THE INFLUENCE OF INSTRUCTION IN A TWO-CHOICE PROBABILISTIC LEARNING TASK UNDER PARTIAL REINFORCEMENT

BY

H. C. VAN DER MEER

Psychological Laboratory, University of Utrecht

Humphreys' (23) "verbal conditioning experiment", now a classic, set off a long series of investigations into behavior in a situation in which a S must make a large number of predictions as to which of two or more alternative events will take place.

If we limit ourselves to a situation with two alternative events, the general pattern of these experiments is that a signal stimulus is repeated a great many times. After each signal the S must predict which of two events, E_1 and E_2, (e.g. whether a lamp to the left or to the right is switched on) will occur. These events occur in a random sequence, but with a fixed probability of π and $1-\pi$, respectively. Usually the probability π of E_1 occurring is fixed, not only for the whole series, but also within each block of 20 trials. E_1 and E_2 are mutually exclusive, though one of the two always occurs.

There are two theoretical models, in particular. which predict behavior in this type of situation. The one is a stochastic model, the other a game-theoretic model. Two versions of the stochastic model exist: the statistical learning model of Estes (8, 9, 10, 11, 12) and the linear operator model of Bush & Mosteller (4, 5). Both these models predict that, after a great many trials, the S learns to match his response ratios with the objective probabilities in which both events occur, i.e. the relative frequencies of his predictions of both events finally correspond with the objective probabilities of these events.

Estes' *set-theoretical model* assumes that there is a population of stimulus elements. The state of each element of this population is such that it tends to be conditioned to either A_1 (prediction of E_1) or to A_2 (prediction of E_2). On each trial, the signal to respond activates a random sample containing the mean proportion θ of the population of stimulus elements. θ is the proportion of the elements in this population constituting the effective sample on any one trial. At the end of each trial, all elements in the sample are conditioned to the response class which was, in fact, correct on that trial, independent of S's actual response. The sample is then returned to the population and the process is repeated with a new random sample on the next trial. It is assumed that all samples from the population of stimulus elements are statistically independent.

If p_n is the probability of A_1 after n trials, then p_n is the proportion of elements in the population conditioned to A_1, and $1-p_n$ is the proportion of elements in the population conditioned to A_2.

If, after trial n, E_1 has occurred, the change in the response probability is expressed by

$$p_{n+1}=p_n+\theta(1-p_n)=(1-\theta)p_n+\theta \text{ (increase in probability of } A_1).$$

If, however, E_2 has occurred, then the change in the response probability is

$$p_{n+1}=p_n-\theta p_n=(1-\theta)p_n \text{ (decrease in probability of } A_1).$$

The average probability of A_1 after $n+1$ trials is given by the relation

$$\bar{p}_{n+1}=\pi[(1-\theta)\bar{p}_n+\theta]+(1-\pi)(1-\theta)\bar{p}_n=(1-\theta)\bar{p}_n+\theta\pi.$$

If, on each trial, one of the alternative responses is reinforced with the probability π, and the other with the probability $1-\pi$, then it can be shown by mathematical induction that p_n, the expected probability of A_1 after n trials, can be expressed by the formula:

$$p_n=\pi-(\pi-p_1)(1-\theta)^{n-1}.$$

This is the equation of a monotonic increasing, negatively accelerated function, which, as $n\to\infty$, has the value π as asymptote.

The above equation is that of a learning curve in which the parameter θ denotes the learning rate. If n is large, p_n, the probability of A_1, approaches π, which means that the probability that the S predicts E_1 equals the probability that E_1 occurs.

The *linear operator model* of Bush & Mosteller (4, 5) results in the same predictions as the model of Estes. Bush & Mosteller believe that their general model can be applied to any given learning situation in which reinforcement or non-reinforcement produces a change in response probability. This change is expressed by the linear transformation:

$$Qp_n=\alpha p_n+(1-\alpha)\lambda.$$

In a two-choice situation the two operators are:

$$Q_1p_n = \alpha_1 p_n + (1-\alpha_1)\lambda_1$$
$$Q_2p_n = \alpha_2 p_n + (1-\alpha_2)\lambda_2,$$

expressing the effect of reinforcement and non-reinforcement, respectively.

In the situation described above, in which a choice must be made between two alternative, mutually exclusive events that occur with a probability of π and $1-\pi$ respectively and that are non-contingent to the choice of the S, certain restrictions must be applied to the parameters of these equations. Since, in this type of experiments, 100 % reinforcement and 100 % non-reinforcement of a response lead to asymptotic probabilities of 1 and 0 respectively, the restrictions obtaining are: $\lambda_1 = 1$ and $\lambda_2 = 0$, in which λ is the parameter expressing the limit of the operator.

The parameter α indicates the learning rate and, taking into account the experimental results regarding the asymptotic value in the case of partial reinforcement, a further restriction obtains, viz. $\alpha_1 = \alpha_2 = \alpha$. These restrictions imply that the situation is symmetrical, that reinforcement and non-reinforcement have an equal but opposite effect on behavior and that the reinforcement of A_1 has the same effect on p as the reinforcement of A_2 on $q = 1-p$.

