Int J Game Theory (2002) 31:223–228

International Journal of Game Theory

Notes on a stochastic game with information structure

Thomas Ferguson[1], **Lloyd Shapley**[1] **and Robert Weber**[2]

[1] University of California, Dept. of Mathematics, Hilgard Av. 405, 90024-1555 Los Angeles, CA, USA
[2] Northwestern University, Kellogg School of Management, 2001 Sheridan Rd., Evanston, IL 60208, USA

Received: February 2002

The game described below appeared in conversations after a talk given by Lloyd Shapley at the Second International Workshop in Game Theory in Berkeley, Summer 1970. This was one of the problems circulating around the conference to which several participants contributed.

The general problem is to study two-person zero-sum games with stochastic movement among subgames in which the subgame being played is not precisely known, but in which information about this aspect may be gathered or partially revealed by the strategy choices of the players, the actual game being played, and chance. Without stochastic movement among the subgames, such a game would be a repeated game of incomplete information defined and studied by the Mathematica Inc. group at Princeton 1966–1968; see the book of Aumann and Maschler, "Repeated Games with Incomplete Information", MIT Press, 1995. Granted that the problem is interesting, one searches for the simplest non-trivial example of the class and is led to the following game.

Subgames. There are two basic subgames, G_1 with matrix $(0 \quad 1)$, and G_2 with matrix $\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$, where $a > 0$. Player I is the row chooser, Player II is the column chooser, and the entries of the matrix represent as usual the winnings of Player I and the losses of Player II.

Movement. Movement between the subgames takes place in a Markov fashion. If G_1 is played, then with probability π the next subgame is G_1 and with probability $1 - \pi$ the next subgame is G_2, independent of the past. If G_2 is played, then G_2 is played as long as Player I chooses the top row; but if Player I chooses the second row, then G_1 is played next.

Information. Player I is given perfect information; in particular, he is always aware of which subgame is being played. Player II is told only the choice of row of Player I; this means that Player II learns the transitions from

G_2 to G_1. Player II is not informed of the outcomes of the subgames as play proceeds.

Payoff. Let β, $0 < \beta < 1$, denote the discount factor; the Game payoff is the discounted sum of all the subgame payoffs.

When Player II makes the choice of a column, the only relevant information he possesses is K, the number of subgames played since the last transition from G_2 to G_1. A stationary strategy for II is a sequence, $\mathbf{y} = (y_0, y_1, y_2, \ldots)$, where y_j denotes the probability of choosing column 1 when $K = j$. Clearly, any optimal choice of $\mathbf{y}$ will have $y_0 = 1$. Therefore we assume that $y_0 = 1$ throughout. A stationary strategy for I is a sequence $\mathbf{x} = (x_1, x_2, \ldots)$, where x_j denotes the probability of choosing row 1 when $K = j$ and subgame G_2 is being played.

For a given pair of strategies, $\mathbf{x}$ and $\mathbf{y}$, let $V_j = V_j(\mathbf{x}, \mathbf{y})$ denote the expected discounted total return to Player I when $K = j$ and G_1 is to be played, and let $W_j = W_j(\mathbf{x}, \mathbf{y})$ denote the expected discounted total return to Player I when $K = j$ and G_2 is to be played.

These quantities are related by the following equations.

$$\begin{aligned} V_j &= (1 - y_j) + \beta[\pi V_{j+1} + (1 - \pi) W_{j+1}] \quad j = 0, 1, 2, \ldots \\ W_j &= a x_j y_j + \beta[x_j W_{j+1} + (1 - x_j) V_0] \quad j = 1, 2, \ldots \end{aligned} \tag{1}$$

Lemma 1. *There is a unique solution to equations (1) subject to the boundary conditions, $\beta^n W_n \to 0$ and $(\pi\beta)^n V_n \to 0$ as $n \to \infty$. The solution for V_0 is*

$$V_0 = \frac{\sum_{j=0}^{\infty} (\pi\beta)^j (1 - y_j) + a \sum_{i=1}^{\infty} \beta^i x_i y_i Z_i(\mathbf{x})}{1 - \beta \sum_{i=1}^{\infty} \beta^i (1 - x_i) Z_i(\mathbf{x})}. \tag{2}$$

where

$$Z_i(\mathbf{x}) = (1 - \pi) \sum_{j=1}^{i} \pi^{j-1} \prod_{\ell=j}^{i-1} x_\ell. \tag{3}$$

