Psychological Reports, 1966, 19, 311-324. © Southern Universities Press 1966

PATTERN DISCRIMINATION LEARNING WITH RHESUS MONKEYS[1]

SANDRA R. BLEHERT

Stanford University

Summary.—Rhesus monkeys were trained to criterion on a 2-stimulus and a 5-stimulus pattern discrimination task. The probabilities of response to the various stimuli throughout learning are examined for individual *Ss*, and it is found that *Ss* exhibit consistency in the order and manner in which incorrect stimuli are eliminated. This suggests a simple mathematical description of the process, which is used to deepen the analysis of the data, permitting estimation of individual learning parameters and construction of more meaningful summaries of the group data.

There are very few published data on pattern discrimination in primates that would allow one to choose among several alternative descriptions of the learning process. For example, can such learning be described by models that predict a gradual divergence in the probabilities of correct and incorrect responses (e.g., Burke & Estes, 1957) or is there an appreciable period of random choice among the alternatives before any divergence is observed (Atkinson & Estes, 1963)? When more than two alternatives are present, do all of the incorrect responses begin to drop out together (possibly at different rates) or are they eliminated consecutively?

Using retarded children, Zeaman and House (1961) found that the rise in the probability of a correct response on a two-alternative discrimination was preceded by a period of random responding and that the length of this period depended on the mental age of *Ss* (the rate of the rise once it began seemed to be relatively independent of mental age). Unpublished observations from K. H. Pribram's laboratory suggest that similar results are found with subhuman primates, at least where the discrimination is of appreciable difficulty. No evidence seems to be available concerning the second question.

The data reported here are from Rhesus monkeys trained on a two-alternative and on a five-alternative pattern discrimination. The results for individual *Ss* are presented and are analyzed with the help of a simple mathematical model which permits estimation of learning rate parameters and the trials on which stimulus discrimination begins.

[1]This research was supported by USPHS Grants MH-03732, MH-15, 214 HD-00918-04 and U. S. Army Contract DA-49-193-MD 2328. The author wishes to thank R. C. Atkinson and K. H. Pribram for their assistance in the preparation of this manuscript. The experiment was conducted according to the APA statement of "Guiding Principles for the Humane Care and Use of Animals," December 15, 1962.

METHOD

Subjects

*S*s were eight Rhesus monkeys 12 to 18 mo. old at the start of training. They comprised the Normal (Nos. 160, 162, 165, and 170) and Crosshatch (Nos. 158, 159, 161, and 166) groups of a study dealing with the effects on discrimination ability of two types of lesions in the inferotemporal neocortex (Pribram, Blehert, & Spinelli, 1966). Prior to discrimination training the Crosshatch *S*s underwent surgery in which a net of small vertical subpial cuts was made bilaterally in the inferotemporal cortex. No differences could be detected between the Crosshatch and Normal groups on any of the tasks, and they will be treated together in this report (see Results and footnote 5).

Apparatus

All testing was carried out in the DADTA machine (Pribram, Gardner, Pressman, & Bagshaw, 1962), an apparatus for programmed presentation of stimulus patterns and automatic response recording. *S* sat in a travelling cage facing an array of 16 plastic panels (4×4 array), onto which the stimulus patterns were projected from the rear. When *S* pressed any panel on which a pattern appeared, a microswitch was activated and the discriminanda disappeared for 6 sec., after which they reappeared on another randomly chosen set of panels. If the stimulus chosen was the correct one, a peanut was delivered into a cup in the center of the array. If an incorrect stimulus was pressed, nothing happened during the intertrial interval. Responses to panels on which no pattern was displayed produced no change in the display. Presentation of stimuli was controlled automatically from an adjacent room and the responses (position and identity of the stimulus chosen) were recorded on punched tape. An overhead light provided illumination in the testing cage, and the noise of a blower masked extraneous sounds.