Hence, the two operators become:

$$Q_1p_n = \alpha p_n + 1 - \alpha$$
$$Q_2p_n = \alpha p_n.$$

If we assume that $\alpha = 1-\theta$, then these become:

$$Q_1p_n = (1-\theta)p_n + \theta$$
$$Q_2p_n = (1-\theta)p_n.$$

These formulae for the changes in probability of A_1, when preceded by E_1 and E_2 respectively, are therefore identical with those of Estes. By analogy:

$$p_n = \pi - (\pi - p_0)\alpha^n.$$

Both models thus predict that, after a great number of trials, the probability of A_1 (the prediction of E_1) equals the probability that E_1 occurs. Numerous investigations (12, 16, 20, 21, 23, 24, 25, 27, 29, 33) confirm the predictions of these models.

This "probability matching" behavior was considered by many game theorists as being irrational and thus not readily acceptable. According to the game-theoretical model formalized by Von Neumann & Morgenstern (28), the S will learn to maximize the expected frequency of correct predictions. This can be achieved by always choosing the more frequent event. If, for instance, the probability of E_1 occurring is 0.75 and that of E_2 is 0.25, the probability of a correct prediction is 0.75 using this pure strategy, If, on the other hand, a mixed strategy is applied and E_1 is chosen in 75 % of the trials and E_2 in 25 %, the expectation of a correct prediction is only $0.75(0.75)+0.25(0.25)=0.625$.

Simon (32) has already drawn attention to the difference between subjective and objective rational behavior. The experimenter who knows that E_1 and E_2 occur in a random sequence but with a constant probability might think it irrational if the S does not always choose E_1. The S, on the contrary, does not know for certain that the probabilities of E_1 and E_2 remain constant and his aim is to obtain as high a score as possible.

Flood (13) disputed the game-theoretical arguments by pointing out that, if the S is aiming at maximizing his score and not his expectation, he will not always try to apply a pure strategy. Flood's second argument was that the Von Neumann–Morgenstern game theory is inapplicable in this situation unless the organism can assume safely that the experimental stimulus is generated by a stationary stochastic process, i.e. that E_1 and E_2 occur with a constant probability in a random sequence. If the S believes that there may be some pattern (non-stationarity) over time in the stimulus, then a mixed strategy would appear more rational to him than a pure strategy, for the latter would give him no way to discover any pattern effect. The instructions of Estes (12) were such that no attempt was made to suggest to his Ss that they were confronted with a stationary process.

Simon (32) demonstrated that the model of Estes agrees with the game-theoretical conceptions if it is assumed that the S does not maximize his expectation but minimizes his regret, since he does not know the reward probabilities nor that they are constant.

Nevertheless, there are a number of investigations (7, 15, 17, 25, 31), which support, to some extent, the game-theoretical conceptions. Here it concerns investigations in which each correct prediction was rewarded. In these cases the S is more motivated to make a correct prediction than when his efforts are unrewarded and evidently, in these circumstances, he learns to choose E_1, albeit not for 100 %, yet with a relative frequency that is higher than that predicted by the stochastic models which is equal to π.

Let us return to Flood. Flood thus predicts that the S will behave according to Estes' model if he expects a system in the sequence of E_1 and E_2; if, on the other hand, he knows that the sequence of E_1 and E_2 is random, he will choose the more frequent event in 100 % of the trials.

One is inclined to wonder whether Flood's argumentation, notwithstanding his grasp of the subjective elements in the "rational" behavior, is not too mathematical and accords too little with the psychological way of thinking.

If the S is instructed along certain lines, the manner in which he observes the situation is also organized to some extent. In this way he is given a frame of reference for his perceptions. If the S is instructed that the sequence in which E_1 and E_2 occur is random, this implies that the S knows he is confronted with a situation in which variability and lack of system prevail. How will he respond to this situation? Will it be with the variability according to the situation evoked by the instruction?

If, on the other hand, the S is instructed that there is a certain system in the sequence in which E_1 and E_2 occur and that his task is to discover that system, will the S not have more motivation than a subject working under the former instruction? Will the utility of a correct prediction not be greater for him in this case than in the former case?

Contrary to Flood's expectation, the alternative hypothesis which we should like to postulate is that, if E_1 occurs with the greatest frequency, the proportion of A_1 responses is larger when the subject is instructed to search for a system than when he is instructed that there is no system. This hypothesis will be tested in a 75:25 situation.

METHOD

Apparatus. Three lamps were placed in front of the S on eye-level; the middle lamp was the signal light. When the signal light switched on, the S had to predict whether the lamp to the right or to the left would then be switched on.

Behind a screen, out of sight of the S, there was an apparatus (a rotating drum) on which the sequence of the two events, E_1 and E_2, had been programmed. When the drum rotated, at intervals of

5 seconds the signal lamp was switched on for 1 second; 2 seconds later, the lamp either to the right or to the left was switched on according to the programme.

