GENERAL THREE-STATE MARKOV LEARNING MODELS

by

Gordon H. Bower

TECHNICAL REPORT NO. 41

September 26, 1961

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

Applied Mathematics and Statistics Laboratories

STANFORD UNIVERSITY

Stanford, California

GENERAL THREE-STATE MARKOV LEARNING MODELS

by

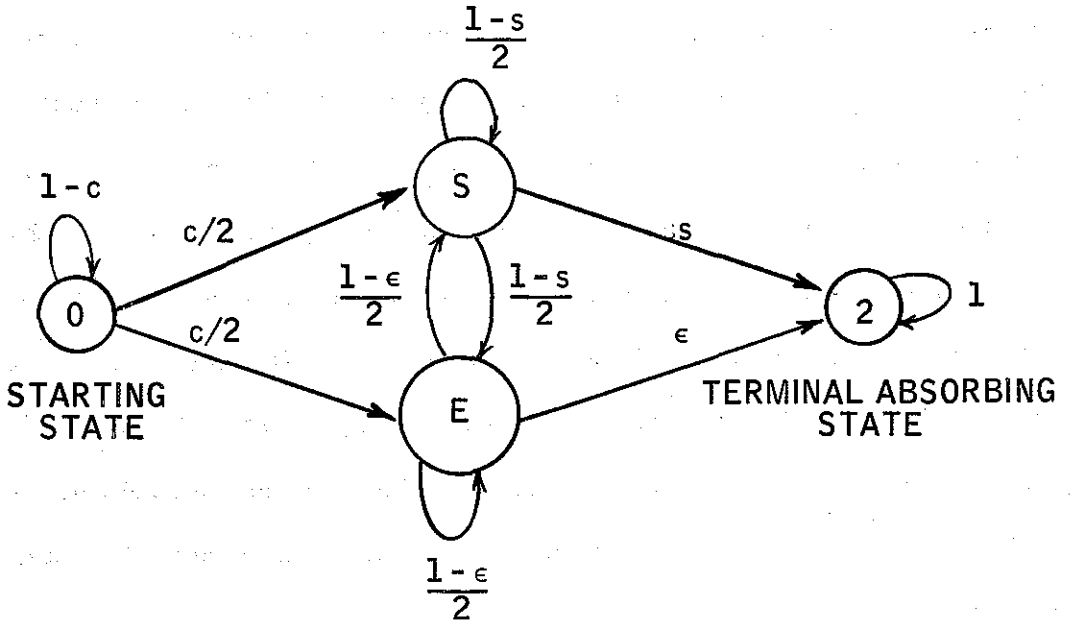
Gordon H. Bower

This note is a sequel to Theios' technical report [4] "A three state Markov model for simple avoidance learning in rats," and we will assume the reader's familiarity with that paper. Here it will be shown that the stimulus sampling model developed by Theios is the best possible three-state model for the extensive experimental results he reported. We do this by first developing the general class of three-state Markov models having certain qualitative features, and then showing that the only sensible parameter values for the Theios data are precisely those dictated by stimulus sampling theory. This is a powerful result since it is rare indeed that one can demonstrate that a particular model is the best possible out of a large set of alternative models.

The three states of the Markov model refer to the three possible values (0, 1/2, and 1) of the probability of the avoidance response. All rats begin at 0 and eventually end up at 1 after going through the intermediate state with response probability at 1/2. The evidence for the intermediate value of 1/2 is direct: Analysis of those trials between the first success and the last failure shows that the response sequences may be characterized as an independent Bernoulli trials process with probability 1/2 of a successful avoidance response.

Because of this striking qualitative feature of the data, we restrict ourselves here to three state Markov models with response probabilities of 0, 1/2, and 1. Any alternative model not having this qualitative feature simply will not give a realistic account of the data.

The random walk process envisaged by these models is shown in figure 1.



The animal begins in state 0, taking one step per trial, each step with the indicated probabilities. States S and E are the intermediate states representing a success and failure (error), respectively. The stimulus sampling model developed by Theios dictates that $s = 0$ and $\epsilon = c$. Here we investigate the general case in which c, s, and ϵ are unrestricted.