Pretraining

*S*s were trained to enter travelling cages from their home cages and their behavior was gradually shaped in the testing apparatus to press any panel on which an illuminated pattern appeared. The pattern used for pretraining was the number 1. The number of lighted 1s was gradually reduced from 12 to 1, and shaping continued until *S* responded about 60 times over 2 consecutive days to presentation of a single 1.

Throughout training (except just prior to surgery and during the 2 wk. allowed for recovery) *S*s were fed 8 to 10 standard lab pellets per day and an occasional orange in addition to the peanuts they obtained during testing. Surgery for the Crosshatch *S*s followed pretraining. After recovery, retention of the pretraining responses was checked for all of the animals and then training began.

Training

Fifty trials were given per day. The only exceptions were the 30 trials of the first day and the few times when *S* refused to test. All discrimination training was continued until *S* reached a criterion of 90% correct on 2 consecutive days. The discriminations which *Ss* learned were as follows.

Two-stimulus discrimination.—The discriminanda were the numbers 3 and 8. The positive stimulus was 3. On each trial the following sequence of events occurred. The 3 and 8 appeared on 2 panels chosen randomly from among the 16 panels available. *S* made a choice by pressing one of the 2 panels; if the choice was correct he was rewarded. In the event of either a correct or an incorrect response the stimuli disappeared for the 6-sec. intertrial interval and then reappeared on two other randomly chosen panels.[2]

Multiple-stimulus discrimination.—The 5 stimuli for this discrimination were the capital letters A, H, K, N, and M. The M was the rewarded stimulus pattern. The sequence of events on each trial was identical to that of the previous discrimination except for the number of patterns displayed.

Results and Discussion

The trials to criterion, N_c, on the two discriminations (Table 1) indicate that the 3-8 discrimination was considerably easier than the multiple discrimination.[3]

The individual learning curves of Fig. 1 show that the probability of a response to the correct cue in the 3-8 discrimination, Pr(3), does not rise grad-

TABLE 1
TOTAL TRIALS TO CRITERION, N_c, FOR EACH *S* ON TWO-STIMULUS AND MULTIPLE-STIMULUS DISCRIMINATIONS

S	N_c 3 vs 8	N_c Multiple
N-160	380	800
N-162	280	400
N-165	380	550
N-170	450	700
C-158	480	400
C-159	280	575
C-161	680	750
C-166	230	450
M, including Subject 170	395.0	578.1
M	387.2	560.7

[2]A very simple discrimination between a Red and a Green circle was presented between the 2-alternative and the 5-alternative discriminations. The results are not relevant to this paper. They can be obtained from the report of the inferotemporal lesions.

[3]Subject 170 has been dropped from the rest of the analysis. The taped records of his performance on 3 days of the multiple discrimination were incomplete and only the number of correct responses could be determined.

