Journal of Experimental Psychology
1961, Vol. 62, No. 4, 385–394

PROBABILITY LEARNING IN 1000 TRIALS [1]

WARD EDWARDS

Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan

This paper reports a simple two-alternative noncontingent probability learning experiment with an unconventional feature: each *S* made 1000 consecutive predictions, making possible very detailed analysis of responses which occurred after learning was essentially completed.

Some abbreviations will be useful. The *S* predicts either L or R; after each prediction he observes either l or r. The probability that *S* will make prediction L on trial $t+1$ will be called p_t. The probability of l on any trial is a constant for any given *S*; it will be called π. The occurrence of a prediction will be called a *response*; the occurrence of a display of an event following a response will be called an *outcome*. An outcome follows each response; the nature of the outcome is independent of the nature of the response.

The interpretation of this experiment will focus on three issues:

1. *The probability matching hypothesis.* The probability matching hypothesis (PMH) asserts that the asymptotic probability of choice, p_∞ ($p_\infty = \lim_{t \to \infty} p_t$; it is assumed that this limit exists) equals π. It was originally proposed by Grant, Hake, and Hornseth (1951), is predicted by the Estes learning model (Estes, 1950, 1957; Estes & Burke, 1953; Estes & Straughan, 1954) and by the equal-alpha case of the Bush-Mosteller learning model (Bush & Mosteller, 1955), and has been supported by a number of experiments, though not by others.

2. *The extreme-asymptote generalization.* In 1956 I reported an experiment which argued against PMH and in favor of a theory about p_∞ which I call the extreme-asymptote generalization. That generalization asserts that p_∞ is more extreme than π, and as the absolute value of the difference between π and 0.5 increases the difference between p_∞ and π also increases until p_∞ becomes 1 or 0. As stated, this hypothesis makes only ordinal predictions; a way of making it yield ratio scale predictions (and also of applying it to situations in which amount of payoff is varied) is discussed in Edwards (1956) and applied later in this paper.

3. *Sequential dependencies, the gambler's fallacy, and path independence.* Stochastic learning theories often assume that the effects of events prior to a given trial are summarized in a set of probabilities for the responses available on that trial; this assumption is known as the path independence assumption (for a better definition, see Bush & Mosteller, 1955, p. 17). Contradictory to this is the common observation that if a flipped coin comes up heads eight or nine times in a row, *S* is likely to decide that tails is "due" and so predict or bet on it on the next toss. This and similar sequential effects have been called the gambler's fallacy; they have been demonstrated experimentally by many *Es*. Other hypotheses about sequential effects in probability learning have also

[1] This research was supported in part by the Operator Laboratory, Air Force Personnel and Training Research Center, and in part by Project MICHIGAN under Department of the Army Contract DA-36-039-SC-78801, administered by the United States Army Signal Corps. Reproduction for any purpose of the United States Government is permitted. I am grateful to L. W. Dildy and J. L. McDade for collecting and processing data, to W. H. Evans for writing computer programs, and to Carol J. Orwant and G. C. Morningstar for additional data processing.

been proposed, with varying degrees of empirical support.

Adequate study of each of these issues depends on long experiments; the reasons why will be examined in the discussion section. Both PMH and sequential dependencies are harder to examine at more extreme probabilities than at less extreme ones. So this experiment used only the probabilities 0.5, 0.6, and 0.7 and their complements.

Method

Apparatus.—Each *S* was given a tray containing 13 IBM mark sense multiple choice answer sheets. On top of the stack of sheets was a covering board with 80 pairs of holes in it, each hole filled by an ordinary cork. Each hole exposed two adjacent spaces where a mark could be made on the answer sheet. The mark sense sheets were prepared in advance by filling in one of the two mark spaces under the right-hand hole of each pair.

Subjects.—The *S*s were 120 basic airmen, trainees at Lackland Air Force Base. They were unselected except that no *S* who fell in Category 4 (the lowest category) of the Armed Forces Qualification Test, a paper-and-pencil test of general intelligence, was used. But the population of basic airmen includes relatively few college level men. The *S*s used in this experiment, therefore, are selected from a population which has almost no overlap with the college population from which *S*s have been selected for other probability learning experiments, except those by Neimark and Shuford (1959) and Nicks (1959), who also used basic airmen.