First, we find quantities $w_{i,n}$ representing the probability that the system is in state i at the beginning on trial n given that it started off in state 0. Recursive equations for the $w_{i,n}$ are

$$
\begin{aligned}
w_{o,n+1} &= (1-c)w_{o,n} \\
w_{S,n+1} &= \frac{(1-s)}{2}w_{S,n} + \frac{(1-\epsilon)}{2}w_{E,n} + \frac{c}{2}w_{o,n} \\
w_{E,n+1} &= \frac{(1-s)}{2}w_{S,n} + \frac{(1-\epsilon)}{2}w_{E,n} + \frac{c}{2}w_{o,n} \\
w_{2,n+1} &= w_{2,n} + sw_{S,n} + \epsilon w_{E,n}
\end{aligned}
\tag{1}
$$

the solution of the first equation for $w_{o,n}$ gives

$$w_{o,n} = (1-c)^{n-1} . \tag{2}$$

The second and third equations in (1) imply that $w_{S,n} = w_{E,n}$ for all n. Solve one of these equations for, say, $w_{E,n}$.

$$w_{E,n+1} = \frac{(2-s-\epsilon)}{2}w_{E,n} + \frac{c}{2}w_{o,n} . \tag{3}$$

Substitution into (3) of our result in (2) yields, after some simplification,

$$w_{E,n} = \frac{c}{2(\theta-c)}\left[(1-c)^{n-1} - (1-\theta)^{n-1}\right] \tag{4}$$

where $\theta = \frac{s+\epsilon}{2}$ is the average learning rate in the intermediate state.

We define for each subject a sequence of response random variables, x_n, taking the value 1 if an error (non avoidance) occurs on trial n and taking the value 0 if a success occurs on trial n. We let q_n represent the average probability of an error on trial n; it is

$$q_n = P(x_n = 1) = w_{o,n} + w_{E,n} \cdot \tag{5}$$

We define $T = \sum_{n=1}^{\infty} x_n$ as the total errors to the criterion of learning; in the Theios data the criterion was 20 consecutive successes (avoidance responses). The mean total errors (shocks) will be

$$E(T) = \sum_{n=1}^{\infty} q_n = \frac{1}{c} + \frac{1}{s+\epsilon} \tag{6}$$

We define J_k to be the number of errors between the $k^{\underline{th}}$ and $(k+1)^{\underline{st}}$ success, and let J_o be the errors before the first success. The distribution of J_o is

$$P(J_o = i) = \frac{c}{2}(i-c)^{i-1} + \frac{c(1+\epsilon)}{4\left[\frac{1-\epsilon}{2} - (1-c)\right]}\left[\left(\frac{1-\epsilon}{2}\right)^{i-1} - (1-c)^{i-1}\right] \tag{7}$$

(for $i \geq 1$).

having mean

$$E(J_o) = \frac{1}{c} + \frac{1}{1+\epsilon} \tag{8}$$

To obtain the distribution of J_k we require g_k, the probability that the k^{th} success occurs with the subject in the intermediate S state (rather than in state 2). The results here are

$$g_1 = 1 - \sum_{i=0}^{\infty} \frac{\epsilon}{2} \left(\frac{1-\epsilon}{2}\right)^i = \frac{1}{1+\epsilon}$$

(9)

$$g_k = g_{k-1} \left(\frac{1-s}{2}\right) \sum_{i=0}^{\infty} \left(\frac{i-\epsilon}{2}\right)^i = \frac{(1-s)}{1+\epsilon} g_{k-1} \ .$$

The solution of this last difference equation for g_k is

$$g_k = \frac{1}{1-s} \left(\frac{1-s}{1+\epsilon}\right)^k \qquad \text{for} \quad k = 1, 2, \ldots \tag{10}$$

This quantity enters importantly into the distribution of J_k since errors can occur following the k^{th} success only if the k^{th} success occurred in state S.

We can write the distribution of J_k as

$$P(J_k = i) = \begin{cases} 1 - g_k\left(\frac{1-s}{2}\right) & \text{for} \quad i = 0 \\ g_k\left(\frac{1-s}{2}\right)\left(\frac{1-\epsilon}{2}\right)^{i-1} \left(\epsilon + \frac{1-\epsilon}{2}\right) & \text{for} \quad i \geq 1 \end{cases} \tag{11}$$

having mean value

$$E(J_k) = \left(\frac{1-s}{1+\epsilon}\right) g_k = \left(\frac{1}{(1-s)}\right)\left(\frac{1-s}{1+\epsilon}\right)^{k+1} \tag{12}$$

We then define L as the number of successes occurring before the last failure. The distribution of L is

$$P(L = i) = \begin{cases} 1 - \dfrac{(1-s)}{(1+s)(1+\epsilon)} & \text{for} \quad i = 0 \\ \dfrac{\epsilon+s}{(1+s)(1+\epsilon)} \left[\dfrac{1-s}{1+\epsilon}\right]^i & \text{for} \quad i \geq 1 \end{cases} \tag{13}$$

having mean value

$$E(L) = \frac{1-s}{(1+s)(\epsilon+s)} \tag{14}$$

We now seek the distribution of n', the trial of the last failure. For these purposes it is convenient to first find f, the probability that the subject makes no further errors after he starts on some particular trial in state S. This probability is calculated as follows:

$$f = s \sum_{j=0}^{\infty} \left(\frac{1-s}{2}\right)^j = \frac{2s}{1+s} \tag{15}$$

