ON GAME-LEARNING THEORY AND SOME DECISION-MAKING EXPERIMENTS

Merrill M. Flood

P-346

17 November 1952

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

TABLE OF CONTENTS

TABLES

ON GAME-LEARNING THEORY AND SOME DECISION-MAKING EXPERIMENTS

Merrill M. Flood
The RAND Corporation

Summary: This paper reports on games in which a player "learns" to improve his strategy during the course of a sequence of plays. The fusion model developed by Bush and Mosteller to explain observed behavior of rats in experimental learning situations was used as the basis for both theoretical and experimental investigation of the efficiency of this type of stochastic process in learning to play games. The experiments reported here were with human subjects. Their game-learning performance was compared with that of the "stat-rat", represented by the fusion model with numerical values of the parameters estimated to fit experimental data for rats. The theoretical models accept basic assumptions of von Neumann-Morgenstern game theory and Bush-Mosteller learning theory. The theoretical and experimental results are directly relevant for any situation in which a sequence of decisions is made.

1. Introduction

The theory of games [1]* provides a general mathematical model that may sometimes be used to approximate a real situation. Usually, in real cases, the situation is much too complicated to permit its formulation even conceptually as a formal game. And in the few cases that can be so formulated, it is almost always impractical to attempt gathering the necessary data or to do the elaborate calculations required for a solution.

The non-constant-sum case, even with two players, remains unsolved in the sense of von Neumann-Morgenstern. There are theoretical proposals that dispose reasonably well of the two-person case, and of many other broad special cases; I have discussed some of these in another paper [2].

---

* Bracketed numbers refer to the bibliography at the end of the paper.

I am indebted to Dr. D. R. Fulkerson for a careful reading of the manuscript, and for many helpful comments during the course of the work.

In this paper I investigate game-like situations in which the players are limited biologically in their choices of moves. These limitations are reflected in the method of play of the formal game and stem from the notion that animal organisms seem to learn by some sort of conditioning process that alters the probability that some one of several mutually exclusive alternatives will be selected in each new instance.

This approach was suggested to me by the work of R. F. Bales and A. S. Householder [3] on the group interaction process, and is closely connected with the work of R. R. Bush and C. F. Mosteller [4] on mathematical models for learning. I have profited from discussions with all four of these men. There is also an interesting philosophical discussion of stochastic learning models in a recent paper by D. M. MacKay [5], and a stimulating essay by E. G. Boring [6] on "robotology"; both of these latter papers seem to me to support the methodological viewpoint that I have adopted.

2. Generalities

The approach used in this paper is applicable to situations involving more than two organisms, but I shall concentrate somewhat on the two-player case. A player could in fact be a group of people, or a component of personality within one individual, but I shall only touch upon such interpretations occasionally. Since my main object is to treat some one case of real behavior, I shall usually be content with a discussion in terms of a special real-life situation, leaving broader interpretations to the reader.

The connection with game theory is the correspondence between the notion of choice of a strategy for a game in normal

[redacted] and the notion of individual choice of course of action in biological activity. A very fundamental case, and perhaps the simplest one, is the problem of choosing whether or not to act in a situation where there appears to be only one choice: acting or not acting. For example, in experiments like those of B. F. Skinner [7] with rats, the choice at some moment is whether or not to press a bar. For a human example, the choice might be whether or not to accept a particular offer for a new position. In these examples, and in most real-life situations, the organism somehow reduces its range of alternatives to a relatively small number from which it feels it must choose;* it is this recognized field of choices; whether they are considered to be conscious or unconscious alternatives, that corresponds to the set of strategies listed in the normal form of the formal game.

In what follows, I shall try to use such terms as game, strategy, move, and player from game theory only with the meaning attached formally by von Neumann and Morgenstern [1]; for other purposes I shall use alternative words such as situation, plan, act, and subject.

3. The problem

The games in which we shall be interested are defined in terms of expectation functions:

$$v_i^j \equiv v_{i_1 i_2 \cdots i_n}^j \quad \text{for} \quad (i_k = 1, 2, \cdots, m_k;\ j = 1, 2, \cdots, n).$$

---

* This process has been most systematized for the art of decision by military commanders [8].

A play of the game consists of simultaneous independent choices of specific values i_k^o for the i_k by the n players; the quantity $V^j_{i^o}$ is the expectation for player j, where the units for $V^j_{i^o}$ relate to a measure of the utility attached by player j to the payments he receives. The functions V_i^j are real-valued.

The actual payment to player j, when the choice of pure strategies is i on a play, is a quantity x given by the distribution function $P_i^j(x)$ whose mean value is V_i^j; of course $|x|$ is bounded, so that

$$P_i^j(x) = 0 \text{ if } |x| > \bar{b}_i^j .$$

Our problem is to select a good method of play that can be used by player ℓ when his information about the structure of the game is knowledge only of

$$m_\ell \text{ and a bound } \bar{b}^\ell, \geq \max_i \bar{b}_i^\ell > 0,$$

and where his information about the distribution functions $P_i^j(x)$ is gained entirely from his experience while playing the game.

It is assumed, in the process of passing from the normal to the extended form of a game, that only the mean values V_i^j of these distributions affect the situation [1]; we can assume without essential loss of generality, therefore, that the variance of $P_i^j(x)$ is zero so that x only assumes the value V_i^j. Furthermore, since the problem is essentially unchanged if the utility measure x is subjected to a linear transformation [1] we may take $\bar{b}^j=1$ and suppose that $P_i^j(x)=0$ if $x<0$ or $x>1$.

The experience gained by player j in N plays of the game consists of a record of his own choices $i_j(t)$, and receipts

$r_j(t)$ for t=1, 2, ···, N. The central problem is to find a rule of play that will tend to maximize total receipts in a sequence of plays where the player is given some information about the number of plays before he starts on a sequence; I am intentionally vague at this point in the paper about the exact nature of the rule of play and about the advance information concerning the length of sequence.

