The determinants of random choice*

DONNA BRYANT† and RUSSELL M. CHURCH

Brown University, Providence, Rhode Island 02912

The purpose of the present experiment was to describe a pattern of reinforcement sufficient to produce an unpredictable pattern of choice response by rats. On each trial two levers were inserted into an experimental chamber. If the reinforcement was always contingent upon single alternation (a simple pattern), Ss learned to alternate at significantly better than chance level; if reinforcement was contingent upon alternation on 50% of the trials (an insoluble pattern), Ss developed a position preference. To produce apparently random responding, the less preferred response (an alternation) was differentially reinforced on 75% of the trials. A simple stochastic model adequately described the results.

There are situations in which the success of an animal depends upon its ability to perform in an unpredictable manner. For example, in certain competitive situations the likelihood that an animal will achieve a goal is greater if its behavior cannot be predicted by another animal. The purpose of the present experiment was to identify a reinforcement schedule which would produce completely unpredictable choices between two opposing spatial responses. Such behavior can be characterized as random.

Various patterns of reinforcement generate highly predictable sequences of responses. For instance, if a simple pattern of responses is reinforced (e.g., single alternation), rats will learn the pattern. Even if the pattern is fairly complex (e.g., a Gellerman series), a rat may learn some of the regularities and receive significantly more than 50% of the reinforcements (Terman, 1970). For the purpose of producing random choices by rats, no reinforcement pattern could be selected that would yield more than 50% reinforcements.

An insoluble problem would serve to guarantee that the rat would be correct about half the time, but it would create other difficulties. If the pattern of reinforcement is random (an insoluble problem), an animal in a noncorrection procedure typically adopts a systematic strategy (Brody, 1965; Meyer, 1960; Weinstock et al, 1965). For the purpose of producing random choices by rats, no reinforcement pattern could be selected that would lead the rat to adopt a systematic pattern of responses.

A reinforcement pattern that was successful in maintaining a stable balance between opposing response tendencies might generate a highly unpredictable sequence of responses. The general procedure adopted to produce random choice was to reinforce the less likely response (alternation) with a higher probability than the more likely response (stay).

*This investigation was supported by Research Grant MH-19704 from the National Institute of Mental Health. Requests for reprints should be sent to R. M. Church, Department of Psychology, Brown University, Providence, Rhode Island 02912.

†Now at the University of North Carolina, Chapel Hill, North Carolina.

Linear stochastic models have been used successfully to predict the sequence of choice responses of animals (Bush & Mosteller, 1955). The theoretical purpose of the present experiment was to determine whether the same parameters of a stochastic model that account for the development of patterned behavior could also account for the development of random behavior.

METHOD

Subjects

The Ss were 12 naive male Norway albino rats (Charles River CD). They arrived from the breeding laboratory at 42 days of age, and they were housed in individual cages. During the first 13 days, they were given ad lib food and water; for the next 11 days preceding testing, they were given a daily ration of 12 g of ground Purina chow mixed with water. The Ss were handled daily.

Apparatus

Two leverboxes (23.5 x 20.3 x 19.4 cm high) were enclosed in sound-attenuating boxes. The two side walls and hinged lid were acrylic, and the front and back walls were aluminum. Two retractable stainless steel levers (5.1 x 1.3 cm) could project 5.1 cm through the front panel 5.1 cm above the floor. A distance of 7.0 cm separated the closest points of the two levers, and a food tray was located between the two levers. A force of about 30 g on the lever was required to operate the microswitch. A pellet dispenser delivered 45-mg Noyes food tablets to the tray. Air blowers ventilated the chamber. Control of the experimental procedure and recording of data was implemented by a time-shared PDP-12 computer located in an adjacent room.

Procedure

Pretraining. On the first session the levers were retracted and one food pellet was delivered each minute for 30 min. On the second session both levers were inserted into the box and each response was reinforced. When a particular lever had been pressed 30 times, it was retracted; when both levers had been retracted, the session was terminated. On the next six sessions, either the left or right lever was randomly selected on a given trial for insertion into the box. If a response was made within 5 sec, the lever was withdrawn immediately and a reinforcer was delivered at the end of the 5-sec interval; if a response was not made within 5 sec, the lever was withdrawn without reinforcement. One hundred trials per session were given, with a 15-sec intertrial interval (25-min session). On the next eight sessions the same training procedure was employed, except that only a random 66% of the responses were reinforced.