Procedure. The entire series consisted of 300 trials. In 75 % of the trials the one lamp (E_1) was switched on, and in 25 % the other (E_2). For half of the Ss E_1 was the left-hand lamp, and for the other half of the Ss it was the right-hand lamp. The sequence of E_1 and E_2 had been drawn up with the aid of a table of random numbers, with this restriction, that in each block of 20 trials the percentages of E_1 and E_2 remained constant, i.e. 75 % and 25 % respectively.

Experiment I. As soon as the signal light flashed on, the S had to press down a transmitting key placed to the left of him if he predicted that the left-hand lamp would be switched on; or a transmitting key to the right if he expected the right-hand lamp to go on. According to this procedure, the choices of the more frequent event (A_1 responses) in each block of 20 trials were automatically registered.

Experiment II. As it appeared that too much information was lost with automatic registration, a second experiment was designed in which the S had to say "left" or "right" as soon as the signal light flashed on. The experimenter recorded the answer. In this experiment, moreover, there was the restriction that the event E_1 would not occur more then 5 times consecutively.

The experimental series of 300 choices was preceded by a short test-run of 10 choices, in which E_1 and E_2 both occurred for 50 %, to enable S to get acquainted with the method. Then, if the S required no further information, the whole series of 300 trials was carried out without interruption.

Instructions. After the Ss had been informed about the working method and their task, half of them received instruction A, and the other half instruction B.

Instruction A. The lamps will be switched on according to a fixed system, and you must try to discover what that system is. Perhaps you may often guess wrongly in the beginning, but, once you have discovered the system you can always guess the right answer. In any case you should try to guess correctly as often as possible.

Instruction B. The lamps will be switched on in a completely random sequence, in which there is no system. Nevertheless, you must try to guess correctly, as often as possible, which light will be switched on.

Subjects. The subjects were students. Both in experiment I and in experiment II, 20 Ss participated with instruction A and 20 Ss with instruction B. In order to cancel positional effects, E_1 was the switching on of the left-hand light for half of the Ss and of the right-hand light for the other half, both in group A and in group B. The subjects of exp. I and exp. II were run by the students Miss L. M. A. Wever and Miss A. Tai–A–Pin respectively.

Results

Table 1 gives the results of exp. I. If, like Estes, we take the mean percentage A_1 responses over the last 40 choices as the asymptotic value, then we find that for group A (instruction with system) this is 81.5 % and for group B (instruction without system) 78.25 %. In both cases the asymptotic value of 75 % predicted by the models of Estes and Bush & Mosteller is exceeded. In the case of instruction A

Table 1

Observed and predicted proportions of the A_1 responses per 20-trial block

Blocks of 20 trials	exp. I				exp. II			
	instr. A		instr. B		instr. A		instr. B	
	empir.	theor.	empir.	theor.	empir.	theor.	empir.	theor.
1	.485	—	.438	.527	.568	.640	.465	.465
2	.780	—	680	.684	.633	.750	.585	.591
3	.730	—	.693	.730	.788	.750	.703	.661
4	.790	—	.793	.744	.750	.750	.673	.700
5	.730	—	.685	.748	.718	.750	.665	.722
6	.738	—	.708	.749	.795	.750	.770	.734
7	.688	—	.718	.750	.705	.750	.738	.741
8	.795	—	.765	.750	.785	.750	.750	.745
9	.775	—	.753	.750	.713	.750	.718	.747
10	.768	—	.775	.750	.728	.750	.695	.748
11	.795	—	.788	.750	.808	.750	.795	.749
12	.800	—	.783	.750	.768	.750	.725	.750
13	.798	—	.793	.750	.750	.750	.783	.750
14	.828	—	.768	.750	.790	.750	.750	.750
15	.805	—	.798	.750	.850	.750	.795	.750
$\hat{\theta}$	—		0.059		0.114		0.029	
$\hat{p}_0$	—		0.376		0.500		0.376	

the difference with regard to this value is significant at the 0.001 level ($t=4.36$), in the case of instruction B the difference is not significant ($t=1.88$).

Contrary to Flood's expectation, the percentage A_1 responses with instruction A – where a non-stationary process is suggested to the S – is greater than in the case of instruction B – in which the stationary character has been emphasized.

Estes' model does not make any provision for the asymptotic value π being exceeded, and, for this reason alone, this model might be considered to be inadequate. Nevertheless, we shall examine whether the learning curves obtained from the results may be considered as functions of the form

$$\bar{p}_n = \pi - (\pi - \bar{p}_0)\,(1-\theta)^n.$$

In this formula $\bar{p}_0$ and θ are unknown quantities. If $\overline{P}_m$ is the mean percentage A_1 over the m^{th} block of 20 trials, then:

$$\overline{P}_m = \pi - (\pi - \overline{P}_1)\,(1-\theta)^{20(m-1)}$$

and

$$\sum^{k} \overline{P}_m = k\pi - (\pi - \overline{P}_1)\,\frac{1-(1-\theta)^{20k}}{1-(1-\theta)^{20}}. \qquad (1)$$

If we substitute the observed values $\overline{P}_1$ and $\sum \overline{P}_m$ in this formula, the parameter θ can be estimated by means of a method of successive approximation. In the case of instruction A, the equation cannot be solved for any value of θ with $0 \leqslant \theta \leqslant 1$. The model of Estes and the equal α-model of Bush & Mosteller cannot be applied to the learning curve of group A.

