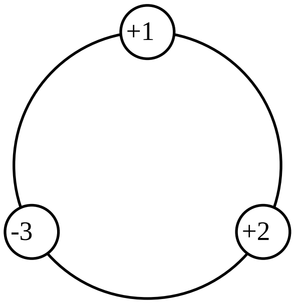
GLEASON'S GAME

T. S. Ferguson and L. S. Shapley
University of California, Los Angeles

Abstract

A two-person zero-sum game invented by Andrew Gleason in the early 1950's has a very simple description and yet turns out to be quite difficult to solve. This game is a stochastic game with an information lag for both players. No strategy with a bounded memory of past moves can be optimal. Yet using the notion of generalized subgames, we show that there exist optimal strategies of a simple nature based on functions easily approximable by standard methods of computation for stochastic games.

1. Description of the game. Two players, Andy and Dave, move a counter around a three node board. The nodes are arranged in a circle and are labeled $+1$, $+2$, and -3. Initially the counter rests on node $+1$ and Andy starts.



Thereafter, the players alternate moves. There is a one move delay in informing the players of the position of the counter, so that, except for the first move, players make their moves only knowing the node from which the opponent has just moved. A move consists of instructing a referee to move the counter either clockwise or counterclockwise to the next node. One is not allowed to leave the counter where it is. After each move is given to the referee, the referee announces the node that the counter just left, and requires Dave to pay Andy the amount on the label of that node. (Thus, Dave wins $+3$ from Andy if this node is -3.) The problem is for Andy to maximize and for Dave to minimize the limiting average payoff[1].

It seems that Andy's first move should be clockwise, from $+1$ to $+2$, and then that Dave's first move should also be clockwise, to -3, giving so far a

[1] Since the limiting average payoff may not exist, to be precise we should use the limsup (or the liminf) of the average payoff.

total payoff of zero. Now Andy should randomize his choice; if he always goes to $+2$ then Dave can always put him back to -3; if he always goes to $+1$, the same thing will happen. So Andy will decide to go counterclockwise (cc.) with a certain probability q and clockwise (cl.) with probability $p = 1-q$, and now Dave will have a similar problem and must also randomize his choice. If Dave goes cc. suspecting that Andy went to $+1$ when Andy actually went to $+2$, Dave will lose a total of three in those two moves. A similar analysis holds if Dave goes cl. We will see later the continuation of this game using optimal play.

Let us give a preliminary analysis of Gleason's game. We call the player who wins the amounts labeled on the nodes Player I and the one who loses these amounts Player II. A strategy in the infinite game is a rule that gives the probability that a player moves cc. as a function of the information received so far in the game. Results of Scarf and Shapley [5] show that the minimax theorem holds in this situation, namely, that the game has a value, V, and that both players have optimal strategies, σ_I for I, and σ_{II} for II. This means that if I uses σ_I, the expected liminf of the average payoff is at least V no matter what II does, and if II uses σ_{II} the expected limsup of the average payoff is at most V no matter what I does.

It turns out that $V < 0$. The game favors Player II. This was known in the 1950's. In fact, there is a simple strategy for II whereby he can keep the limiting average payoff less than or equal to zero. It is as follows.

If last at $+1$, go cl.
If last at $+2$, go cc.
If last at -3, go cc. with probability $1/3$ and cl. with probability $2/3$.

This strategy is simple in that it uses knowledge of the past only through the last known state. After some calculation, it is not hard to see that I's best reply to this strategy only guarantees that he will break even in the long run. One can also show that the above is the best strategy for II that uses only the last known state. It is also fairly easy to see that this rule cannot be optimal for II. If II uses more past information, he is certain to be able to do a little better, so the value must be strictly negative. This is essentially what was known about the game in the 1950's.

2. Better bounds on the value. We may try to extend the above method to obtain better approximations to the value. Consider strategies for Player II that depend on the last two known states. The best among these seems to be the following:

If last at $+1$, go cl.
If last at $+2$, go cc.
If the last two nodes are $+1$ and -3, go cc. with probability .34.
If the last two nodes are $+2$ and -3, go cc. with probability .49.