Instructions.—Each *S* was told to lift the upper left hand cork, and to make a mark in either the left or the right space on the sheet underneath it. A mark in, for instance, the left space was a prediction that the left space under the other cork of the pair would turn out to be filled in. After making the mark, he lifted the other member of the pair of corks, and saw whether his prediction had been correct or incorrect. After this, he replaced both corks, lifted the cork immediately beneath the first one he had lifted, and made his next prediction. When he finished 80 predictions, he removed the covering board, put the finished answer sheet underneath the stack, replaced the covering board, and continued making predictions. All *S*s were instructed: "Your purpose is to get as many predictions correct as possible. You will not be able to get all of them correct at any time during the test. There is no pattern or system you can use which would make it possible to get all of your answers correct. But you will find that you can improve your performance in the test if you pay attention and think what you are doing."

Experimental design.—There were 12 groups of 10 men each; each *S* made 1000 binary predictions in one unbroken session, usually lasting about 3 hr. The *S*s came in groups of 12; each *S* was arbitrarily assigned to one of the experimental groups. Twelve *S*s and *E* sat at a long conference table; *E* monitored continuously to make sure that all *S*s followed instructions and kept at the task. No effort was made or needed to pace *S*s. Each *S* present at a given time was a member of a different experimental group from all others then present, so no *S* could profit from looking at another *S*'s predictions.

Three basic probabilities were used: 0.5, 0.6, and 0.7. These numbers are the probabilities that a prediction of left will be correct. Sequences of 1000 trials embodying these probabilities were prepared in two different ways, which this paper will call *constrained* and *random*. All constrained sequences were prepared as follows. First, the expected number of occurrences of runs of length 1, 2, . . ., *n* for each of the two alternatives was calculated, up to a value of *n* for which that expected number is less than 0.5. All numbers were rounded off to integers. The runs of l were put in one box, the runs of r were put in another, and runs were drawn at random from the two boxes alternately until both were empty. This procedure makes not only run lengths but also conditional probabilities (based on sequences which are short compared with *n*) come out at their expected values. The random sequences were simply chosen from a table of random numbers in accordance with their probabilities, with no constraints at all.

Three probabilities and two ways of preparing sequences require six sequences. Six more sequences, each a mirror image of one of the six original sequences, were also used. The mirror image sequences were prepared by substituting an l for each r and an r for each l. One of these 12 sequences was administered to each of the groups; all *S*s in a group got the same sequence.

Results

Asymptotic probabilities.—Figure 1 shows mean relative frequencies of choice by blocks of trials. Each data point represents 40 binary choices by each of 10 *S*s, or 400 binary choices in all. In each of the eight groups for which the probability of reward is not 0.5 and so for which PMH and the extreme-asymptote generalization make different predictions, the results support the extreme-asymptote generalization. Inspection of the 50–50 groups suggests that there is a bias in favor of the R response (which is surprising, since for a right-handed *S* the L response is a trifle easier to make), but the bias is not large enough to affect the finding.

Inspection of Fig. 1 indicates that *S*s tended to follow local changes in the probability of reward. A local increase in frequency of 1 events produces a local increase in frequency of L predictions, and similarly for decreases. This effect is superimposed on the slower and larger changes

