sequence, conditional on (s_n^i, s_{n-1}), summarized by the distribution of an initial segment of the sequence? This question appears to be open in general. In the case of transitions that do not depend on actions, the answer is known: it is enough to consider the next $2|S^i| + 1$ values of the sequence $(s_{n'}^{-i})_{n' \geq n}$.

At the very least, when types are correlated and the Markov chain exhibits time dependence, it is useful to condition player i's continuation payoff given his report about s_n^i on $-i$'s next private state, s_{n+1}^{-i}. Because this suffices to obtain sufficient conditions analogous to those invoked in the static case, we will limit ourselves to this conditioning.

The game we will consider has the following sequence: In each round $n \geq 1$, timing is as follows:

1. Each player $i \in I$ privately observes his own state $s_n^i \in S^i$;

2. Players simultaneously make reports $(m_n^i)_{i=1}^I \in \times_i M^i$, where M^i is a finite set. These reports are publicly observed;

3. The outcome of a public randomization device (p.r.d.) is observed.

4. Players independently choose actions $a_n^i \in A^i$. Actions taken are not observed;

5. A public signal $y_n \in Y$, a finite set, and the next state profile $s_{n+1} = (s^i_{n+1})_{i \in I}$ are drawn according to some joint distribution $p(\cdot, \cdot \mid s_n, a_n) \in \Delta(S \times Y)$.

VI Three Programs

One of the difficulties with incomplete information is that it is not even clear what the right "benchmark" upper bound on the set of Nash equilibrium payoffs actually is: clearly, there are feasible, individually rational payoffs (leaving aside how to define individual rationality) that *cannot* be equilibrium payoffs, as they would require players to divulge information that goes against their best interest.

Hence, the first task is derive such an upper bound. Then, we will develop a scoring algorithm as in FL in the case in which transitions are not affected by actions and monitoring is perfect, and argue that the folk theorem holds. We then state how it generalizes once actions affect transitions, and monitoring is imperfect.

All this is done with private values and independent types. For the generalization of the program that applies to arbitrary (possibly interdependent) values and not necessarily independent types, see Hörner, Takahashi and Vieille (2015).

First, we focus on *independent private values* (hereafter, IPV). This is defined as the special case in which (i) transitions satisfy

$$p(t, y \mid s, a) = p(y \mid a) \times \times_{i \in I} p^i(t^i \mid s^i, y),$$

as well as

$$\pi_1(s) = \times_{i \in I} \pi^i_1(s^i),$$

for some transitions $\{p^i(\cdot \mid s^i, y)\}_{s^i, y} \subseteq \Delta(S^i)$, and distributions $\{p(\cdot \mid a)\}_a \subseteq \Delta(Y)$, $\pi^i_1 \in \Delta(S^i)$, all $i \in I$, and (ii) rewards satisfy, for all $i \in I$, $s \in S$, $a \in A$, $r^i(s, a) = r^i(s^i, a)$. The first assumption guarantees that beliefs over state profiles are common knowledge throughout the game, on and off path.

We denote by $\mu \in \Delta(S \times S)$ the invariant distribution of two consecutive states (s_n, s_{n+1}). Marginals of μ will also be denoted by μ. Our purpose is to describe explicitly the asymptotic equilibrium payoff set. The feasible (long-run) payoff set is defined as

$$F := \text{co} \left\{ v \in \mathbb{R}^I \mid v = \mathbb{E}_{\mu, \rho}[r(s, a)], \text{ some policy } \rho : S \to A \right\}.$$

When defining feasible payoffs, the restriction to deterministic policies rather than arbitrary strategies is clearly without loss. Given the public randomization device, F is convex.

A An Upper Bound

We start by defining a set of payoffs that includes the (limit) set of Bayes Nash equilibrium payoffs both in the original game and in the revelation game. In addition to assuming IPV, we further assume here that actions do not affect transitions: $p^i(t^i \mid s^i, y) = p^i(t^i \mid s^i)$, for all $i, s^i, t^i \in S$.

Fix some direction $\lambda \in \Lambda$, where $\Lambda := \{\lambda \in \mathbb{R}^n : \|\lambda\| = 1\}$. What is the highest score $\lambda \cdot v$ that can be achieved over all Bayes Nash equilibrium payoff vectors v?

If actions can be dictated, knowing the state profile can only help. But if $\lambda^i < 0$, this information would be used against i's interests. Not surprisingly, player i is unlikely to be forthcoming about this. This suggests distinguishing players in the set $I_+(\lambda) := \{i : \lambda^i > 0\}$ from the others. Suppose that players in $I_+(\lambda)$ truthfully disclose their private state, while the remaining players choose a reporting strategy that is independent of their private state.

Define

$$\bar{k}(\lambda) := \max_{\rho} \mathbb{E}_{\mu,\rho} \left[\lambda \cdot r(s,a)\right],$$

where the maximum is over all policies $\rho : \times_{i \in I_+(\lambda)} S^i \to A$ (with the convention that $\rho \in A$ for $I_+(\lambda) = \emptyset$). Note that $\mathbb{E}_{\mu,\rho} \left[\lambda \cdot r(s,a)\right]$ is the long-run payoff vector when players report truthfully and use the policy ρ. Furthermore, let

$$V^* := \cap_{\lambda \in \Lambda} \left\{v \in \mathbb{R}^n \mid \lambda \cdot v \leq \bar{k}(\lambda)\right\}.$$

We call V^* the set of *incentive-compatible* payoffs. Clearly, $V^* \subseteq F$. Note also that V^* depends on the transition matrix only via the invariant distribution. It turns out that the set V^* is a superset of the set of *all* equilibrium payoff vectors.

Let NE_δ denote the equilibrium payoffs in the original game, given $\delta \in [0,1)$.

Proposition 3. *Assume IPV. The limit set of Bayes Nash equilibrium payoffs is contained in V^*:*

$$\limsup_{\delta \to 1} NE_\delta \subseteq V^*.$$

Proof. Here we provide a sketch (for $\limsup_{\delta \to 1} NE_\delta$) in the case in which the initial belief $(\pi_1^i)_{i \notin I_+(\lambda)}$ is equal to the ergodic distribution $(\mu^i)_{i \notin I_+(\lambda)}$. Fix $\lambda \in \Lambda$. Fix also $\delta < 1$. Consider

the Bayes Nash equilibrium σ of the game (with discount factor δ) with payoff vector v that maximizes $\lambda \cdot v$ among all equilibria (where v^i is the expected payoff of player i given π_1). This equilibrium need not be truthful or in pure strategies. Consider $i \notin I_+(\lambda)$. Along with σ^{-i} and π_1, player i's equilibrium strategy σ^i defines a distribution over histories. Fixing σ^{-i}, let us consider an alternative strategy $\tilde{\sigma}^i$ where player i's reports are replaced by realizations of the public randomization device with the same distribution (round by round, conditional on the realizations so far), and player i's action is determined by the randomization device as well, with the same conditional distribution (given the simulated reports) as would specify if this had been i's report. The new profile $(\sigma^{-i}, \tilde{\sigma}^i)$ need no longer be an equilibrium of the game. Yet, thanks to the IPV assumption, it gives players $-i$ the same payoff as σ and, thanks to the equilibrium property, it gives player i a weakly lower payoff. Most importantly, the strategy profile $(\sigma^{-i}, \tilde{\sigma}^i)$ no longer depends on the history of types of player i. Clearly, this argument can be applied to all players $i \notin I_+(\lambda)$ simultaneously, so that $\lambda \cdot v$ is lower than the maximum inner product achieved over strategies that only depend on the history of types in $I_+(\lambda)$. Maximizing this inner product over such strategies is a standard partially observable Markov decision problem, which admits a solution within the class of deterministic policies (on the state space $\times_{i \in I_+(\lambda)} S^i \times \times_{i \notin I_+(\lambda)} \Delta(S^i)$).

Because transitions do not depend on actions, the belief $p_n \in \times_{i \notin I_+(\lambda)} \Delta(S^i)$ in round n about the states of players in $I \setminus I_+(\lambda)$ remains equal at all times to the ergodic distribution $(\mu^i)_{i \notin I_+(\lambda)}$. This defines a strategy that is only a function of the states $(s^i)_{i \in I_+(\lambda)}$ (the solution of the partially observable Markov decision problem evaluated at the belief $(\mu^i)_{i \notin I_+(\lambda)}$).

Taking $\delta \to 1$ yields that the limit set is included in $\left\{ v \in \mathbb{R}^n \mid \lambda \cdot v \leq \bar{k}(\lambda) \right\}$, and this is true for all $\lambda \in \Lambda$. ■

We may now turn to (set-theoretic) lower bounds to the equilibrium payoff set.

B Perfect Monitoring, Action-Independent Transitions

We first assume that (i) monitoring is perfect, (ii) actions do not affect transitions. We define an algorithm similar to the scoring algorithm, but since there is incomplete information, the one-shot game we consider is now a Bayesian game rather than a complete information game. In what follows, the set of public outcomes in a given round is $\Omega_{\text{pub}} := S \times A$ (where the S-components stand for the reports). Let a Markov strategy (or *policy*) $\rho : S \to \times_{i \in I} \Delta(A^i)$, and a map $x : S \times \Omega_{\text{pub}} \to \mathbb{R}^I$ be given. The vector $x(\bar{s}, \omega_{\text{pub}})$ is to be interpreted as transfers,

contingent on previous reports $\bar{s}$, and on the current public outcome ω_{pub}.[4] Assuming states are truthfully reported and actions chosen according to ρ, the sequence (ω_n) of outcomes is a unichain Markov chain, and so is the sequence of pairs of reports (s_{n-1}, s_n). Let $\theta_{\rho,r+x} : S \times S \to \mathbb{R}^I$ denote the relative values of the players, obtained when applying the ACOE to the latter chain (and to all players).

As FL, we start with an auxiliary one-shot game. We define $\Gamma(\rho, x)$ to be the one-shot Bayesian game with communication where:

(i) first, $(\bar{s}, s) \in S \times S$ is drawn according to μ; each player i is publicly told $\bar{s}$ and privately s^i;

(ii) each player i reports publicly some state $m^i \in S^i$, then chooses an action $a^i \in A^i$.

The payoff vector is $r(s, a) + x(\bar{s}, \omega_{\text{pub}}) + \theta_{\rho,r+x}(m, t)$, where $\omega_{\text{pub}} := (m, a)$ and $t \sim p(\cdot \mid s)$.

Given $\lambda \in \Lambda$, we denote by $\mathcal{P}_0(\lambda)$ the optimization program $\sup \lambda \cdot v$, where the supremum is computed over all payoff vectors $v \in \mathbb{R}^I$, policies $\rho : S \to \times_{i \in I} \Delta(A^i)$ and transfers $x : S \times \Omega_{\text{pub}} \to \mathbb{R}^I$ such that

(a) truth-telling followed by ρ is a PBE outcome of $\Gamma(\rho, x)$, with expected payoff v;

(b) $\lambda \cdot x(\cdot) \leq 0$.

Condition (a) implies that for all $\bar{s}, s \in S$, the mixed profile $\rho(s)$ is a Nash equilibrium in the (complete information) game with payoff function $r(s, a) + x(\bar{s}, (s, a)) + \mathbb{E}_{t \sim p(\cdot|s)} \theta_{\rho,r+x}(t)$. It puts no restriction on equilibrium behavior following a lie at the report step.

The condition that v be the equilibrium payoff in $\Gamma(\rho, x)$ writes

$$v = \mathbb{E}_{(\bar{s},s) \sim \mu, a \sim \rho(s)} \left[r(s, a) + x(\bar{s}, \omega_{\text{pub}}) \right],$$

where $\omega_{\text{pub}} = (s, a)$.

We denote by $k_0(\lambda)$ the value of $\mathcal{P}_0(\lambda)$, and let $\mathcal{H}_0 := \{v \in \mathbb{R}^I, \lambda \cdot v \leq k_0(\lambda) \text{ for all } \lambda \in \Lambda\}$ be the convex set with support function k_0.

Theorem 3 below is the exact analog of FLM and HSTV, yet requires a (rather innocuous) non-degeneracy assumption.

[4] Conceptually, it might make sense to condition transfers on previous actions as well. This extension is not needed when transitions are action-independent.

Two states s^i and $\tilde{s}^i$ of player i are *equivalent* if $r^i(s^i,\cdot) = r^i(\tilde{s}^i,\cdot) + c$ for some $c \in \mathbb{R}$. Assume that there is no player with two distinct, equivalent states.

Theorem 3. *Assume that $\mathcal{H}_0$ has non-empty interior. Then $\mathcal{H}_0$ is included in the limit set of equilibrium payoffs:*

$$\mathcal{H}_0 \subseteq \liminf_{\delta\to 1} E_\delta.$$

We may then prove the folk theorem as in FLM, by computing scores in each direction.

For $i \in I$, define $\underline{v}^i := \min_{a^{-i}\in A^{-i}} \max_{\rho^i : S^i \to A^i} \mathbb{E}_\mu \left[r^i(s^i, (a^{-i}, \rho^i(s^i))) \right]$.

