האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

THE VALUE OF TWO-PERSON ZERO-SUM REPEATED GAMES WITH INCOMPLETE INFORMATION AND UNCERTAIN DURATION

by

ABRAHAM NEYMAN

Discussion Paper # 512   May 2009

מרכז לחקר הרציונליות

**CENTER FOR THE STUDY
OF RATIONALITY**

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135   FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

The Value of Two-Person Zero-Sum Repeated Games with Incomplete Information and Uncertain Duration

Abraham Neyman*

May 27, 2009

Abstract

It is known that the value of a zero-sum infinitely repeated game with incomplete information on both sides need not exist [1]. It is proved that any number between the minmax and the maxmin of the zero-sum infinitely repeated game with incomplete information on both sides is the value of the long finitely repeated game where players' information about the uncertain number of repetitions is asymmetric.

1 Introduction

Two-player repeated games with incomplete information (RGII), introduced by Aumann and Maschler [1][1], model long-term interactions in which players have asymmetric information about the actual one-shot game that is repeatedly played. Modeling the long-term interactions was focused initially on the

*Institute of Mathematics, and Center for the Study of Rationality, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel. *e-mail: aneyman@math.huji.ac.il*
This research was supported in part by Israel Science Foundation grants 382/98, 263/03, and 1123/06.

[1]This book is based on reports by Robert J. Aumann and Michael Maschler which appeared in the sixties in *Report of the U.S. Arms Control and Disarmament Agency.* See "Game theoretic aspects of gradual disarmament" (1966, ST–80, Chapter V, pp. V1–V55), "Repeated games with incomplete information: a survey of recent results" (1967, ST–116, Chapter III, pp. 287–403), and "Repeated games with incomplete information: the zero-sum extensive case" (1968, ST–143, Chapter III, pp. 37–116).

infinitely repeated game Γ_∞ and the finitely repeated game Γ_n. Studying the repeated game Γ_n assumes that the number of repetitions n is known to both players, and moreover that n is common knowledge. These assumptions are difficult to justify in many applications of long-term interactions. [10] studies two-player repeated games where the players have symmetric information about the uncertain number of repetitions. The present paper studies the model of the zero-sum RGII where the players have asymmetric information about the number of repetitions.

In the zero-sum infinitely repeated game Γ_∞, Player 1 (P1) can guarantee v if for every $\varepsilon > 0$ he has a strategy σ such that for any sufficiently large number of repetitions n, for each strategy τ of Player 2 (P2) the expected average per-stage payoff is at least $v - \varepsilon$. Similarly, P2 can guarantee v in Γ_∞ if for every $\varepsilon > 0$ he has a strategy τ such that for any sufficiently large number of repetitions n, for each strategy σ of P1 the expected average per-stage payoff is at most $v + \varepsilon$. The game Γ_∞ has a uniform value v if each player can guarantee v. The definition of the uniform value implies that whenever the uniform value exists, e.g., RGII on one side (RGIIOS) [1] or stochastic games [4], the limit of the values of the finitely repeated games (where payoffs are the average per-stage payoffs) converge to the uniform value as the number (or the expected number in the model with uncertain duration) of repetitions goes to infinity.

The uniform value need not exist in RGII on both sides (RGIIBS) [1, Section 4.3]. Nonetheless, v_n, the value of the n-stage RGIIBS (with state-independent signaling) converges to a limit as $n \to \infty$ [6, 3], and more generally, v_θ, the value of the finitely RGIIBS (with state-independent signaling) with a random number of repetitions θ and where the players have symmetric information about θ, converges to a limit as the expectation of the number θ of repetitions goes to infinity [10]. The present paper characterizes the limit points of v_θ as $E(\theta) \to \infty$ and players' information about the number of repetitions θ is asymmetric.

In RGII, one of finitely many one-shot games is repeatedly played and each player has only partial information about the one-shot game that is being repeated. The RGII (denoted Γ) is described as follows. There is a finite set of normal form games G^m, $m \in M$, with finite action sets I for P1 and J for P2. The state $m \in M$ is chosen at random according to a publicly known probability p, and each player receives partial information about m. The partial information of the players is defined by two functions, $c : M \to C$ and $d : M \to D$; P1 observes $c = c(m)$ and P2 observes $d = d(m)$. In addition, after each stage the players obtain some further information about the previous choice of moves.[2] This is represented by a map Q from $I \times J$ to probabilities on $A \times B$. At stage t, given the state m and the moves (i_t, j_t), a pair (a_t, b_t) is chosen at random according to the distribution $Q(i_t, j_t)$.[3] A play of the game is thus a sequence $m, i_1, j_1, a_1, b_1, \ldots, i_t, j_t, a_t, b_t, \ldots$, while the information to P1 before his play at stage t is $c(m), i_1, a_1, \ldots, i_{t-1}, a_{t-1}$, and the information to P2 before his play at stage t is $d(m), i_1, b_1, \ldots, j_{t-1}, b_{t-1}$. The repeated game is thus represented by the tuple $\Gamma = \langle M, p, M^1, M^2, I, J, G, Q, A, B \rangle$, where M^1 is the partition of M defined by the values of c and M^2 is the partition of M defined by the values of d.

The payoff at stage t of the repeated game, $g_t := G^m_{i_t, j_t}$, depends on the chosen state m and the action pair (i_t, j_t) at stage t. A pair of strategies σ of P1 and τ of P2 in the repeated game Γ defines a probability distribution $P_{\sigma,\tau}$ on the space of plays, and thus a probability distribution on the stream of payoffs $g_1, g_2, \ldots$. The value of the n-stage zero-sum game, v_n, where P1 maximizes the (expectation of the) average $\bar{g}_n := (g_1 + \ldots + g_n)/n$ of the payoffs in the first n stages, exists and equals $\max_\sigma \min_\tau E_{\sigma,\tau} \bar{g}_n$ (where the max is over all strategies σ of P1 and the min is over all strategies τ of P2, and $E_{\sigma,\tau}$ stands for the expectation w.r.t. the probability $P_{\sigma,\tau}$), which by the minmax theorem is equal to $\min_\tau \max_\sigma E_{\sigma,\tau} \bar{g}_n$.