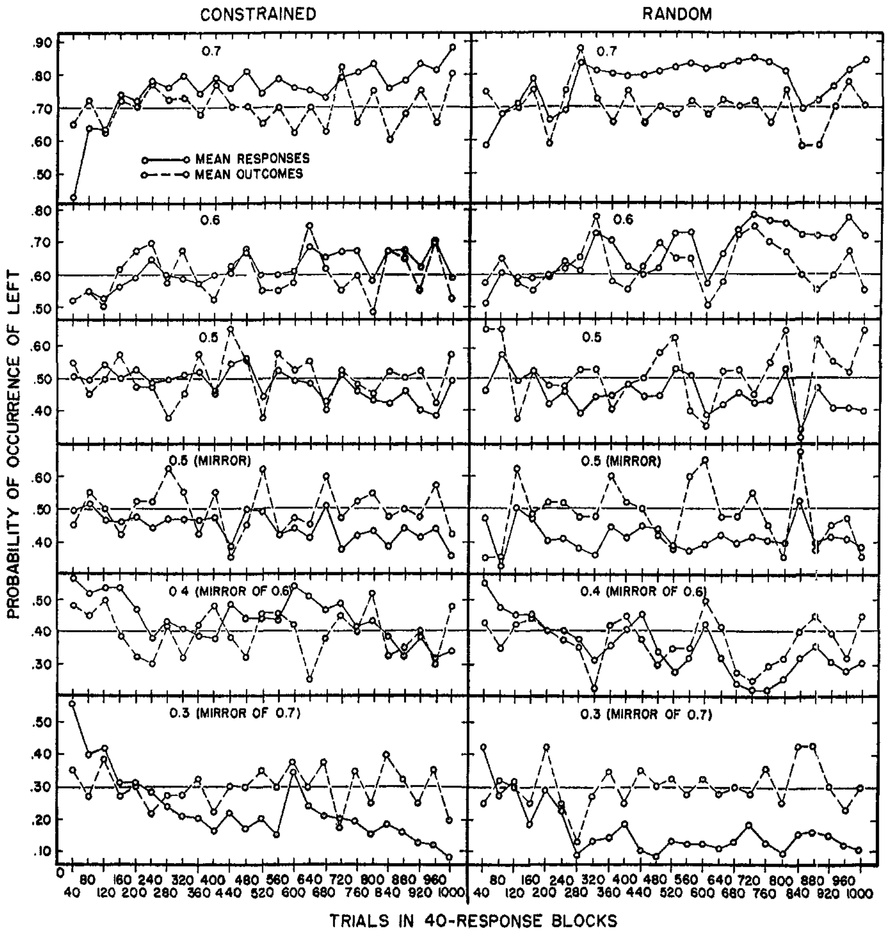

FIG. 1. Probability of left response in 40-trial blocks. (The number at the top of each rectangle is the π for that group, as is also the thin horizontal line within each rectangle. Each data point connected with solid lines is the relative frequency of prediction of left on a given block of trials; each point is based on 400 binary choices. Each point connected with dashed lines is the relative frequency with which the left event actually occurred in that block of 40 trials.)

TABLE 1

PERCENTAGE OF PREDICTIONS OF LEFT ON LAST 80 TRIALS FOR EACH *S* NOT IN A 50-50 GROUP

$\pi = 0.7$		$\pi = 0.6$		$\pi = 0.4$ (Mirrors of 0.6 Groups)		$\pi = 0.3$ (Mirrors of 0.7 Groups)	
Constrained	Random	Constrained	Random	Constrained	Random	Constrained	Random
100%	100%	76%	91%	48%	49%	26%	26%
97	93	74	90	47	43	20	20
96	88	70	85	46	40	17	13
95	88	69	81	43	35	12	13
91	87	66	77	43	31	11	13
85	85	64	74	31	26	8	13
80	80	64	71	29	21	4	11
75	65	61	63	22	16	4	8
70	60	59	61	20	15	0	0
58	56	46	56	0	11	0	0

Note.—The actual relative frequencies of outcomes in the last 80 trials deviated slightly from the theoretical probabilities. They were 0.73 for the 0.7 constrained group, 0.74 for the 0.7 random group, 0.61 for the 0.6 constrained and random groups, 0.28 for the 0.3 constrained group, 0.26 for the 0.3 random group, and 0.39 for the 0.4 constrained and random groups. If these rather than the theoretical probabilities are used in the nonparametric test discussed in the text, no change in conclusions results.

in prediction with which PMH and the extreme-asymptote generalization are concerned.