Training. The 12 Ss were randomly partitioned into three

Table 1
Proportion of Left and Right Responses on Trial N + 1 as a
Function of Responses and Reinforcements
on Trial N (Sessions 9-16)

	Trial N	Reinforcement on Trial N		Trial N	Nonreinforcement on Trial N	
		Left	Right		Left	Right
Group 100:0	Left	.087	.248	Left	.056	.089
	Right	.227	.108	Right	.113	.072
Group 75:25	Left	.133	.118	Left	.143	.126
	Right	.126	.119	Right	.119	.116
Group 50:50	Left	.181	.062	Left	.201	.067
	Right	.054	.187	Right	.074	.174

groups of four Ss each. As in pretraining, all Ss received 16 sessions of 100 trials per session, with a 15-sec intertrial interval (25-min sessions). On each trial *both* levers were inserted into the box. When a response was made, both levers were withdrawn and, if a reinforcement was to be delivered, it came at the end of the 5-sec interval. If no response was made, the levers were withdrawn after 5 sec. The first response of a session was reinforced on a random 50% schedule. Subsequent reinforcement depended upon whether or not an alternation was made. An alternation was defined as a response on Trial N that was opposite the previous response (i.e., a "left" response following a "right" response or a "right" response following a "left" response). A stay response was defined as a response on Trial N that was the same as the previous response. The three conditions were as follows: Group 100:0—all alternation responses were reinforced; no stay responses were reinforced. Group 75:25—a random 75% of the alternation responses were reinforced, and a random 25% of the stay responses were reinforced. Group 50:50—a random 50% of the alternation responses were reinforced, and a random 50% of the stay responses were reinforced.

RESULTS

During pretraining, the percentage of trials on which the inserted lever was pressed increased to 93% of the final session. On the first session of training, 40% of the responses were alternations. During the subsequent sessions, the percentage of alternation responses reflected the contingencies of reinforcement (see Fig. 1). The method for generating the straight line segments is described in the Discussion section.

The Ss reinforced for all alternation responses (100:0) gradually increased the percentage of alternation responses on successive sessions. After 1600 trials they alternated their responses on 69% of the trials. The Ss reinforced for three-fourths of their alternation responses (75:25) reached an asymptote of about 49% alternation, and the Ss reinforced for half of their alternation responses (50:50) reached an asymptote of about 26% alternation. The mean percentage of alternation responses on the final eight sessions of training was related to the probability of reinforcement of an alternation, and there was no overlap among the three distributions.

Table 1 describes the mean proportion of left and right responses on Trial N + 1 as a function of the response and reinforcement on Trial N for Ss in the

three groups. The three-dimensional contingency tables for each S were analyzed with chi-square tests of nonhierarchical hypotheses based upon logarithms of adjusted cell frequencies (Goodman, 1971; Shaffer, 1973). The results described below were individually significant at about the .001 level ($z = 3.0$), so the nominal level of significance for the set of 84 comparisons is approximately .01.

Three of the four Ss in Group 100:0 developed the alternation response and, therefore, they were reinforced on more than half of the trials. This alternation tendency was greater following a reinforcement than following a nonreinforcement for two of the Ss, one of which also had a position preference.

All of the Ss in Group 50:50 developed position preferences. (Two Ss responded predominantly on the left lever and two on the right.) Three of the four Ss also had a significant tendency to stay on a given lever, although in some cases this could have been due to a change in the position preference.

If the Ss in a group were randomly choosing the left or right response on each trial, the proportions in each of the eight cells of Table 1 would be .125. An overall Pearson chi-square test based on nearly 100 responses per session for the last eight sessions could not reject the chance hypothesis for three of the four Ss in Group 75:25. Individual chi-square tests were consistent with the hypothesis that three of the four Ss made about 50% of their responses to each of the levers and that all Ss alternated about 50% of the time. The outcome of the response on Trial N (reinforcement or nonreinforcement) did not have a reliable effect on the choice of response on Trial N + 1 for any of the Ss.

A test of the number of runs of left and right responses on the last eight sessions was performed as a further test of the hypothesis that Ss in Group 75:25 were choosing their responses at random and with equal

Fig. 1. The percentage of alternation responses as a function of sessions of training under three patterns of reinforcement. The number preceding the colon is the percentage of alternation responses that were reinforced; the number following the colon is the percentage of stay responses that were reinforced.

probability. A separate test was performed for each S on each of the last eight sessions (32 tests). None of these runs tests could reject the chance hypothesis at the .01 level of confidence.

DISCUSSION

The major conclusion from the present experiment is that certain patterns of reinforcement in a two-choice situation can lead an animal to make a sequence of choice responses that is virtually unpredictable and which appears to be random. The response choice of three of four Ss in Group 75:25 could not be reliably predicted on the basis of any of the tests employed. These Ss had no bias for the left or right lever, and they had no bias to alternate or stay. They received a reinforcement on about half of the trials, and the reinforcement did not reliably influence the next response. The number of runs of responses to one lever or the other was not reliably different from the number expected by a random generator. Undoubtedly there are some subtle regularities in the sequence of response of these animals that could be identified with even more extensive analysis of these data. The extent to which a choice response could be predicted, however, is bound to be small.

To produce apparently random responding, the less preferred response (an alternation) was differentially reinforced on 75% of the trials. This procedure is related to Blough's (1966) procedure of reinforcing the least frequent interresponse times relative to an exponential distribution that would have been generated by a random process. Pigeons trained under such a schedule of reinforcement had interresponse time distributions that were approximately exponential, although responding was not random, since there were sequential dependencies of interresponse times.