Proposition 4. *For every $\lambda \neq -e^i$, $k_0(\lambda) = \bar{k}(\lambda)$.*

For $\lambda = -e^i$, $k_0(-e^i) \geq -\underline{v}^i$.

Set $V^{**} := \{v \in V^*, v \geq \underline{v}\}$. By Proposition 4, $V^{**} \subseteq \mathcal{H}_0$. Hence Theorem 3 implies the following.

Corollary 4. *Assume that V^{**} has non-empty interior. Then*

$$\liminf_{\delta\to 1} E_\delta \supseteq V^{**}.$$

C Imperfect Monitoring, Action-Dependent Transitions

As the title suggests, we now drop the two assumptions of perfect monitoring and action-independent transitions from the last section. We maintain IPV.

D The Superset Revisited

Example 2. There are two players. Incomplete information is one-sided: player 2 might be in state $s = 0, 1$. Player 2 has a single action, while player 1 chooses action $a = 0, 1$. Transitions are given by $p(s_{n+1} = a \mid s_n = s, a_n = a) = 1/3$, for all $s = 0, 1$. That is, the state is twice as likely to differ from the previous action chosen by player 1 as it is to coincide with this choice. As for rewards, $r^2(s, a) = -1$ if $s = a$, $= 0$ otherwise. Suppose that the objective is to minimize player 2's payoff. We note that any constant strategy (*i.e.*, $a = 0$ or $a = 1$ in all periods) yields a payoff of $-1/3$, while a strategy that alternates deterministically between actions has a payoff that tends to $-2/3$ as $\delta \to 1$.

This example demonstrates that constant action choices no longer suffice to minimize or maximize a player's payoff, when his state is unknown to others and he fails to reveal it, even

as $\delta \to 1$. Plainly, in the example, player 1's belief about the state of player 2 matters for the choice of an optimal action, and the chosen action matters for player 1's next belief. Hence, if we wish to describe player 1's choice as a (Markov) policy, we must augment the state space to account for player 1's belief. In the previous example, there is a binary sufficient statistic for this belief, namely, the last action chosen by player 1. Yet in general, the role of the belief is not summarized by such a simple statistic. It is necessary to augment the state space by (at least) an arbitrary summary statistic, which follows a Markov chain as well. The next result establishes that finite representations suffice, under our assumptions.

We need to generalize the notion of a policy. Let a finite set K, and a map $\phi : K \times Y \to K$ be given. Together with ϕ, any map $\rho : S \times K \to \Delta(A)$ induces a Markov chain (s_n, k_n, a_n, y_n) over $S \times K \times A \times Y$. We refer to such a triple $\rho_{\text{ext}} = (\rho, K, \phi)$ as an *extended* policy. An extended policy is thus a policy that is possibly contingent on a public, extraneous and payoff irrelevant variable k whose evolution is dictated by y. The extended policy ρ_{ext} is *irreducible* if the latter chain is irreducible. We then denote by $\mu_{\rho_{\text{ext}}} \in \Delta((S \times K \times A \times Y)^2)$ the invariant distribution of successive states, actions and signals. Again, we will still denote by $\mu_{\rho_{\text{ext}}}$ various marginals of $\mu_{\rho_{\text{ext}}}$.

Given a direction $\lambda \in \Lambda$, let as before $I_+(\lambda) = \{i \in I, \lambda^i > 0\}$. We then set $\bar{k}_1(\lambda) := \sup_{\rho_{\text{ext}}} \mathbb{E}_{\mu_{\rho_{\text{ext}}}} [\lambda \cdot r(s,a)]$, where the supremum is taken over all pure irreducible extended policies $\rho_{\text{ext}} = (\rho, K, \phi)$ such that $\rho : S \times K \to A$ depends on s only through its components s^i, $i \in I_+(\lambda)$.

Let then $V_1^* := \{v \in \mathbb{R}^I, \lambda \cdot v \leq \bar{k}_1(\lambda) \text{ for all } \lambda \in \Lambda\}$, and denote by $NE_\delta(\pi_1)$ the set of Nash equilibrium equilibrium payoffs of the game with discount factor δ, as a function of the initial distribution π_1.

Proposition 5. *Assume IPV. Then* $\limsup_{\delta \to 1} NE_\delta(\pi_1) \subseteq V_1^*$, *for all* π_1.[5]

Given an irreducible extended policy $\rho_{\text{ext}} = (\rho, K, \phi)$, the relevant set of public outcomes is $\Omega_{\text{pub}} = S \times K \times Y$, where elements of S have to be interpreted as reports. Let a map $x_{\text{ext}} : \Omega_{\text{pub}} \times \Omega_{\text{pub}} \to \mathbb{R}^I$ be given. The vector $x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}})$ is interpreted as transfers, contingent on the public outcomes in the previous and current rounds. Relative values associated with the pair $(\rho_{\text{ext}}, x_{\text{ext}})$ are thus maps $\theta_{\rho_{\text{ext}}, r + x_{\text{ext}}} : \Omega_{\text{pub}} \times S \times K \to \mathbb{R}^I$.

We then define $\Gamma(\rho_{\text{ext}}, x_{\text{ext}})$ to be the one-shot Bayesian game with communication where (i) $(\bar{\omega}_{\text{pub}}, s, k) \in \Omega_{\text{pub}} \times S \times K$ is first drawn according to $\mu_{\rho_{\text{ext}}}$, (ii) each player i is publicly told $\bar{\omega}_{\text{pub}}$ (from which he deduces $k = \phi(\bar{k}, \bar{y})$) and privately told s^i, publicly reports some state $m^i \in S^i$,

[5] A more precise statement holds. For each $\eta > 0$, there is $\bar{\delta} < 1$ such that, for each discount factor $\delta \geq \bar{\delta}$ and each initial distribution $\pi_1 \in \times_{i \in I} \Delta(S^i)$, $NE_\delta(\pi_1)$ is included in the η-neighborhood $V_{1,\eta}^*$ of V_1^*.

then chooses an action $a^i \in A^i$, and the payoff vector is

$$r(s,a) + x_{\text{ext}}(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}) + \mathbb{E}_{(y,t)\sim p(\cdot|s,a)}\theta_{\rho_{\text{ext}}, r+x_{\text{ext}}}(\omega_{\text{pub}}, t),$$

with $\omega_{\text{pub}} = (m, k, y)$.

Given $\lambda \in \Lambda$, we denote by $\mathcal{P}_1(\lambda)$ the optimization program $\sup \lambda \cdot v$, where the supremum is over payoffs $v \in \mathbb{R}^I$, extended policies $\rho_{\text{ext}} = (\rho, K, \phi)$ and transfers $x_{\text{ext}} : \Omega_{\text{pub}} \times \Omega_{\text{pub}} \to \mathbb{R}^I$, such that

(a) truth-telling followed by ρ is a perfect Bayesian outcome of $\Gamma(\rho_{\text{ext}}, x_{\text{ext}})$ with expected payoff v;

(b) $\lambda \cdot x_{\text{ext}}(\cdot) \leq 0$.

We denote by $k_1(\lambda)$ the value of $\mathcal{P}_1(\lambda)$.

As in the case of action-independent transitions and perfect monitoring, we prove our characterization result, Theorem 5 below, under a non-degeneracy assumption on payoffs, which we now introduce.

Given an action profile $a \in A$, let $\vec{a}$ be the policy which plays a in each state profile $s \in S$. Observe that for $i \in I$ and $s \in S$, the relative value $\theta^i_{\vec{a},r}(s)$ is independent of s^{-i} under IPV.

A1 For all $i \in I$, $s^i \neq \tilde{s}^i \in S^i$, there exist action profiles $a, b \in A$, such that

$$\theta^i_{\vec{a},r}(s^i) - \theta^i_{\vec{b},r}(s^i) \neq \theta^i_{\vec{a},r}(\tilde{s}^i) - \theta^i_{\vec{b},r}(\tilde{s}^i). \tag{9}$$

When successive states are *i.i.d.*, **A1** is equivalent to the assumption of no-two-equivalent states made under perfect monitoring. However, when **A1** is specialized to the case of action-independent states, it neither implies nor is implied by this assumption.

In addition, we require the usual identifiability condition. In **A2**, p refers to the marginal distribution over signals $y \in Y$ only. Let $Q^i(a) := \{p(\cdot \mid \hat{a}^i, a^{-i}) : \hat{a}^i \neq a^i\}$ be the distributions over signals y induced by a unilateral deviation by i at the action step, whether or not the reported state s^i corresponds to the true state $\hat{s}^i$ or not.

A2 For all $a \in A$,

1. For all $i \neq j$, $p(\cdot \mid a) \notin \operatorname{co}\{Q^i(a) \cup Q^j(a)\}$.

2. For all $i \neq j$, $\text{co}\,(p(\cdot \mid a) \cup Q^i(a)) \cap \text{co}\,(p(\cdot \mid a) \cup Q^j(a)) = \{p(\cdot \mid a)\}$.

For $i \in I$, we set $\underline{v}^i := \min_{a^{-i} \in A^{-i}} \max_{\rho^i : S^i \to A^i} \mathbb{E}_{(s,a) \sim \mu_{(\rho^i, a^{-i})}} [r^i(s,a)]$. Proposition 6 and Theorem 5 then parallel those from perfect monitoring.

Proposition 6. *Assume IPV. Under **A2**, $k_1(-e^i) \geq -\underline{v}^i$ and $k_1(\lambda) = \bar{k}_1(\lambda)$ for all $\lambda \neq -e^i$.*

Theorem 5 (Folk theorem)**.** *Assume that IPV and Assumption **A1** and **A2** hold, and V_1^{**} has non-empty interior, then*

$$\liminf_{\delta \to 1} E_\delta(\pi_1) \supseteq V_1^{**}.$$

VII Literature

The example in Section I.A is due to Renault, J., E. Solan and N. Vieille (2013), "Dynamic Sender-Receiver Games," *Journal of Economic Theory*, **148**, 502–534. Using review phases or blocks is not new: in the context of incomplete information, it goes back to Jackson, M.O. and H.F. Sonnenschein (2007), "Overcoming Incentive Constraints by Linking Decision," *Econometrica*, **75**, 241–258, as well as Fang, H. and P. Norman (2006), "To Bundle or Not To Bundle," *RAND Journal of Economics*, **37**, 946–963. In the context of imperfect monitoring, it goes back to Radner, R. (1986), "Repeated Partnership Games with Imperfect Monitoring and No Discounting," *Review of Economic Studies*, **53**, 43–57, among others.

The first example in Section II is due to Escobar, P. and J. Toikka (2013), "Efficiency in Games with Markovian Private Information," *Econometrica*, **81**, 1887–1934 who provide an analysis of Bayesian games under some assumptions: monitoring is perfect, actions do not affect transitions, and types are independent across players. Results in the last section come from Hörner, J., S. Takahashi and N. Vieille (2015), "Truthful Equilibria in Dynamic Bayesian Games," *Econometrica*, forthcoming.

The second example in Section III has been introduced by Renault, J. (2006), "The Value of Markov Chain Games with Lack of Information on One Side," *Mathematics of Operations Research*, **31**, 490–512, building on a famous example of Aumann and Maschler. For $p \leq 2/3$, the value has been solved by Hörner, J., D. Rosenberg, E. Solan and N. Vieille (2010), "On a Markov Game with One-Sided Incomplete Information," *Operations Research*, **58**, 1107–1115. The results have been recently extended by Bressaud, X. and A. Quas (2014), "Asymmetric Warfare," http://arxiv.org/abs/1403.1385, to $p \leq .719$, using methods from thermodynamics. Peski and Toikka (private communication) have shown that the value is decreasing in p.

See also Peski, M. and T. Wiseman (2015), "A Folk Theorem for Stochastic Games with Infrequent State Changes," *Theoretical Economics*, **10**, 131–173, for an interesting asymptotic analysis when it is the time interval between successive periods that is taken to zero (rather than the discount factor to 1).

Repeated Games, Part V: Incomplete Information

Lecture Notes, Yale 2015, Johannes Hörner

September 3, 2015

I Motivation

In many applications of repeated games, there is incomplete information regarding the players' preferences. For instance:

1. A trader might have private information about the liquidation value of the asset that he is repeatedly trading.

2. A firm privately knows its production cost.

3. A patient of an insurance company has private information regarding his risk type.

These three examples illustrate the richness of the problems raised. Take the third example: In some cases, the private information pertains to characteristics that are fixed once and for all (genetic diseases, for instance); in others, these characteristics evolve over time (fitness-related diseases). In the first case, types are (perfectly)*persistent*. In the second, types are imperfectly persistent. The second case can be further divided along two important dimensions: first, whether the players' actions affect the evolution of the characteristics or not. Fitness depends on one's lifestyle. Production costs depend on the firms' productive investments. On the other hand, weather-related fluctuations in the productivity of crops is not controlled (yet). Second, whether there are some types that, once "reached," cannot be "left" again. As it turns out, irreversible states substantially complicate the analysis, especially when actions affect transitions.