Special subclasses of RGII are defined by the signaling structure and the initial information about the state. The classical case of standard signaling corresponds to $A = J$, $B = I$, and to $Q(i, j)$ being the Dirac measure on (j, i), or equivalently, to $A = B = I \times J$ and to $Q(i, j)$ being the Dirac measure on $((i, j), (i, j))$. RGIIOS corresponds to the case where $c(m) = m$ and $d(m)$ is a constant, or equivalently, only P1 receives a signal about m. Deterministic signaling corresponds to $Q(i, j)$ (respectively, $Q(m, i, j)$ in the state-dependent signaling) being a Dirac measure; in this case we can think of the signal to a player as a deterministic function of (i, j) (respectively, (m, i, j)).

The independent case corresponds to an initial probability p such that the probability defined on $C \times D$ by $p(c, d) = p(\{m : c(m) = c \text{ and } d(m) = d\})$ is a product probability. In this case we may assume without loss of

[2] This is called state-independent signaling. In more general state-dependent signaling, the players obtain further information about the previous choice of moves and the state.

[3] In state-dependent signaling, Q is a map from $I \times J \times M$ to probabilities on $A \times B$, and at stage t, given the state m and the moves (i_t, j_t), the pair (a_t, b_t) is chosen at random according to the distribution $Q(m, i_t, j_t)$.

generality that $M = C \times D$ and that the initial probability distribution is a product probability $p \otimes q$ where p is a probability on C and q is a probability distribution on D. [1, Section 4.2] shows that each game with incomplete information in the dependent case is equivalent to a game with incomplete information in the independent case. Therefore, it is sufficient for our main result to handle the independent case, where the statement and the proof of the main result are simplified.

In this paper we study the asymptotic behavior of the value of zero-sum repeated games with an uncertain number of repetitions θ. θ is an integer-valued random variable on a probability space $(\Omega, \mathcal{B}, \mu)$ with finite expectation and each player observing partial information about θ. The normalized value is denoted v_θ. We prove that any value between the max min (the maximal payoff that P1 can guarantee) and the min max (the minimal payoff that P2 can guarantee) of Γ_∞ can be obtained as the value v_θ for an asymmetric uncertainty about the number of repetitions θ with arbitrarily large expected duration $E(\theta)$. As any limit point of v_θ as $E(\theta) \to \infty$ is in the interval $[\max\min \Gamma_\infty, \min\max \Gamma_\infty]$, the result characterizes the set of limit points of v_θ as $E(\theta) \to \infty$.

2 The Game Model

RGIIBS is defined in the standard signaling and the independent case by the tuple $\langle C, D, p, q, I, J, G \rangle$, where C, D, I, J are finite sets, p and q are probability distributions on C and D respectively, and G is a list of $I \times J$ two-person zero-sum games $G^{c,d}$, $c \in C$ and $d \in D$. The repeated game proceeds in stages. In stage 0, nature chooses a pair (c, d) with probability $p(c)q(d)$. P1 is informed of c and P2 is informed of d. At stage $t \geq 1$, P1 is first informed of j_{t-1} and then chooses $i_t \in I$, and simultaneously P2 is first informed of i_{t-1} and then chooses $j_t \in J$. The payoff (from P2 to P1) in stage t is $g_t = G^{c,d}_{i_t,j_t}$.

The repeated game is denoted Γ for short, or $\Gamma(p, q)$ to emphasize the dependence on the probability distributions p and q and the fixing of the other parameters C, D, I, J, G that define the repeated game.

A behavioral strategy of P1 in Γ is a map $\sigma : C \times (I \times J)^* \to \Delta(I)$, where $(I \times J)^*$ stands for all finite strings of $I \times J$ elements, namely, $(I \times J)^* = \cup_{t \geq 0} (I \times J)^t$, and $\Delta(X)$ stands for all probability distributions on X, and a behavioral strategy of P2 is a map $\tau : D \times (I \times J)^* \to \Delta(J)$.

A pair of behavioral strategies, σ of P1 and τ of P2, defines a probability distribution $P_{\sigma,\tau}$ on the space of plays $(c, d, i_1, j_1, i_2, j_2, \dots)$ by $P_{\sigma,\tau}(c, d) = p(c)q(d)$, $P_{\sigma,\tau}(c, d, i_1, j_1) = p(c)q(d)\sigma(c)[i_1]\tau(d)[j_1]$, and by induction on t

$$P_{\sigma,\tau}(c, d, h_t, i_t, j_t) = P_{\sigma,\tau}(c, d, h_t)\,\sigma(c, h_t)[i_1]\ \tau(d, h_t)[j_1]$$

for $h_t = (i_1, j_1, \dots, i_{t-1}, j_{t-1}) \in (I \times J)^{t-1}$.

The uncertainty of the number of repetitions θ is modeled as follows. The number of repetitions θ is an integer-valued random variable θ defined on a probability space $(\Omega, \mathcal{B}, \mu)$ and with finite expectation. Before the start of the repeated game the players receive partial information about the value of θ; P1 observes $s^1(\omega) \in S^1$ and P2 observes $s^2(\omega) \in S^2$, where S^1 and S^2 are finite sets. The interpretation is that at stage 0, nature chooses $\omega \in \Omega$ according to the probability μ, and independently of the choices of nature in the repeated game Γ, the number of repetitions is $\theta(\omega)$, and P1 and P2 are informed of $s^1(\omega)$ and $s^2(\omega)$ respectively. The joint distribution of (θ, s^1, s^2) is assumed to be independent[4] of the state (c, d).