We shall be interested in what follows, then, only in the game whose normal form has the expectation functions

$$0 \leq V_i^j \leq 1.$$

We shall be especially concerned with one extended form of this game in which there is one chance move for each player and the actual payments are always unity or zero, whence the probability of a unit-payment to player j is V_i^j if the players choose the pure strategies i. We have noted that any game can be reduced to this form by suitable linear transformations on the utility measures of the individual players, provided only that there are known finite bounds on the possible payments; use was also made of the assumption that games in extended form are equivalent if their normal forms are identical.

4. The game-learning model

The type of rule of play, that is investigated here, is represented for player j by the relation

$$p^j(t+1) = M^{jk} p^j(t) \text{ for } k = 0, 1, 2, \cdots, 2m_j ,$$

where $p^j(t)$ is a vector of probabilities and the M^{jk} are matrices.*

---

* The M^{jk} are necessarily stochastic matrices such that the elements are non-negative and, in any column, sum to unity.

The components $p_i^j(t)$, for $i=0, 1, 2, \cdots, m_j$, are the probabilities that player j selects value i for i_j on play t; $p_0^j(t)$ has a special significance to be discussed later. The elements $M^{jk}_{\alpha\beta}$ of M^{jk}, for $\alpha, \beta = 1, 2, \cdots, m_j$, are given real numbers. The probability that M^{jk} is applied after play t depends only upon $p^j(t)$, and certain constants to be discussed, and so the procedure is a Markov process.

We now describe how the appropriate operat . M^{jk} is chosen after play t. The matrices M^{jk} are first separated into two classes of m_j+1 members each, denoted R^{jk} and P^{jk} for $k=0, 1, 2, \cdots, m_j$, and either $R^{j\,i_j(t)}$ or $P^{j\,i_j(t)}$ is selected; the choice between these two matrices is made with probabil'.y $r_j(t)$ in favor of $R^{j\,i_j(t)}$.

We have now defined a method of play that can be used by any player after he has made his initial strategic choice $p^j(0)$, and after specific values have been assigned for the elements of M^{jk}; in actual practice he will need to know m_j and S^j also.

5. The fusion game-learning model

We shall now consider a special parametric form for the matrices M^{jk}. For convenience, we shall omit the designation of the player when this leads to no ambiguity.

We set:

$$R^i_{\alpha\beta} = (1-b^i-c^i)\,\delta_{\alpha\beta} + c^i\delta_{\alpha 1} + b^i\delta_{\alpha 0}, \text{ and}$$

$$P^i_{\alpha\beta} = (1-a^i)\delta_{\alpha\beta} + a^i\delta_{\alpha 1},$$

for $(i, \alpha, \beta = 0, 1, 2, \cdots, m)$, where* a^i, b^i, and c^i are in the

---

* $\delta_{\alpha\beta}$ is the well-known Kronecher delta, and is unity or zero according as α and β are or are not equal.

closed interval (0, 1). This special form of the more general Markov game-learning process was developed by Bush and Mosteller [4c] so as to fit data obtained in a number of learning experiments with rats; they have named it the "fusion model."

We shall be interested in the case in which all but one player, say number 1, choose constant strategies p^j, but where player 1 uses the fusion model; this means that the probability that player $j>1$ will select the value i for i_j is p_i^j on each play, and for player 1 it is $p_i^1(t)$. Since these choices are all made independently it follows immediately that the expectation of player 1, if he chooses the value x for i_1 on play t, is:

$$G_x = \sum_{i_2=1}^{m_2} \cdots\cdots\cdots \sum_{i_n=1}^{m_n} V_{xi_2\ldots i_n}\, p_{i_2}^2\, p_{i_3}^3 \cdots\cdots p_{i_n}^n .$$

We may suppose, without loss of generality, that:

$$G_{m_1} \geq G_{m_1-1} \geq \cdots \geq G_1 .$$

It follows that player 1 could do no better than to choose

$$p_i^1(t) = \delta_{im_1}$$

so that his expectancy on each play is G_{m_1}, and we shall be interested in comparing his expectation when he makes use of the fusion model with this maximum possible expectation G_{m_1}; of course, his success with the fusion model may also depend upon his choice of the starting vector $p^1(0)$.

The component $p_0^j(t)$ is interpreted in the game situation as the probability that no choice will be made by player j at time t,

and this is called "thinking." This feature can be introduced mathematically into the game model by defining:

$$V_i^j = \quad \text{if any component of } i \text{ is zero,}$$

where the range of i_k has been extended to include zero. This augmentation of the original game problem will be used whenever the fusion game-learning model is under discussion.

6. A special fusion model

We shall now consider a specialization of the fusion model in which a^i, b^i, and c^i are positive and independent of i; we denote their common values a, b, and c. We shall also suppose that the quantities G_x, for x=1, 2, ···, m. are distinct and non-zero, and that $G_0=0$.

The expected value of p^{t+1}, given p^t, is:

$$\left\{ \sum_{\alpha=0}^{m} p_\alpha \left(G_\alpha R^\alpha + \left[1-G_\alpha \right] P^\alpha \right) \right\} p^t .$$

This can be rewritten to yield the following relation for the k^{th} component of the expected value of p^{t+1}:

$$E(p_k^{t+1}) = p_k^t + (a-b-c) \sum_{\alpha=0}^{m} p_\alpha^t G_\alpha p_k^t + (c-a) G_k p_k^t + b \sum_{\alpha=0}^{m} p_\alpha^t G_\alpha \delta_{k0},$$

for k=0, 1, 2, ···, m.