In the first two blocks of 20 trials there occurred a few long runs of E_1 which influenced the results and led to high percentages of A_1 responses in the second block. This gave rise to difficulties in estimating the parameter θ in the case of group B. Estimation of θ from formula (1) gives the value $\theta=0.187$. This value is obviously too large. If it is used to estimate $\bar{p}_0$, a negative value is found. Alternatively, θ may be estimated from the formula:

$$\bar{p}_n = 300\,\pi - (\pi - \bar{p}_0)\,\frac{1-(1-\theta)^{300}}{\theta} \qquad (2)$$

if $\bar{p}_0$ can be estimated. Since we have no information at our disposal, in this experiment, about the individual choices, we cannot estimate $\bar{p}_0$ from the experimental data. The most obvious value $\bar{p}_0=0.50$ is too large. As an estimation of $\bar{p}_0$ we have used the same value as in exp. IIB, and in this way we find $\theta = 0.059$.

In exp. IIA the difference between the mean percentage A_1 responses over the last 40 trials, namely 82 %, and the asymptotic value of 75 %, is significant at the 0.001 level ($t = 4.42$). In exp. IIB the difference between the observed asymptotic value 77.25 % and the theoretical value of 75 % is not significant ($t = 1.16$). The results of exp. II are given in Table 1.

The difference between the asymptotic values of group A and group B is significant at the 0.05 level.

In exp. IIA an estimation of θ from formula (1) did not give a value $0 \leqslant \theta \leqslant 1$. Therefore, $\bar{p}_0$ was estimated from the first 5 trials and assumed to be 0.50; thereafter θ was calculated from formula (2). It is evident that the calculated theoretical curve does not cover the observed results (see Table 1).

In exp. IIB θ was estimated from formula (1) and thereafter $\bar{p}_0$ from

$$\bar{p}_0 = \pi - (20\pi - \bar{T}) \frac{\theta}{1-(1-\theta)^{20}},$$

in which $\bar{T}$ is the mean number A_1 responses over the first 20 trials. The theoretical curve corresponds comparatively well with the results.

From the above it may be concluded that the behavior of the groups A and B is not identical. We may express the difference more precisely by applying Grant's (19) extension of Alexander's (1) test for trend.

If $p(A_1)$ is a function of the number of reinforced trials, we can express this function in the form:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots .$$

This formula may also be written in the form

$$y = A_0 + A_1 Q_1(x) + A_2 Q_2(x) + A_3 Q_3(x) + \dots,$$

in which $Q_i(x)$ is the orthogonal polynomial of the i^{th} degree. The terms A_i can be tested independently and thus independent tests are available for the existence of significant differences between various groups as regards the linear, quadratic, cubic, etc. components of the trend.

In Table 2 the results of the trend analysis for exp. II are summarized over the mean percentages A_1 responses for each block of 60 trials.

TABLE 2

Trend analysis of the frequencies of the A_1 response per 60-trial block for groups II A and II B

Source	*df*	SS	MS	Error term	*F*	*P*
A. Over-all trend . . .	(4)	(2167.13)	(541.78)	E	(36.86)	0.001
1. linear	1	1780.84	1780.84	E. 1	84.32	0.001
2. quadratic	1	171.61	171.61	E. 2	18.45	0.001
3. cubic	1	191.82	191.82	E. 3	13.74	0.001
4. quartic	1	22.86	22.86	E. 4	1.64	—
B. Between group means	1	237.62	237.62	D	5.37	0.03
C. Between group trends	(4)	(130.03)	(32.51)	E	(2.21)	—
1. linear	1	64.00	64.00	E. 1	3.03	—
2. quadratic	1	38.06	38.06	E. 2	4.09	0.05
3. cubic	1	0.56	0.56	E. 3	0.04	—
4. quartic	1	27.40	27.40	E. 4	1.21	—
D. Between indiv. means	38	1679.96	44.21	E	3.01	0.001
E. Between indiv. trends	(152)	(2235.24)	(14.70)			
1. linear	38	802.66	21.12			
2. quadratic	38	353.47	9.30			
3. cubic	38	530.62	13.96			
4. quartic	38	548.49	14.43			
F. Total	199	6449.98	—			

Since the groups IA and IIA, IB and IIB respectively, did not differ essentially from each other, the results of both experiments have been combined, an analysis for trend has been applied to them and the consequent results have been summarized in Table 3. According to Bartlett's test the variances were not homogeneous in this case. As, however, the numbers of Ss in both groups are equal and the distribution is normal, the effect of heterogeneity of variances is very slight (2, 22), and thus a trend analysis was made. In marginal cases, however, where the computed F-ratio approaches the critical ratio of the point of confidence, the results should be interpreted with some reserve.