The best that Player I can do against this strategy gives him a limiting average return of about $-.01316$. One might suspect that we are close to the value of Gleason's game since adding one more stage of memory to II's strategy reduces the limiting average return by only .01316. Unfortunately, the corresponding upper bound to the value, requiring Player I to use a strategy that uses only the last known state, turns out to be $-.17859$. So we have $-.17859 < V < -.01316$. There is quite a gap to narrow.

From an analysis of this sort, we can see why the problem seems easy but is actually hard. When the referee announces the state just vacated, both players know the history of the game up to that point. Indeed, this information is common knowledge (both players know the other knows, both know the other knows he knows, etc.). At first sight, it might be thought that one need not remember back past that point in choosing a strategy. This is not so because when the opponent made his last move, he had to choose it not knowing the actual state, and you should be able to take advantage of that. And the opponent chose his strategy trying to take advantage of your lack of knowledge of the previous state, so that should be taken into account, and so on. As we shall see, no strategy that remembers only a bounded number of past moves can be optimal.

Still, we might try to find as an approximation to the optimal strategy the best strategy that takes into account only the last k states. The trouble is that this requires considering 3 times $2k-1$ parameters in $[0,1]$. Even for $k=3$, the asymptotic Markov chain analysis required for this is exceedingly complex. However, the point is moot since there exists an optimal strategy that uses knowledge only of the last two known states and of one parameter which summarizes all the past history.

3. Generalized subgames. Without the information lag feature, this game would be a stochastic game, so that methods that work for stochastic games might be expected to be of use. A *stochastic game* G is a finite collection of matrix games, $G = \{G^1, \ldots, G^m\}$, together with a set of transition probability matrices, $P = \{P_1, \ldots, P_m\}$, jointly controlled by the pure strategy choices of the players. If game $G^i = (g^i_{rc})$ is being played, then simultaneously I chooses a row r and II chooses a column c. Then there is an immediate payoff of g^i_{rc} from II to I, and the next game played is G^j with probability $P_i(j|r,c)$. There may be a discount, $0 < \beta < 1$, in which case the total payoff is the present value of the infinite stream of payoffs, $\sum_1^\infty \beta^n Z_n$, where Z_n represents the payoff at stage n. In problems without a discount ($\beta = 1$), the limiting average payoff is used, say $\liminf_{n\to\infty}(Z_n/n)$. The recent book of Filar and Vrieze [3] contains a thorough and careful exposition of the area.

Stochastic games were introduced by Shapley [6] who gave a proof of existence of the value and optimal strategies in the discounted case. He q p

also gave a method of approximating the solution by looking at *subgames* G_i which represent the game G when the initial game played is G^i. If V_i represents the value of G_i, then

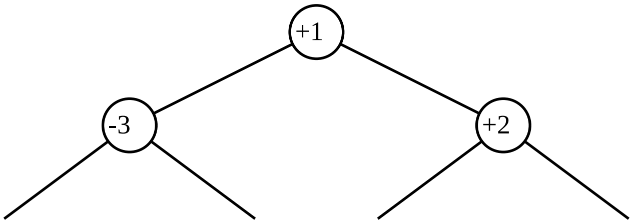
$$V_i = \mathrm{Val}(g^i_{rc} + \beta \sum_{j=1}^{m} P_i(j|r,c)V_j) \quad \text{for} \quad i = 1, \ldots, m. \tag{1}$$

One may choose an arbitrary set of initial values for the V_i and iterate this set of equations to approximate the solution. This method of approximating the solution is called Shapley iteration.

Gleason's game is not a stochastic game because of the information lag. However, a theory of games with information lag was developed by Scarf and Shapley [5]. The notion of a *generalized subgame*, introduced there, may be used to reduce Gleason's game to a stochastic game. The cost of this reduction is that the state space and the strategy spaces become infinite.