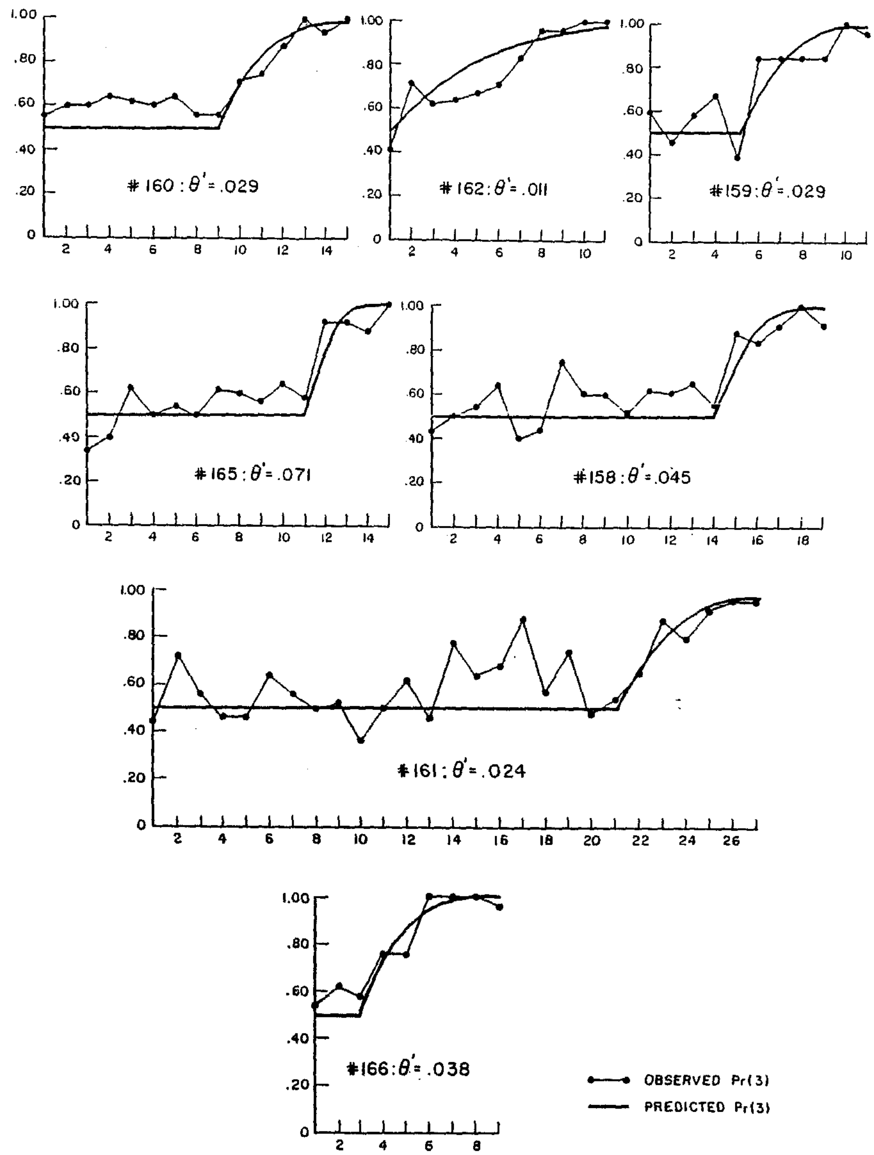

FIG. 1. Observed and predicted proportion correct responses for individual *S*s on the two-stimulus discrimination. Each point is the mean for a 25-trial block (the first point has 30 trials). The values of N' and θ' used in constructing the theoretical curves are those given in Table 2.

ually from the beginning of training. For all *S*s except one there are *at least* two days during which no change occurs in Pr(3). This period is succeeded by a gradual increase in $Pr(3)$ to the criterion level. The one exception, Subject 162, starts to discriminate by the second block of 25 trials. In contrast to these individual learning curves, the mean learning curve (obtained by assuming each S to be responding perfectly after criterion) rises gradually throughout with only a small initial negative acceleration to hint at what is occurring in the individual S's data (Fig. 2a).

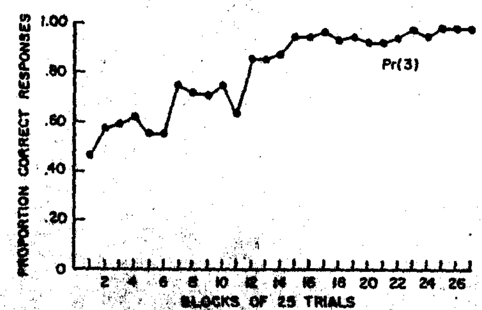

FIG. 2a. Mean proportion correct responses, $Pr(3)$, on two-stimulus discrimination. Each point is the mean for a 25-trial block and includes all of the Ss by assuming no errors after criterion ($N = 7$).