From line A (see Tables 2 and 3) it appears that, when the learning curves A and B are considered in conjunction, they possess a significant linear, quadratic and cubic component.

From line B it appears that the mean percentages A_1 responses over the 300 trials for the groups A and B differ significantly.

From line C it appears, as regards exp. II only (Table 2), that the groups A and B differ in the quadratic component of the trend. A combination of the results of exp. I and exp. II gives a significant difference in over-all trend and in the linear and quadratic components.

The linear component indicates the slope of the curve and represents the learning rate. The most important difference in trend between the groups A and B, resulting in a difference in asymptotic value, is, however, formed by the quadratic component.

TABLE 3

Trend analysis of the frequencies of the A_1 response per 60-trial block for groups A and B of exp. I and II considered together

Source	*df*	SS	MS	Error term	*F*	*P*
A. Over-all trend . . .	(4)	(4535.54)	(1133.88)	E	(85.01)	0.001
1. linear	1	3767.12	3767.12	E. 1	200.08	0.001
2. quadratic	1	437.50	437.50	E. 2	37.84	0.001
3. cubic	1	262.21	262.21	E. 3	24.77	0.001
4. quartic	1	68.71	68.71	E. 4	5.55	0.025
B. Between group means	1	334.89	334.89	D	7.24	0.01
C. Between group trends	(4)	(184.54)	(46.13)	E	(3.45)	0.01
1. linear	1	92.48	92.48	E. 1	4.91	0.05
2. quadratic	1	84.70	84.70	E. 2	7.33	0.01
3. cubic	1	0.32	0.32	E. 3	—	—
4. quartic	1	7.04	7.04	E. 4	—	—
D. Between indiv. means	78	3609.47	46.28	E	3.47	0.001
E. Between indiv. trends	(312)	(4161.53)	(13.34)			
1. linear	78	1468.60	18.83			
2. quadratic	78	901.80	11.56			
2. cubic	78	825.58	10.58			
4. quartic	78	965.56	12.38			
F. Total	399	12825.96	—			

In the introduction it was assumed that the variability in the case of instruction B would be greater than in the case of instruction A. This can manifest itself in various ways. The inter-individual variability, of which the variance is a measure, could be greater in the case of instruction B. If this were so, and if it were a result of a difference in set evoked by the instruction, then we should expect this difference mainly in the beginning of the conditioning process. For the first 100 trials the variance in exp. I, with instruction B is significantly larger at the 0.05 level of confidence than with instruction A. In exp. IIB the variance over the first 100 trials is, indeed, larger than in exp. IIA, but the difference is not significant. If the results are combined, the difference is significant at the 0.01 level. There is no difference in variance between the two groups over the last 100 trials.

A measure for the intra-individual variability can be found in the number of runs per S (a run may be defined as a series of successive A_1 or A_2 responses). If the Mann–Whitney test is applied, the number of runs for the first 100 trials in exp. IIB is found to be significantly greater at the 0.05 level of confidence than in exp. IIA. The mean numbers of runs for the last 100 trials are the same for both groups. Under both conditions the intra-individual variability decreases regularly. If a non-paramatric test for trend is applied, the gradual decrease for group A is significant at the 0.01 level, for group B at the 0.001 level.

Both the inter- and the intra-individual variability for the first 100 trials are greater in the case of instruction B than in that of instruction A. There is no difference in variability for the last 100 trials. There is a significant difference at the 0.02 level between the average lengths of run of A_1 and A_2 responses according to the Mann–Whitney test. The average length of run of A_1 responses is larger and of A_2 responses smaller for group A than for group B.

TABLE 4

Conditional frequencies of A_1 and A_2 responses per 60-trial block for exp. II

$A_1 \mid E_1$		$A_1 \mid E_2$		$A_2 \mid E_1$		$A_2 \mid E_2$	
A	B	A	B	A	B	A	B
535	478	248	215	345	402	52	85
635	602	263	241	265	298	37	59
601	615	280	267	299	285	20	33
630	598	291	287	270	302	9	13
671	658	285	273	229	242	15	27
$\varepsilon\ (A_1 \mid E_1) = 675$		$\varepsilon\ (A_1 \mid E_2) = 225$		$\varepsilon\ (A_2 \mid E_1) = 225$		$\varepsilon\ (A_2 \mid E_2) = 75$	

Table 4 includes the frequencies of A_1 and A_2 responses under the conditions E_1 and E_2, summarized over blocks of 60 trials. Also included in this table are the asymptotic frequencies expected according to the models of Estes and Bush & Mosteller. The table shows that, after E_1, group A behaves according to the expectations of the stochastic model. The frequency of $A_1|E_1$ for group B is slightly less than the expectation. After E_2, the behavior of both groups is contrary to the expectation; this applies especially to group A. The frequencies of $A_2|E_2$ are considerably less than the predictions of the model. Both after E_1 and E_2, over 10 blocks of 30 trials, the difference in behavior between the two groups A and B is significant at the 0.01 level when tested with the Wilcoxon test for paired replicates.