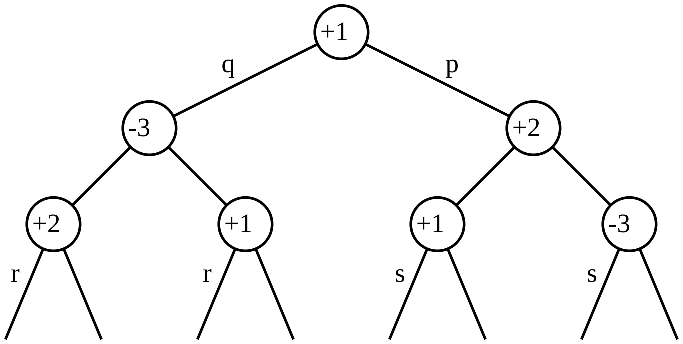
Recall that even if I is required to announce his mixed strategy, the value of the game will still be V. Thus, if we change the game by requiring I to tell II what mixed strategies he will be using in the future, the resulting game will have the same value and the same optimal strategy for Player I as the original game. Suppose that I announces that his first move is cc. with probability q, and cl. with probability $p = 1 - q$. We refer to this game as the generalized subgame for Player I starting at $+1$ with probability q counterclockwise, and denote it by $G(1, q)$. Similarly, $G(2, q)$ and $G(3, q)$ represent the generalized subgames with I moving cc. with probability q at nodes $+2$ and -3 respectively. The situation for Player II is then:

The Generalized Subgame, G(1,q).



There will be an immediate payoff to I of $P_\beta(1, q) = 1 + \beta(-3q + 2p)$ in the discounted case, and $P(1, q) = 1 - 3q + 2p$ in the limiting average case. Now II must decide cl. or cc. knowing only that he is at state -3 with probability q and at state $+2$ with probability $1 - q$. But he asks I the probabilities I will use on the next round. Player I will say something like "If I went to -3, I will next go cc. with probability r; otherwise I will go cc. with probability s". The tree expands to:

The Expanded Subgame.



The collection of generalized subgames, $G(k, q)$, forms a stochastic game with state space, $\{(k, q) : k \in \{1, 2, 3\}, q \in [0, 1]\}$. This game is played as follows. Suppose we start at game $G(k, q)$. Player I, knowing the state (k, q), must choose $r \in [0, 1]$ and $s \in [0, 1]$. Player II, knowing the state and I's strategy choice (r, s), must choose to go cc. or cl. Then there is an immediate payoff of $P_\beta(k, q)$, and transition to a new generalized subgame occurs according to the following probabilities. If II chose cc., the next game played is $G(k+1, r)$ with probability q and $G(k, s)$ with probability $1-q$. If II chose cl., the next game played is $G(k, r)$ with probability q and $G(k+2, s)$ with probability $1-q$. (In these expressions, $k+1$ and $k+2$ are understood to be read modulo 3.)

There are certain noteworthy features of these games. First note that Player I may as well be informed of the state when he makes his choice. Secondly, we may assume the players move alternately, with Player I going first, and that the players have full information of past moves. In other words, this is a stochastic game of perfect information! Thirdly, the immediate payoffs, $P_\beta(1, q) = 1 + \beta(-3q + 2p)$, $P_\beta(2, q) = 2 + \beta(q - 3p)$ and $P_\beta(3, q) = -3 + \beta(2q + p)$, depend only on the state and not on the strategy choices of the players.

Let $V_\beta(1, q)$ (resp. $V_\beta(2, q)$ and $V_\beta(3, q)$) denote the value of the β-discounted subgame starting at node $+1$ (resp. $+2$ and -3) when q is announced. Then, I chooses r and s to maximize his future expected return and II will similarly minimize, so that

$$\begin{aligned} V_\beta(k, q) &= P_\beta(k, q) \\ &\quad + \beta^2 \max_{r,s} \min\{qV_\beta(k+1, r) + pV_\beta(k, s), qV_\beta(k, r) + pV_\beta(k+2, s)\} \end{aligned} \tag{2}$$

for $k = 1, 2, 3$. This gives three simultaneous functional equations to be solved for the three value functions, $V_\beta(k, q)$. If β is not too close to one, iteration techniques of discounted stochastic games such as Shapley iteration may be employed.