The repeated game with uncertain duration Γ_θ is the repeated game Γ, where the choice of P1's (respectively, P2's) action at stage t, i_t (respectively, j_t), may depend in addition on $s^1(\omega)$ (respectively, $s^2(\omega)$). Therefore a strategy σ of P1 in Γ_θ is in fact a list of strategies σ^s ($s \in S^1$) in Γ, and a strategy τ of P2 in Γ_θ is in fact a list of strategies τ^s ($s \in S^2$) in Γ.

The un-normalized payoff in Γ_θ is $\sum_{t=1}^{\theta} g_t$ ($:= \sum_{t\geq 1} g_t I(t \leq \theta)$ where I stands for the indicator function) and the normalized one is $\frac{1}{E(\theta)} \sum_{t=1}^{\theta} g_t$. The value of Γ_θ (with the normalized payoff) exists, is denoted v_θ, and equals $\max_\sigma \min_\tau E_{\sigma,\tau,\mu} \frac{1}{E(\theta)} \sum_{t=1}^{\theta} g_t$ (where the max is over all strategies σ of P1 in Γ_θ, the min is over all strategies τ of P1 in Γ, and $E_{\sigma,\tau,\mu}$ stands for the expectation with respect to the probability $P_{\sigma,\tau,\mu}$ induced on the joint probability of the number of repetitions θ and the play by σ, τ, μ). We are interested in the asymptotic behavior of v_θ as the expected duration $E(\theta)$ goes to ∞.

Given $p \in \Delta(C)$ and $q \in \Delta(D)$ we denote by $G^{p,q}$ the $I \times J$ stage-payoff matrix $\sum_{c,d} p(c)q(d)G^{c,d}$ and by $u(p, q)$ its minmax value. For $x \in \Delta(I)$, $y \in \Delta(J)$, and an $I \times J$ matrix G we denote by xGy the sum $\sum_i \sum_j x(i)G_{i,j}y(j)$. This is the classical notation that corresponds to matrix multiplication, viewing x as an I row vector and y as a J column vector.

[4]The more general model, where the duration depends on the state, is obviously of interest. However, restrictive assumptions on the uncertain duration make our main result stronger.

Given a compact convex set Y and a bounded function $u : Y \to \mathbb{R}$ we denote by $\text{cav}_y u$ the smallest concave function from Y to $\mathbb{R}$ that is $\geq u$ and by $\text{vex}_y u$ the largest convex function from Y to $\mathbb{R}$ that is $\leq u$. If $u : \Delta(C) \times \Delta(D) \to \mathbb{R}$ we denote by $\text{cav}_p u$ the smallest function on $\Delta(C) \times \Delta(D)$ that is concave in p and is not smaller than u at each point (p, q). Similarly, $\text{vex}_q\, u$ is the largest function on $\Delta(C) \times \Delta(D)$ that is convex in q and is not larger than u at each point (p, q). Note that cav_p and vex_q are operators on bounded functions on $\Delta(C) \times \Delta(D)$, and thus can be iterated. The value of the function $\text{vex}_q \text{cav}_p u$, respectively $\text{cav}_p \text{vex}_q u$, at the point (p, q) is denoted $\text{vex}_q \text{cav}_p u\ (p, q)$, respectively $\text{cav}_p \text{vex}_q u\ (p, q)$.

P2 can guarantee v in $\Gamma_\infty(p, q)$ if for every $\varepsilon > 0$ there is a strategy τ of P2 and a positive integer N such that for every $n \geq N$ and every strategy σ of P1 we have

$$E_{\sigma,\tau} \frac{1}{n} \sum_{t=1}^{n} g_t \leq v + \varepsilon$$

Similarly, P1 can guarantee v in $\Gamma_\infty(p, q)$ if for every $\varepsilon > 0$ there is a strategy σ of P1 and a positive integer N such that for every $n \geq N$ and every strategy τ of P2 we have

$$E_{\sigma,\tau} \frac{1}{n} \sum_{t=1}^{n} g_t \geq v - \varepsilon$$

It follows that if P2, respectively P1, can guarantee v in $\Gamma_\infty(p, q)$, then for every $\varepsilon > 0$ there is N such that for every uncertain duration with $E(\theta) > N$ we have $v_\theta \leq v + \varepsilon$, respectively $v_\theta \geq v - \varepsilon$. If each player can guarantee v in $\Gamma_\infty(p, q)$, then v is called the uniform value, or for short a value, of $\Gamma_\infty(p, q)$, and is denoted $v_\infty(p, q)$.

Aumann and Maschler [1], respectively Stearns [1, Theorem 4.11], proved that P1 can guarantee, respectively cannot guarantee more than, $\text{cav}_p \text{vex}_q u\ (p, q)$ and that P2 can guarantee, respectively cannot guarantee more than, $\text{vex}_q \text{cav}_p\ u(p, q)$, and therefore $\Gamma_\infty(p, q)$ has a uniform value iff

$$\text{cav}_p \text{vex}_q u\ (p, q) = \text{vex}_q \text{cav}_p u\ (p, q)$$

There are games for which

$$\text{cav}_p \text{vex}_q u\ (p, q) > \text{vex}_q \text{cav}_p u\ (p, q);$$

see [1].

3 Preliminary results

3.1 The Posteriors and Conditional Payoffs

In this section we review a few classical tools used in the analysis of RGII. The space of plays of a RGIIBS with standard signaling is the space of sequences $(c, d, i_1, j_1, i_2, j_2, \ldots)$ with the minimal σ-algebra for which all functions $(c, d, i_1, j_1, i_2, j_2, \ldots) \mapsto (c, d, i_1, j_1, i_2, j_2, \ldots, i_t, j_t)$, $t \geq 0$, are measurable. $\mathcal{H}_t$ denotes the minimal σ-algebra (in fact, an algebra) for which the function $(c, d, i_1, j_1, i_2, j_2, \ldots) \mapsto h_t := (i_1, j_1, i_2, j_2, \ldots, i_{t-1}, j_{t-1})$ is measurable.