Second, a player's private information might only affect his own payoff, or be relevant to all players. In the second example, it is plausible to assume *private values*: that is, a firm is not concerned about its rivals' production cost, for a given action profile. This does not mean that knowing this information would not be valuable to the firm, as it would help in predictions

these actions. But it is not directly payoff-relevant. In the first and third example, information is directly payoff-relevant: other traders would like to know the liquidation value (in fact, they might not care about one particular trader's actions *per se*, and only the information it contains might matter!) Insurance companies would like to know which patients have low risk, to better select them. In the literature on repeated games, private values is often referred to as *known-own payoffs*. When values are not private, they are *interdependent*. Common values obtains when all players' care in the same way about the information, as is plausible in the first example.

Third, the players' information need not be statistically independent; and there is no reason to assume that they collectively know everything there is to know. For instance, the feasibility of a particular venture, the profitability of a certain new product are likely to be only imperfectly known by the players. This means that players do not learn about each others' information from the actions that they might observe, but they might also learn about the uncertainty from the signals that result from these actions. This makes experimentation valuable, whereby players' choice of actions is directly motivated by the possibility of learning. This motive makes the analysis of such games interesting and non-trivial even if only one player is present.

Very few of the possibilities sketched above are understood. There is by now a vast literature in the case in which types are privately known, types are persistent and values are private. Almost all this literature is devoted to the very special case of *reputation*: by and large, this refers to the case in which only one player holds private information, and in fact, except for one particular type of his, his strategy (in the repeated game) for any other type he might have is exogenously specified; implicitly, this assumes that for all but one of his types, the player has a strictly dominant strategy in the repeated game –an extreme assumption which has allowed to make tremendous progress. See the notes on reputations.

The case in which types are changing over time requires a first extension of the theory examined so far: namely, we must understand whether and how the techniques from repeated games can be adapted to situations in which there is a state variable, such as the vector of production costs (of all firms), *under perfect information*. Formally speaking, such games are no longer repeated games, but stochastic games. See the notes on stochastic games.

I will not cover the cases of experimentation here.

II Zero-Sum

To focus on those issues, rather that on the issue of how to sustain cooperation that comes up in repeated games, it is good to understand first a simple case, namely when players have diametrically opposed interests. Clearly, this restricts us to two-player games. These are called **zero-sum games**. Furthermore, we will start by considering the case of one-sided incomplete information. Such games have very convenient properties. First, all Nash equilibria of such games yield the same payoff. Hence, we can unambiguously make predictions about this payoff (called the **value** v of the game). Second, Nash equilibria of such games are **optimal strategies** in a strong sense: an optimal strategy σ_i guarantees the payoff v for player i not only against the other player's optimal strategy σ_{-i}, but against **all** strategies of player $-i$: even if player $-i$ deviates, player i can only gain from it. Clearly, this need not be the ca! se for Nash equilibria in non-zero-sum games.

Furthermore, this implies that, if we find a strategy that gives player i at least a given payoff v, no matter what $-i$ does, and similarly player $-i$ has a strategy that guarantees that player i's payoff does not exceed v, no matter what i does, then these two strategies form an equilibrium, and v is the value of this game (viewed from i's perspective). This simplifies a lot the problem of finding equilibria, as we no longer need to find "fixed-points" of the best-reply correspondences, but can proceed player by player, as we shall see.

Finally, zero-sum repeated games with one-sided incomplete information have one more desirable property: the equilibrium payoff of the discounted game is known to converge to the equilibrium payoff of the undiscounted game (we have not discussed what it means with an infinite horizon, but we shall do so presently). Because the latter turns out to be more convenient to analyze, we shall consider it instead of the usual discounted case.

We assume perfect monitoring throughout: that is, actions by both players are perfectly observed.

A Examples

A.1 When Revealing Information is Bad

Consider the following two stage games (see Figure 1).

Because we are considering zero-sum games, we are only indicating player 1's reward from a given action profile. Player 2's reward is the opposite (we can think of this entry as a payment from player 2 to player 1).

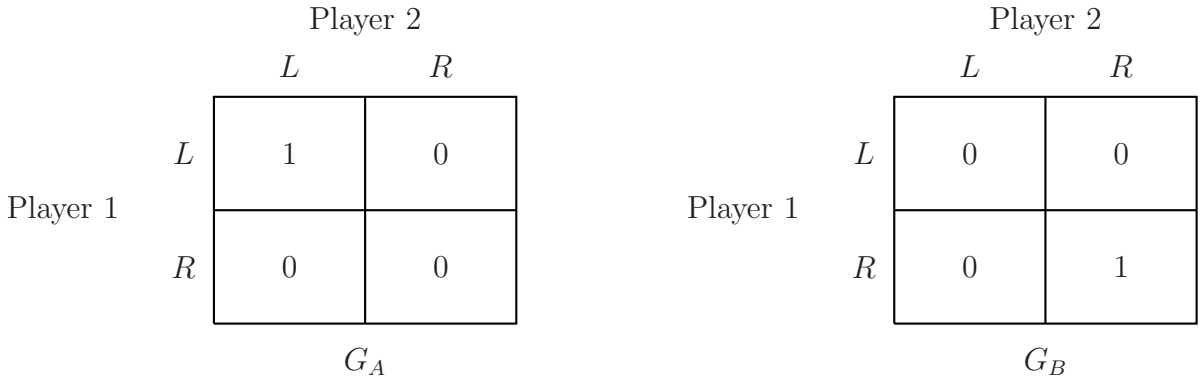


Figure 1: Rewards in the two states

If the game G_A is played, the value (or unique Nash equilibrium payoff, if you prefer) is $w_A = 0$. Obviously, repeating it indefinitely changes nothing to the game: player 1 can secure 0 by playing R, and player 2 can guarantee that player 1's payoff does not exceed 0, by playing R as well. Similarly, the value of the game G_B, whether it is repeated or not, is $w_B = 0$ as well.

Let us now consider the case in which player 1 is informed of which is the true game, G_A or G_B, while player 2 is not. All player 2 knows is that each of the two games (or **states**) is equally likely. I shall refer to 1_A and 1_B as player 1 in the event that the game is G_A or G_B, respectively.

If the game is played once, it is quite clear what player 1 should do: play L if G_A, R if G_B. Player 1 can assure himself an (expected) payoff of 1/2. Given that player 2 is indifferent between both actions, 1_A and 1_B's payoffs depend on the probability β with which 2 plays L. Namely, $v_A = \beta$ and $v_B = 1 - \beta$.

What if this game is repeated infinitely often, but the game is determined at the beginning, once and for all? Denote this game $G_\infty(p)$, where p is the prior distribution on the underlying game that is being played. Clearly, we cannot recommend that player 1 plays in every period as if this was the one-shot game, because player 2 would be able to infer what game is being played at the end of the first period, and could make sure that player 1 never gets more than 0 again. That is, unless player 1 is very impatient, but here we assume that he is infinitely patient. More precisely, we shall seek to identify the following value.

Definition 1 *A number v is called the **value** for the infinitely repeated game $G_\infty(p)$ if there exists a strategy profile $\sigma \in \Sigma$ (the **optimal strategies**) such that, for every $\varepsilon > 0$, there exists $T_0 \in \mathbb{N}$, $\forall T \geq T_0$, and all strategy profiles $\sigma' \in \Sigma$,*

$$v_T(\sigma_1, \sigma_2') + \varepsilon \geq v \geq v_T(\sigma_1', \sigma_2) - \varepsilon.$$

Here, $v_T(\sigma)$ is the payoff in the T-finitely repeated game, given strategy profile $\sigma \in \Sigma$. That is, v is the value if player 1 can secure arbitrarily close to v by playing σ_1, provided the horizon is long enough, and similarly for player 2. This means, of course, that our players are arbitrarily patient. In general, the existence of the value is a non-trivial problem, but it is not in the games that we will consider.

Going back to our example, what should player 1 do? If he reveals the state through his first-period action, he cannot secure more than 0 in G_∞. But player 1 could play as if he did not know the outcome of the move by Nature either, and play both actions with probability 1/2 in each period. In that case, he can secure a payoff of 1/4. (In non-zero-sum games, it is in general not enough to exhibit a strategy that guarantees a given payoff to conclude that the corresponding player must get at least as much in every sequential equilibrium, because of the lack of commitment. That this is the case with zero-sum games is precisely what makes them so tractable.)

Could he do better? The surprising answer is that he cannot. The proof of this is rather complicated and will be sketched later on. But it is indeed optimal for player 1 to randomize in every period, as if he did not know the true game being played. On the other hand, there is no simple optimal strategy for player 2: randomizing equally in every period, *independently* of what player 1 has done in the past, is clearly not an optimal strategy, because player 1 could take advantage of this strategy by playing always L or R, depending on the true game.

A.2 When Revealing Information is Good

Consider now the example in Figure 2. The information structure is as before: player 1 is informed of the game that is being played, G_A or G_B, while player 2 is not, and assigns equal probability to both of them. Here again, if the game were known, the value would be zero: $w_A = w_B = 0$. Note that this is the best payoff that player 1 can hope for in either case.

If player 1 were to hide his information, and play as if he did not know the state, then his payoff would be $-1/4$. If instead, he repeatedly plays R if the game is G_A, and L if the game is G_B, then he would secure 0. Because this yields the highest possible payoff, independently of what player 2 does, it is an optimal strategy. Of course, by doing so, player 2 will find out what the true game being played is after the first period, but so what?

Note that, from player 2's point of view, any strategy is optimal.

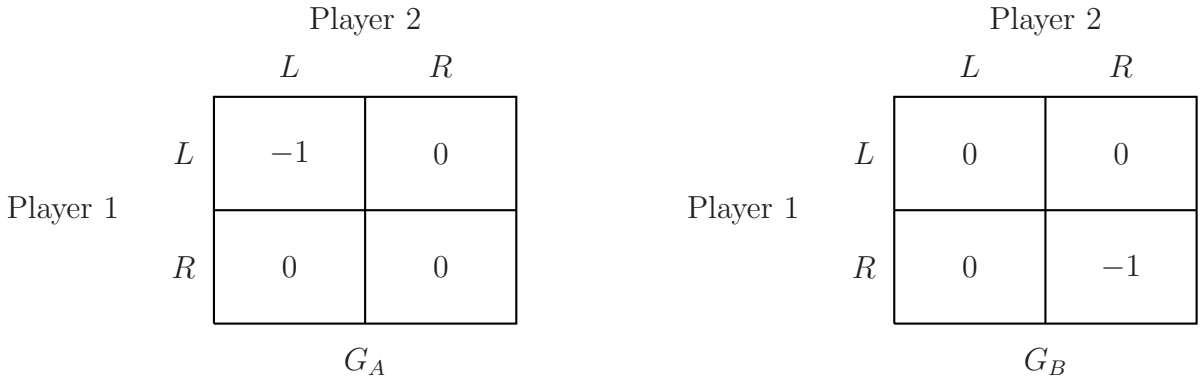


Figure 2: Rewards in the two states

A.3 When Partial Disclosure is Optimal

Consider finally the two stage games given in Figure 3. Again, both games are equally likely, player 1 is informed, while player 2 is not. It is easy to check that the values of the games are $w_A = 2/3$ and $w_B = -2/3$. Note also that, if player 1 were not informed of the game played either, the value of the game would be zero, with the optimal strategies being R for player 1 and L for player 2.

Therefore, player 1 faces a dilemma. He would prefer to disclose his information, *i.e.* play the optimal strategy in G_A in every period, if the stage game is G_A (as $2/3 > 0$), but hide it if the stage game is G_B (as $-2/3 < 0$), by acting as though he himself did not know the game being played. Of course, this is not possible, because player 2 knows that player 1 knows the game being played, and would infer from the second course of action that the stage game is actually G_B, leading to a payoff no larger than $-2/3$ to player 1. This would mean that, on average, player 1 gets no more than 0. Can he do better?

Let us start with an upper bound on what player 1 can get. If player 2 were to play L at each stage with probability $2/3$, randomizing independently in each period, player 1_B could do no better than choosing R at every stage, so that, independently of what player 1_A does, player 1's payoff would be at most

$$\frac{1}{2}\left(\frac{2}{3}\right) + \frac{1}{2}\left(-\frac{1}{3}\right) = \frac{1}{6}.$$

In order to better understand what player 1 can guarantee, it is useful to start with a careful examination of the one-shot game in which player 2's belief p that the stage game is G_A is treated as a parameter. The normal form of this game (call it $\Gamma(p)$) is given in the left panel of Figure 4 (the pair of actions of player 1 correspond to what he does in game G_A and G_B respectively).