Hence, as reinforcer, E_2 has not the positive recency effect that accords with the stochastic models (i.e. the response probability is predicted to change each trial in the direction of the most recently occurring event) but predominantly a negative recency effect. After E_2, group A shows significantly more negative recency than group B, which behaves more in conformity with the prediction of the model. This may be expressed in other terms. The average uncertainty of the predictions without knowledge of the preceding events E is, for group A, 0.76 bit and for group B, 0.80 bit (see Table 5), calculated over the last 120 trials. Knowledge of preceding events influences the uncertainty, defined as

$$H_i(j) = -\sum_i \sum_j p(i) p_i(j) \log_2 p_i(j),$$

in which $H_i(j)$ is the uncertainty in bits per response, provided series of i events precede the response; $p_i(j)$ is the probability that response j follows sequence i; $p(i)$ is the probability that sequence i will occur (14, 26).

TABLE 5

The mean uncertainties of the response resulting when the previous events are known

Mean uncertainties	Group A	Group B
$H(y)$	0.76	0.80
$H_x(y)$	0.76	0.76
$H_{x,x}(y)$	0.71	0.69
$H_{x,x,x}(y)$	0.67	0.65
$H_{x,x,x,x}(y)$	0.63	0.63

Knowledge of event E preceding A does not decrease the uncertainty for group A – $H_x(y)$ remains 0.76 – but it does so for group B. Decrease in the uncertainty of 0.04 bit is statistically significant at the 0.01 level. Knowledge of the two preceding events results in a greater decrease in uncertainty for group B than for group A. Knowledge of the four preceding events E gives for both groups the same value for the 5th order approximation of the uncertainty, viz. 0.63 bit. For group A the total decrease in uncertainty is 0.13 bit, for group B 0.17 bit.

Discussion

It is evident from what has been discussed above that the mode of Estes and the equal α-model of Bush & Mosteller do not provide an adequate prediction of the behavior of group A. Both models are based on the principle of positive recency, i.e. each occurrence of E_1 increases the probability that the next answer will be A_1 and each occurrence of E_2 decreases the probability of A_1 and increases the probability of A_2. A strong negative recency effect appears, however, after E_2 in both groups, but especially in group A. Literature on the subject has drawn attention to the existence of negative recency (3, 18, 20, 23, 24). It was believed, however, that the average results over blocks of 20 trials corresponded with the predictions of the stochastic models, even though the behavior of the Ss did not agree with the principle of positive recency in a number of predictions.

Estes & Straughan (12) also observed that the facts did not entirely agree with the theoretical predictions. The parameter θ can be estimated in two ways: from the experimental data, as we have done above, and from the formula:

$$\theta = \bar{p}(A_1|E_1) - \bar{p}(A_1|E_2),$$

in which $\bar{p}(A_1|E_1)$ is the average probability per response that E_1 is followed by A_1. Theoretically, both estimations should lead to the same value, but in practice they reveal great discrepancy. With the latter method, Estes found considerably greater positive values for θ. We, on the other hand, find negative values for both groups; for group A, $\theta = -0.227$, and for group B, $\theta = -0.199$, which indicates a strong negative recency effect.

In group A the asymptotic values of $p(A_2|E_2)$ approaches 0. If p_1 is the probability of A_1 following E_1 and p_2 the probability of A_1 following E_2, then the asymptotic value is $p_1\pi + p_2(1-\pi)$. For the model of Estes this is $0.75(0.75)+0.75(0.25)=0.75$. If, based on the results of exp. II, we assume an average positive recency effect after E_1 corresponding to the expectations of the Estes' model, and a negative recency effect after E_2, then we find that the asymptotic value is

$$p_\infty = 0.75(0.75) + 1(0.25) = 0.8125.$$

The asymptotic value 0.82 estimated from the last 40 choices for group IIA agrees reasonably well with this. The asymptotic value of group IIB differs significantly from this at the 0.05 level of confidence ($t = 2.3$).

This result might be considered an artifact occasioned by the restrictions obtaining in the composition of the sequence of E_1 and E_2, viz. a maximum homogeneous length of run of 5. Exp. I may be used as a check; the maximum length of run was 11 in that case. The uncertainties of the E-series in exp. I and II are mentioned in Table 6.

TABLE 6

The mean uncertainties of the event series

Mean uncertainties	Exp. I	Exp. II
$H(x)$	0.81	0.81
$H_x(x)$	0.80	0.77
$H_{x,x}(x)$	0.78	0.70
$H_{x,x,x}(x)$	0.76	0.68

In exp. IA we find a corresponding asymptotic value, viz. 0.816.

If the results of exp. I and II are combined, the asymptotic value of group A does not differ from 0.8125 ($t=0.37$); for group B the difference is significant at the 0.01 level ($t=2.73$).