In the following notations and observations we assume the independent case. Let τ be a behavioral strategy of P2 in Γ. We define the functions q_t, $t \geq 1$, from plays to $\Delta(D)$ (called posteriors) by induction on t as follows. $q_1 = q$, and

$$q_{t+1}(d) = \frac{q_t(d)\tau(d, h_t)[j_t]}{\sum_d q_t(d)\tau(d, h_t)[j_t]} \tag{1}$$

Note that q_t is $\mathcal{H}_t$-measurable.

Lemma 1 *For every strategy σ of P1 and every c, h_t such that $P_{\sigma,\tau}(c, h_t) > 0$, the conditional probability*

$$P_{\sigma,\tau}(d \mid c, h_t) = q_t(h_t)[d] \tag{2}$$

and thus (= $P_{\sigma,\tau}(d \mid h_t)$ and) is independent of the strategy σ of P1.

The next lemma is a classical tool in the study of games with incomplete information. It is presented here for completeness. Note that if P is the joint distribution of $(d, j) \in D \times J$, then $\sum_j P(j) \sum_d |P(d \mid j) - P(d)| = \sum_{d,j} |P(d, j) - P(d)P(j)| = \sum_d P(d) \sum_j |P(j \mid d) - P(j)|$. Therefore, if we set

$$y_t^d(h_t) = \tau(d, h_t), \quad y_t(h_t) = \sum_d q_t(d)\tau(d, h_t) \;\text{ and }\; \|y_t^d - y_t\| = \sum_j |y_t^d(j) - y_t(j)|$$

and apply the above equalities to the conditional distribution of (d, j_t) given $\mathcal{H}_t^1$ – the algebra spanned by (c, h_t) – we have

Lemma 2

$$E_{\sigma,\tau}(\|q_{t+1} - q_t\| \mid \mathcal{H}_t^1) = E_{\sigma,\tau}(\|q_{t+1} - q_t\| \mid \mathcal{H}_t) = \sum_d q_t(d)\|y_t^d - y_t\|$$

3.2 Information-theoretic tools

Given two probabilities P and Q on a finite set A, we have[5] $\|P - Q\|^2 \leq 2D(P\|Q)$, where $\|P - Q\| = \sum_a |P(a) - Q(a)|$ and $D(P\|Q) = \sum_a P(a) \log \frac{P(a)}{Q(a)}$ (where log denotes the natural logarithm and $0 \log 0 = 0$), e.g., [2, p. 300]. Let P be a probability distribution on the product of two sets A_1 and A_2, denote the marginal of P on A_i by P_i, and let (x, y) be a random variable having distribution P ($x \sim P_1$ and $y \sim P_2$). Then (a straightforward computation yields) $D(P\|P_1 \times P_2) = H(x) - H(x \mid y)$ and thus

$$\|P - P_1 \otimes P_2\| \leq \sqrt{2}\sqrt{H(x) - H(x \mid y)} \tag{3}$$

where H is the entropy function ($H(x) = -\sum_{a \in A_1} P_1(a) \log P_1(a)$, and $H(x \mid y) = H(x, y) - H(y) = -\sum_{(a,b) \in A_1 \times A_2} P(a, b) \log P(a, b) - H(y)$). As the square root is a concave function we have

Lemma 3 *Let P be a probability distribution on $A_1 \times A_2 \times A_3$ and let (x, y, z) be a random variable with distribution P. Then if P^z denotes the conditional distribution of (x, y) given z and P_i^z denotes its marginal on A_i, then*

$$E_P\|P^z - P_1^z \otimes P_2^z\| \leq \sqrt{2}\sqrt{H(x \mid z) - H(x \mid y, z)}$$

3.3 The Variation of Martingales of Probabilities

Lemma 4 *Let q_t, $t = 1, \ldots, K + 1$ be a martingale with values in $\Delta(D)$ where D is a finite set. Then*

$$E \sum_{t=1}^{K} \|q_{t+1} - q_t\| \leq \sqrt{K} \min(\sqrt{2 \log |D|}, \sqrt{|D| - 1}) \tag{4}$$

where $\|q_{t+1} - q_t\| = \sum_{d \in D} |q_{t+1}(d) - q_t(d)|$ and $|D|$ stands for the number of elements of D.

Proof. The bound $\sqrt{K}\sqrt{|D| - 1}$ is classical (see, e.g., [1]), and the bound $\sqrt{2k \log d}$ is proved in [9]. For completeness we here reproduce the proofs.

[5] This inequality is often called Pinsker's inequality. The constant 2 appearing in the equality is actually an improvement on the one obtained by Pinsker.

W.l.o.g. we assume that q_1 is a constant. For every $d \in D$, $q_t(d)$, $t = 1, \ldots, K+1$, is a real-valued martingale. By the Cauchy-Schwarz inequality we have

$$\sum_{t=1}^{K} |q_{t+1}(d) - q_t(d)| \leq \sqrt{K} \left(\sum_{t=1}^{K} (q_{t+1}(d) - q_t(d))^2 \right)^{1/2}$$

and as the square root is a concave function we have by Jensen's inequality

$$E \left(\sum_{t=1}^{K} (q_{t+1}(d) - q_t(d))^2 \right)^{1/2} \leq \left(E \sum_{t=1}^{K} (q_{t+1}(d) - q_t(d))^2 \right)^{1/2}$$

As q_t is a martingale, $E(q_{t+1}(d) - q_t(d))^2 = E(q_{t+1}(d))^2 - E(q_t(d))^2$, thus

$$E \sum_{t=1}^{K} (q_{t+1}(d) - q_t(d))^2 \leq E(q_{K+1}(d))^2 - (q_1(d))^2 \leq q_1(d) - (q_1(d))^2$$

where the last inequality uses the inequality $x \geq x^2$ for $0 \leq x \leq 1$ and the martingale property that implies $Eq_{K+1}(d) = q_1(d)$. Therefore

$$E \sum_{t=1}^{K} \|q_{t+1} - q_t\| \leq \sqrt{K} \sum_d \sqrt{q_1(d) - (q_1(d))^2}$$