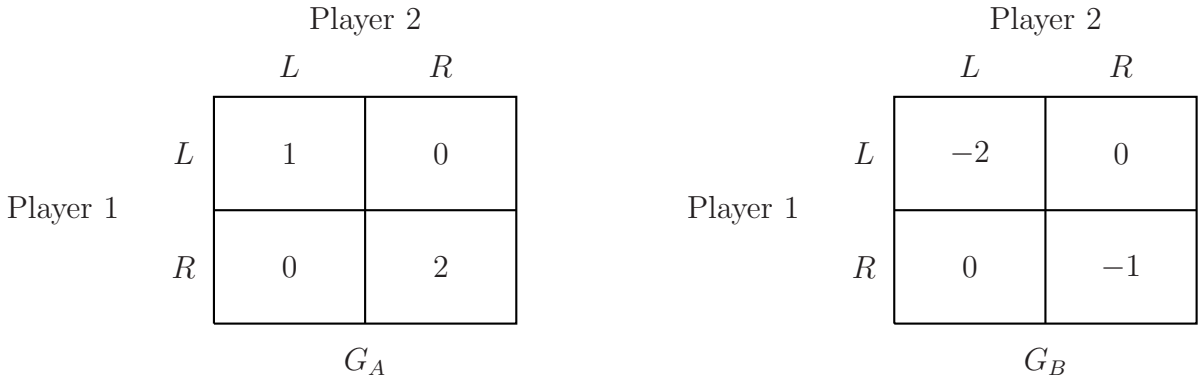


Figure 3: Rewards in the two states

The solution of this game depends on p:

1. $p \in (0, 1/3)$: player 1 should randomize between RL and RR only, with probabilities $(1-3p)/3(1-p)$ on RL, and the value is

$$v(p) = -\frac{2}{3}(1-3p).$$

2. $p \in (1/3, 1)$: player 1 should randomizes between LR and RR only, with probability $(3p-1)/3p$ on LR, and the value is

$$v(p) = \frac{1}{3}(3p-1).$$

3. $p = 1/3$: player 1 should play RR with probability 1, and the value is $v(1/3) = 0$.

4. $p \in \{0, 1\}$: the values are $-2/3$ and 0 respectively.

The value $v(p)$ is shown in the right panel of Figure 4.

Returning to the infinite-horizon, we now take advantage of the fact that this is a zero-sum game: because σ_1 must be optimal *for every* strategy of player 2, it must be optimal against the strategy that 2 would have been played *had he known* what exact strategy player 1 ends up playing. So we may assume that player 1 is committed to σ_1, and that player 2 knows both σ_1 and that player 1 is committed to it. Knowing σ_1, player 2 can update p_t in every period, and it is plausible (and indeed correct, if it not immediate) that a best-reply for player 2 consists in playing his optimal strategy from $\Gamma(p_t)$ in each period. (Note that this does not mean that this strategy is optimal for player 2, because we do not know that it is a best-reply to *all* strategies σ_1 simultaneously).

Player 1 \ Player 2	L	R
LL	$3p-2$	0
LR	p	$p-1$
RL	$2p-2$	$2p$
RR	0	$3p-1$

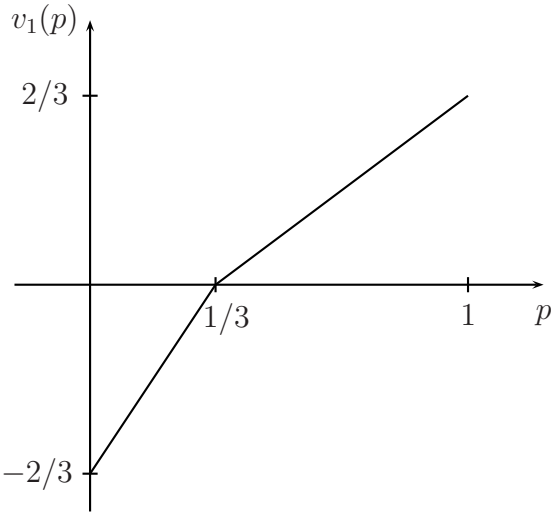


Figure 4: The normal form of $\Gamma(p)$ and the graph of its value

Returning to player 1, we can now describe an optimal strategy for player 1 (as we shall show). Namely, *player 1 should play as if the game he is facing is* $\Gamma(p_t)$ and play accordingly.

For instance, he starts playing as in $\Gamma(1/2)$. That is he plays L with probability $\alpha_A = 1/3$ and $\alpha_B = 0$ according to whether he knows that the state is A or B. If the realized action is L, player 2 updates to $p = 1$, and play proceeds from that point on as in $\Gamma(1)$, *i.e.* player 1 plays L with probability $2/3$ (of course this only happens if the game is G_A). If the realized action is R, player 2 correctly updates to the belief $p_1 = 2/5$.

More generally (for $p > 1/3$), a choice of L leads to $p_{t+1} = 1$ and future play as in $\Gamma(1)$. On the other hand, a choice of R leads to

$$p_{t+1} = \frac{p_t \cdot \frac{1}{3p_t}}{p_t \cdot \frac{1}{3p_t} + (1-p_t)\cdot 1} = \frac{1}{4-3p_t},$$

whose solution (given $p_0 = 1/2$) is

$$p_t = \frac{3^t + 1}{3^{t+1} + 1},$$

which asymptotically tends to $1/3$. The probability q_t that the belief has not jumped to 1 by (the beginning of) stage $t \geq 1$ is given by

$$q_t = \times_{\tau=0}^{t-1} \left(p_\tau \cdot \frac{1}{3p_\tau} + (1-p_\tau)\cdot 1 \right) = \frac{3}{4}(1 + 3^{-(t+1)}),$$

so that the expected reward collected by player 1 in period t is given by

$$\left(1-\frac{3}{4}(1+3^{-(t+1)})\right)v(1)+\frac{3}{4}(1+3^{-(t+1)})v\left(\frac{3^t+1}{3^{t+1}+1}\right)\geq\frac{1}{6}-\frac{3^{-(t+1)}}{2},$$

where we use that $v(1)=2/3$ and $v(p)\geq 0$ for $p\geq 1/3$. Therefore, this strategy guarantees player 1 a payoff of $1/6$ (recall that the payoff is the limit of the means). Because we have seen that player 2 can guarantee $1/6$ as well, this must be the value of the game.

As mentioned, player 1's strategy is optimal: even when player 2 best-replies, player 1 secures $1/6$, the maximum he can hope for. On the other hand, player 2 strategy is "only" a best-reply, not an optimal strategy. Still, this suffices to argue that the strategy profile that we have described is a Nash equilibrium, and that all Nash equilibrium must give player 1 a payoff of $1/6$.

Note that the probability of the game is eventually disclosed is only $1/4$, as the probability that the asymptotic belief is $1/3$ is $3/4$.

Here is a simpler way to think about what player 1 could do, if he had access to coins. Player 1 prepares two non-symmetric coins. One coin always turns Heads, the other is equally likely to turn Heads and Tails. Player 2 does not get to see what coin Player 2 picks, but he gets to see the outcome of the coin flip. Here is a strategy available to player 1: if the game is G_B, he picks the coin that always turns Heads, while he picks the other coin if the game is G_A. As a result, player 2 updates his beliefs to either 1 if he sees Tails, and to $1/3$ if he sees Heads. After the coin has been tossed, player 1 plays from that point as if he did not know the true game, that is, he discloses no further information. Note that $v(1/3)=0$. In that fashion, he secures a payoff of

$$\frac{1}{4}\left(\frac{2}{3}\right)=\frac{1}{6},$$

the weighted average of the values $v(1)$ and $v(1/3)$, with weights given by the probabilities of the two events.

B A General Result for Zero-Sum Games

The previous examples illustrated that partial, full or no disclosure will occur depending on the underlying game. Is there any way to determine what the optimal disclosure is, and what the value of the information is, without having to use as much ingenuity as we had to?

It turns out that there is a very simple way to do so. As in our examples, though, deter-

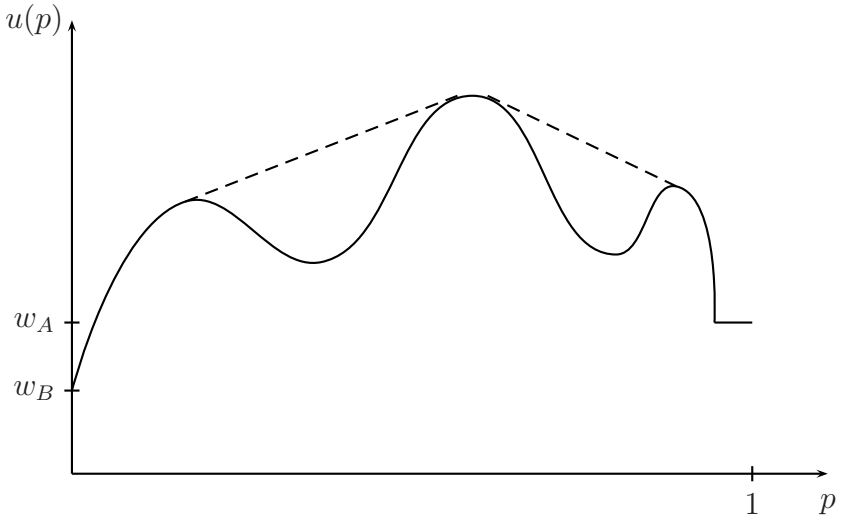


Figure 5: The concavification of $u(p)$

mining the optimal strategies can be quite complicated. Fix a game with one-sided incomplete information. That is, player 1 is informed of which one of finitely many stage games is the one that is actually being played, while player 2 has a prior belief p_0 over this set. Consider (as in our last example) the one-shot game in which player 2's belief is given by p (player 1 being informed) and let $u(p)$ denote the value of this game. It can be shown quite easily that u is continuous in p. In general, however, it isn't a concave function, and it is useful to define the function cav $u(p)$ as the smallest concave function larger than or equal to u. See Figure 5. This is also called the **concavification** of u.

Aumann and Maschler have established the following remarkable result.

Theorem 1 *The value of the infinitely repeated game with incomplete information exists, and is equal to cav $u(p_0)$, where cav u is the concavification of u.*

Proof. One of the inequalities is straightforward. Given the K possible stage games, let $(p_e)_{e\in E}$ be finitely many points in $\triangle K$, and let $\lambda = (\lambda_e)_{e\in E} \in \triangle E$ be such that $\sum_{e\in E} \lambda_e p_e = p$. That is, p is a convex combination of the different p_e's. Suppose that we compare the following two games. In both games, Nature first draws $e \in E$ according to λ, then $k \in K$ is chosen according to p_e. Player 1 is informed of both the realized e and k. In the first version, player 2 is informed of e; in the second, he is not. Clearly, player 2 is better off in the first, because he can always use in the first version the strategy that he would pick in the second.[1] Therefore,

[1]Note that this argument again only applies to zero-sum games: to prove that the value is no more than a

player 1 is better off in the second, and so, if he can secure a given $f(p)$, he can also guarantee cav $f(p)$. Furthermore, given player 2's belief p, player 1 can secure $u(p)$ by playing i.i.d. the optimal strategy of the average game $\Gamma(p) = \sum_k p_k G_k$. Combining both observations, player 1 can secure cav $u(p)$.

The other direction requires some more work. As usual, let $H^t := (A_1 \times A_2)^t$ denote the set of histories of length t (recall that monitoring is perfect), $H^\infty = (A_1 \times A_2)^\infty$ denote the set of infinite histories, $H = \cup_t H^t$ the set of all histories. Let $\mathcal{H}^t$ denote the σ-algebra generated by the cylinders above H^t, and set $\mathcal{H}^\infty = \bigvee_t \mathcal{H}^t$. A (behavior) strategy profile σ induces a distribution over the measurable space $(K \times H^\infty, 2^K \otimes \mathcal{H}^\infty)$, our reference probability space. We can then define, for all $k \in K$,

$$p_k^t = \mathbb{E}[k \mid \mathcal{H}^t]$$

as the posterior probability distribution over K of player 2 at stage t, in short, his belief. We set $p_0 = p$, the prior belief. By definition, the sequence (p^t) is a $\mathcal{H}^t$-martingale, which means that, for all $t \geq 0$,

$$\mathbb{E}[p^{t+1} \mid \mathcal{H}^t] = p^t,$$

and in particular $\mathbb{E}[p^t] = p$. It is a well-known application of the Cauchy-Schwartz inequality that such a martingale satisfies

$$\frac{1}{t+1} \sum_{s=0}^{t} \mathbb{E}\|p^{s+1} - p^s\| \leq \sum_{k \in K} \sqrt{\frac{p_k(1-p_k)}{t+1}}. \tag{1}$$

Note that the right-hand side is bounded above by $\sqrt{|K|-1}/(t+1)$ (the term $\sum_k \sqrt{p^k(1-p^k)}$ is maximized by setting $p_k = 1/|K|$). Therefore, "beliefs cannot vary much."