It follows easily that a vector V^h satisfies the equation $MV=V$ if and only if it is the unit vector e_0 or has the following form:

$$V_0^h = \frac{b}{b+c-a}, \text{ and } V_i^h = (1-V_0^h)\,\delta_{hi} \text{ for } (h, i = 1, 2, \cdots, m).$$

Bush and Mosteller [4b] have called the matrix Q the "expected operator" and they have made use of the vectors V^h in discussing the asymptotic behavior of p^t.

Rather little is known concerning this asymptotic behavior, and still less is known about methods for estimating the parameters in the fusion model from experimental data, so we shall have to resort to Monte Carlo computational methods in our exploration of the properties of the fusion model. For this purpose, we shall turn to some numerical examples.

7. A rat experiment

The rat Su must choose one of two rooms. Five seconds after a warning bell the rat is either rewarded (fed) or punished (shocked) by the experimenter Ex, and Su does not see what the result would have been had Su chosen the other room on that particular trial.

Ex rewards or punishes according to a rule prescribed in advance. The experimental situation may be summarily described as a two-person game. The payoff matrix for Su is:

Su \ Ex		Places food in: Room 1	Room 2
Sits in:	Room 1	1	-y
	Room 2	-y	1

At the moment, I am not interested in the payoff matrix for Ex but shall simply suppose that Ex has chosen a strategy (π_1, π_2) such that Ex places the food in Room 1 $100\pi_1$ per cent of the time.

The expected payoff for Su is then $V(p_1)=p_1(2\pi_1-1)(1+y)+(1-\pi_1-y\pi_1)$, where y is non-negative and p_1 denotes the proportion of the time Su sits in Room 1. Now in this situation, even if Su had superior human intelligence, game theory would give Su little real help in choosing a strategy because there is no meaning to attach to the notion of payoff matrix for Ex; even if there were a payoff matrix for Ex, the game would probably be non-constant sum and the value of y would: vary from Su to Su, vary from time to time, and be difficult to estimate. Nevertheless, a rat or a human found in this situation does behave in some fashion, and our scientific problem is to explain and predict actual behavior as well as possible.

Before turning to the game-learning theory approach, it may be instructive to discuss the situation in the usual manner, from the standpoint of rationality. For example, if $\pi_1=1/2$, it does not matter what Su does, since the result is independent of his choices.* If $\pi_1\neq 1/2$, then Su should choose $p_1=1$ or $p_1=0$ according as $\pi_1 > 1/2$ or $\pi_1 < 1/2$. On the other hand, if Su feels that its past behavior (including its biological characteristics) may be analyzed intelligently by an Ex that strives to minimize the payoff to Su by choosing a time-dependent strategy in terms of past behavior of Su, then Su should somehow protect against this unwanted result by concealing its pattern of behavior from

---

* Perhaps this is a way out of the so-called "free-will dilemma." It leaves lots of room for the conscious conviction of choice with no requirement that main trends be affected thereby! From a game-theoretic standpoint this corresponds to the case in which all available pure strategies are included in a solution.

Ex (perhaps by randomization as proposed in the theory of games). These dynamic cases are entirely outside the scope of present formal game theory, of course, and are the principal cases of interest here.

The basic assumption from learning theory is that Su varies its behavior according to the pattern of its past experience. The special mathematical form assumed here for this effect is the special fusion model of § 6, with m=2. The matrices R^{11} and P^{11} are, therefore:

$$R^{11} = \begin{pmatrix} 1-c & b & b \\ c & 1-b & c \\ 0 & 0 & 1-b-c \end{pmatrix}, \quad R^{12} = \begin{pmatrix} 1-c & b & b \\ 0 & 1-b-c & 0 \\ c & c & 1-b \end{pmatrix}$$

$$P^{11} = \begin{pmatrix} 1-a & 0 & 0 \\ a & 1 & a \\ 0 & 0 & 1-a \end{pmatrix} \qquad P^{12} = \begin{pmatrix} 1-a & 0 & 0 \\ 0 & 1-a & 0 \\ a & a & 1 \end{pmatrix}$$

$$P^{10} = \begin{pmatrix} a & a & 1 \\ 0 & 1-a & 0 \\ 0 & 0 & 1-a \end{pmatrix}$$

R^1 is used after Room 1 is chosen when the food was placed there, and P^1 is used after Room 1 is chosen when the food was not placed there. Of course, Ex, as player 2, may be represented by the relations:

$$p^2(0) = \begin{pmatrix} 0 \\ \pi_1 \\ 1-\pi_1 \end{pmatrix}, \text{ and } [illegible] = P^{21} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The π_1 correspond exactly to the G_1 of § 5. In this case, Ted Harris has shown [Appendix A] that the asymptotic value of $p^1(t)$ is

$$e_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{if } 0 < \pi_1 < 1$$

whatever the value of $p^1(0)$; indeed, the probability is one that there will eventually be an unbroken sequence of applications of P^0 that terminates the process.

Monte Carlo computations were made for this model with numerical values for the parameters, chosen in agreement with estimates by Bush and Mosteller [4c] on the basis of data from learning experiments with rats, as follows:

$$a=b=d=0.01, \quad c=0.10.$$

These computations were made for a rather careless assortment of values for $p^1(0)$ and π_1, and the main results are shown in Tables 1A–1G. All the computed cases show a strong tendency for p_0 to seek an equilibrium near the value 0.1, and it is interesting also that p_2 seemed always to go to zero when $\pi_1 > 0.5$; this constitutes a tendency toward optimal behavior since only choice of Rooms 1 and 2 represent actual decisions by the rat under our interpretation of the fusion model. There is not enough data in Table 1 to permit any real analysis of the moments of the distribution of p^N.