Finally, inspection of Fig. 1 indicates that the difference between constrained and random sequences is relatively unimportant except for the fact that constrained sequences come out more nearly to the expected number of l's and r's in each block of trials, and so provide slightly less scope for the probability following phenomenon discussed above to become visible.

A significance test for the difference between the estimated p_∞ and π is desirable. Table 1 exhibits the percentage of choices of L on the last 80 trials for each *S*, omitting 50–50 groups. Only 16 *S*s out of 80 have estimated p_∞ equal to or less extreme than π. If PMH were correct, at least half the *S*s should have estimated p_∞ equal to or less extreme than π. The difference is significant beyond the .0001 level. Table 1 also makes it clear that the distribution of estimated p_∞ is not bimodal; indeed, it looks relatively normal. That fact permits the use of more sensitive parametric tests of significance—but the results of the nonparametric test given above makes the use of more sensitive tests unnecessary.

Since so many data were collected, a number of the variables and interactions not mentioned here were in fact statistically significant; this discussion has dealt with all which are believed to be also intelligible and important. All subsequent statistics will combine corresponding random and constrained groups and will combine all 50–50 groups. Each statistic was calculated separately for each of the 12 groups; in no case does the combining average numbers or functions which appeared dissimilar.

Sequential effects: *Information analysis.*—To study the determiners of responses in a more specific way than Fig. 1 permits, detailed examination of sequences of responses and outcomes is necessary. For this purpose, multivariate information transmission analysis (Garner & McGill, 1956; McGill, 1954) is exceptionally convenient. The model underlying the use of this statistic assumes stable conditional probabilities; the analysis avoided basing calculations on changing overall probabilities by using only the last 480 trials. Special attention to the nonorthogonality of predictor variables and to the choice of proper degrees of freedom for the Miller-Madow (1954) bias correction and significance test was necessary; for a discussion of these issues and related ones concerning the application of information statistics to sequences of responses, see Edwards (1954; in press).

Figure 2 shows the effect of taking increasingly remote predictor variables into account in predicting responses in the last 480 trials. (In all information calculations, no differences worth noting existed between original and mirror groups, so they are combined in Fig. 2 and 3.) Note that the *y* axis is the percentage of information in the responses *not* accounted for by the predictor variables considered. It is evident that although increasing numbers of predictor variables improve predictions

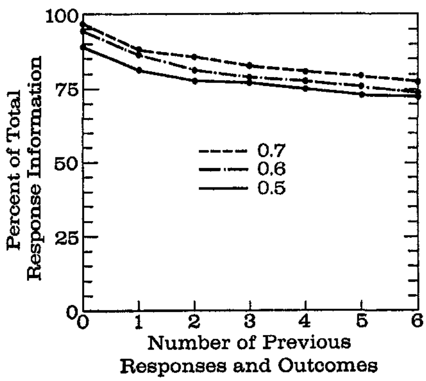

Fig. 2. Percentage of total response information unexplained by various predictor variables. (The *x* axis is cumulative. At Step 0, only *S*s are used as predictor variables. At Step 1, *S*s and the immediately preceding outcome are used. At Step 2, the variables already listed and also the immediately preceding response are used. At Step 3, the variables already listed and also the second preceding outcome are used. And so on. Only the last 480 trials were used.)

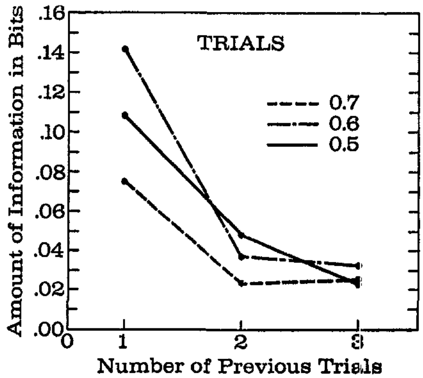

Fig. 3. Amount of information in bits transmitted by preceding trials to the present response. (The *S*s and trials which intervene between the predictor trial and the predicted response are held constant. Only the last 480 trials were used.)