As the square root and the function $x \mapsto x - x^2$ are concave functions and $\sum_d q_1(d) = 1$, we deduce from Jensen's inequality that

$$E \sum_{t=1}^{K} \|q_{t+1} - q_t\| \leq \sqrt{K}|D| \sqrt{\frac{1}{|D|}(1 - \frac{1}{|D|})} \leq \sqrt{K(|D| - 1)}$$

We now present a proof of the bound $\sqrt{2K}\sqrt{\log |D|}$ (which is sharper for $|D| > 4$, and significantly sharper for large $|D|$; $log|D| = o(|D| - 1)$ as $|D| \to \infty$), using information-theoretic tools. Let (q_t) be $(\mathcal{H}_t)_t$-adapted, that is, q_t is measurable w.r.t. $\mathcal{H}_t$. W.l.o.g. we can assume[6] that $\mathcal{H}_t$ are finite (namely, algebras). In this case we can assume that P is a probability

[6] If q_t is measurable w.r.t. the σ-algebra $\mathcal{H}_t \subset \mathcal{H}_{t+1}$, one replaces $\mathcal{H}_t$ with an algebra $\mathcal{H}^*_t \subset \mathcal{H}^*_{t+1}$ and so that $\|E(q_t \mid \mathcal{H}^*_t) - q_t\| \leq \varepsilon/K$, and replaces q_t with $\hat{q}_t := E(q_t \mid \mathcal{H}^*_t)$ $(= E(q_{K+1} \mid \mathcal{H}^*_t)$. Note that $\sum_{t=1}^{K} \|\hat{q}_{t+1} - \hat{q}_t\| + 2\varepsilon \geq \sum_{t+1}^{K} \|q_{t+1} - q_t\|$.

on the product $D \times (\times_{t=1}^K A_t)$, where A_t are finite sets (e.g., the atoms of the algebra $\mathcal{H}_t$), that $(d, x_1, \ldots, x_K)$ is a vector of random variables having distribution P, and that q_{t+1} is the posterior of d given $x_1, \ldots, x_t$. Let P_t be the conditional (joint) distribution of (d, x_t) given $h_t := x_1, \ldots, x_{t-1}$, P_{tD} its marginal on D, and P_{tA_t} its marginal on A_t. By Lemma 3 we have

$$E_P \| P_t - P_{tD} \otimes P_{tA_t} \| \leq \sqrt{2} \sqrt{H(d \mid x_1, \ldots, x_{t-1}) - H(d \mid x_1, \ldots, x_t)}$$

As $E(\|q_{t+1} - q_t\| \mid h_t) = \sum_{a \in A_t} P_{tA_t}(a) \sum_d |\frac{P_t(d,a)}{P_{tA_t}(a)} - P_{tD}(d)| = \sum_{a \in A_t} \sum_d |P_t(d,a) - P_{tA_t}(a) P_{tD}(d)| = \|P_t - P_{tD} \otimes P_{tA_t}\|$ we have

$$E_P \|P_t - P_{tD} \otimes P_{tA_t}\| = E_P \|q_{t+1} - q_t\|$$

As the square root is a concave function, we have (using Jensen's inequality and the inequality $\sum_{t=1}^K (H(d \mid x_1, \ldots, x_{t-1}) - H(d \mid x_1, \ldots, x_t) = H(d) - H(d \mid x_1, \ldots, x_K) \leq H(d))$

$$E \sum_{t=1}^{K} \|q_{t+1} - q_t\| \leq \sqrt{2K} \sqrt{H(d)} \leq \sqrt{2K} \sqrt{\log |D|}$$

□

The tightness of the order of magnitude of the bound $\sqrt{2K}\sqrt{\log |D|}$ is demonstrated in [9], by proving that there is a constant $C > 0$ such that 1) for every d and K there is a martingale $p_0, \ldots, p_K$ of probabilities on a set D with d elements with total variation $E \sum_{t=1}^K \|p_t - p_{t-1}\| \geq C\sqrt{K \log d}$, and moreover, 2) for every d there is a RGIIOS $\Gamma = \langle M, p, I, J, G \rangle$ with $|M| = d$ with $v_k \geq v_\infty + C\sqrt{\log d}/\sqrt{k}$ for every k.

3.4 A Strategy of the Informed Player in RGIIOS

We present here a result of [1] that is used in the proof of our main result.

Lemma 5 ([1]) *There is a strategy σ in $\Gamma(p)$ such that for every t and every strategy τ we have*

$$E_{\sigma,\tau}(G^m_{i_t, j_t} \mid \mathcal{H}_t) \geq cav_p u\ (p)$$

The following implication of this result is used in our analysis of RGIIBS with uncertain duration. Fix a sequence $n_1 < n_2 < \ldots < n_K$ and a vector of independent random variables $\overrightarrow{c} = c_1, \ldots, c_K$, each c_k distributed according to p_k (e.g., as in our application $p_k = p$), whose realization is private information of P1, e.g., generated by a secret lottery performed by P1. Then, for every strategy τ of P2 in $\Gamma(p,q)$ and every sequence $\hat{q}_k \in \Delta(D)$ where $\hat{q}_k$ is measurable w.r.t. $\mathcal{H}_{n_{k-1}+1}$ (e.g., as in our application, the posteriors of d before the play at stage $n_{k-1}+1$), there is a strategy σ of P1 such that for every $n_{k-1} < t \leq n_k$ we have

$$E_{\sigma,\tau} G^{c_k,\hat{q}_k}_{i_t,j_t} \geq \mathrm{cav}_p u\ (p_k, \hat{q}_k)$$