TABLE 1

Stat–rat Strategies

1A: $\pi_1=0.5$

t	p_0	p_1	p_2
0	100	450	450
5	95	555	350
10	94	450	456
30	91	512	397
50	108	446	446

1C: $\pi_1=0.5$

t	p_0	p_1	p_2
0	900	50	50
5	895	58	47
10	900	53	46
30	900	52	48
60	692	281	25
90	355	374	271
120	193	472	335
150	111	589	299
180	139	722	39
210	97	762	141
240	98	866	36

[illegible] $\pi_1=0.5$

t	p_0	p_1	p_2
0	0	500	500
5	19	579	402
10	46	617	337
30	85	[illegible]	241
60	106	[illegible]	90
90	97	869	33
115	97	897	4

1D: $\pi_1=0.51$

t	p_0	p_1	p_2
0	800	100	100
5	717	196	87
10	717	202	80
30	398	337	265
60	337	497	165
90	148	606	247
120	116	815	69
150	98	893	9
180	115	623	262
210	98	371	31
240	88	906	6

TABLE 1 (Continued)

1E: $\pi_1 = 0.55$

t	p_0	p_1	p_2
0	800	100	100
5	721	193	86
10	656	262	81
30	411	439	151
60	239	715	45
90	118	875	7
120	98	901	1

1F: $\pi_1 = 0.9$

t	p_0	p_1	p_2
0	900	100	0
5	735	265	0
10	744	256	0
30	593	406	0
60	126	874	0
90	106	894	0
120	90	908	0
150	111	889	0
180	92	908	0
210	98	902	0
240	101	898	0

1G: $\pi_1 = 1.0$

t	p_0	p_1	p_2
0	114	100	786
5	109	96	795
10	104	179	715
30	107	716	177
60	99	892	8
90	107	892	1

8. Human subjects.

All that has been said about the game-learning model is applicable in the analysis of experimental data with human subjects. There may be a considerable advantage in using human subjects since the conditions of the experiment can be explained to them easily, and because their choices are made quite rapidly. Some very tentative trials were run in order to gain some experience with the experimental situation, as a first step toward a design of more conclusive trials.

In the first series the subject Su was asked to call "head" or "tail" in an attempt to match the random choice made by Ex with fixed probability π_1. The success of Su was compared with that of the special fusion model (stat-rat) used in §7, where random numbers were used to yield $\pi_1=0.55$. The number of trials was too small to permit any quantitative conclusion to be drawn, but it seemed likely that a more extensive series of trials would be worthwhile. The scheme was discontinued in favor of more promising ones to be discussed next.

The general 3x3 zero-sum symmetric game, which has no pure strategy as a solution, is represented by a three-parameter payoff matrix for its normal form:

$$V = \begin{pmatrix} 0 & u & -v \\ -u & 0 & w \\ v & -w & 0 \end{pmatrix}$$

where u, v, and w are positive. The solution of this game is the unique mixed strategy:

$$\frac{1}{u+v+w}\,(w,\ v,\ u).$$

Three trials were made in which a subject played a 3x3 zero-sum symmetric game against the stat-rat, as defined numerically for the special fusion model in §7, where:

(a) The absolute values for u, v, and w were taken directly from a table of random numbers,

(b) The subject was not told the payoff matrix but was told the exact method used to select it,

(c) The subject was told that he was playing against a rat in mathematical form.

The three payoff functions were:

Subject	u	v	w	Solution
RF	-6	6	8	Col. 2
MD	6	-8	7	Col. 1
MF	6	10	-4	Col. 3

These three games each have a pure strategy for a solution, as shown in the final column of the table just above. The results of the three trials are summarized in Table 2. The trend in the stat-rat's mixed strategy is shown for each game in Table 2, along with an estimate of the mixed strategy in use by the subject based on the average of his ten choices centered at the play listed. Again the data are too few to justify careful analysis, or quantitative conclusions, and this type of trial was discontinued in favor of a more promising one to be discussed next.

TABLE 2

3x3 Symmetric Games

	RP game		MD game		MF game	
	RF	Stat-rat	MD	Stat-rat	MF	Stat-rat
No. of plays	19	19	25	25	20	20
No. of wins	6	13	5	20	10	10
Percentage wins	32	68	20	80	50	50
Through Play	Frequency of Use of Solution*					
1	—	.33	—	.33	—	.33
6	.3	.4	.4	.61	.8	.55
10	.4	.58	.5	.68	1.0	.66
15	.8	.72	.6	.70	1.0	.70
20	—	.76	.7	.74	—	.80
25	—	—	—	.74	—	—

* For the stat-rat the frequency through play x is computed as the ratio of

$$p^{x}_{Sol} + (1-p^{x}_{o}),$$

and for the subjects it is one-tenth the number of wins in the ten plays centered at play x.

There are really two rather different types of problems involved thus far in our discussions of game-learning situations in which the payoff functions are unknown:

(a) (Static) Those situations in which it is assumed that the opponents of the main player choose and use a fixed mixed strategy for the duration of a sequence of plays,

(b) (Dynamic) Those situations in which it is assumed that the opponents of the main player may vary their strategic behavior during the sequence in a manner that somehow takes account of the results they obtain on earlier plays.

The static case, with known payoff functions, is the usual one considered [1] whereas our interest centers here on the dynamic case with essentially unknown payoff functions. The static case always reduces to one in which there is a set of numbers G_x, in the closed interval (0, 1), that represents the payment expectation if our main player chooses pure strategy x on a given play—and the numbers G_x remain constant for the sequence of plays; the dynamic case takes the same form except that the numbers G_x may vary in some manner that is dependent upon the choices made by our main player in preceding plays. The game-learning model is equally applicable in either the static or the dynamic case.