An action of player 1 in stage t is summarized by $\alpha_k^t \in \triangle A_1$, where k refers to the true state G_k, and given $\bar{\alpha}^t := \sum_k p_k^t \alpha_k^t$, Bayes' rule yields that

$$\mathbb{P}[k \mid \mathcal{H}^t, a_1^t = a_1] = \frac{p_k^t \alpha_k^t(a_1)}{\bar{\alpha}^t(a_1)},$$

whenever $\bar{\alpha}^t(a_1) > 0$. Close strategies imply close posterior beliefs, since it follows from this

given v, it suffices to show that player 2 has some strategy that guarantees it, not that this strategy is actually optimal for player 2.

equation that, for any α,

$$\mathbb{E}\left(\|\alpha^t - \bar{\alpha}^t\| | \mathcal{H}^t\right) = \mathbb{E}\left(\|p^{t+1} - p^t\| | \mathcal{H}^t\right).$$

Finally, we relate the distance between payoffs and strategies. Given σ and a period t, let $\tilde{\sigma}_1^t$ denote player 1's strategy that coincides with σ_1 except in stage t, where $\tilde{\sigma}_1^t$ specifies $\bar{\sigma}_1^t$. Then the random payoff $u^t(\sigma)$ conditional on $\mathcal{H}^t$ that obtains in period t given σ satisfies

$$|u^t(\sigma) - u^t(\tilde{\sigma}_1^t, \sigma_2)| \leq M \cdot \sum_k p_k^t \|\sigma_k^t - \bar{\sigma}^t\| = M \cdot \mathbb{E}\left(\|\sigma^t - \bar{\sigma}^t\| | \mathcal{H}^t\right),$$

where M is a bound on all payoffs of the stage games G_k.

Consider now the game repeated $t+1$ times, $\Gamma^t(p)$. We claim that, for any strategy σ_1 of player 1, player 2 has a strategy σ_2 that guarantees an expected payoff in period t no larger than

$$\text{cav } u(p) + \frac{M}{\sqrt{t+1}} \sum_k \sqrt{p_k(1-p_k)}.$$

The result then follows (formally, first as an application of the minmax theorem, then) by taking $t \to \infty$. To see that player 2 can guarantee this, consider the following strategy of player 2. In period t, compute p^t and play an action α_2^t that is optimal in the one-shot game $\Gamma(p^t)$. From our previous remarks, we have

$$u^t(\sigma) \leq u^t(\tilde{\sigma}_1^t, \sigma_2) + M \cdot \mathbb{E}\left(\|p^{t+1} - p^t\| \mid \mathcal{H}^t\right).$$

Because $\tilde{\sigma}^t$ specifies a constant (state-independent) action in period t, and given that σ_2 is optimal in $\Gamma(p^t)$, it follows that

$$u^t(\tilde{\sigma}_1^t, \sigma_2) \leq u(p^t) \leq \text{cav } u(p^t).$$

Adding up over $s = 0, \ldots, t$ and taking expectation (applying Jensen's inequality to the term involving cav $u(p)$),

$$v^t(\sigma) \leq \text{cav } u(p) + \frac{M}{t+1} \sum_{s=0}^{t} \mathbb{E}\|p^{t+1} - p^t\|.$$

Finally, apply 1. ■

As a final remark, the Cauchy-Schwartz also implies

$$\sum_{t=0}^{\infty} \delta^t (1-\delta) \|p_k^{t+1} - p_k^t\| \leq \sqrt{\frac{1-\delta}{1+\delta}} \sqrt{p_k(1-p_k)},$$

from which it follows that the value v_δ of the discounted game satisfies

$$0 \leq v_\delta(p) - \text{cav } u(p) \leq M \sum_k \sqrt{p_k(1-p_k)} \sqrt{1-\delta}.$$

Zero-sum games with incomplete information have been extensively studied, and the characterization of the value extended to the case of imperfect (public or private) monitoring.

III Non-Zero Sum Games with Known-Own Payoffs

A Belief-Free Equilibria, Perfect Monitoring

Uncertainty is capture by a *state*. The set of states is $K := \{1, \ldots, K\}$, finite. As usual, Player i chooses action a_i from A_i, finite, and $a \in A := \prod_i A_i$ is an action profile.

Player i's reward is a map $u_i : K \times A \to \mathbb{R}$. Let $M := \max_{i=1,\ldots,n, k \in K, a \in A} ||u_i(k,a)||$. A reward profile is denoted $u := (u_1, \ldots, u_n)$. Mixed actions of player i are denoted α_i. The definition of rewards is extended to mixed, possibly correlated, action profiles $\mu \in \triangle A$ in the usual way.

At the beginning of the game, each player receives once and for all a signal that allows her to narrow down the set of possible states of nature. Without loss of generality, this process can be represented by an information structure $\mathcal{I} := (\mathcal{I}_1, \ldots, \mathcal{I}_n)$, where $\mathcal{I}_i$ denotes player i's information partition of K. We let $I_i(k)$ denote the element of $\mathcal{I}_i$ containing k. We refer to $I_i(k) =: \theta_i \in \Theta_i$ as player i's *type*, and write $\Theta := \prod_i \Theta_i$, and $\Theta_{-i} := \prod_{j \neq i} \Theta_j$. Given $\theta \in \Theta$, $\kappa(\theta) := \bigcap_{i \in N} \theta_i$ denote the set of states that are consistent with type profile θ. Also, for $\theta_{-i} \in \Theta_{-i}$, we write $\kappa(\theta_{-i}) := \bigcap_{j \neq i} \theta_j$ for the set of states that are consistent with a type profile of all players but i. We do not require that $\kappa(\theta) \neq \varnothing$: it might be that some type profile cannot arise. Similarly, it might be that $|\kappa(\theta)| > 1$, so that the join of the players' information partitions need not reduce to the state: that is, the state of nature need not be distributed knowledge. Without loss, assume there are no redundant states: if $k, k' \in \kappa(\theta)$, then $u(k, \cdot) \neq u(k', \cdot)$. The information partitions are common knowledge, but the realized signal is private information.

The game is infinitely repeated, with periods $t = 0, 1, 2, \ldots$ A history of length t is a vector

$h^t \in H^t := A^t$ ($H^0 := \{\varnothing\}$). An outcome is an infinite history $h \in H := A^\infty$. Neither mixed actions nor realized payoffs are observed. On the other hand, realized actions are perfectly observed. A behavior strategy for player i's type θ_i is a mapping $\sigma_{i,\theta_i} : \cup_{t\in\mathbb{N}} H^t \to \triangle A_i$. We write $\sigma_i := \{\sigma_{i,\theta_i}\}_{\theta_i\in\Theta_i}$ for player i's strategy, and $\sigma := (\sigma_1, \ldots, \sigma_N)$ for a strategy profile.

Players use a common discount factor $\delta < 1$. The *payoff* of player i in state k is the expected average discounted sum of rewards, where the expectation is taken with respect to mixed action profiles. That is, given some outcome $h = (a_0, \ldots, a_t, \ldots)$, player i's payoff in state k is

$$\sum_{t\geq 0}(1-\delta)\delta^t u_i(k, a_t).$$

As usual, the domain of rewards is extended to mixed action profiles and strategy profiles. Given a strategy profile σ, let $\mu_k \in \triangle A$ denote the *occupation measure* over action profiles induced by σ when the state is k, that is, for every $a \in A$,

$$\mu_k(a) := (1-\delta)\,\mathbb{E}_\sigma\left[\sum_{t\geq 0}\delta^t 1_{\{a_t=a\}}\right].$$

Let $u(k, \mu_k) \in \mathbb{R}^n$ denote the players' payoff vector in state k under the occupation measure μ_k:

$$u(k, \mu_k) := \sum_{a\in A}\mu_k(a)u(k, a).$$

Definition: A *belief-free equilibrium* (hereafter, an equilibrium) is a strategy profile σ such that, for every state k, σ is a subgame-perfect Nash equilibrium of the game with rewards $u(k, \cdot)$. A vector $v \in \mathbb{R}^{nK}$ is an *equilibrium payoff vector* if there exists an equilibrium σ such that $v = u(\sigma)$.

We write v^k for the payoff vector in state k, and B_δ for the set of belief free equilibrium (BFE) payoff vectors of the δ-discounted game. The goal is to characterize $\lim_{\delta\to 1} B_\delta$ and establish conditions under which this set is non-empty.

A.1 Necessary Conditions

We first derive necessary conditions for a vector $v \in \mathbb{R}^{nK}$ to be an equilibrium payoff vector. These are of three types: feasibility, incentive compatibility, and (individual and joint) rationality.

Feasibility First, a payoff vector must obviously be feasible.

Definition: The payoff vector $v \in \mathbb{R}^{nK}$ is *feasible* if there exists $(\mu_k)_{k\in K} \in (\triangle A)^K$ such that

1. $\forall k \in K : v^k = u(k, \mu_k)$;

2. $\forall k, k'$: $I_i(k) = I_i(k')$ $\forall i = 1, \ldots, n \Rightarrow \mu_k = \mu_{k'}$.

The first condition is obvious: there must exist an occupation measure μ_k that yields the payoff vector v^k. The second is a measurability restriction. If players cannot collectively distinguish two states, then the equilibrium occupation measures over action profiles must be the same in both states. Given the second condition, we may write μ_θ for the occupation measure. Conversely, the notation $(\mu_\theta)_{\theta\in\Theta}$ will imply that the set $(\mu_k)_{k\in K}$ satisfies the second condition.

Incentive Compatibility If signals θ_i and θ_i' are consistent with the profile θ_{-i}, it must be that player i prefers the occupation measure $\mu_{\theta_i,\theta_{-i}}$ to $\mu_{\theta_i',\theta_{-i}}$ in every possible state given (θ_i, θ_{-i}). Thus, if v is an equilibrium payoff vector, it must be feasible for some distributions satisfying a set of incentive compatibility conditions. Define UD_i (for unilateral deviation) as the set of triples $(\theta_i, \theta_i', \theta_{-i}) \in \Theta_i \times \Theta_i \times \Theta_{-i}$ such that $\kappa(\theta_i, \theta_{-i}) \neq \varnothing$ and $\kappa(\theta_i', \theta_{-i}) \neq \varnothing$. The incentive compatibility conditions are

$$\forall i, (\theta_i, \theta_i', \theta_{-i}) \in UD_i, k \in \kappa(\theta_i, \theta_{-i}) : u_i(k, \mu_{\theta_i,\theta_{-i}}) \geq u_i(k, \mu_{\theta_i',\theta_{-i}}). \ (IC(i, \theta_i, \theta_i', \theta_{-i}))$$

Lemma: If $v \in B_\delta$, then v is feasible for some $(\mu_\theta)_{\theta\in\Theta}$ that satisfy $IC(i, \theta_i, \theta_i', \theta_{-i})$ for all $i = 1, \ldots, n$ and $(\theta_i, \theta_i', \theta_{-i}) \in UD_i$.

Proof: Suppose for the sake of contradiction that for some $i \in N$ and $(\theta_i, \theta_i', \theta_{-i}) \in UD_i$, the reverse inequality holds. Consider now the game of complete information in which the state is k, and consider player i of type θ_i. By playing as if her type were θ_i', player i can guarantee $u_i(k, \mu_{\theta_i',\theta_{-i}})$, which exceeds her equilibrium payoff $u_i(k, \mu_{\theta_i,\theta_{-i}})$. This is a profitable deviation. □

Individual and Joint Rationality A deviating player might be easy to identify or not. For instance, if player i chooses an action that is inconsistent with all her types' equilibrium strategies, then it is common knowledge that i deviated. Since we seek to find a necessary condition that player i's equilibrium payoff vector must satisfy, the more effective the punishment, the weaker the condition. Thus, we start by assuming that, if player i deviates, all other players commonly know the information distributed among them, as these are the most favorable conditions for a

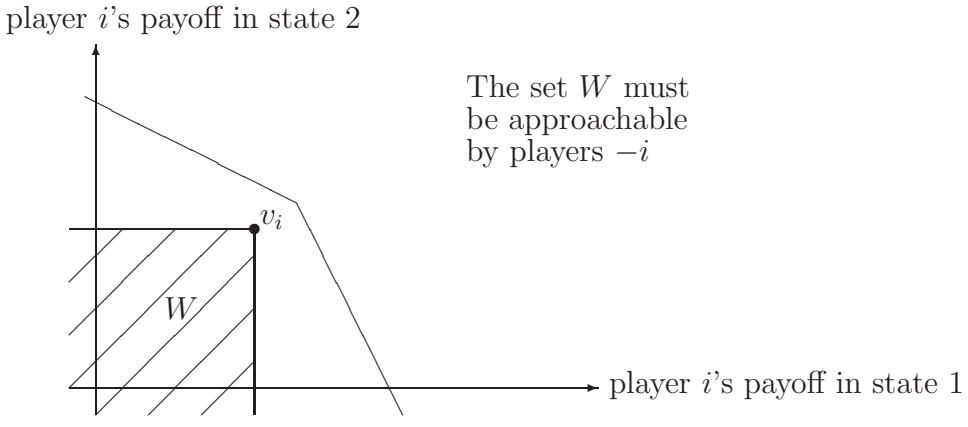


Figure 6: Players $-i$ must have a strategy that guarantees that i's payoff lies in W.

punishment. Similarly, we may assume that player i's deviation is common knowledge, even if, for some deviations, this need not be.

Still, if the set of states $\kappa(\theta_{-i})$ is not a singleton, players $-i$ cannot tailor the punishment strategy to the state of the world. Suppose, for instance, that $\kappa(\theta_{-i}) = \{1, 2\}$, as in Figure 6. After a deviation, player $-i$'s strategy must be effective in both games of complete information simultaneously, and guarantee that player i's payoff is lower than v_i in both coordinates, independently of i's strategy. Note that it does not matter whether i can distinguish these states.