(a mathematical necessity), the asymptotic level of predictive effectiveness leaves about 75% of the response information unexplained. If these numbers were variances, this would seem like a very large amount of unexplained variance. But they are not variances; they are ratios of bits of information. Users of multivariate information transmission analysis always report large percentages of unexplained response information; in fact, experiments in which as much as 25% of response information is explained by predictor variables are very rare (except in psychophysical scaling). No formal discussion of this common finding is known, but an obvious hypothesis is that the logarithmic nature of the information measure accounts for this difference between information and variance analyses.

Figure 2 shows how much prediction can be done, but does not show how to do it. In order to get a better idea about that, consider Fig. 3. It shows the amount of information (in bits, not a ratio) transmitted to the present response by the preceding three trials (calculations for the second and third preceding trials hold what happened in intervening ones constant). Again calculations are for the last 480 trials only. It is apparent that the most information is transmitted by the immediately preceding trial, and lesser amounts by trials prior to that. All amounts of information in Fig. 3 are significantly different from zero by the Miller-Madow test (1954).

What is doing the transmitting from each trial to the present response? It could be responses, outcomes, interactions between them, or any combination of these three factors. Unfortunately, the interactions between responses and outcomes are not directly interpretable because of the nonorthogonality of the predictor variables. Figure 2 is based on a definition of outcomes as being l or r; call this *noncontingent* coding. It would also be possible to define outcomes as + or − (meaning in agreement or disagreement with the preceding prediction); call this *contingent* coding. Further analysis of the data using noncontingent coding shows that almost all information transmitted by a trial is transmitted by its outcome; the amount of information transmitted by responses is trivial (though significant; because of the large numbers of responses involved, just about all differences which are observable at all are significant in this experiment). The implication, a sensible one, is that *S*s pay little or no attention to their own previous responses and instead concentrate on the previous set of outcomes in determining their present response.

Of course, similar analysis applied to contingently coded data shows that almost all information transmitted by a trial is transmitted by its response; this is an inevitable consequence of the fact that a + or − is meaningless as a predictor variable unless the preceding response which defines it is also considered. So two different methods of coding the data lead to two different interpretations of the results. A decision between these interpretations would require examination of the interactions, and nonorthogonality rules out the obvious ways of doing so. But a stab at it is available. If only the trial immediately preceding a response is considered, then the information transmitted from the response and information transmitted from the outcome should be orthogonal to each other. The information transmitted from the response can be calculated two different ways: with the effect of the outcome partialled out, or with the effect of the outcome uncontrolled. Table 2 presents the results of these two methods of calculation for each method of coding. No substantial difference between methods of calculation appears unless the method of coding forces it to appear by making the outcome variable taken by itself meaningless. For that reason, this paper used the noncontingent method of coding, and will accept the conclusion that *S*s are much more concerned with previous outcomes than with their own previous responses. Conclusive resolution of the dilemma, however, would require a three-alternative experiment, in which case

TABLE 2

INFORMATION IN BITS TRANSMITTED FROM PREVIOUS RESPONSE TO PRESENT RESPONSE BY THREE METHODS OF CALCULATION

Analysis	π		
	0.5	0.6	0.7
Intervening outcome ignored	.024	.040	.016
Intervening outcome held constant, noncontingent coding	.031	.045	.020
Intervening outcome held constant, contingent coding	.104	.131	.058

contingent and noncontingent coding would not in general lead to the same amounts of information transmission.