The ability to respond in an unpredictable manner has considerable adaptive value, particularly in situations involving inter-S competition (Suppes & Atkinson, 1960). The fact that an animal can be trained to perform in an unpredictable manner does not conflict with a deterministic view of animal behavior. Recursive formulae that serve as generators of pseudorandom numbers may serve as a useful analogy. Any given number is determined, but a long sequence of numbers may meet all the standard tests of randomness. Similarly, even if a given response of the rat is determined, the sequence of choice may meet the tests of randomness.

The results are consistent with a simple stochastic model in which reinforcement of an alternation increases the probability of an alternation, and reinforcement of a stay decreases the probability of an alternation. More explicitly, let P_n equal the probability of an alternation on Trial n , let λ_a and λ_b be asymptotic values, and "a"

and "b" be learning rate parameters. Then, if an alternation is reinforced:

$$P_{n+1} = P_n + a(\lambda_a - P_n) \quad (1)$$

If a stay response is reinforced:

$$P_{n+1} = P_n + b(\lambda_b - P_n) \quad (2)$$

We will assume that the animals would eventually alternate on all trials if alternation responses were consistently reinforced ($\lambda_a = 1$) and that the animals would eventually stay on all trials if all stay responses were reinforced ($\lambda_b = 0$). We will also assume that nonreinforcement leaves the alternation probability unchanged ($P_{n+1} = P_n$). This is the simplest assumption, and it is consistent with data described by Brody (1965). Although nonreinforcement may be a factor, it is small relative to reinforcement; consequently, it can be neglected for our purposes.

The remaining three parameters must be estimated from the data. The initial probability of an alternation on the first session was approximately .4. The value of a , the learning rate parameter for a reinforced alternation, was estimated from the 100:0 group by a method developed by Bush and Mosteller (1955, p. 181). The value is approximately .001.

The value of b , the learning rate parameter for a reinforced stay response, was estimated from the 50:50 group. The asymptote of the expected operator is $a/(a+b)$, and the asymptotic probability of an alternation is approximately .25. Thus, $b = 3a$, a value which is approximately .003. It is also possible to assume that $a = b = .003$, but that a reinforced alternation is effective only on about one-third of the reinforced alternation trials. The greater effectiveness of reinforcement of a stay response than an alternation response is not surprising. To respond consistently to a given lever, the animal merely needs to retrieve the reinforced alternative from long-term storage; to alternate consistently, the animal must recall its response on the previous trial, as well as the general alternation rule. With a 15-sec intertrial interval and 100 trials per session, it is reasonable to assume that the animal could not always recall which of the two responses it had made on the previous trial. As a result, the magnitude of the change in probability was greater when a stay response was reinforced than when an alternation response was reinforced.

The straight line segments of Fig. 1 represent the mean performance of 10 stat-rats. The mean probability of alternation was simulated by repeated applications of Eq. 1 (for reinforced alternations), Eq. 2 (for reinforced stay responses), and the identity operator (for nonreinforced responses). The same values of the parameters were used under all three experimental

conditions ($\lambda_a = 1$, $\lambda_b = 0$, $P_0 = .001$, $a = .001$, $b = .003$). This incremental model adequately describes the data of the present experiment (see Fig. 1).

None of the parameters were estimated from the 75:25 group, yet the model is as appropriate for this group as for the others. The asymptote of the expected operator for the 75:25 group is $.75a/ (.75a + .25b)$. If $b = 3a$, as previously estimated, then the asymptotic probability of an alternation would be $.5$. This was approximately the level observed. Furthermore, if each event has only a small impact on the probability of alternation (and $.001$ and $.003$ are the estimated values), then the sequence of responses would be virtually indistinguishable from a random series.

REFERENCES

- Blough, D. S. The reinforcement of least-frequent interresponse times. *Journal of the Experimental Analysis of Behavior*, 1966, 9, 581-591.
- Brody, A. L. Nonreinforcement in a noncorrection T maze. *Journal of Comparative & Physiological Psychology*, 1965, 60, 428-431.
- Bush, R. R., & Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.
- Goodman, L. A. Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. *Journal of the American Statistical Association*, 1971, 66, 339-344.
- Meyer, D. R. The effects of differential probabilities of reinforcement on discrimination learning by monkeys. *Journal of Comparative & Physiological Psychology*, 1960, 53, 173-175.
- Shaffer, J. P. Defining and testing hypotheses in multidimensional contingency tables. *Psychological Bulletin*, 1973, 79, 127-141.
- Suppes, P., & Atkinson, R. C. *Markov learning models for multiperson interactions*. Stanford: Stanford University Press, 1960.
- Terman, M. Discrimination of auditory intensities by rats. *Journal of the Experimental Analysis of Behavior*, 1970, 13, 145-160.
- Weinstock, S., North, A. J., Brody, A. L., & LoGuidice, J. Probability learning in the T maze with noncorrection. *Journal of Comparative & Physiological Psychology*, 1965, 60, 76-81.

(Received for publication February 4, 1974;
accepted April 22, 1974.)