Determining for which values of v_i players $-i$ have such a strategy may appear a formidable task, but as is well-known, this is equivalent (at least in the undiscounted case) to the orthant $W := \{v_i\} - \mathbb{R}^2_+$ being an approachable set. Necessary and sufficient conditions are given by Blackwell (1956). Define, for $\theta_{-i} \in \Theta_{-i}$,

$$\varphi_{i,\theta}(q) := \min_{\alpha_{-i} \in \prod_{j \neq i} \triangle A_j} \max_{a_i \in A_i} \sum_{k \in \kappa(\theta_{-i})} q(k) u_i(k, \alpha_{-i}, a_i).$$

For each player i and each $\theta_{-i} \in \Theta_{-i}$, consider the set of inequalities

$$\forall q \in \triangle \kappa(\theta_{-i}) : \quad \sum_{k \in \kappa(\theta_{-i})} q(k) v_i^k \geq \varphi_{i,\theta}(q). \qquad (IR(i, \theta_{-i}))$$

These inequalities are generalizations of individual rationality for two players. If $\kappa(\theta_{-i}) = \varnothing$, they are vacuously satisfied. If $\kappa(\theta_{-i}) = \{k\}$, they reduce to the definition of individual rationality under complete information: $v_i^k \geq \operatorname{val} u_i(k, \cdot)$, where $\operatorname{val} u_i(k, \cdot)$ is i's minmax payoff in state k.

Lemma: If $v \in B_\delta$, it satisfies the inequalities $(IR(i, \theta_{-i}))$ for each player i and θ_{-i}.

Proof: If one condition is violated, there exists i, a type profile θ_{-i} and $q \in \triangle\kappa(\theta_{-i})$ such that the reverse inequality holds. This implies that for every α_{-i}, there exists $a_i(\alpha_{-i}) \in A_i$ such that

$$\sum_{k\in\kappa(\theta_{-i})} q(k)u_i(k, \alpha_{-i}, a_i(\alpha_{-i})) > \sum_{k\in\kappa(\theta_{-i})} q(k)v_i^k. \tag{2}$$

Assume that v is in B_δ and let σ be the corresponding equilibrium. Note that players $-i$ play the same strategy in each state $k \in \kappa(\theta_{-i})$. Consider the strategy τ_i of player i that plays $a_i(\alpha_{-i})$ after a history h^t such that $\sigma_{-i}(h^t) = \alpha_{-i}$. The reward of i under (τ_i, σ_{-i}) satisfies (2) and thus, so does the payoff. Thus, there exists a state $k \in \kappa(\theta_{-i})$ at which τ is a profitable deviation. □

Under these conditions, following Blackwell (1956), players $-i$ can devise a punishing strategy against player i. Given θ_{-i}, and any payoff vector v that satisfies these inequalities strictly, there exists $\varepsilon > 0$ and a strategy profile $\widehat{s}_{-i}^{\theta_{-i}}$ for players $-i$ such that, if players $-i$ use $\widehat{s}_{-i}^{\theta_{-i}}$, then player i's undiscounted payoff in any state k that is consistent with θ_{-i} is less than $v_i^k - \varepsilon$ in any sufficiently long finite-horizon version of the game, no matter i's strategy. By continuity, this also holds true for sufficiently long finite-horizon versions of the game when payoffs are discounted, provided the discount factor is high enough, fixing the length of the game. When players $-i$ use $\widehat{s}_{-i}^{\theta_{-i}}$, they *minmax* player i; player i is the *punished* player, and players $-i$ are the *punishing* players.

While individual rationality is a necessary condition, it is not the only one. There are other conceivable deviations, leading to additional conditions. Even if a deviation gets detected, it might not be possible to identify the deviator: i's action might be consistent with some of her types' strategies, and so might player j's action, but no pair of types for which both actions would be both consistent might exist. Then it is common knowledge that some player deviated, but not necessarily whether it is player i or j, unless there are only two players. Let D be the set of type profiles that are inconsistent, but could arise if there was a unilateral deviation. That is, θ is in D if $\kappa(\theta) = \varnothing$ and $\Omega_\theta := \{(i, \theta_i') \mid i = 1, \ldots, n,\ \kappa(\theta_i', \theta_{-i}) \neq \varnothing\} \neq \varnothing$. If players were to report their types, and the reported profile was in D, players would know that someone has lied. The set Ω_θ is the set of pairs (player, type) that might have caused the announcement θ. For each $\theta \in D$, consider

$$\exists \mu \in \triangle A,\ \forall (i, \theta_i') \in \Omega_\theta, \forall k \in \kappa\left(\theta_i', \theta_{-i}\right) :\ v_i^k \geq u_i(k, \mu). \qquad (JR\left(\theta\right))$$

These inequalities are called Joint Rationality (JR), since they involve payoffs of different players

simultaneously. Note that joint rationality does not imply individual rationality (there is no requirement that player i's action be a best-reply), nor is it implied by it.

Lemma 1 *Every $v \in B_\delta$ satisfies all constraints $(JR(\theta))_{\theta \in D}$.*

Proof: Let $v \in B_\delta$ be an equilibrium payoff vector and σ be the corresponding equilibrium. Let $\theta = (\theta_i)_i \in D$ and consider for each $(i, \theta_i') \in \Omega_\theta$ the deviation τ^i by i such that, if her type is θ_i', she plays as if she were of type θ_i, *i.e.* $\tau_{i,\theta_i'} = \sigma_{i,\theta_i}$, and which coincides with σ_i for all other types. Take two elements (i, θ_i') and (j, θ_j') in Ω_θ. The distribution over outcomes under $(\tau_{i,\theta_i'}, \sigma_{-i,\theta_{-i}})$ and $(\tau_{j,\theta_j'}, \sigma_{-j,\theta_{-j}})$ are the same, *i.e.* this is the distribution under $\sigma_\theta = (\sigma_{l,\theta_l})_{l=1,\ldots,n}$. In words, there is no way to distinguish the situation in which player i consistently mimics type θ_i and the one in which player j consistently mimics type θ_j. Let $\mu \in \triangle A$ denote the occupation measure generated by σ_θ. If $JR(\theta)$ is violated, there exists a player i and a state $k \in \kappa(\theta_{-i})$ such that player i's equilibrium payoff in state k, v_i^k, is strictly lower than her payoff if she were to follow σ_{θ_i}, a contradiction. □

Note that the conditions $JR(\theta)$ are closely related to the conditions $IR(i, \theta)$. Indeed, using the minmax theorem, we may write those inequalities as

$$\forall q \in \triangle\{(i,k) :\ k \in \kappa(\theta_{-i})\} \ :\ \sum_{i,k} q(i,k) v_i^k \geq \min_{a \in A} \sum_{i,k} q(i,k)\, u_i(k,a),$$

which suggests interpreting the identity of the deviator as part of the uncertainty itself. For the sake of brevity, we often refer to each type of condition simply as IC, IR, or JR.

A.2 Sufficient Conditions

Let $V^* \subset \mathbb{R}^{nK}$ denote the set of feasible payoff vectors that satisfy IC, IR, and JR. We show that this set characterizes the set of belief-free equilibrium payoff vectors, up to its boundary. Let $\hat{K} := \left\{k \in K : \bigcap_{i=1,\ldots,n} I_i(k) \neq \{k\}\right\}$ be the set of states that cannot be distinguished by the join of the players' information partitions. Let $\hat{u}$ be the $(n \times |\hat{K}|, |A|)$-matrix $(u_i^k(a))$, where $k \in \hat{K}$. The reward u is *generic* if the matrix $\hat{u}$ has rank $n \times |\hat{K}|$. Viewing any such matrix as an element of $\mathbb{R}^{n|\hat{K}||A|}$, this condition is generically satisfied whenever $|A| \geq n|\hat{K}|$. A generic reward function guarantees that there are occupation measures providing each player with appropriate incentives even in states that cannot be distinguished. The next result characterizes the limit set of BFE payoffs.

Theorem 2 *If $v \in \text{int } V^*$ and u is generic, there exists $\bar{\delta} < 1$, $\forall \delta \in (\bar{\delta}, 1)$, $v \in B_\delta$.*

The proof is a variation on the construction of Fudenberg and Maskin (1986).

A.3 Existence

Write $V^*(\mathcal{I}, u)$ for V^* when the information structure $\mathcal{I}$ and the reward function is u. When is $V^*(\mathcal{I}, u) \neq \varnothing$? Assume that this is a game of known-own payoff.

Definition: The game has *known-own payoffs* (KOP) if the reward function of each player i depends only on the action profile and her type: $\forall a \in A$, $\forall k, k' \in K$:

$$I_i(k) = I_i(k') \Longrightarrow u_i(k, a) = u_i(k', a).$$

Let $\mathcal{S}_\mathcal{I}$ be the set of KOP reward functions when the information structure is $\mathcal{I}$.

Note that the definition of known-own payoff implies that $\bigcap_{i \in N} I_i(k) = \{k\}$ (recall that there are no redundant states). It follows from the results below that in two-player games with KOP, existence obtains whenever information is one-sided, that is, whenever player 1 has more information than player 2. However, the following example shows that having one fully informed player is not sufficient to ensure existence with more than two players.

Example 3 *There are three states k, k', k''. The information of player 1 is $I_1(k) = \{k, k''\}$, $I_1(k') = \{k'\}$. The information of player 2 is $I_2(k) = \{k, k'\}$, $I_2(k'') = \{k''\}$. Player 3 knows the state. The payoff matrix is as follows.*

	L	R
T	$3,1,0$	$0,0,0$
B	$0,0,0$	$1,3,0$

state k

	L	R
T	$3,0,3$	$0,1,3$
B	$0,0,3$	$1,1,0$

state k''

	L	R
T	$1,1,0$	$1,0,3$
B	$0,0,3$	$0,3,3$

state k'

In this game, $V^* = \varnothing$. Assume instead that there exists $v \in V^*$. Individual rationality of players 1 and 2 imply that in k', T is always played, and (T, R) is played with a (discounted) frequency no greater than $1/4$. Player 3's payoff in k' is thus $v_3^{k'} \leq 3/4$. Similarly, in k'', R is always played, and (T, R) with frequency no greater than $1/4$. Player 3's payoff in k'' is thus $v_3^{k''} \leq 3/4$. Consider now the inconsistent reports in which player 1 claims that the state is k', while player

3 claims it is k. Continuation play must "punish" player 1 in k, and player 3 in k′. Note that, for all a, $u_1^k(a) + u_3^{k'}(a) \geq 3$. *Assume that player 1's payoff in k is* $v_1^k \leq \frac{11}{16}3$. *Then*

$$v_1^k + v_3^{k'} \leq \frac{11}{16}3 + 3/4 = 45/16 < 3,$$

which is impossible: from JR, there must exist a distribution α *such that* $v_1^k \geq u_1^k(\alpha)$ *and* $v_3^{k'} \geq u_3^{k'}(\alpha)$ *and* $u_1^k(\alpha) + u_3^{k'}(\alpha) \geq 3$. *So* $v_1^k > \frac{11}{16}3$. *A similar argument (considering the case in which player 2 claims that the state is* k'' *and player 3 claims it is k) yields* $v_2^k > \frac{11}{16}3$. *So* $v_1^k + v_2^k > 66/16 = 4 + 1/8$, *which is impossible, since no action profile in state k yields* $u_1^k + u_2^k > 4$.

To attempt a characterization of the information structures for which existence obtains, we first reduce the problem by distinguishing among subsets of states that can be made common knowledge among players even under unilateral deviations. This defines a partition over the set of states K. An element of this partition is a *majority component*: if the true state k belongs to the majority component A, then under strategies that ask players to report whether the state is in A or not, it becomes common knowledge that the state lies in A once reports are made, even if a player unilaterally deviates. This requires that, for all $k'' \in K \setminus A$, at least three players know that the state is not k'', so that, if one of them deviates, at least two players' reports rule out k''. Conversely, if two states k and k' belong to the same majority component A, then, for some player's report, there are no two other players who could, by reporting truthfully, distinguish between k and k'.

Definition:

- For each pair of states k, k', let $\nu(k, k')$ be the number of players who distinguish k from k'. Define the binary relation R by kRk' iff $\nu(k, k') \leq \min\{2, n-1\}$.

- Let $k \sim k'$ iff there is a chain of states $k = k_1, k_2, \ldots, k_l = k'$ such that $k_m R k_{m+1}$ for each m. A *majority component* of K is an equivalence class of this relation.

Note that R is symmetric but not necessarily transitive, and $\sim$ is the transitive closure of R (*i.e.* the smallest transitive extension of R), thus it is an equivalence relation. If A, B are two distinct majority components of K and $n \geq 3$, then for each $k \in A$ and each k' in B, $\nu(k, k') \geq 3$. Otherwise, there would exist a link (for the relation R) between some point in A and some point in B, and thus a chain linking any point in A to any point in B. For 2-player games, two states belong to the same majority component only if they can be distinguished by at most one player.