3.5 Mixing Uncertain Durations

Lemma 6 *For every two uncertain durations Θ_1 and Θ_2 and $0 \leq \beta \leq 1$ there is an uncertain duration Θ such that $E(\theta) \geq \min(E(\theta_1), E(\theta_2))$ and*

$$v_\theta = \beta v_{\theta_1} + (1-\beta) v_{\theta_2}$$

Proof. Let $\Theta_1 = \langle(\Omega_1, \mathcal{B}_1, \mu_1), \theta_1, s^1_1, s^2_1\rangle$ and $\Theta_2 = \langle(\Omega_2, \mathcal{B}_2, \mu_2), \theta_2, s^1_2, s^2_2\rangle$ be two uncertain durations. W.l.o.g. we can assume that Ω_1 and Ω_2 are disjoint and that $S^i_1 = s^i(\Omega_1)$ and $S^i_2 = s^i(\Omega_2)$ are disjoint. For every $0 \leq \alpha \leq 1$, we define the uncertain duration $\alpha\Theta_1 + (1-\alpha)\Theta_2$ as the uncertain duration $\Theta = \langle(\Omega, \mathcal{B}, \mu), \theta, s^1, s^2\rangle$, where Ω is the disjoint union of Ω_1 and Ω_2, the restriction of s^j $(j = 1,2)$, respectively θ, to Ω_i $(i = 1,2)$ is s^j_i, respectively θ_i, $\mathcal{B}$ consists of all unions $B_1 \cup B_2$ where $B_i \in \mathcal{B}_i$, and $\mu(B_1 \cup B_2) = \alpha\mu_1(B_1) + (1-\alpha)\mu_2(B_2)$. Then

$$E(\theta) = \alpha E_{\mu_1}(\theta_1) + (1-\alpha) E_{\mu_2}(\theta_2) \geq \min(E_{\mu_1}(\theta_1), E_{\mu_2}(\theta_2))$$

and

$$v_\theta = \frac{\alpha E(\theta_1) v_{\theta_1} + (1-\alpha) E(\theta_2) v_{\theta_2}}{E(\theta)} = \beta v_{\theta_1} + (1-\beta) v_{\theta_2}$$

and note that as α ranges over [0,1] so does $\beta = \beta(\alpha)$. □

4 The main result

Theorem 1 *For every repeated game with incomplete information on both sides, every $\varepsilon > 0$, and every $vex_q cav_p u\ (p,q) \geq v \geq cav_p vex_q u\ (p,q)$, there is an uncertain duration Θ with $E(\theta) > 1/\varepsilon$ and such that*

$$|v_\theta - v| < \varepsilon$$

Proof. It suffices to prove that

$$\forall \varepsilon > 0 \ \ \exists \Theta \text{ with } E(\theta) > 1/\varepsilon \text{ and } v_\theta \geq \text{vex}_q \text{cav}_p u \ (p, q) - \varepsilon \tag{5}$$

Explicitly, for every $\varepsilon > 0$ there is an uncertain duration $\Theta = \langle(\Omega, \mathcal{B}, \mu), \theta, s^1, s^2\rangle$ such that $E(\theta) > 1/\varepsilon$ and $v_\theta \geq \text{vex}_q \text{cav}_p u \ (p, q) - \varepsilon$. Indeed, (5) implies by duality[7] that

$$\forall \varepsilon > 0 \ \ \exists \Theta \text{ with } E(\theta) > 1/\varepsilon \text{ and } v_\theta \leq \text{cav}_p \text{vex}_q u \ (p, q) + \varepsilon \tag{6}$$

The conclusion of the theorem follows from Lemma 6 together with (5) and (6).

We now turn to the proof of (5). Without loss of generality assume that $\max_{c,d,i,j} |G^{c,d}_{i,j}| \leq 1$. Fix $\varepsilon > 0$.

Let K be sufficiently large so that $B := \min(\sqrt{2 \log |D|}, \sqrt{|D| - 1}) \leq \varepsilon\sqrt{K}/3$. Fix a sequence $n_0 = 0 < n_1 < n_2 < \ldots < n_K$ with $n_{k-1} \leq \varepsilon n_k/2$. Set $\ell_k = n_k - n_{k-1}$. Let $\mu(\theta = n_k) = \frac{1}{n_k \sum_{k=1}^K 1/n_k}$, $k = 1, \ldots, K$. P1 is informed of the value of θ, P2 is not informed of θ. Note that

$$\forall k \leq K, \quad n_k \mu(\theta = n_k) = \frac{1}{\sum_{k=1}^K 1/n_k} = \frac{E_\mu(\theta)}{K} \tag{7}$$

We prove that for every strategy τ in $\Gamma_\theta(p, q)$ ($p \in \Delta(C)$ and $q \in \Delta(D)$) there is a strategy $\sigma = \sigma(\tau)$ such that

$$g_\theta(\sigma, \tau) \geq \text{vex}_q \text{cav}_p u(p, q) - 3\varepsilon$$

Let τ be a strategy of P2 in Γ_θ. As P2 has no information about the realized value of θ, τ is a strategy in Γ. Let q_t be the posterior of d before the play at stage t. Let $\hat{q}_k := q_{n_{k-1}+1}$ (the posterior of d before the play at stage $n_{k-1}+1$). Note that $\hat{q}_k$ is a function of the strategy τ and the sequence of actions $\hat{h}_k := h_{n_{k-1}+1} = (i_1, j_1, \ldots, i_{n_{k-1}}, j_{n_{k-1}})$.

We now define a strategy σ of player 1. Let $\overrightarrow{c} = c_1, c_2, \ldots, c_K$ be a sequence of C-valued random variables such that conditional on the value of θ they are independent, c_k has distribution p, and for k such that $\theta = n_k$ we have $c_k = c$.

[7] Namely, by reversing the roles of P1 and P2 so that P2 is the maximizer and P1 the minimizer with stage payoff $-g$.