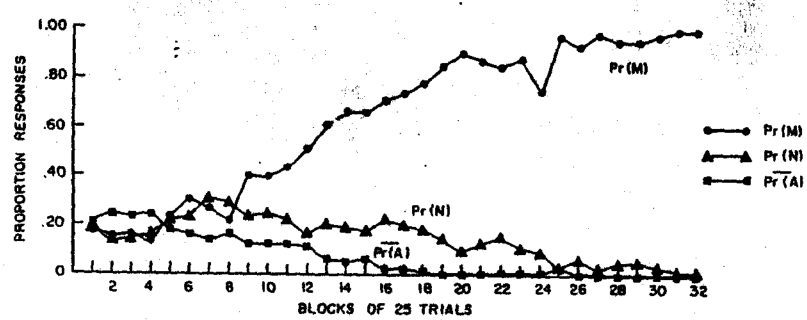

FIG. 2b. Mean proportion responses to each stimulus on multiple discrimination in 25-trial blocks ($N = 7$)

That this long initial period during which there is no change in the response probabilities is not peculiar to the two-choice situation and is not due to the monkey's lack of familiarity with the discrimination procedure, is demonstrated by the individual learning data from the multiple stimulus problem (Fig. 3). All of the cues are chosen with equal probability before the three cues, A, H, and K, begin to drop out simultaneously. No discrimination among these cues was evidenced—in all cases choices of *each* cue dropped suddenly to about 1 or 2 in 25, hence the average value, $\overline{Pr(A)} = \frac{1}{3}[Pr(A) + Pr(H) + Pr(K)]$, was plotted to minimize confusion of the various curves. As these cues