However, Gleason's Game as originally stated is a limiting average payoff game. To obtain the corresponding equations for the limiting average payoff case, we may apply a standard approximation used in Markov decision processes and stochastic games by writing

$$V_\beta(k,q) = V/(1-\beta) + W(k,q) + o(1-\beta) \quad \text{for} \quad k = 1,2,3. \tag{3}$$

Here, V is the overall limiting average payoff, independent of the initial state, and $W(k,q)$ represents the amount above or below some standard that being in state (k,q) confers upon I (on the average). The actual values of the $W(k,q)$ are only fixed when we fix the standard; otherwise only the differences $W(k,q)-W(k',q')$ are fixed. For example, we may fix $W(1,0) = 0$ for the standard, or we may fix $\sum_k \int W(k,2q)\,dq = 0$. If we replace $V_\beta(k,q)$ by its approximation and let $\beta \to 1$, we find

$$\begin{aligned} &2V + W(k,q) = \\ &\quad P(k,q) + \max_{r,s} \min\{qW(k+1,r) + pW(k,s), qW(k,r) + pW(k+2,s)\} \end{aligned} \tag{4}$$

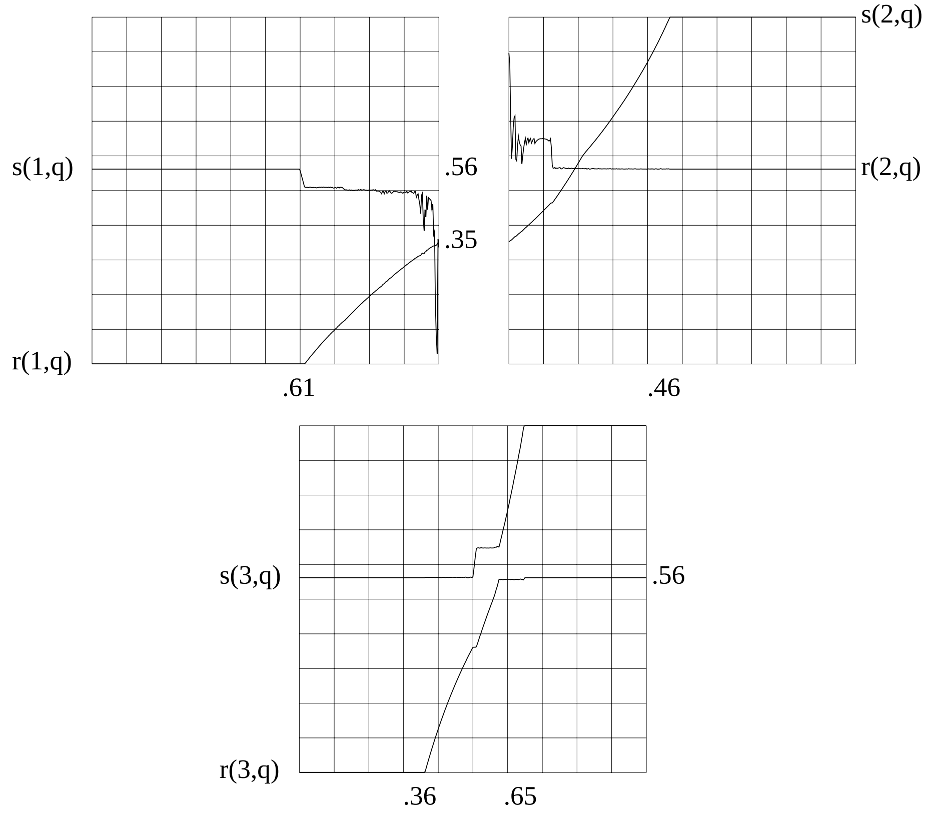
for $k = 1,2,3$, where $P(k,q)$ represents $P_\beta(k,q)$ for $\beta = 1$.

The value of V that satisfies these equations is the value of Gleason's game. A stationary strategy for I is a set of six functions, $r(k,q)$ and $s(k,q)$, for $k = 1,2,3$. The optimal stationary strategy for I can be found as the values of r and s that achieve the maximum in the above equation. These are the optimal strategies for I in the original Gleason's game. However, this method gives very little information about II's optimal strategy in the original game. To find II's optimal strategy, we must view the generalized subgames from his point of view and derive and solve the corresponding set of equations.