Sequential effects: Run analyses.—The information statistics presented above examine sequential effects in a manner which assumes that the extent of sequential dependency is independent of the particular sequence considered. Clearly that assumption can be only a first approximation. The literature suggests that one kind of past history is especially likely to lead to sequential effects: homogeneous runs of previous outcomes. Rather than examine such runs by information methods, it is easier to examine conditional probabilities based on them directly. Figure 4, again based on the last 480 trials only, shows the conditional probability (multiplied by 100) that L will be predicted given each possible preceding homogeneous outcome run of length eight or less. The data do not permit these probabilities to be estimated for longer runs with acceptable accuracy. An example may make the interpretation of the *x* axis easier. The value 4 on the right run side of the *x* axis means, for example, that the points plotted above it are conditional probabilities of predicting L given that the last five outcomes preceding the prediction were lrrrr. (Note that one actually knows five preceding outcomes, not four, since the outcome preceding a homogeneous outcome run of r must necessarily be l, and vice versa.) Figure 4 justifies the conclusion that outcome runs of length up to four certainly influence responses, and so indicates that for at least some past histories the extent of sequential dependencies is longer into the past than the information analysis taken alone would suggest. But the nature of the dependencies is that the longer an outcome run gets, the more likely *S* is to predict that outcome. What happened to the gambler's fallacy?

Most experiments which have found gambler's fallacies used fewer trials than this one. Perhaps the gambler's fallacy is a phenomenon of early trials and vanishes later. If so, strictly speaking no run curves like those in Fig. 4 are appropriate to use in studying it during early trials, while response probabilities are changing rapidly. But it is reasonable to assume as a first approximation that

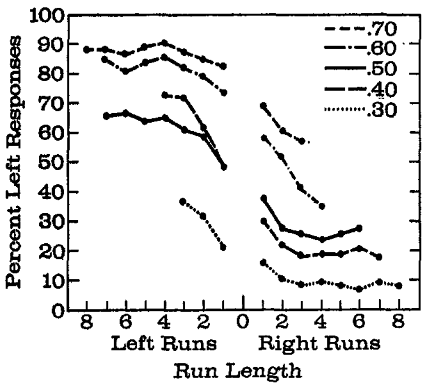

FIG. 4. Percentage of left responses following homogeneous outcome runs. (The *x* axis indicates the number of left or right outcomes included in the run for further explanation, see text. Only the last 480 trials were used.)

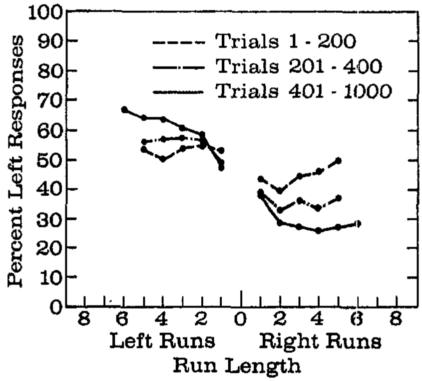

FIG. 5. Percentage of left responses following homogeneous outcome runs for 50–50 group *S*s only. (The axes have the same meaning as those in Fig. 4.)

at least for the 50–50 groups the overall probabilities are not changing very fast, and so curves like those in Fig. 4 can be based on early trials for those groups. Figure 5 presents such curves for Trials 1–200, 201–400, and 401–1000 for all 50–50 *S*s. A small gambler's fallacy, much smaller than any previously reported, appears in the first 200 trials; thereafter the pattern of run effects systematically shifts in the direction of those found in Fig. 4.

Discussion

Probability matching.—In 1956 I reviewed all experiments relevant to a narrow definition of PMH published up to that time (Edwards, 1956, pp. 184–185). Only experimental groups in which the two outcomes were mutually exclusive and exhaustive, in which successive outcomes were independent, in which π was not 0, 0.5, or 1, and in which *S* had had no previous experimental experience with a different value of π were considered. Of 11 groups meeting these conditions, only 1 had an estimated p_∞ which was equal to or less extreme than π. In the other 10 groups, p_∞ was always more extreme than π. The differences were small, but they were all in the same direction.

Of experiments containing relevant groups published since then, those by Gardner (1957), Cotton and Rechtschaffen (1958), and Nicks (1959) are inconsistent with PMH; those by Neimark (1956), Engler (1958), Neimark and Shuford (1959), and Rubinstein (1959) support PMH. No probability learning experiments (as here narrowly defined) reviewed in 1956 or published since then used more than 300 trials at a fixed probability except those by Gardner (1957), Cotton and Rechtschaffen (1958), and Nicks (1959), all three of which are inconsistent with PMH. Figure 1 indicates that in this experiment probabilities of choice were still becoming more extreme at Trial 300 and beyond. Longer experiments at fixed π values might perhaps have produced fewer acceptances of PMH.