The strategy σ will collate a sequence of strategies σ_k, $k = 1, \ldots, K$, by following σ_k in stages $n_{k-1} < t \leq n_k$. The strategy σ_k will depend on $\hat{h}_k := h_{n_{k-1}+1}$ by being a function of $\hat{q}_k$. By Lemma 5, we can select σ_k to be a strategy of P1 in the repeated game $\Gamma^{\hat{q}_k}(p)$ such that for every strategy $\bar{\tau}$ of P2 in $\Gamma^{\hat{q}_k}(p)$, and every $1 \leq t \leq \ell_k$, we have

$$E_{\sigma_k,\bar{\tau}}\ (g_t \mid \mathcal{H}_t) \geq \mathrm{cav}_p\ u(p, \hat{q}_k) \tag{8}$$

In stage $n_{k-1} + t \leq n_k$ of the repeated game $\Gamma(p, q)$, the behavioral strategy σ of P1 plays the mixed action

$$\sigma_k(c_k, i_{n_{k-1}+1}, j_{n_{k-1}+1}, \ldots, i_{n_{k-1}+t-1}, j_{n_{k-1}+t-1})$$

We define the auxiliary stage payoffs g_t^* as follows. For $n_{k-1} < t \leq n_k$ we set

$$g_t^* = G^{c_k,d}_{i_t,j_t}$$

Recall that $2n_{k-1} \leq \varepsilon n_k$, and note that on $\theta = n_k$ we have $g_t^* = g_t := G^{c,d}_{i_t,j_t}$ for $n_{k-1} < t \leq n_k$. Therefore,

$$\sum_t g_t I(t \leq \theta) \geq \sum_t g_t^* I(t \leq \theta) - \varepsilon\theta$$

and thus

$$E_{\sigma,\tau,\mu} \sum_t g_t I(t \leq \theta) \geq E_{\sigma,\tau,\mu} \sum_t g_t^* I(t \leq \theta) - \varepsilon E(\theta) \tag{9}$$

The definition of σ implies that the conditional distribution of $g_1^*, g_2^*, \ldots$, given θ, is independent of θ.

The definition of σ implies that for every $n_{k-1} < t \leq n_k$ we have

$$E_{\sigma,\tau}\ (G^{c_k,\hat{q}_k}_{i_t,j_t} \mid \mathcal{H}_t) \geq \mathrm{cav}_p u\ (p, \hat{q}_k) \tag{10}$$

For every $1 \leq t$ we set

$$y_t^d = \tau(d, h_t) \text{ and } y_t = \sum_d q_t(d) y_t^d$$

Recall that $y_t^d = \tau(d, h_t)$ and that $y_t = \sum_d q_t(d) y_t^d$ is measurable w.r.t. $\mathcal{H}_t$. The play of the strategy σ depends on the realization of $\overrightarrow{c}$. Its play in stages $n_{k-1} < t \leq n_k$ depends only on the value of c_k (which need not be equal to the actual value of c) and therefore (by abuse of notation) we denote

$$x_t^{c_k} = \sigma(c_k, h_t)$$

and for every $n_{k-1} < t \leq n_k$ we denote by p_t the posterior given h_t of c_k.

The definitions of σ, p_t, q_t, y_t^d, and y_t (all as a function of the given strategy τ of P2), together with property (8), imply that for every $n_{k-1} < t \leq n_k$ we have

$$\begin{aligned}
E_{\sigma,\tau}\left(g_t^* \mid \mathcal{H}_t\right) &= \sum_c p_t(c) \sum_d q_t(d) x_t^c G^{c,d}(y_t + y_t^d - y_t) \\
&\geq \sum_c p_t(c) \sum_d (\hat{q}_k(d) + q_t(d) - \hat{q}_k(d)) x_t^c G^{c,d} y_t - \sum_d q_t(d) \|y_t^d - y_t\| \\
&\geq \sum_c p_t(c) \sum_d \hat{q}_k(d) x_t^c G^{c,d} y_t - \|q_t - \hat{q}_k\| - E_{\sigma,\tau}(\|q_{t+1} - q_t\| \mid \mathcal{H}_t) \\
&\geq \sum_c p_t(c) x_t^c G^{c,\hat{q}_k} y_t - \|q_t - \hat{q}_k\| - E_{\sigma,\tau}(\|q_{t+1} - q_t\| \mid \mathcal{H}_t) \\
&\geq \text{cav}_p u\,(p, \hat{q}_k) - \|q_t - \hat{q}_k\| - E_{\sigma,\tau}(\|q_{t+1} - q_t\| \mid \mathcal{H}_t)
\end{aligned}$$

where the second inequality uses Lemma 2 and the last inequality uses inequality (10). Therefore, as $E_{\sigma,\tau}\hat{q}_k = q$ and $\text{vex}_q\text{cav}_p u$ is convex in q and $\leq \text{cav}_p u$,

$$E_{\sigma,\tau}(g_t^*) \;\geq\; \text{vex}_q\text{cav}_p\, u(p,q) - E_{\sigma,\tau}\, \|q_{t+1} - q_t\| - E_{\sigma,\tau}\, \|q_t - \hat{q}_k\|$$

By the triangle inequality (or equivalently, the convexity of the norm) we have $E_{\sigma,\tau}\, \|q_t - \hat{q}_k\| \leq E_{\sigma,\tau}\, \|\hat{q}_{k+1} - \hat{q}_k\|$ and $\|q_{t+1} - q_t\| \leq \|q_{t+1} - \hat{q}_k\| + \|q_t - \hat{q}_k\|$, and therefore, by setting $\eta_k = E_{\sigma,\tau}\, \|\hat{q}_{k+1} - \hat{q}_k\|$, we have

$$\sum_{n_{k-1} < t \leq n_k} E_{\sigma,\tau}(g_t^*) \geq \ell_k \text{vex}_q\text{cav}_p\, u(p,q) - 3\eta_k \ell_k$$

and therefore

$$\sum_{1 \leq t \leq n_k} E_{\sigma,\tau}(g_t^*) \geq n_k \text{vex}_q\text{cav}_p\, u(p,q) - \varepsilon n_k - 3\eta_k n_k$$