4. Numerical Approximations. To approximate the solution of the generalized subgames, we discretize the strategy space of Player I by restricting him to use probabilities in some discrete set, say $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$ for some large value of n. Restricting I in this way, yields a game whose value is a lower bound to the value of the original game. For each n, the class of approximating generalized subgames becomes a stochastic game with finite state and action spaces and limiting average payoff.

It is known that Shapley iteration does not ordinarily converge for stochastic games with limiting average payoff. However, the method of Hoffman and Karp [4] can be applied to this problem. (See Algorithm 5.1.1 in [3].) This is an iterative method that involves alternately solving a finite game and a Markov decision problem with limiting average payoff. Solving the

Figure 1: Optimal Strategy Diagrams for Player I



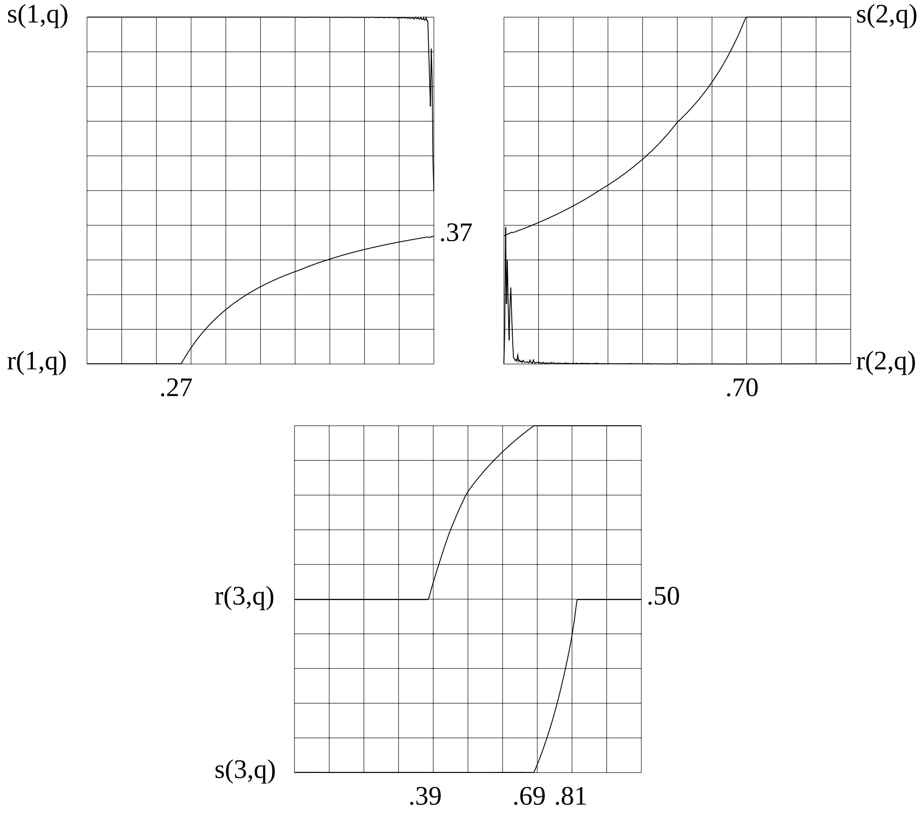
finite game is easy because it is a game of perfect information; so one does not have to resort to mixed strategies.

This method was applied to solve the discrete version of equation (4) using $n = 1600$. (I's strategy set consists of about $6 \times 1600 = 9600$ variables, each capable of taking values in a set of size 1600. Thus, I has about 1600^{9600} pure strategies!) Approximate computations have been performed which show that the value of Gleason's game is approximately $V = -.0933$ (actually, $-.09336 < V < -.09323$).

Of great interest, of course, are the optimal strategies of Player I in this version since they are the same as for the original game. Thus, we are ultimately interested in the values of r and s that achieve the maximum in the above equations. We denote these by $r(k, q)$ and $s(k, q)$. Knowledge of these functions tell Player I how to play Gleason's game optimally. The numerical approximation to these functions is displayed in Figure 1.