If the above mentioned formula for the determination of the asymptotic value in the case of instruction A, in its general form expressed as:

$$p_\infty = \pi^2 + 1(1-\pi),$$

could be applied on a broader scale – which would have to be checked by further experiments – then the situation in which $\pi = 0.75$ is probably one of the most ideal situations in which a difference in instruction can be found. Should the difference between π and $1-\pi$ be greater, then the deviation of the predicted asymptotic value with regard to π is smaller, e.g. the prediction is 0.91 for $\pi=0.90$, and the prediction is 0.84 for $\pi=0.80$. In the case that the difference between π and $1-\pi$ is smaller, e.g. $\pi=0.60$, it is not likely that a mixed strategy will be used in which $p(A_2|E_2) \rightarrow 0$.

The nature of the differences between group A and group B has been analysed in detail above. The way in which the instruction brought about these differences has not yet been demonstrated, however.

One might expect that instruction B would lead to a greater variability than instruction A. A significant difference in variability does, indeed, exist in the first 100 trials. This difference disappears later, and in the last 100 trials there is no difference in variability whatsoever. This might be explained by the influence of the conditioning process and by the fact that, in the long run group B tends to develop certain hypothesis about the expected events, in spite of the instruction (see Table 5). For that matter, the variability factor is not the main one leading to differences in the asymptotic values for groups A and B. A difference in intra-individual variability could lead to a difference in learning rate. The analysis for trend (Tables 2 and 3) has shown that there is no, or only a very slight difference. The main difference between group A and group B is formed by the quadratic component of the learning curves, which results in different asymptotic values. The reason for this must be sought in differences in strategy determined by the instruction.

A possible difference in motivation is yet another factor which figures in the instructions given to the groups A and B. Group A is told that there is a system which can be found and once found, it can be used to make correct predictions; group B is informed that the sequence is an entirely random one, without any system. A S in group A may be worried about a faulty prediction, since it shows he has not yet found the system. A S in group B cannot be blamed for a faulty prediction because the sequence is a random one. To use a term borrowed from the decision-making theory (6), one might assume that the utility of a correct prediction is greater for group A than for group B. In publications on this subject (7, 15, 17, 25, 31) it has been pointed out that the stochastic models are inadequate and that the asymptotic values are greater than those predicted by these models if each correct prediction is rewarded and the utility of a correct prediction thus increases.

Siegel (30) has evolved three models based on the decision-making theory. Model I of these may be applied to our data.

Let

π = the probability of E_1,
p = the proportion of A_1 responses,
a = the marginal utility of a correct prediction,
b = the marginal utility of varying one's responses.

The expectation, E_x, that a prediction will be correct, is

$$E_x = p\pi + (1-p)(1-\pi) = (1-\pi) + p(2\pi - 1).$$

Then the expected utility of a correct prediction, U_r, is

$$U_r = aE_x = a[(1-\pi) + p(2\pi-1)].$$

If U_v = the utility of varying one's responses $= f(p)$, then

$$U_v = bp(1-p).$$

The total expected utility of a given strategy is then

$$U(p) = U_r + U_v = a[(1-\pi) + p(2\pi-1)] + bp(1-p).$$

The strategy p which maximizes expected utility is at

$$\frac{dU(p)}{dp} = 0$$

$$\frac{dU(p)}{dp} = a(2\pi - 1) + b - 2bp = 0.$$

$$p = \frac{a(2\pi-1)}{2b} + 0.5.$$

If $\alpha = a/b$, then

$$p = \alpha(\pi - 0.5) + 0.5.$$

Only when $\alpha = 1$ and thus $a = b$, $p = \pi$. This means, that the stochastic learning models only lead to a correct prediction of the asymptotic values if the marginal utility of a correct prediction equals the marginal utility of varying one's responses. The Estes model makes no provision for predictions in cases there is an inequality between these two.

From the above mentioned formula it follows that we can estimate α from

$$\hat{\alpha} = \frac{p - 0.5}{\pi - 0.5}.$$

If we estimate α from the last 40 choices, the difference between group A and group B is significant at the 0.01 level, when tested with the Mann–Whitney U-test.

It has already been shown that the variability – the mean number of runs in the last 100 trials – was equal for both groups; the same applies for the last 40 trials. From this it may be concluded that the utility of varying one's responses is equal for both groups. The utility of a correct prediction is, therefore, significantly greater for group A than for group B.

The final conclusion is, thus, that the utility of a correct prediction is greater for group A than for group B since there is more motivation for a correct prediction with instruction A than with instruction B. This, in turn, leads to a higher asymptotic value of A_1 responses in the case of instruction A.

Summary

Flood (13) assumed that the more frequent event would be chosen in 100 % of the trials—in accordance with the game-theoretical conceptions that one learns to maximize the expectation of a correct prediction—if the S is convinced that a sequence of alternative events is generated by a stationary process. If, on the other hand, the S believes that the process is non-stationary, he would apply a mixed strategy in which the more frequent event would be chosen in a proportion equal to the objective probability with which the event occurs.

Our counter-argument was that a difference in instructions as described above would influence behavior with regard to at least two factors, viz. a difference in variability in the predictions, and a difference in motivation. Thus the alternative hypothesis put forwards was, that the instruction suggesting a non-stationary process (instruction A) would lead to a higher asymptotic value of the predictions of the more frequent event than the instruction that the process was stationary (instruction B). This hypothesis was tested in a non-contingent two-choice situation, in which $\pi = 0.75$.