The study of belief-free equilibria can be made on each majority component separately. Given $A \subseteq K$, let $\mathcal{I}_A$ denote the information structure on A induced by $\mathcal{I}$:

$$I_{A,i}(k) = I_i(k) \cap A, \ \forall i = 1, \ldots, n, \ \forall k \in A.$$

By definition, a BFE given K and $\mathcal{I}$ must induce a BFE given A and $\mathcal{I}_A$ since the equilibrium must be free of beliefs concentrated on A. Conversely, if we have a BFE on each majority component A, we also have a global BFE since A can be made common knowledge by truthful announcements (under any unilateral deviation). The next lemma summarizes this discussion.

Lemma: $V^*(u, \mathcal{I}) \neq \varnothing$ iff for each majority component A, $V^*(u, \mathcal{I}_A) \neq \varnothing$.

As explained below, if V^* is nonempty in all games with KOP, then for each state k, first, there exists a player i who is as well informed as all others at that state, and second, either no player can distinguish any two states for which she is not the best informed player (if she ever is), or there is a second player $j \neq i$ who is as well informed as all players but i at that state. In this latter case, V^* is nonempty in all games with KOP. More formally, *player i has more information than player j* if player i can deduce player j's type from her own type, *i.e.* if player i's information partition is finer than player j's partition: $I_i(k) \subseteq I_j(k)$ for each k.

Definition:

1. The information structure is *locally weakly embedded* (LWE) if for each state k, there exists a pair of players i, j, such that player i has more information than any other player, and player j has more information than any player other than i (if any).

2. The information structure has the *all-or-nothing property* if there exists a partition of K, $K = \cup_{i=1,\ldots,n} K_i$ with K_i possibly empty, such that for each i, $I_i(k) = \{k\}$ if $k \in K_i$, $I_i(k) = K \setminus K_i$ otherwise.

Note that in the above definition of LWE, the pair of players i, j (called the leaders) may depend on the state. In fact, this pair is the same for all states in the same majority component. Note that two-player games with only one informed player necessarily satisfy LWE. The next lemma further describes LWE information structures (proof omitted).

Lemma: If the information structure is LWE, then on each majority component A, the pair of leaders (i, j) is the same for all states in A, and the information $I_{A,l}$ is trivial for each other player l ($I_l(k) \cap A = A$, for $k \in A$, $l \neq i, j$).

Recall that attention is restricted, without loss, to a single majority component. One can show the following:

Theorem: If $V^*(\mathcal{I}, u) \neq \varnothing$, $\forall u \in \mathcal{S}_\mathcal{I}$, then the information structure is locally weakly embedded, or has the all-or-nothing property. Further, if the information structure is locally weakly embedded, then $V^*(\mathcal{I}, u) \neq \varnothing$, $\forall u \in \mathcal{S}_\mathcal{I}$.

The question remains open for information structures satisfying the all-or-nothing property. Numerical simulations suggest the following.

Conjecture: The set $V^*(\mathcal{I}, u)$ is non-empty for all $u \in \mathcal{S}_\mathcal{I}$ if and only if the information structure is locally weakly embedded, or has the all-or-nothing property.

B Belief-Free Equilibria, Imperfect Public Monitoring

Here we extend the analysis to imperfect public monitoring. Belief-free equilibria are here as well particularly useful, since they allow an extension of the scoring algorithm from imperfect information to incomplete information.

The notation is the same as before. However, monitoring being imperfect, we re-introduce public signals Y and distributions $\{\pi^k(\cdot \mid a)\}$. Note that the monitoring structure is indexed by the state k, allowing signals to reveal some information about the state even when players do not know it. Hence, $\pi^k(y \mid a)$ is the probability of observing the public signal y when the action profile is a and the state is k. As before, we denote by $\theta_i = I_i(k)$ the type of player i, that is, the information he holds at the start of the game about the state of the world.

As with public monitoring, we restrict attention to equilibria in which players' strategies are measurable with respect to the public history, given their private information at the start. That is, a strategy σ_i is ("type-contingent") public if $\sigma_i(\theta_i, h_i^t) = \sigma_i(\theta_i, \tilde{h}_i^t)$ for all private histories $h_i^t, \tilde{h}_i^t \in H_i^t = (A_i \times Y)^t$ for which the public signals observed are the same. A public history (of length t) is denoted $h^t \in H^t := Y^t$, as before. Write $\sigma|_{\theta,h^t} := (\sigma_i|_{\theta_i,h^t})_{i=1}^n$. We may then extend the notion of belief-free equilibrium as follows.

Definition: A public strategy profile σ is a BFE if for any k, t, and $h^t \in H^t$, $\sigma|_{\theta(k),h^t}$ is a Nash equilibrium of the game with state k.

The set of BFE payoffs given δ is still denoted B_δ. We can then extend the scoring algorithm as follows. Given $\lambda \in \mathbb{R}^{nK}$, with $\|\lambda\| = 1$, let

$$k(\lambda) = \sup \lambda \cdot v$$

over $v \in \mathbb{R}^{nK}$, $x : Y \to \mathbb{R}^{nK}$, $\alpha_i : \Theta_i \to \triangle A_i$, $i = 1, \ldots, n$, such that, for all $i = 1, \ldots, n$, $k = 1, \ldots, K$,

$$v_i^k = u_i(k, \alpha) + \sum_y \pi_i^k(y \mid \alpha^{\theta(k)}) x_i^k(y);$$

for all $a_i' \in A_i$,

$$v_i^k \geq u_i(k, a_i', \alpha_{-i}^{\theta_{-i}(k)}) + \sum_y \pi_i^k(y \mid a_i', \alpha_{-i}^{\theta_{-i}(k)}) x_i^k(y),$$

and for all $y \in Y$,

$$\lambda \cdot x(y) \leq 0.$$

As before, we let $\mathcal{H}(\lambda) := \{v \in \mathbb{R}^{nK} : \lambda \cdot v \leq k(\lambda)\}$, and define $\mathcal{H} := \cap_\lambda \mathcal{H}(\lambda)$. One can show:
Theorem: If int $\mathcal{H} \neq \varnothing$, then

$$\lim_{\delta \to 1} B_\delta = \mathcal{H}.$$

No description of the left-hand side limit (assuming it exists) is known when the interiority assumption fails.

C Known-Own Payoffs

The results so far were frustratingly restrictive on two accounts: First, belief-free equilibria may fail to exist. Second, when they do, they typically fail to exhaust the set of all equilibrium payoffs.

In this section, the second shortcoming is addressed in the following sense. When belief-free equilibria exist, what is the set of all Bayes Nash equilibrium payoffs? Surprisingly, it turns out that existence of BFE greatly facilitates the characterization of all Nash equilibrium payoffs, whether belief-free or not.

Throughout, monitoring is assumed to be perfect. Known-own payoffs is assumed throughout, so that we write $u_i(\theta_i, a)$ rather than $u_i(k, a)$. Assume $|\Theta_i| \leq |A_i|$. We also write $\Theta := \cup_{i=1,\ldots,n} \Theta_i$.

Because we are no longer interested in payoffs conditional on the state of the world, but rather expected payoffs, payoffs are now identified with vectors $v = (v_i(\theta_i))_{i,\theta_i} \in \mathbb{R}^\Theta$, with the interpretation that $v_i(\theta_i)$ is the expected payoff of type θ_i of player i. The vector $v_i \in \mathbb{R}^{\Theta_i}$ are the coordinates of v that specify player i's expected payoffs according to his type.

Each player starts the game with some prior belief $\mu_i^{\theta_i}(\cdot) \in \triangle \Theta_{-i}$ about the types of the other players. Assume that this belief is drawn from some common prior $\mu \in \triangle(\times_i \Theta_i)$. His expected

average payoff is then, given strategies $\sigma_i : \Theta \times \cup_t H_t \to \triangle A_i$,

$$v_i^{\theta_i,\mu,\delta}(\sigma) = \sum_{\theta_{-i}\in\Theta_{-i}} \mu_i^{\theta_i}(\theta_{-i})\mathbb{E}_{\sigma(\theta_i,\theta_{-i})}\left[(1-\delta)\sum_{t\geq 0}\delta^t u_i(\theta_i, a_t)\right] \in \mathbb{R},$$

and write $v^{\mu,\delta} \in \mathbb{R}^\Theta$ for the corresponding payoff vector $(v_i^{\theta_i,\mu,\delta}(\sigma))_{i,\theta_i}$. Let $E^\delta(\mu)$ denote the (Bayes) Nash equilibrium payoff set.

Feasible and individually rational payoffs: We first need to define feasible payoffs. Given $a \in A$, let $u(a) = (u_i(\theta_i, a))_{i,\theta_i}$ be the payoff vector obtained when each player i plays a_i independently of his type. The corresponding set

$$V^{NR} := \text{co}\,\{u(a) : a \in A\} \subseteq \mathbb{R}^\Theta$$

is the set of *feasible non-revealing payoffs*. It is a subset of the set of all feasible payoffs when actions vary with types.

Individual rationality is defined as before (although with known-own payoffs we can ignore the types of other players): for each player i, define

$$\varphi_i(q) := \min_{\alpha_{-i}\in\prod_{j\neq i}\triangle A_j} \max_{a_i\in A_i} \sum_{\theta_i\in\Theta_{-i}} q(\theta_i)u_i(\theta_i, \alpha_{-i}, a_i).$$

over $q \in \triangle\Theta_i$, and let

$$V^{IR} = \{v \in \mathbb{R}^\Theta : \forall q \in \triangle\Theta_i,\ \sum_{\theta_i\in\Theta_{-i}} q(\theta_i)v_i^{\theta_i} \geq \varphi_i(q)\}.$$

The candidate equilibrium payoff set: Given any $\mu \in \triangle(\times_i\Theta_i)$, and any set $D \subseteq \mathbb{R}^\Theta$, let

$$D^{\mu+} := \{v' \in \mathbb{R}^\Theta : \exists v \in D, \forall(i,\theta_i), v_i'(\theta_i) \geq v_i(\theta_i) \text{ with equality if } \mu(\theta_i \times \Theta_{-i}) > 0\}.$$

In words, the payoffs in $D^{\mu+}$ (the *enhancement* of D) are the same as in D, except that types assigned probability zero under μ can get more. Given a correspondence $F : \triangle(\times_i\Theta_i) \rightrightarrows \mathbb{R}^\Theta$, define $F^+ : \triangle(\times_i\Theta_i) \rightrightarrows \mathbb{R}^\Theta$ by $F^+(\mu) = (F(\mu))^{\mu+}$. In a sense, defining a set of payoffs or its enhancement is equivalent, because payoffs of zero-probability types are irrelevant. But it is often more convenient to work with the enhancement.

We define two operators on correspondences: First, *Averaging*: Given $F : \triangle(\times_i \Theta_i) \rightrightarrows \mathbb{R}^\Theta$, define $\mathcal{A}F : \triangle(\times_i \Theta_i) \rightrightarrows \mathbb{R}^\Theta$ by

$$\mathcal{A}F(\mu) = \text{int}\left(V^{IR} \cap \text{co}\left(F(\mu) \cup V^{NR}\right)\right).$$

Elements of $\mathcal{A}F(\mu)$ are individually rational payoff vectors u that can be written as averages of a payoff u' in $F(\mu)$ and a vector v in the non-revealing payoff set. Intuitively, we can think of playing for t periods the non-revealing action profile yielding v as a reward before playing according to the strategy that gives u': the payoff obtained is then an average with weight $\beta := \delta^t$ on u'.

Second, *Belief Splitting*: Given $\mu \in \triangle(\times_i \Theta_i)$, let $L(\mu) = \{(\alpha, u) : \alpha_i : \Theta_i \to \triangle A_i, u : A \to \mathbb{R}^\Theta\}$ such that, for all $a_i \in A_i$ such that $\alpha_i(a_i \mid \theta_i') > 0$ for some θ_i' (we say that *a has pos. prob.*):

1. there exists θ_i with $\mu(\theta_i \times \Theta_{-i}) > 0$ such that $\alpha_i(a_i \mid \theta_i) > 0$: that is, actions a which have positive probability under α are actually taken by some type who is assigned positive probability under μ, so that Bayes rule can be used to compute the posterior belief given a, and $l := (\alpha, u)$,
$$p^{\mu,l}(a) = \left(p_i^{\mu,l,\theta_i}(a)\right)_{i,\theta_i}.$$

2. it holds that, for all θ_i,
$$\mathbb{E}_{\mu_i^{\theta^i}}\left[u_i(\theta_i, a_i, \alpha_{-i}(\theta_{-i}))\right] = \mathbb{E}_{\mu_i^{\theta^i}}\left[u_i(\theta_i, \alpha(\theta))\right],$$
that is, type θ_i is willing to play any action that is in the range of α_i if others play according to α_{-i}, and the reward is given by the function u. This is stronger than the usual incentive compatibility condition, as a_i need not be assigned positive probability by $\alpha_i(\theta_i)$. This will not matter, as continuation payoffs after actions that are not supposed to be taken by some types (so that they are assigned zero probability after such an action is observed) can be "enhanced" to make them just barely indifferent between taking them or not. (This is where the enhancement operation is very useful.)