We see in these diagrams a strange behavior of $s(1, q)$ for q near 1 and $r(2, q)$ for q near 0. This occurs for two reasons. First, these values play only a very small role in the value of the game. In fact, $s(1, 1)$ and $r(2, 0)$

Figure 2: Optimal Strategy Diagrams for Player II



play no role at all; they may be set arbitrarily. Second, the use of discrete probabilities in the approximation plays a role in creating artificial small oscillations in the functions. Based on mesh size 1/1600, it is difficult to guess how the true function behaves in the corresponding regions, though we suspect it is still monotone there.

The use of these diagrams may be explained as follows. Suppose that on his last move I announced that he was going cc. from +1 with probability q. Then if I has actually gone cc. to -3, he would next go cc. with probability $r(1,q)$, while if I had actually gone cl. to +2, he would next go cc. with probability $s(1,q)$.

We may find II's optimal strategy by viewing the generalized subgames from II's point of view and deriving a similar set of equations. When this is done, we find corresponding optimal strategy diagrams for Player II as displayed in Figure 2.

5. Final Remarks. As an illustration of the use of these figures, let us continue the game between Andy and Dave. Since Andy's first move was cl. from +1 to +2 with probability one (q=0), on his next move he should go cc. with probability $s(1,0) = .562$ from Fig. 1. Dave doesn't know if the counter is on +1 or +2, but since on his first move he went cl. from +2 with probability one ($q = 0$), he should now go cc. with probability $s(2,0) = .369$ from Fig. 2.

Let us continue this game for two more steps. Suppose the outcome of Andy's random move from −3 was cl. to +1. Then his next move should be cc. with probability $s(3,.562) = .648$ using Fig. 1. (If he had moved cc. from −3 to +2, his next move would have been cc. with probability $r(3,.562) = .510$.) Suppose in addition that the outcome of Dave's random move from +1 was cc. to −3. Then his next move should be cc. with probability $r(1,.369) = .129$ using Fig. 2, and so forth.

Using an analysis of this sort, one can see that no strategy that remembers only a bounded number of past moves can be optimal. This may be done by exhibiting a sequence of positions with positive probability for each n under an optimal strategy for one of the players such that all positions differ. This is perhaps most easily done for Player II, starting at $(2,.0)$ and alternating between (cl., cl.) and (cc., cc.) by the players. This leads to a sequence of positions $(2,.0)$, $(1,.369)$, $(2,.129)$, $(1,.421)$, $(2,.173)$, $(1,.442)$, $(2,.188)$, $(1,.449)$, $(2,.193)$, (1.452), …, alternating between the functions $s(2,q)$ and $r(1,q)$ of Fig. 2.

In closing, we would like to mention a slightly simpler form of Gleason's game that arises by modifying the numbers on the nodes, and hence the payoffs, to read +1, +1, and −2. By changing location and scale of the payoffs, we may define an equivalent game with nodes numbered 0, 0 and +1. This version is simpler than the original Gleason's game because it is symmetric in simultaneously interchanging the 0 nodes and interchanging cc. and cl. From this symmetry we may conclude that the optimal strategies are also symmetric in the sense that $r(1,q) = 1 - s(2,1-q)$, $r(2,q) = 1 - s(2,1-q)$, and $r(3,q) = 1 - s(3,1-q)$, where 3 represents the node with payoff +1. The value of this game has been evaluated by the above methods to be $.35575 \pm .000014$.

REFERENCES

References

[1] Blackwell, David (1955). The Prediction of Sequences. RAND Report RM1570, RAND corporation, Santa Monica.

[2] Blackwell D. and Ferguson, T. S. (1968). The Big Match. *Ann. Math. Statist.* **39** 159-163.

[3] Filar, J. and Vrieze, K. (1997). *Competitive Markov Decision Processes.* Springer-Verlag, New York.

[4] Hoffman, A. J. and Karp, R. M. (1966). On Non-Terminating Stochastic Games. *Management Science* **12** 359-370.

[5] Scarf, H. E. and Shapley, L. S. (1957). Games with Partial Information. *Contribution to the Theory of Games, III, Ann. Math. Studies 39* 213-229.

[6] Shapley, L. S. (1953). Stochastic Games. *Proc. Nat. Acad. Sci. U.S.A.* **39** 327-332.