Proof: Let $X_j = a x_j y_j + \beta(1 - x_j) V_0$. Then, from (1),

$$\begin{aligned} W_j &= X_j + \beta x_j X_{j+1} + \cdots + \beta^n \left(\prod_{\ell=j}^{j+n-1} x_\ell \right) X_{j+n} + \beta^{n+1} \left(\prod_{\ell=j}^{j+n} x_\ell \right) W_{j+n+1} \\ &\to \sum_{i=j}^{\infty} \beta^{i-j} X_i \prod_{\ell=j}^{i-1} x_\ell \end{aligned}$$

using $\beta^n W_n \to 0$. Similarly, let $Y_j = (1 - y_j) + (1 - \pi)\beta W_{j+1}$, and find

$$\begin{aligned} V_0 &= Y_0 + \pi\beta Y_1 + \cdots + (\pi\beta)^n Y_n + (\pi\beta)^{n+1} V_{n+1} \\ &\to \sum_{j=0}^{\infty} (\pi\beta)^j Y_j \end{aligned}$$

using $(\pi\beta)^n V_n \to 0$. Hence,

$$\begin{aligned}
V_0 &= \sum_{j=0}^{\infty} (\pi\beta)^j((1-y_j) + (1-\pi)\beta W_{j+1}) \\
&= \sum_{j=0}^{\infty} (\pi\beta)^j(1-y_j) + (1-\pi)\beta \sum_{j=1}^{\infty} (\pi\beta)^{j-1} \sum_{i=j}^{\infty} \beta^{i-j} X_i \prod_{\ell=j}^{i-1} x_\ell \\
&= \sum_{j=0}^{\infty} (\pi\beta)^j(1-y_j) + (1-\pi) \sum_{i=1}^{\infty} \sum_{j=1}^{i} \pi^{j-1}\beta^i a x_i y_i \prod_{\ell=j}^{i-1} x_\ell \\
&\quad + V_0(1-\pi)\beta \sum_{i=1}^{\infty} \sum_{j=1}^{i} \pi^{j-1}\beta^i(1-x_i) \prod_{\ell=j}^{i-1} x_\ell \\
&= \sum_{j=0}^{\infty} (\pi\beta)^j(1-y_j) + a\sum_{i=1}^{\infty} \beta^i x_i y_i Z_i(\mathbf{x}) + V_0\beta \sum_{i=1}^{\infty} \beta^i(1-x_i) Z_i(\mathbf{x}).
\end{aligned}$$

Solving for V_0 gives equation (2). ■

Since the conditions, $\beta^n W_n \to 0$ and $(\pi\beta)^n V_n \to 0$ as $n \to \infty$, are automatically satisfied in our game, the problem reduces to a game in strategic form with payoff $V_0(\mathbf{x}, \mathbf{y})$ given by (2). However, another form of this function is more useful for deriving the value and optimal strategies.

Lemma 2.

$$(1-\beta)V_0 = \frac{\sum_{i=0}^{\infty}(\pi\beta)^i(1-y_i) + a\sum_{i=1}^{\infty}\beta^i x_i y_i Z_i(\mathbf{x})}{(1+\beta-\pi\beta)(1-\pi\beta)^{-1} + \beta\sum_{i=1}^{\infty}\beta^i x_i Z_i(\mathbf{x})}, \tag{4}$$

Proof: To derive (4) we must transform the denominator of (2). First note that $Z_i(\mathbf{x})$ of (3) has a recursive definition: $Z_1(\mathbf{x}) = 1-\pi$, and for $i \geq 1$, $Z_{i+1}(\mathbf{x}) = x_i Z_i(\mathbf{x}) + (1-\pi)\pi^i$. Therefore,

$$\begin{aligned}
\sum_{i=1}^{\infty} \beta^i Z_i(\mathbf{x}) &= \beta Z_1(\mathbf{x}) + \sum_{i=1}^{\infty} \beta^{i+1} Z_{i+1}(\mathbf{x}) \\
&= \beta(1-\pi) + \beta\sum_{i=1}^{\infty} \beta^i x_i Z_i(\mathbf{x}) + \beta(1-\pi)\sum_{i=1}^{\infty}(\pi\beta)^i \\
&= \frac{\beta(1-\pi)}{(1-\pi\beta)} + \beta\sum_{i=1}^{\infty} \beta^i x_i Z_i(\mathbf{x}).
\end{aligned}$$