The probability that the subject makes his last error on trial k is

$$P(n' = k) = w_{o,k} \frac{c}{2} f + w_{E,k} \left[\epsilon + \frac{(1-\epsilon)}{2} f\right] \tag{16}$$

where $w_{o,k}$ and $w_{E,k}$ were defined in equations (2) and (4). The mean trial of the last failure will be

(17) $$E(n') = \frac{1}{c}\,\frac{2}{(1+s)(\epsilon+s)}$$

The distribution of total errors, T, is given by

(18) $$P(T=k) = (1-c)^{k-1}\,\frac{cf}{2} + \sum_{i=1}^{k-1} (1-c)^{i-1}\,\frac{c}{2} \sum_{x=0}^{\infty} \binom{k-i-1+x}{x} \left(\frac{1-\epsilon}{2}\right)^{k-i-1} \left(\frac{1-s}{2}\right)^{x} \left[\epsilon + \frac{(1-\epsilon)f}{2}\right]$$

The first term is the probability of making all k errors in state 0, then moving to state S and having no subsequent errors. The second term is the sum over $i \leq k-1$ and all x of the joint probability of (a) getting i errors in state 0, then moving to state E or S, (b) getting k-i-1 errors and x successes in the intermediate state, and (c) the $k^{\underline{th}}$ error is followed by no more errors. The series in equation (18) can be solved to give

(19) $$P(T=k) = (1-c)^{k-1}\,\frac{cf}{2} + \frac{c(\epsilon+s)}{(1+s)\,[c-\epsilon-s(1-c)]} \left[\left(\frac{1-\epsilon}{1+s}\right)^{k-1} - (1-c)^{k-1}\right]$$

for $k \geq 1$

The mean value of this random variable is

(20) $$E(T) = \frac{1}{c} + \frac{1}{\epsilon+s}$$

which is the same as that derived in equation (6).

Sequential features of the data may be examined in terms of runs and autocorrelations of errors. To obtain error-run statistics, it is useful to first work with j-tuples of errors. Formally define $u_{j,n}$ as

$$u_{j,n} = x_n \, x_{n+1} \cdots x_{n+j-1} \, .$$

To have $u_{j,n}=1$ we can either (a) be in state E on trial n and stay there for j-1 trials, or (b) start in state 0 on trial n and either stay there all j-1 trials or else move to state E at some point and stay in state E for the remainder of the j-1 trials. The probability of event (a) is

$$w_{E,n} \left(\frac{1-\epsilon}{2}\right)^{j-1}$$

the probability of event (b) is

$$w_{0,n} \left[(1-c)^{j-1} + \sum_{i=1}^{j-1} (1-c)^{i-1} \, \frac{c}{2} \left(\frac{1-\epsilon}{2}\right)^{j-1-i}\right] .$$

The expectation of $u_{j,n}$ is the sum of these two terms. If we now define u_j as the sum over all trial of $u_{j,n}$, it has expectation

$$(21) \quad E(u_j) \; E\left[\sum_{n=1}^{\infty} u_{j,n}\right] = \frac{(1+\epsilon-c)}{1+\epsilon-2c} (1-c)^{j-1} + \frac{(1-s-2c)}{(\epsilon+s)(1+\epsilon-2c)} \left(\frac{1-\epsilon}{2}\right)^{j-1} .$$

The u_j values are related to error runs of length j (r_j) by the expressions

$$r_j = u_j - 2u_{j+1} + u_{j+2}$$

(22)

$$R = \sum_{j=1}^{\infty} r_j = u_1 - u_2 \ .$$

We define a statistic $c_{k,n} = x_n x_{n+k}$ which counts errors occurring k trials apart without regard to the intervening responses. The expectation of $c_{k,n}$ is

$$(23) \qquad E(c_{k,n}) = w_{o,n}\left[w_{o,k+1} + w_{E,k+1}\right] + w_{E,n}(1-\epsilon)\left(\frac{1-\theta}{2}\right)^{k-1} \ .$$