   Write $v^{\mu,l} = (v_i^{\mu,l}(\theta_i))_{i,\theta_i}$ for this payoff vector.

Given $F : \triangle(\times_i \Theta_i) \rightrightarrows \mathbb{R}^\Theta$, define $\mathcal{B}F : \triangle(\times_i \Theta_i) \rightrightarrows \mathbb{R}^\Theta$ by

$$\mathcal{B}F(\mu) = \left\{v^{\mu,l} \in \mathbb{R}^\Theta : l \in L(\mu) \text{ with } u(a) \in F(p^{\mu,l}(a)) \text{ for each pos. prob. } a\right\}.$$

We can now define a sequence of correspondences. First, let $F_0 : \triangle(\times_i \Theta_i) \rightrightarrows \mathbb{R}^\Theta$ by

$$F_0(\mu) = \text{int}\left(V^{IR} \cap (V^{NR})^{\mu+}\right).$$

Elements of $F_0(\mu)$ are the (strictly) individually rational payoffs (or rather the enhancement of these payoffs) that obtain when players do not condition play on their types. This set might be empty, although it is not if BFE exists.

Correspondences can be (partially) ordered: $F \subseteq F'$ if $F(\mu) \subseteq F'(\mu)$ for all $\mu \in \triangle(\times_i \Theta_i)$.

Lemma: There exists a smallest correspondence $F : \triangle(\times_i \Theta_i) \rightrightarrows \mathbb{R}^\Theta$ with $F_0 \subseteq F^*$, and $\mathcal{A}F^* \subseteq F^*$, $\mathcal{B}F^* \subseteq F^*$. In addition, $F^* = \cup_n F_n$, where for all $n \geq 1$, $F_n = \mathcal{B}\mathcal{A}F_{n-1}$.

It is relatively easy to construct explicitly strategies (as in Fudenberg and Maskin, 1986) to prove the following.

Lemma: It holds that, for all $\mu \in \triangle(\times_i \Theta_i)$,

$$F^*(\mu) \subseteq \liminf_{\delta \to 1} E^\delta(\mu).$$

What is remarkable is that the converse holds when BFE exist (up to closure). More precisely, define an *open thread* as a map $u^* : \times_i \Theta_i \to \mathbb{R}^\Theta$ such that

1. For all $\theta \in \times_i \Theta_i$, and all $\mu^\theta \in \triangle(\times_i \Theta_i)$

$$u^*[\theta] \in F_0(\mu^\theta),$$

2. For all i, $\theta_i, \theta_i' \in \Theta_i$, $\theta_{-i} \in \Theta_{-i}$,

$$u_i^*[\theta](\theta_i) = u_i^*[\theta_i', \theta_{-i}](\theta_i).$$

That is, the (i, θ_i)-th coordinate of $u_i^*[\theta]$ is only a function of θ_{-i}. To put it differently, if we interpret the domain of u_i^* as messages, and the range as payoffs that obtain in subsequent play, this condition states that type θ_i's payoff is independent of the message that he sends, conditional on the other players' messages.

Theorem: If an open thread exists, then, for all $\mu \in \triangle(\times_i \Theta_i)$,

$$\text{cl } (\limsup_{\delta \to 1} E^\delta)^+(\mu) = \text{cl } F^*(\mu)$$

The trick of the proof is to perturb an arbitrary equilibrium to drive the continuation payoff vector closer and closer to the open thread. By definition of the open thread, it holds that for some small $\varepsilon > 0$,

$$B(u^*(\theta), \varepsilon) \subset F_0(\mu^\theta)$$

for all $\theta \in \times_i \Theta_i$ (here, $B(x, \varepsilon)$ is the ball around x of radius x). Once the payoff reaches this set, a continuation play can be found –namely, a belief-free equilibrium. Furthermore, because this is achieved in finitely many steps, the payoff we have started with must be in the iterates of the operators $\mathcal{A}, \mathcal{B}$, starting from F_0.

Driving the perturbed payoff vector closer and closer to the open thread requires also perturbing the discount factor, which is fine since the objective is to characterize the limiting set. To get some intuition, suppose that we start with a payoff vector

$$v = (1-\delta)u(a) + \delta w,$$

where a is the (non-revealing) action profile that must be played in the first period. Suppose we perturb the payoff v and consider instead the weighted average $v' := \gamma v + (1-\gamma)u^*(\mu)$ for some $\gamma \in (0,1)$ (close to 1). If we insist on the play of a in the initial period, can we find $\delta' \in (\delta, 1)$ and $\gamma' \in [0,1]$ so that using as continuation payoff $w' := \gamma' w + (1-\gamma')u^*(\mu)$, we indeed obtain v' as the payoff (and the correct incentives to play a)? We must have

$$\begin{aligned} v' &= \gamma v + (1-\gamma)u^*(\mu) \\ &= \gamma((1-\delta)u(a) + \delta w) + (1-\gamma)u^*(\mu) \\ &= (1-\delta')u(a) + \delta'(\gamma' w + (1-\gamma')u^*(\mu)), \end{aligned}$$

which is indeed solvable, with $1-\delta' = (1-\delta)\gamma < 0$ (so: we make the players more patient), and $\frac{1-\gamma}{1-\gamma'} = 1 - \gamma(1-\delta) < 1$, so that $\gamma' < \gamma$: the weight on the open thread is higher for the continuation payoff than for the payoff itself; continuing recursively, the distance to the open thread can be reduced (provided players are made more patient).

There is still some work: the argument above is for periods in which play is non-revealing. It

must be also applied to the periods in which beliefs potentially change. The argument is similar and omitted, but the key here is that, because the set of BFE is independent of beliefs, the expected payoffs that this set span is a linear subspace (relative to feasible individually rational payoffs), so that, even when beliefs change, the distance to the closest point in that set when payoffs are perturbed still shrinks.

D Unknown Payoffs

Interesting, if only partial, results have been obtained for the case in which the game G_k is not known by any player. There are n players, K states, over which players share a common prior $p \in \triangle K$, and an imperfect signal of the state: each period, a signal $y \in Y$, a subset of a Euclidean space, is drawn according to $\pi_k(dy|a)$. It is assumed that, for each $k, k' \in K$, there exists a such that $\pi_k(\cdot|a) \neq \pi_{k'}(\cdot|a)$, so if players cooperate, they can eventually learn the state.

Monitoring of actions is perfect, so (public) histories are elements $h^t \in H^t := (A \times Y)^t$. Given such a history, players share a belief $p^t = p(h^t) \in \triangle K$ about the state. Let V_k denote the convex hull of the set of feasible payoffs in state k. Write $\underline{v}_{i,k}$, $\underline{\alpha}_{i,k}$ for player i's minmax payoff and action profile achieving this minmax payoff, and define $\underline{V}_k = \{v \in V_k : v_i > \underline{v}_i, \forall i\}$ as the set of feasible and (strictly) individual rational payoffs in state k. This set is assumed to have non-empty interior for all k. Wiseman proves:

Theorem 4 *For any $\varepsilon > 0$, $v_k \in int\ \underline{V}_k$, $k = 1, \ldots, K$, and any prior $p \in int\ \Delta K$, there exists $\underline{\delta} < 1$, for all $\delta \in (\underline{\delta}, 1)$, there is a (sequential) equilibrium such that, when the realized state is k, the expected payoff is within ε of v_k.*

The proof is based on an elegant extension of APS and FLM, using the common belief of the players as an additional state variable.

This sounds very much like a folk theorem, but it is correctly called a partial folk theorem. To understand this, consider the two-player, two-state, two-action game represented in Figure 7. Assume that actions are perfectly observed and that the prior on G_A is 1/2: both states are equally likely.

Note that, in G_A, player 1 can secure 3, so that the only equilibrium payoff vector is $(3, -3)$. Similarly, the only equilibrium payoff vector in G_B is $(-3, 3)$. As a result, given the prior, the unique equilibrium expected payoff vector, if learning occurs "fast," is $(0, 0)$. But players would be better off playing (L, L), not learning the state, and getting a higher payoff. If either player deviates, learning instantly occurs, and play reverts to the unique equilibrium action profile in

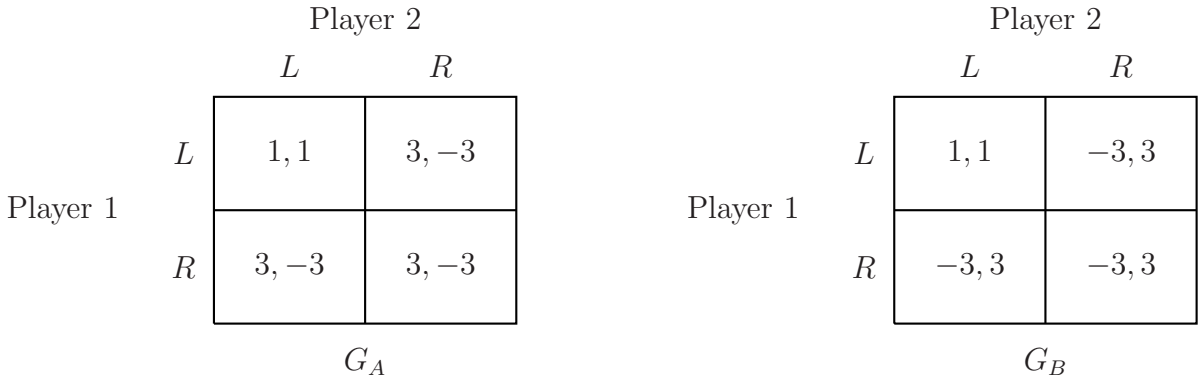


Figure 7: Rewards in the two states

the corresponding game. Note that players have no incentive to deviate, provided that they are patient enough, as this would yield an expected payoff of approximately 0. Learning is not always beneficial, as the minmax payoff in a given state might restrict the set of individual rational payoffs in that state sufficiently, so as to yield as unique candidates for the theorem payoff vectors whose expected value is Pareto-dominated by other feasible payoff vectors.

As a result, we have no characterization yet of the set of equilibrium payoff vectors for this game. On the other hand, the partial folk theorem has been extended to richer environments by now (in particular, to games with private monitoring).

IV Literature

The literature on repeated games with incomplete information was really started by Aumann and Maschler, who developed the examples in Section I, as well as Theorem 1 from Section II. See Aumann, Robert J. and Michael B. Maschler (1995), *Repeated Games with Incomplete Information*, MIT Press, Cambridge, Massachusetts (with the collaboration of Richard E. Stearns). The extension to two-sided incomplete information without discounting is due to Hart, Sergiu (1985), "Nonzero-sum two-person repeated games with incomplete information," *Mathematics of Operations Research*, 10, 117–153. Discounting poses no conceptual difficulty, but the value function (as a function of the belief) has been known to be badly behaved (e.g., to be non-differentiable) early on, see Mayberry, J.P. (1967), "Discounted Repeated Games with Incomplete Information," in *Reports of the U.S. Arms Control and Disarmament Agency ST-116*, Washington, D.C., Chapter V, 435–461.

In the non-zero sum discounted case, there is a nice paper that was not discussed here, by

Cripps, Martin W. and Jonathan P. Thomas (2003), "Some asymptotic results in discounted repeated games of one-sided incomplete information," *Mathematics of Operations Research*, 28, 433–462. This paper looks at the limit correspondence of payoffs when the probability of one of the payoff types is close to 1, and establishes some folk theorem in this case.

The belief-free approach has been introduced to repeated games with incomplete information by Hörner, J. and S. Lovo, "Belief-Free Equilibria in Games With Incomplete Information," *Econometrica*, 77, 453–487. This paper deals only with two players and perfect monitoring. The extension to n players is in Hörner, J., S. Lovo and T. Tomala, "Belief-free equilibria in games with incomplete information: Characterization and existence," *Journal of Economic Theory*, 146, 1770–1795, while the extension to public monitoring is due to Fudenberg, D. and Y. Yamamoto, "Repeated games where the payoffs and monitoring structure are unknown," *Econometrica*, 78,1673–1710; See also Fudenberg, D. and Y. Yamamoto, "Learning from private information in noisy repeated games," *Journal of Economic Theory*, 146, 1733–1769.

The analysis in Section C is based on Peski, M., 2014, "Repeated Games with Incomplete Information and Discounting," *Theoretical Economics*, which extends his earlier work: Peski, M. 2008, "Repeated games with incomplete information on one side," *Theoretical Economics*, 3, 29–84. See Forges, F. and A. Salomon, "Bayesian Repeated Games and Reputations," (2014, working paper) for a construction that does not rely on the belief-free assumption in a class of examples.

The Section D is based on Wiseman, T., 2005, "A Partial Folk Theorem for Games with Unknown Payoff Distributions," *Econometrica*, 73, 629–645. Wiseman's result has been generalized by a more recent paper of his ("A Partial Folk Theorem for Games with Private Learning," *Theoretical Economics*, forthcoming), but that paper does not provide a more complete folk theorem than the previous one; rather it provides a partial folk theorem along the same lines for more general (in particular, to some extent private) monitoring structures.