Hence, the denominator of (2) is

$$1 - \beta\sum_{i=1}^{\infty}\beta^i(1-x_i)Z_i(\mathbf{x}) = 1 - \beta\sum_{i=1}^{\infty}\beta^i Z_i(\mathbf{x}) + \beta\sum_{i=1}^{\infty}\beta^i x_i Z_i(\mathbf{x})$$

$$= \frac{(1-\beta)(1+\beta-\pi\beta)}{1-\pi\beta} + (1-\beta)\beta\sum_{i=1}^{\infty}\beta^i x_i Z_i(\mathbf{x}).$$

Substitution of this into (2) gives (4). ■

Theorem 1. *Let $k \geq 1$ be the integer such that $\pi^{k-1} > a/(1+a)$ and $\pi^k \leq a/(1+a)$. Then the value of the game is V^*, where*

$$(1-\beta)V^*/a = \frac{a(1-\beta^k)(1-\pi\beta) - (1-\beta)(a-(a+1)(\pi\beta)^k)}{a(1-\beta^{k+1})(1-\pi\beta) + (1-\beta)\beta(a+1)(\pi\beta)^k} \tag{5}$$

An optimal strategy for Player I is

$$x_j^* = \begin{cases} 1 & \text{for } j < k \\ \pi^k/(a(1-\pi^k)) & \text{for } j = k \\ \pi/(1+a(1-\pi)) & \text{for } j > k \end{cases} \tag{6}$$

An optimal strategy for Player II is

$$y_j^* = \begin{cases} 1 & \text{for } j < k \\ \beta(1-\beta)V^*/a & \text{for } j \geq k \end{cases} \tag{7}$$

It is interesting to note that Player I's optimal strategy is independent of β.

Proof: We show that each conjectured optimal strategy is a best reply to the other. But we use a proof that shows how the optimal strategies are derived using the principle of indifference.

First we show how Player I's strategy is derived. Note that V_0 depends on $\mathbf{y}$ only through the numerator of (4) (or (2)) and is linear in the y_i. Player I's strategy is chosen to make Player II indifferent as to his choice of strategy for $i \geq k$, that is to make the coefficient of y_i equal to zero for $i \geq k$. For $i \geq 1$, the coefficient of y_i in this numerator is $-(\pi\beta)^i + a\beta^i Z_i(\mathbf{x})$. Since $Z_i(\mathbf{x})$ depends on $\mathbf{x}$ only through $x_1, \ldots, x_{i-1}$, we can set the coefficient of y_i to zero and solve for x_i. This gives

$$x_i = \frac{\pi^i}{aZ_i(\mathbf{x})}. \tag{8}$$

This is clearly positive, but it must also be no greater than 1. For $i < k$ this cannot be achieved, so we assume $x_i = 1$ for $i < k$. If this is so, we may compute inductively that $Z_i(\mathbf{x}) = 1 - \pi^i$ for $i \leq k$. Then for $i = k$, (8) gives

$$x_k = \frac{\pi^k}{a(1-\pi^k)},$$

which is less than or equal to 1 by the definition of k. Substituting this into the formula for Z_{k+1} gives $Z_{k+1}(\mathbf{x}) = \pi^k(1+a-a\pi)/a$, from which we may compute $x_{k+1} = \pi/(1+a-a\pi)$. Then $Z_{k+2}(\mathbf{x}) = \pi^{k+1}(1+a-a\pi)/a$, and $x_{k+2} = \pi/(1+a-a\pi)$, and so by induction,

$$x_i = \pi/(1+a-a\pi) \quad \text{and} \quad Z_i(\mathbf{x}) = \pi^{i-1}(1+a-a\pi)/a$$

for all $i > k$. Thus the strategy (6) makes V_0 independent of y_i for $i \geq k$. The coefficient of y_i for $i < k$ is $\beta^i(a-(a+1)\pi^i)$ which is negative. This requires $y_i = 1$ for $i < k$ as a best reply.

To compute $\max_{\mathbf{y}} V_0(\mathbf{x}^*, \mathbf{y})$, we may take $y_i = 1$ for all i. This involves computing

$$\begin{aligned}\sum_{i=1}^{\infty} \beta^i x_i^* Z_i(\mathbf{x}^*) &= \sum_0^{k-1} \beta^i(1-\pi^i) + \beta^k\pi^k/a + \sum_{k+1}^{\infty} \beta^i\pi^i/a \\ &= \frac{1-\beta^k}{1-\beta} - \frac{a-(a+1)(\pi\beta)^k}{a(1-\pi\beta)} \end{aligned} \tag{9}$$

When this is substituted into expression (4) for $V_0(\mathbf{x}^*, \mathbf{y})$ with y_i identically 1, the V^* of expression (5) emerges. Thus, Player I, using $\mathbf{x}^*$, is guaranteed an expected return of V^*.