The game of Morra is a convenient one for our purposes both because it is of handy size (9x9) and because it has been completely solved.* [7] The static case was examined experimentally

* Morra is discussed in Appendix B.

for Morra by having two subjects and the stat-rat play the game knowing only that it was 9x9 and symmetric, and that their opponent would not be using a game-theoretic solution. Subject BC has no knowledge of game theory and Subject RB is a mathematician who is expert in game and decision theory. Actually, a fixed pure strategy, not in the solution mixture, was used in opposing the subjects and the stat-rat; it is represented by the following set of values for G_x that were used against BC and the stat-rat, those for RB being 2/3 as great:

G_x = (.500, .500, .833, .250, 250, .500, .500, .500, 1.000).

The results of play are summarized in Table 3. The data are still too skimpy to permit any conclusions to be drawn. There is no particular reason, in the static case at least, why the experimental values chosen for G_x should come from a game that has a well-known extended form. Consequently, we have come to the following type of experiment as the most promising one to use in obtaining data on the behavior of human subjects to be used in estimating parameters in the fusion models and thus eventually to test the hypothesis that this mathematical model represents human learning behavior. The G_x are chosen from a random-number table and the subject then is asked to play a number of times fixed in advance in an effort to maximize his total number of wins; RB and AM, both experts in the relevant mathematical theories, served as subjects for trials in which there were 1,000 plays and:

G_x = (.04, .03, .25, .61, .64, .33, .44, .44, .75, .41).

TABLE 3

Static Morra

	RB	BC	Stat-rat 1	Stat-rat 2
No. of plays	67*	372*	29	43
No. of wins	24	213	14	21
Percentage wins	40	57	48	49
Through Play	Frequency of Use of Solution			
1	—	—	.100	.100
5	.1	.2	.077	.186
10	.6	.1	.058	.155
20	.5	.1	.030	.260
30	—	.1	.017	.304
60	—	.1	——	——
90	—	.1	——	——
120	—	.2	——	——
150	—	.2	——	——
210	—	.2	——	——
270	—	.1	——	——
330	—	.2	——	——
360	—	.0	——	——

* The subject announced at this point that he would continue playing pure strategy No. 9 indefinitely, and so had in effect "solved" the game. RB made this tentative decision after 32 plays and used the other 3[illegible] simply to confirm his decision.

The results are given in Table 4; it seems unlikely that the stat-rat would match this performance, but it would probably take a good many trials to give a statistically significant test of this conjecture.

One static-nine game has been played by the stat-rat with ten replications. In this play, $p_\alpha^0 = 0.1$ for $\alpha = 0.1, \cdots, 9$, and the values for G_i chosen from the random-number table were

$$G_i = (.097, .510, .433, .274, .442, .364, .503, .929, .256)$$

This was the first IBM machine run and the computations were carried to eight decimal places for two hundred steps each. In nine out of the ten cases, the value of p^{200} was essentially such that $p_8^{200}=0.9$, $p_\alpha^{200}=0.1$; in other words, the stat-rat had reached the optimum strategy in 200 trials. In the tenth case the value of p^{200}, if rounded off at the third decimal place, was essentially $p_6^{200}=0.9$, $p_0^{200}=0.1$; in other words, the stat-rat had reached a very poor strategy in 200 trials. Other details of this run of static-nine are given in Table 5, including the "winning rate" w^t which represents the expectation on the first decision after time t, where:

$$w^t = \frac{\sum_{i=1}^{9} p_i^t G_i}{\sum_{i=1}^{9} p_i^t} .$$

The dynamic case is perhaps the most interesting one experimentally, especially where a fusion model is pitted against a

TABLE 4

Static-10-Game

	RB	AM
No. of plays	1,000	1,000
No. of wins*	719	715
Percentage wins	71.9	71.5
No. of last play before deciding permanently on strategy 9	133	204

* Based on expectation after decision to play strategy 9

TABLE 5

Static-9 Game

Item	Average excl. Game 9	Game Number 1	2	3	4	5	6	7	8	9	10
Time at which winning rate first exceeded:											
.84	59	41	29	58	88	98	65	72	58	—	26
.90	81	97	41	88	99	111	76	92	68	—	36
Winning rate at time:											
50	70	87	84	73	44	58	62	59	74	36	92
100	91	90	93	92	90	82	93	91	93	36	93
200	93	93	93	93	93	93	93	93	93	36	93
Percentage wins in:											
100 decisions	71	63	85	72	57	57	75	65	75	38	86
200 decisions	82	80	89	83	75	75	83	80	83	38	91
p_o^{100}	.100	.106	.091	.092	.106	.101	.093	.104	.096	.095	.110
p_o^{200}	.097	.090	.100	.092	.105	.105	.100	.092	.091	.093	.098
No. of thinking steps for 200 decisions:	23	23	25	15	29	27	18	22	21	14	29

human subject. Before going too far with such a program, it will be necessary to develop a better mathematical understanding of the models in order to design the experiments so as to permit statistical significance tests to be applied; this point has been discussed by Bush and Mosteller [4b], and they and others are gradually developing some of the mathematical tools that are needed. We have some of these experiments under way with human subjects, using Morra and other games of about this complexity for the purpose. It would be interesting also to run some trials of exactly the same sort with real rats (e.g., playing Morra against the stat-rat).

9. Against the Stat-rat

We have seen how the stat-rat is able to play any game with known bounds on the payments, even though we have not been able to settle the question concerning its degree of skill at games. We shall now be interested in how best to exploit this knowledge of the procedure used by the stat-rat in playing games when we are its opponent. Of course, if we know the expectation functions and have computed the solution of the game, then we can guarantee at least a certain minimum result by choosing the game-theoretic solution; our object is to do better than this safe solution guarantees, and we also should like to know how to play when we do not know the expectation functions or the theoretical solution.