The last term in this expression is the probability of being in state **E** on trial n, not becoming conditioned on that trial, staying in the intermediate states for $k-1$ trials, then ending on trial $n+k$ in state **E** with probability $1/2$. The expectation of the overall trial sum of $c_{k,n}$ is

$$(24) \quad c_k = E\left[\sum_{n=1}^{\infty} c_{k,n}\right] = \frac{1}{c}\left[w_{o,k+1} + w_{E,k+1}\right] + \left(\frac{1}{\epsilon+s}\right)\left[\frac{1-\epsilon}{2}(1-\theta)^{k-1}\right].$$

Analogous to c_k, we may derive a statistic representing the number of times a success is followed by an error k trials later. Define $d_{k,n} = (1-x_n)x_{n+k}$ which counts a success then error k trials later. The expectation of $d_{k,n}$ is

$$(25) \qquad E(d_{k,n}) = w_{S,n}\left(\frac{1-s}{2}\right)(1-\theta)^{k-1}$$

the mean of the overall trial sum of $d_{k,n}$ is

(26) $$d_k = E \sum_{n=1}^{\infty} d_{k,n} = \frac{(1-s)}{2(\epsilon+s)}(1-\theta)^{k-1} .$$

II Estimation of Parameters.

Because of the complexity of the model and especially because the states of the Markov chain are not observable, maximum likelihood estimates of the parameters can not be found. The alternative estimation scheme used here is the method of moments; this consists of taking some moment (usually the mean) of a random variable (e.g., total errors), equating the observed value to the parametric function derived from the model, and then solving for the unknown parameters. In the general model considered here, there are three parameters; therefore, three different data statistics will be required for estimation purposes. These estimation statistics must be chosen with some care to obtain some natural separation in the parameters. For these reasons, the three data statistics used were $E(T)$, c_1, and d_1. The equations derived from the model for c_1 and d_1 are

(27)
$$c_1 = \frac{1-c}{c} + \frac{1-s}{2(\epsilon+s)}$$
$$d_1 = \frac{1-s}{2(\epsilon+s)} .$$

Inspection of these equations reveals why they were chosen for estimation purposes, since their difference is only a function of c. Therefore, our c estimate is

(28) $$\hat{c} = \frac{1}{1+c_1-d_1}$$

In the Theios data, $c_1 = 2.50$ and $d_1 = 1.18$. Therefore, $\hat{c} = .431$. Next we use $E(T)$ to estimate the sum of $\epsilon + s$, viz.,

$$E(T) = \frac{1}{c} + \frac{1}{\epsilon+s} \quad \text{or,} \tag{29}$$

$$\epsilon + s \triangleq \frac{1}{E(T) - \frac{1}{\hat{c}}} .$$

The value of $E(T)$ was 4.68 in the data, yielding an estimate of $\epsilon + s \triangleq .424$. Finally, this estimate of $\epsilon + s$ is used in d_1 to obtain an estimate of s.

$$d_1 = \frac{1-s}{2(\epsilon+s)} \quad \text{or,} \tag{30}$$

$$\hat{s} = 1-2d_1(\epsilon+s) = 1-.848\ d_1$$

Substituting the observed value of $d_1 = 1.18$, the estimate is $\hat{s} = 0$. Therefore, the three parameter estimates are

$$\begin{aligned} \hat{c} &= .431 \\ \hat{\epsilon} &= .424 \\ \hat{s} &= 0 \end{aligned} \tag{31}$$

Other estimation procedures have been tried, but they gave s estimates that were either impossible (e.g., -.004) or zero.

The impressive feature of the estimates in (31) is that they are practically identical with the values obtained by the two-element

sampling model Theios applied to these data. The sampling theory implies that $s = 0$ and that $c = \epsilon$, and c estimate for the sampling model was .427 for these data. Our general estimates here of $c = .431$ and $\epsilon = .424$ have a mean of .427 and the small difference between the estimates can safely be attributed to sampling and rounding error. Hence, we have shown that the stimulus sampling model is the best possible three-state model for Theios' data on simple avoidance learning.

III Some General Considerations.

The fact that $s = 0$ in these data indicates that the rats effectively were not learning on successful avoidance trials, where "learning" is defined here in terms of moving into the terminal, fully conditioned state. This finding provides no support for anxiety-reduction theories according to which avoidance responses are reinforced and strengthened by the reduction in anxiety ensuing when the anticipatory response terminates the warning signal. Rather the results support the view that the shock (UCS) is serving as the effective reinforcing agent; the performance of avoidance responses results from learning that occurs on shock trials, but these avoidance responses per se do not conbribute to the maintenance of avoidance responding.