Now turn to Player II. If Player II uses $y_i = 1$ for $i < k$, then certainly Player I's best response is to use $x_i = 1$ for $i < k$. Player II would like to choose y_i for $i \geq k$ to make Player I indifferent for choosing x_i for $i \geq k$, that is, to make V_0 independent of x_i for $i \geq k$. This may in fact be done taking y_i constant for $i \geq k$, say $y_i = \theta$. To see this, note that with $x_i = y_i = 1$ for $i < k$ and $y_i = \theta$ for $i \geq k$, the V_0 of (4) as a function of x_i, $i \geq k$, has the form

$$\begin{aligned}(1-\beta)V_0 &= \frac{c_1 + c_2\theta + a\theta\sum_k^\infty \beta^i x_i Z_i(\mathbf{x})}{c_3 + \beta\sum_k^\infty \beta^i x_i Z_i(\mathbf{x})} \\ &= \frac{a\theta}{\beta}\left[\frac{\frac{c_1+c_2\theta}{a\theta} + \sum_k^\infty \beta^i x_i Z_i(\mathbf{x})}{\frac{c_3}{\beta} + \sum_k^\infty \beta^i x_i Z_i(\mathbf{x})}\right]\end{aligned}$$

for some constants c_1, c_2 and c_3. This can be made independent of $\mathbf{x}$ by choosing θ so that $\frac{c_1+c_2\theta}{a\theta} = \frac{c_3}{\beta}$. When this is done, the expression in square brackets is 1, so $(1-\beta)V_0 = a\theta/\beta$. Moreover, the resulting V_0 must be V^* since V^* is obtained when $\mathbf{x} = \mathbf{x}^*$ whatever be θ. Therefore, $\theta = \beta(1-\beta)V^*/a$ as in (7). To show that $\theta < 1$ is straightforward. ■

It is interesting to take the limit as $\beta \to 1$ since this generally gives the solution for limiting average payoff.

$$(1-\beta)V^* \to a\frac{ak(1-\pi) - a + (a+1)\pi^k}{a(k+1)(1-\pi) + (a+1)\pi^k} = V_\infty$$

$$y_j^* \to \begin{cases} 1 & \text{for } j < k \\ V_\infty/a & \text{for } j \geq k. \end{cases}$$

A Beliefs-Based Derivation of the Optimal Strategies. Here is an alternate way of deriving the optimal strategies. For II to randomize in round j, he must be indifferent between columns 1 and 2. This can only be the case if he believes there is probability $b_j = ax_j/(1+ax_j)$ that the game is in state G_1. A round later, his beliefs will have updated to

$$b_{j+1} = \frac{b_j \pi}{b_j + (1-b_j)x_j} = \frac{a\pi}{1+a}.$$

(The pleasant surprise, of course, is that x_j drops out of the formula.) For him to still be indifferent, this must equal $ax_{j+1}/(1+ax_{j+1})$. If he is to be indifferent for all rounds $j \geq k$, then for all $j > k$ we must have $x_j \equiv x$, and

$$\frac{ax}{1+ax} = \frac{a\pi}{1+a} \quad \text{or} \quad x = \frac{\pi}{1+a-a\pi}.$$

The first round in which I randomizes, k, is the first in which II's posterior belief that the state is still G_1, π^{k+1}, would otherwise drop below $a\pi/(1+a)$, and x_k must be just large enough to hold II's beliefs up to that level. In other words, we must have

$$b_{k+1} = \frac{\pi^k \cdot \pi}{\pi^k + (1-\pi^k)x_k} = \frac{a\pi}{1+a} \quad \text{or} \quad x_k = \frac{\pi^k}{a(1-\pi^k)}.$$

For I to be indifferent from round k onwards, his loss from postponing row 2 from round j to the next round, $\beta^{j+1}V^* - \beta^{j+2}V^* = \beta^{j+1}(1-\beta)V^*$, must equal his gain, $\beta^j(ay_j)$. Hence, $y_j = \beta(1-\beta)V^*/a$, for all $j \geq k$. Then V^* may be obtained through direct summation, as in (9).