Why did PMH, at best dubiously supported by experimental data, achieve such widespread acceptance as a well-established truth? Three reasons seem plausible. First, it is a good first approximation to the truth. It is more nearly correct than the assertion that $p_\infty = 0.5$ for any value of π, or that $p_\infty = 1$ whenever π is greater than 0.5. Furthermore, it is predicted by some (not all) stochastic learning models, which themselves are good first approximations to the truth. Secondly, few experiments have run enough trials to obtain a reasonable estimate of p_∞. Inclusion of trials on which p_t is still changing substantially as a function of t in estimates of p_∞ will, of course, produce estimates of p_∞ which are less extreme than they should be, and so come closer to supporting PMH than they should. (The use of cumulative relative frequency as an estimator of p_∞, as in Estes [1957], will of course bias the estimates in favor of PMH still more.) Finally, the custom of obtaining an estimate of p_∞ and testing the null hypothesis that that estimate is not significantly different from π is widespread in the probability learning literature (and was done in this paper). Such a procedure constitutes attempting to prove a null hypothesis; the smaller the amount of data or the greater its variability, the more likely it is that such a procedure will "confirm" PMH. This is why the small but consistent disagreements with PMH revealed by most probability learning experiments have not been noticed.

The RELM rule.—The extreme-asymptote generalization is not very specific. The data from the previous experiment and from this one are consistent with a much more specific hypothesis called the Relative Expected Loss Minimization (RELM) rule (Edwards, 1956, pp. 182–185). That rule includes but goes beyond the extreme-asymptote generalization, and is applicable to a wide variety of experiments. For this kind of experiment, the linear form of that rule predicts that $p_\infty = 0.5 + K(4\pi - 2)$, where

K is a fitted constant greater than 0.25. The size of K presumably varies with motivational and other characteristics of the experimental design. A least squares fit shows that for the data obtained in this experiment $K = 0.395$.

Sequential effects.—The surprise in this experiment is the weakness of the gambler's fallacy found, and its disappearance in later trials. Nicks (1959), Anderson (1960), and Anderson and Whalen (1960) found much larger gambler's fallacies in appropriate groups; in fact, Anderson found gambler's fallacy effects even when his sequences were designed so that the probability of an outcome repetition was higher than it would have been had successive outcomes been independent. (Jarvik [1951] also found large gambler's fallacies, but his experiment was so designed that they were not at all fallacious.) But this experiment does not stand alone; Feldman (1959b) found no gambler's fallacy at all in his 200-trial experiment.

No real explanation of this divergence in presumably similar experiments is apparent. It is possible, however, that the relative inconvenience of the responses in this experiment served to increase the monotony of what was in any case an exceedingly monotonous task. The gambler's fallacy is in a sense a highly intellectual response. The S must have some idea of what probabilities are and also must to some degree keep track of several preceding outcomes in order to exhibit it. For this non-college population boredom may reduce the amount of intellectual effort applied to the task below the level necessary to sustain a gambler's fallacy.

The gambler's fallacy is important because it is inconsistent with most reinforcement theories. Bush and Morlock[2] have formulated a general conditioning axiom which in effect asserts that gambler's fallacies cannot occur. They have proposed a procedure for examining run effects different from that used in Fig. 4 and 5; they examine only the responses and outcomes included in outcome runs of a specified length (or longer). These data were analysed by their method for Run Lengths 5 and 7. The results were essentially similar to those in Fig. 4 and 5, but the greatly decreased number of observations per point resulted in a considerable decrease in stability. The evidence about the general conditioning axiom from this experiment remains ambiguous.