Recall (7) and that the distribution of g_t^* is independent of θ. Therefore

$$E_{\sigma,\tau,\mu}\sum_t g_t^* I(t \le \theta) \ge E(\theta)\mathrm{vex}_q\mathrm{cav}_p u\ (p,q) - \varepsilon E(\theta) - \frac{3E(\theta)}{K}\sum_k \eta_k \quad (11)$$

By Lemma 4, $\sum_k \eta_k \le B\sqrt{K}$, where $B = \min(\sqrt{2\log|D|}, \sqrt{|D|-1})$. Therefore, as K is sufficiently large so that $3B\sqrt{K} \le \varepsilon K$, we have

$$E_{\sigma,\tau,\mu}\frac{1}{E(\theta)}\sum_t g_t^* I(t \le \theta) \ge \mathrm{vex}_q\mathrm{cav}_p u\ (p,q) - 2\varepsilon$$

which together with (9) completes the proof of (5). □

5 Remarks

A natural question that arises is whether we can characterize the asymptotic conditions on the distribution of θ (with finite expectation) so that independently of players' signals about θ we will have $v_\theta \to \lim v_n$.

A simple sufficient condition is that $E(\theta) \to \infty$ and $E(|\theta - E(\theta)| + 1)/E(\theta) \to 0$. Indeed, if $n(\theta)$ is the integer part of $E(\theta)$ then $|\sum_t g_t I(t \le \theta) - \sum_{t \le n(\theta)} g_t| \le |\theta - n(\theta)| \le |\theta - E(\theta)| + 1$. Therefore, if $\|G\| := 2\max_{c,d,i,j}|G_{i,j}^{c,d}|$, an optimal strategy of P1 in $\Gamma_{n(\theta)}$ guarantees in Γ_θ a payoff of at least $v_{n(\theta)} - \frac{\|G\|E(|\theta-E(\theta)|+1)}{E(\theta)} \to \lim v_n$ as $\frac{E(|\theta-E(\theta)|+1)}{E(\theta)} \to 0$.

Another natural question that arises is the asymptotic characterization of the distributions μ of the number θ of repetitions that when P1 is informed of θ and P2 is not, then the value v_θ is close to the minmax ($\mathrm{vex}_q\mathrm{cav}_p u\ (p,q)$) of the repeated game $\Gamma(p,q)$. A close look at the proof of the main result reveals a sufficient condition. Given a distribution μ of the uncertain number of repetitions θ and $0 \le \beta \le 1$ we define $\theta(\beta)$ to be

$$\inf\{\beta : E_\mu(\theta I(\theta \le \beta)) \ge \beta E_\mu(\theta)$$

Note that $\theta(\beta)$ is monotonic nondecreasing in β and that the distribution μ constructed in our proof obeys $\theta(k/K) = n_k$. We have the following result: for every $\varepsilon > 0$ there is $\delta > 0$ such that for an uncertainty structure where P1 is informed of the value of θ and P2 is not, if $E(\theta) > 1/\delta$ and for every $\beta < 1 - \delta$ we have $\theta(\beta + \delta) > \theta(\beta)/\delta$, then

$$v_\theta \ge \mathrm{vex}_q\mathrm{cav}_p u\ (p,q) - \varepsilon$$

It is also of interest to find out the limit behavior of v_θ for specific classes of asymmetric uncertain durations. Two suggestive examples are when θ is uniformly distributed on $\{1, 2, \ldots, n\}$ and when θ has the distribution $P(\theta = n) = (1 - \lambda)\lambda^{n-1}$, and P1 is informed and P2 is not informed of the value of θ. Denote the normalized values by v_{n^*} and v_{λ^*}. What are the limits, if they exist, of v_{n^*} as $n \to \infty$ and of v_{λ^*} as $\lambda \to 1-$.

It is also of interest to study the payoff outcomes of repeated games with incomplete information and uncertain duration where the number of repetitions is known to both players, but not commonly known. A study of such non-zero-sum repeated games with complete information is presented in [8].

References

[1] Aumann R. J. and M. Maschler (1995), *Repeated Games with Incomplete Information*, with the collaboration of R. Stearns, MIT Press.

[2] Cover T. M. and J. A. Thomas (1991), *Elements of Information Theory*, Wiley.

[3] Mertens J.-F. (1971), The Value of Two-Person Zero-Sum Repeated Games: The Extensive Case, *International Journal of Game Theory*, **1**, 217–227.

[4] Mertens J.-F. and A. Neyman (1981), Stochastic Games, *International Journal of Game Theory*, **10**, 53–66.

[5] Mertens J.-F., S. Sorin and S. Zamir (1994), *Repeated Games*, C.O.R.E. D.P. 9420, 9421, 9422.

[6] Mertens J.-F. and S. Zamir (1971), The Value of Two-Person Zero-Sum Repeated Games with Lack of Information on Both Sides, *International Journal of Game Theory*, **1**, 39–64.

[7] Mertens J.-F. and S. Zamir (1977), The Maximal Variation of a Bounded Martingale, *Israel Journal of Mathematics*, **27**, 252–276.

[8] Neyman A. (1999), Cooperation in Repeated Games when the Number of Stages is Not Commonly Known, *Econometrica*, **67**, 45–64.

[9] Neyman A. (2009), The Maximal Variation of Martingales of Probabilities and Repeated Games with Incomplete Information. Preprint.

[10] Neyman A. and S. Sorin (2009), Repeated Games with Public Uncertain Duration Processes. Mimeo.

[11] Zamir, S. (1972), On the Relation between Finitely and Infinitely Repeated Games with Incomplete Information, *International Journal of Game Theory*, **1**, 179–198.