As a special case, consider the ordinary game of matching pennies. We start with the reasonable assumption that the initial vector for the stat-rat is:

$$p^1(0) = \begin{pmatrix} .45 \\ .45 \\ .1 \end{pmatrix} .$$

To make the game quite definite, in our usual notation, we note that the game is usually represented by the expectation functions:

$$V^1_{i_1 i_2} = 2\delta_{i_1 i_2} - 1, \text{ and } V^2_{i_1 i_2} = 1 - 2\delta_{i_1 i_2}$$

We transform this game into an equivalent one, in the sense that linear transformations on the individual utility functions leave the solutions invariant, by setting:

$$V^1_{i_1 i_2} = \frac{1}{2}(V^1_{i_1 i_2} + 1) = \delta_{i_1 i_2}, \text{ and}$$

$$V^2_{i_1 i_2} = \frac{1}{2}(V^2_{i_1 i_2} + 1) = 1 - \delta_{i_1 i_2}.$$

No chance move is really needed, in this special case, since the values of $V^j_{i_1 i_2}$ are all zero or one. Finally, we specify that there is to be a sequence of N plays, and our problem is to choose a method of play that will maximize our expected payments against the stat-rat. Since we can compute the $p^1(t)$ for the stat-rat at each stage, except for the thinking steps when p^1_0 has effect, it is not difficult to find a method of play that gives us an average expectation in excess of that obtained if we play strategies 1 and 2 with equal frequencies. Such a good strategy would be for us always to play the strategy that is less likely to be chosen by the stat-rat. I shall not pursue this very simple example further, except to note that it becomes immediately more difficult if we do not know $p^1(0)$ or if the expectation functions are represented in the equivalent form:

$$\Theta V^j_{i_1 i_2} \quad \text{where } 0 < \Theta < 1.$$

10. A modified fusion model

It is inconvenient to have the situation met in the special fusion model, where there is at [illegible] a position [illegible] bounded away from zero so that the [illegible] tually be applied in an unbroken sequence. To avoid this difficulty, and some others, we shall consider the following slight modification in the special fusion model:

(a) The operator P^0 is never applied.

(b) The probability of selecting pure strategy i at time t is $p_i^t \Big/ \sum_{j=1}^{m} p_j^t$, for i=1, 2, ···, m.

The expected value of the payment on one trial, starting with p^t, is:

$$S_1(p^t) = \frac{\sum_{j=1}^{m} p_j^t G_j}{\sum_{j=1}^{m} p_j^t} \equiv \frac{A_1}{B} \text{ , where } A_k \equiv \sum_{j=1}^{m} p_j^t G_j^k .$$

We shall now be interested in comparing this expectation on the first trial with the corresponding expectation $S_2(p^t)$ on the second trial. For this, we have:

$$S_2(p^t) = \frac{1}{B} \sum_{i=1}^{m} \left[p_i^t G_i S_1(R^i p^t) + p_i^t (1-G_i) S_1(P^i p^t) \right] .$$

If e_α denotes the column vector with unity as its α^{th} component and zeros elsewhere, and where a prime is always used to denote the transpose of a matrix (or vector), then:

$$S_1(R^i p^t) = \frac{\sum_{\alpha=1}^{m} e'_\alpha R^i p^t q_\alpha}{\sum_{\alpha=1}^{m} e'_\alpha R^i p^t} = \frac{\sum_{\alpha=1}^{m} \left[r p_\alpha^t + c\delta_{\alpha 1} + b\delta_{\alpha o} \right] q_\alpha}{\sum_{\alpha=1}^{m} \left[r p_\alpha^t + c\delta_{\alpha 1} + b\delta_{\alpha o} \right]}$$

$$= \frac{rA_1 + cG_1}{rB + c},$$

and

$$S_1(P^i p^t) = \frac{uA_1 + aG_1}{uB + a},$$

where r=1-b-c and u=1-a. So

$$S_2(p^t) = \frac{1}{B} \sum_i \left\{ p_i^t G_i \left[\frac{rA_1 + cG_i}{rB + c} \right] + p_i^t (1-G_i) \left[\frac{uA_1 + aG_i}{uB + a} \right] \right\},$$

and, after reduction, this becomes

$$S_2(p^t) = \frac{A_1}{B} + \frac{(ra-uc)(A_1^2 - BA_2)}{B(rB+c)(uB+a)}.$$

The quantities in the denominator are all positive and so the algebraic sign of the second term depends on its numerator only, in which (ra-uc) depends only on the constant parameters. We note

$$A_1^2 - BA_2 = \left(\sum_{i=1}^{n} p_i^t G_i \right)^2 - \left(\sum_{i=1}^{n} p_i^t \right) \left(\sum_{i=1}^{n} p_i^t G_i^2 \right) = -\sum_{i<j} p_i p_j (G_i - G_j)^2 \leq 0.$$

It follows that

$$S_2(p^t) \geq S_1(p^t) \quad \text{if} \quad (ra-uc) \leq 0.$$

This is an important feature for our game-learning model to have in order that the expectation increase with each play in the static case.

It is obvious that

$$S_i(p^t) \leq G_{max} \equiv G_m.$$

It is likely that p^t can converge asymptotically only to a vector of the form $V^\alpha = \theta e_\alpha + (1-\theta)e_o$, where $0 \leq \theta \leq 1$. One important unsolved problem is to find the probability that p^t converge to V^α when the initial vector p^o is given; it is reasonable to hope that the probability is high that p^t converges to V^m when $p_i^o = 1/m+1$, in which case the modified special fusion model represents a game-learning process that tends asymptotically to find the optimal pure strategy in playing a static game. If these conjectures prove to be well founded, as our Monte Carlo Computations with the special fusion model seem to indicate they may be, there will still be some questions in the degenerate cases in which the G_i are not all distinct or some of the p_i^o are taken to be zero.

11. Summary

This is a very preliminary paper. In it we have shown how a player can "learn" during the course of a sequence of plays of a game to improve his strategy. The fusion model developed by Bush and Mosteller to explain observed behavior of rats in experimental learning situations was used as the basis for both a theoretical and experimental investigation of the efficiency of this type of learning process in learning to play games; the experiments discussed here were with human subjects, and their game-learning performance was compared with that of the "stat-rat"

represented by the fusion model with numerical values of the parameters estimated to fit experimental data for rats.