The following has appeared from this investigation:

1. Contrary to Flood's expectation, instruction A led to a higher asymptotic value of A_1 responses than instruction B, mainly as a result of a larger negative recency effect after E_2 with instruction A.
2. The stochastic learning model of Estes and the equal α-model of Bush & Mosteller are inadequate in the case of instruction A.
3. The hypothesis was offered that the asymptotic value of A_1 responses in the case of instruction A can be determined from the formula

$$p_{\infty} = \pi^2 + 1(1 - \pi),$$

since $p(A_2|E_2) \to 0$.

4. There was a significant difference in intra-individual variability between the groups A and B for the first 100 trials. For the last 100 trials there was no difference in variability whatsoever.
5. The variability factor does not appear to be the cause of differing asymptotic values.
6. The differences in asymptotic values are mainly determined by a difference in motivation, which makes the utility of a correct prediction greater in the case of instruction A than in that of B.

References

1. Alexander, H. W., A general test for trend. *Psychol. Bull.* 1946, *43*, 533–557.
2. Boneau, C. A., The effects of violations of assumptions underlying the t-test. *Psychol. Bull.* 1960, *57*, 49–64.
3. Brown, W. L. and Overall, J. E., Implications of recency effects for probability learning theories. *J. gen. Psychol.* 1959, *61*, 243–251.
4. Bush, R. R. and Mosteller, F., A mathematical model for simple learning. *Psychol. Rev.* 1951, *58*, 313–323.
5. ———, and ———, *Stochastic models for learning.* New York, Wiley, 1955.
6. Edwards, W., The theory of decision making. *Psychol. Bull.* 1954, *51*, 3[illegible]–417.
7. ———, Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.* 1956, *52*, 177–188.
8. Estes, W. K., Towards a statistical theory of learning. *Psychol. Rev.* 1950, *57*, 94–107.
9. ———, Theory of learning with constant, variable, or contingent probabilities of reinforcement. *Psychometrika* 1957, *22*, 113–132.
10. ———, The statistical approach to learning theory. In Koch, S. (Ed.): *A study of a science* II. New York, McGraw-Hill, 1959.
11. ———, and Burke, C. J., A theory of stimulus variability in learning. *Psychol. Rev.* 1953, *60*, 276–286.
12. ———, and Straughan, J. H., Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.* 1954, *47*, 225–234.
13. Flood, M. M., Environmental non-stationarity in a sequential decision-making experiment. In Thrall, R. M., Coombs, C. H. & Davis R. L. (Eds.): *Decision processes.* New York, Wiley, 1954.
14. Frick, F. C. and Miller, G. A., A statistical description of operant conditioning. *Amer. J. Psychol.* 1951, *64*, 20–36.
15. Galanter, E. and Gerstenhaber, M., On thought: The extrinsic theory. *Psychol. Rev.* 1956, *63*, 218–228.
16. Gardner, R. A., Probability-learning with two and three choices. *Amer. J. Psychol.* 1957, *70*, 174–185.
17. Goodnow, J. J., Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.* 1955, *68*, 106–116.
18. ———, Response-sequences in a pair of two-choice probability situations. *Amer. J. Psychol.* 1955, *68*, 624–630.

19. Grant, D. A., Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.* 1956, *53*, 141–154.
20. ———, Hake, H. W. and Hornseth, J. P., Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.* 1951, *42*, 1–5.
21. Hake, H. W. and Hyman, R., Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.* 1953, *45*, 64–74.
22. Horsnell, G., The effect of unequal variances on the F-test for the homogeneity of group means. *Biometrika* 1953, *40*, 128–136.
23. Humphreys, L. G., Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.* 1939, *25*, 294–301.
24. Jarvik, M. E., Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.* 1951, *41*, 291–297.
25. Messick, S. J. and Solley, C. M., Probability learning in children: some exploratory studies. *J. genet. Psychol.* 1957, *90*, 23–32.
26. Miller, G. A. and Frick, F. C., Statistical behavioristics and sequences of responses. *Psychol. Rev.* 1949, *56*, 311–324.
27. Neimark, E. D., Effect on type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.* 1956, *52*, 209–220.
28. Neumann, J. von and Morgenstern, O., *Theory of games and economic behavior.* New York, Princeton, 1947, *2*.
29. Parducci, A., Alternative measures for the discrimination of shift in reinforcement-ratio. *Amer. J. Psychol.* 1957, *70*, 194–202.
30. Siegel, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two-choice uncertain outcome situation. *Psychometrika* 1959, *21*, 303–316.
31. ———, and Goldstein, D. A., Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.* 1959, *57*, 37–42.
32. Simon, H. A., A comparison of game theory and learning theory. *Psychometrika* 1956, *21*, 267–272.
33. Straughan, J. H., Human escape learning in relation to reinforcement variables and intertrial conditions. *J. exp. Psychol.* 1956, *52*, 1–8.
34. Thrall, R. M., Coombs, C. H. and Davis, R. L., *Decision processes.* New York, Wiley, 1954.