The avoidance procedure is characterized by the fact that the UCS is omitted when the anticipatory response occurs; therefore, it provides no information concerning the possible learning effect of giving the UCS on trials when the response occurs. Under special circumstances, it is likely that s will be greater than zero for such trials. The special circumstances, of course, are that we insure

that the response evoked by the UCS is compatible with the anticipatory (avoidance) response. Sheffield [3] has shown the importance of this factor in an experiment on conditioning of a running response in guinea pigs using an unavoidable shock following a two-second warning signal. With this procedure, there were many occasions on which the animal made an anticipatory response to the signal and then was shocked. By careful analysis of the animal's response to shock, Sheffield was able to show that if the response evoked by shock was further running, then the probability of an anticipatory run to the warning signal on the next trial was substantially higher; contrariwise, if the shock evoked responses incompatible with running, then the probability of an anticipatory run was lower on the next trial.

If the experimental procedure insures that the UCS invariably evokes a response compatible with the class of responses being measured as anticipatory CRs, then superior conditioning should result when the UCS is applied on every trial instead of being omitted on response trials. Within the model, on trials when the response occurs and the UCS is omitted, the conditioning parameter, s, is zero; however, if the UCS is applied following a response, then s will be greater than zero and some learning may occur on such trials. The eyelid conditioning situation is well-suited for this comparison since the UCS (a puff of air to the cornea) invariably evokes a response (lid closure) compatible with the anticipatory CR. Two experiments on eyelid conditioning, one by Logan [2] and one by Kimble, Mann, and Dufort [1], have confirmed this prediction that nonavoidance training leads to better performance than does avoidance training.

IV Two element model with biased sampling.

Suppose within the two-element model developed by Theios that we permit the stimuli to vary in their average sampling probability. Let a and b be the respective sampling probabilities for the two elements, where $a + b = 1$ and we impose the restriction that exactly one element is sampled on each trial. Theios assumed that $a = b = 1/2$ but we investigate here the implications of relaxing this restriction on the sampling probabilities. The conditioning assumption remains the same, namely, that the sampled element becomes conditioned to the reinforced response with probability c if it is not already conditioned.

The average probability of an error on trial n for this general model will be

$$q_n = a(1-ac)^{n-1} + b(1-bc)^{n-1} . \tag{32}$$

The average probability of a successful response on those trials between the first success and last failure will be $a^2 + b^2$; that is, a proportion a of the subjects have the a element conditioned first and thus have probability a of success during the intermediate-state trials, whereas a proportion b of the subjects have the b element conditioned first, thus having success probability b during the intermediate state.

One interesting implication of this model is that the distribution of total errors to criterion is independent of the sampling bias, a, provided a is not one or zero. Learning can occur only on error-trials,

when an unconditioned element is sampled and the response is reinforced. The sampling bias affects the trial-rate at which errors occur in the intermediate state, but does not influence the probability of conditioning a sampled element. If $a > 1/2$, then on the average there will be more successes between adjacent errors than when $a = 1/2$, but the total number of errors (reinforcements) required to move into the absorbing state is the same $(2/c)$ in both cases. Comparing the curves of average success probabilities over trials for $a > 1/2$ and $a = 1/2$, the former will lie above the latter for several trials, then they will cross-over and the $a = 1/2$ curve will reach asymptote sooner. This implies that the average trial of the last error will be larger when there is a sampling bias. This may be seen directly in the expression for the average trial of the last error, $\frac{1-a(1-a)}{ca(1-a)}$, which increases as a increases above $1/2$.

In applying this model to the Theios data, we estimate c from mean total errors, obtaining the estimate $c = .427$. The bias parameter, a, can be estimated in a number of ways. One simple and direct estimate is the mean total number of error runs, R, which is given by

$$R = 1 + \frac{1-2a(1-a)}{c}. \tag{33}$$

Substituting the observed value of $R = 2.18$, we obtain the estimate $\hat{a} = .502$. Other estimation procedures for $\hat{a}$ have given comparable results. This estimate differs negligibly from the $a = .500$ assumption that Theios used in applying the model to his data.

REFERENCES

[1] Kimble, G. A., Mann, L. I., and Dufort, R. H, "Classical and instrumental eyelid conditioning," J. Exp. Psychol., 49 (1955), pp. 407-417.

[2] Logan, F. A., "A comparison of avoidance and non-avoidance eyelid conditioning." J. Exp. Psychol., 42 (1951), pp. 390-393.

[3] Sheffield, F. D., "Avoidance training and the contiguity principle," J. Comp. Physiol. Psychol., 41 (1948), pp. 165-177.

[4] Theios, J., "A three state Markov model for simple avoidance learning in rats," Tech. report No. 40, Psychology Series, Institute for Mathematical Studies in the Social Sciences, Stanford University.