The theoretical models accept basic assumptions of vonNeumann-Morgenstern game theory and Bush-Mosteller learning theory, including:

(a) Games with identical normal forms are equivalent, and this equivalence is independent of the probability distribution functions associated with chance moves.

(b) Games that differ only by linear transformations of the individual payoff functions are equivalent.

(c) Learning is a Markov process.

Equivalence here means that the games have the same solutions.

The experimental results consist of Monte Carlo computations for the stat-rat, contests between stat-rat and a human subject, and comparisons of performance of stat-rat and a human subject when playing the same static game. Very limited data indicate that:

(a) The stat-rat usually learns a good strategy when a constant mixed-strategy is played against him; in Morra and the other games played the stat-rat seemed to settle on essentially the best strategy within 200 trials or so.

(b) A person proficient at games would win against the stat-rat in Morra.

(c) The stat-rat does reasonably well in a static game, in comparison with the human subject, but a statistician would certainly defeat the stat-rat.

The theoretical results are very skimpy, the main result being that one modified fusion model does have a non-decreasing expectancy per trial on successive plays of the game; various open mathematical questions are noted.

More extensive experiments are in progress, and it is hoped that these may provide the data necessary to estimate parameter values for human subjects and eventually to test the adequacy of this type of Markov process for description of human learning. It seems very unlikely now that such a Markov process will be adequate.

APPENDIX A

An Asymptotic Case

Bush and Mosteller [4c] have proposed a mathematical model for learning that fits experimental data for rats quite well; they have called this the "fusion model." I shall discuss only the special three-choice case; the argument is easily extended to their general case.

Define five matrices as follows:

$$M^1 = \begin{pmatrix} 1-b & c & c \\ 0 & 1-b-c & 0 \\ b & b & 1-c \end{pmatrix}, \qquad M^2 = \begin{pmatrix} 1-b-c & 0 & 0 \\ c & 1-b & c \\ b & b & 1-c \end{pmatrix}$$

$$M^3 = \begin{pmatrix} 1 & a & a \\ 0 & 1-a & 0 \\ 0 & 0 & 1-a \end{pmatrix}, \qquad M^4 = \begin{pmatrix} 1-a & 0 & 0 \\ a & 1 & a \\ 0 & 0 & 1-a \end{pmatrix}$$

$$M^5 = \begin{pmatrix} 1-d & 0 & 0 \\ 0 & 1-d & 0 \\ d & d & 1 \end{pmatrix}$$

Consider the Markov chain $p^{t+1} = M^{\alpha} p^t$, where the probability that the transition matrix M^{α} is applied at step $t \geq 1$ is $q_{\alpha}^t = q_{\alpha}(p^t)$, and where:

$$q_1^t = r_1 p_1^t, \quad q_2^t = r_2 p_2^t, \quad q_3^t = r_2 p_1^t, \quad q_4^t = r_1 p_2^t, \quad q_5^t = p_3^t .$$

The quantities a, b, c, d, r_1, and r_2 are given real numbers in the

open interval (0,1). The p_i^t are the three components of the vector p^t, and satisfy the conditions:

$$0 \leq p_i^t \leq 1, \quad \sum_{i=1}^{3} p_i^t = 1.$$

If we let $u_n(M^\alpha)$ denote the probability that the matrix M^α is applied after all $t \geq n$ then we say that the process "concludes with the matrix M^α," if $\lim_{n \to \infty} u_n(M^\alpha) = 1$.

Theorem 1.*

The process $p^{t+1} = M^\alpha p^t$ concludes with the matrix M^5.

Proof: The probability that some one of the three operators M^1, M^2, or M^5 will be applied after play t is:

$$\pi^t = \pi_1 p_1^t + \pi_2 p_2^t + p_3^t \geq \eta > 0,$$

where

$$\eta = \min \{\pi_1, \pi_2\}.$$

The operators M^1, M^2, and M^5, when applied to p^t, yield components p_3^{t+1} as follows:

$$M^1: \quad p_3^{t+1} = (1-b-c)p_3^t + b,$$

$$M^2: \quad p_3^{t+1} = (1-b-c)p_3^t + b,$$

$$M^5: \quad p_3^{t+1} = (1-d)\, p_3^t + d.$$

Thus, in all three cases, we have:

$$p_3^{t+1} \geq \min \{b, 1-c, d\} = \lambda > 0.$$

---

* This theorem and its proof are due to Ted Harris (oral communication).

The probability $f_n(p_s^t)$ that operator M^5 is applied after times t, t+1, ···, t+n-1, is:

$$f_n(p_s^t) = p_s^t \prod_{i=1}^{n-1} \left[1-(1-p_s^t)(1-d)^i\right] .$$

To establish this relationship, note that repeated application of M^5 for n times immediately after time t yields the relation:

$$p_s^{t+n} = u^n p_s^t + (1-u^n)$$

where

$$u \equiv 1-d.$$

Next, note that the probability that M^5 occurs at least n times immediately after time t is, as required:

$$f_n(p_s^t) = \prod_{i=0}^{n-1} p_s^{t+i} .$$

We will be interested in the limiting case as $n \longrightarrow \infty$, and next consider:

$$f(p_s^t) = \lim_{n \longrightarrow \infty} f_n(p_s^t)$$

It follows that

$$f_n(p_s^t) \geq f_n(x), \text{ if } p_s^t \geq x,$$

since each term in the product $f_n(x)$ is less than or equal to the corresponding term in the product $f_n(p_s^t)$. In particular, $f(p_s^t) \geq f(\lambda)$ if p^t was obtained by application of M^1, M^2, or M^5. It is easily seen that the secuence $f_n(\lambda)$ is convergent, and so

$$1 \geq f(\lambda) > 0.$$

We now have shown that the probability $g(p^t)$ that there will be an unbroken sequence of applications of M^5 after time t is not less than

$$g \equiv nf(\lambda) > 0.$$

We next obtain the limiting value of the probability that there will be an unbroken sequence of applications of M^5, starting somewhere within n steps after t=0.