Hypothesis-testing behavior.—Goodnow and her collaborators (e.g., Goodnow, 1955; Goodnow & Postman, 1955), Feldman (1959a), and I (Edwards, 1956) have argued that people base predictions in probability learning on local hypotheses about sequential dependencies. This idea is very attractive; the sequential effects examined in this paper make it more so. Unfortunately, too many hypotheses (most necessarily incorrect) are possible, and they change too fast and too irregularly, to make this an easy idea to use. Feldman, working with verbal statements as well as predictions, has found it necessary to construct one hypothesis per S. This is the end point of any attempt to give a detailed, explicit account of probability learning from a hypothesis-testing point of view. We need higher order models, so that each specific set of hypotheses can be included within some more general classificatory or explanatory scheme. No such models are available at present.

Summary

A probability learning experiment is reported in which each of 120 Ss made a sequence of 1,000 predictions about which of two mutually exclusive events will occur. After each prediction, one of the two events occurs; the probability of occurrence of each event is constant (0.5, 0.6, 0.7 and their mirror images). Sequences were randomized in two different ways. For all relevant groups, the asymptotic probability of prediction was more extreme than the probability of occurrence of the event predicted; probability matching did not occur. The Ss responded to small increases or decreases in the relative frequency of an event in a block of trials by similar small increases or decreases in their predictions of that event in that block; this phenomenon was named probability following.

[2] Bush and Morlock, personal communication.

Examination of sequential dependencies by means of information measures indicates that about 25% of response information can be accounted for by the identity of *S*s and the results of the last three trials. The *S*s apparently pay most attention to previous outcomes, and much less attention to previous responses. Most of the predicting is done by the immediately preceding trial; trials further back contribute only small amounts of additional transmitted information.

Analyses of homogeneous outcome runs on later trials show that the longer the run of occurrences of an event, the more likely *S* is to predict that event. For early trials, however, *S*s show a slight tendency to predict the event less often as its run length increases; this is the gambler's fallacy.

References

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

ANDERSON, N. H., & WHALEN, R. E. Likelihood judgments and sequential effects in a two-choice probability learning situation. *J. exp. Psychol.*, 1960, **60**, 111–120.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

COTTON, J. W., & RECHTSCHAFFEN, A. Replication report: Two- and three-choice verbal-conditioning phenomena. *J. exp. Psychol.*, 1958, **56**, 96.

EDWARDS, W. Methods for computing uncertainties. *Amer. J. Psychol.*, 1954, **67**, 164–170.

EDWARDS, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–188.

EDWARDS, W. Measure of information based on overlapping counts for studying sequential dependencies. *Amer. J. Psychol.*, in press.

ENGLER, J. Marginal and conditional stimulus and response probabilities in verbal conditioning. *J. exp. Psychol.*, 1958, **55**, 303–317.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

ESTES, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617.

ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol*, 1954, **47**, 225–234.

FELDMAN, J. An analysis of predictive behavior in a two-choice situation. Unpublished doctoral dissertation, Carnegie Institute of Technology, 1959. (a)

FELDMAN, J. On the negative recency hypothesis in the prediction of a series of binary symbols. *Amer. J. Psychol.*, 1959, **72**, 597–599. (b)

GARDNER, R. A. Probability-learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.

GARNER, W. R., & MCGILL, W. J. The relation between information and variance analyses. *Psychometrika*, 1956, **21**, 219–228.

GOODNOW, J. J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

GOODNOW, J. J., & POSTMAN, L. Learning in a two-choice probability situation with a problem-solving setting. *J. exp. Psychol.*, 1955, **49**, 16–22.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1951, **42**, 1–5.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

MCGILL, W. J. Multivariate information transmission. *Psychometrika*, 1954, **19**, 97–116.

MILLER, G. A., & MADOW, W. G. On the maximum likelihood estimate of the Shannon-Wiener measure of information. *USAF CRC tech. Rep.*, 1954, No. RE-54-75.

NEIMARK, E. D. Effect of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.

NEIMARK, E. D., & SHUFORD, E. Comparison of predictions and estimates in a probability learning situation. *J. exp. Psychol.*, 1959, **57**, 294–298.

NICKS, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

RUBINSTEIN, I. Some factors in probability matching. *J. exp. Psychol.*, 1959, **57**, 413–416.

(Received August 4, 1960)