For convenience, let B^t be the generic name for the matrix operator on p^t. Let $t_1 < t_2 < t_3 < \cdots\cdots t_{\varphi(N)}$ be all the values of $t \leq N$ for which $B^t = M^5$ and $B^{t+1} \neq M_5$; the process does or does not conclude with M^5 according as $\varphi = \lim_{N \to \infty} \varphi(N)$ is or is not finite, and we say that the process has φ "breaks." If we let $P(b,v|p^o)$ be the probability that the process has at least b breaks and also has $p^{t_b} = v$, and if we let $P(b|p^o)$ be the probability that the process has at least b breaks, then:

$$P(b+1|p_o) = \sum_v P(b,v|p^o)(1-P(0|v)).$$

Now, since

$$\sum_v P(b,v|p^o) = P(b|p^o) \text{ and } P(0|v) \geq g > 0,$$

it follows that

$$P(b+1|p^o) \leq (1-g)P(b|p^o).$$

Hence $\lim_{b \to \infty} P(b|p^o) = 0$, and the theorem follows.

APPENDIX B

Morra

1. Game of "Morra"

"Morra" is an example of a game involving only personal moves. Each player shows one, two, or three fingers and simultaneously calls his guess as to the number of fingers his opponent will show. If just one player guesses correctly, he wins an amount equal to the sum of the fingers shown by himself and his opponent; otherwise the game is a draw.

This game consists of one move for each player. A strategy for each player is a pair of numbers (s,g), where $s=1,2,3$ is the number of fingers he shows and $g=1,2,3$ is his guess of the number of fingers his opponent will show. It is evident that each player has nine strategies, and thus there are 81 different possible plays of the game. With each of these 81 ways of playing the game, there is associated a payment to the players, as described by the rules of the game. These payments are summarized by a payoff matrix. In the following payoff matrix, the entries represent payments to player I. Player II will receive the negative of these payments.

"Morra"

RECEIPTS OF PLAYER I

		Player II's Strategies								
		(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
Player I's Strategies	(1,1)	0	2	2	-3	0	0	-4	0	0
	(1,2)	-2	0	0	0	3	3	-4	0	0
	(1,3)	-2	0	0	-3	0	0	0	4	4
	(2,1)	3	0	3	0	-4	0	0	-5	0
	(2,2)	0	-3	0	4	0	r	0	-5	0
	(2,3)	0	-3	0	0	-4	0	5	0	5
	(3,1)	4	4	0	0	0	-5	0	0	-6
	(3,2)	0	0	-4	5	5	0	0	0	-6
	(3,3)	0	0	-4	0	0	-5	6	6	0

2. **Solutions.**

The four basic solutions of Morra are:

(a) $\| 0, 0, \frac{5}{12}, 0, \frac{4}{12}, 0, \frac{3}{12}, 0, 0 \|$

(b) $\| 0, 0, \frac{16}{32}, 0, \frac{12}{32}, 0, \frac{9}{32}, 0, 0 \|$

(c) $\| 0, 0, \frac{20}{47}, 0, \frac{15}{47}, 0, \frac{12}{47}, 0,00 \|$

(d) $\| 0, 0 \frac{25}{61}, 0, \frac{20}{61}, 0, \frac{16}{61}, 0, 0 \|$

The optimal strategy is:

(e) $\| 0, 0, \frac{23}{55}, 0, \frac{18}{55}, 0, \frac{14}{55}, 0, 0 \|$

If player I uses the optimal strategy (e), he can expect to gain at least 2/55 whenever player II departs from strategies 3, 5, or 7.

BIBLIOGRAPHY

[1] von Neumann, John, and Morgenstern, Oskar. Theory of Games and Economic Behavior. Princeton, Princeton University Press, 1944.

[2] Flood, M. M. Some Experimental Games. RM-789, The RAND Corporation, 15 March 1952.

[3]

a. Bales, R. F., Flood, M. M. and Householder, A. S. Some Group Interaction Models. [illegible]953, The RAND Corporation, 10 October 1952.

b. Bales, Robert F. Interaction Process Analysis, Cambridge, Addison Wesley Press, 1951.

[4]

a. Bush, R. R., and Mosteller, C. F. "A Mathematical Model for Simple Learning," Psych. Rev. (58, 5) September, 1951.

b. ———————————. "A Linear Operator Model for Learning," a paper presented on 27 December 1951 at a meeting of the Institute of Mathematical Statistics.

c. ———————————. "A Comparative Study of Three Models for Simple Learning," Memorandum VI, 21 December 1950 (Unpublished).

[5] MacKay, D. M. "Mindlike Behavior in Artefacts," The British Journal for the Philosophy of Science (2, 6) August, 1951, pp. 105-121.

[6] Boring, E. G. "Mind and Mechanism," Amer. J. Psychology (59, 2) April, 1946.

[7] Skinner, B. F. The Behavior of Organisms. Appleton-Century-Crofts, 1938.

[8]

a. Haywood, O. G., Jr. Military Doctrine of Decision and the von Neumann Theory of Games. RM-528, The RAND Corporation, 20 March 1950.

b. ———————————. "Military Decision and the Mathematical Theory of Games," Air Univ. Quarterly Review (4, 1) Summer 1949, pp. 17-30.

[9] Dresher, Melvin. Theory and Applications of Games of Strategy. (Unpublished) The RAND Corporation,