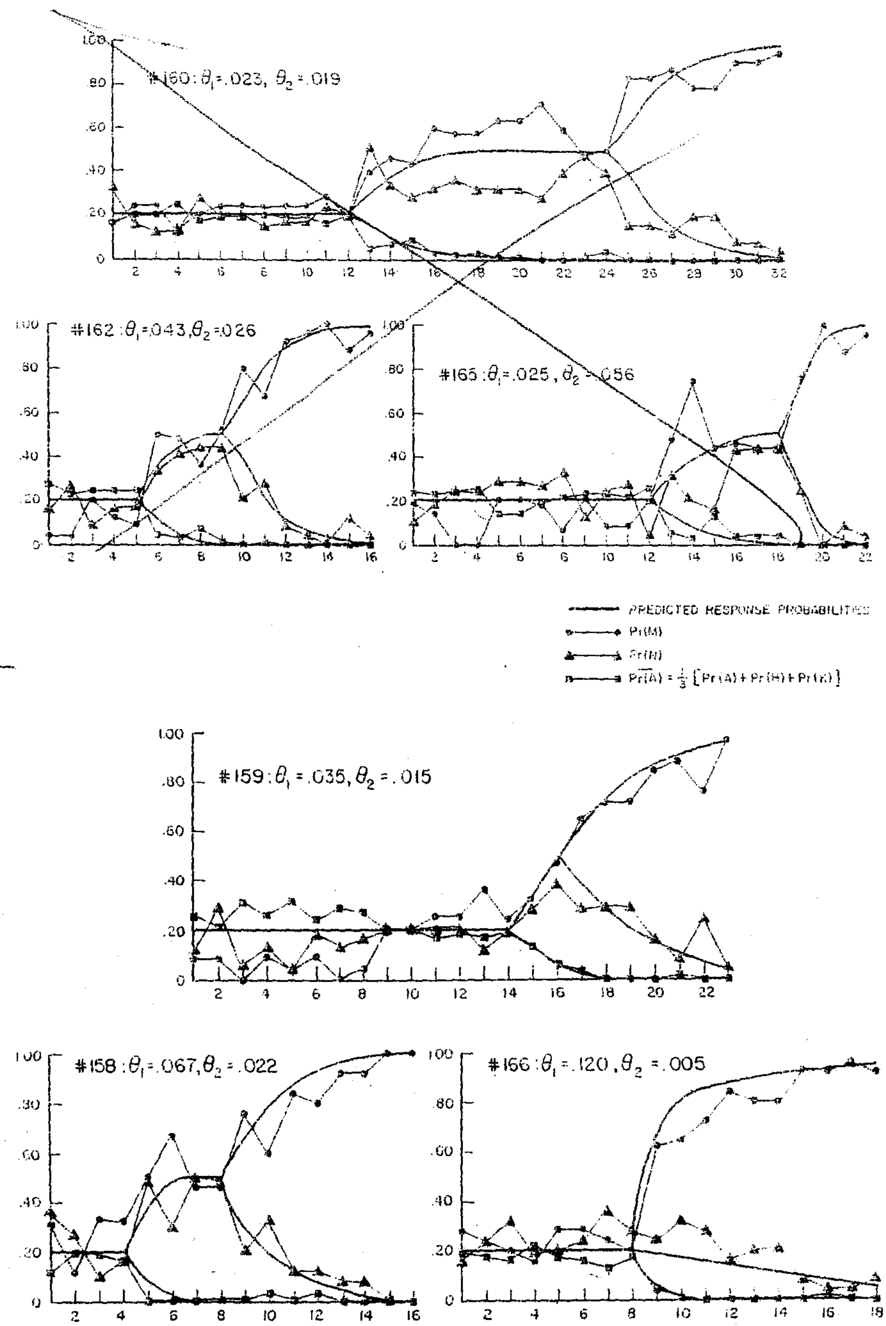

FIG. 5 Observed and predicted proportion responses to each stimulus for individual *Ss* on multiple discrimination in 25-trial blocks.

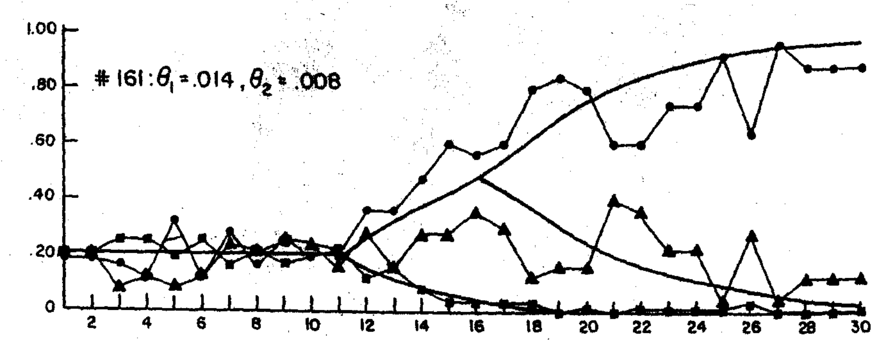

FIG. 3 (Cont'd). Observed and predicted proportion responses to each stimulus for individual *Ss* on multiple discrimination in 25-trial blocks. The values of N_1, N_2, θ_1, and θ_2 used in constructing the theoretical curves are those given in Table 2. $\overline{Pr(A)}$ is the mean proportion responses to the *A*, *H*, and *K* stimuli, i.e., $\overline{Pr(A)} = \frac{1}{3}\,[Pr(A) + Pr(H) + Pr(K)]$.

drop out, choices are divided among the *M* and *N*, whose probabilities thus increase to about ½, until they are discriminated from each other. As in the 3-8 discrimination, the mean learning curve for all *Ss* (Fig. 2b) does not reflect accurately what is happening in the individual curves. The initial period of equal choice among the cues is evident, but then learning seems to be very gradual with stimulus *N* dropping out at a somewhat slower rate than the AHK group.

A more detailed examination of these observations can be achieved by writing a set of equations which describe the response probabilities throughout learning. Consider first the two choice situation. Let n represent the trial number, and let N be the trial on which discrimination between the two stimuli first begins. Then assuming for simplicity that the cues drop out at a rate given by θ':

$$Pr(3_n) = \begin{cases} \frac{1}{2} & \text{, for } n \leq N' \\ 1 - \frac{1}{2}(1-\theta')^{n-N'} & \text{, for } n > N' \end{cases}$$

where of course $Pr(3_n) = 1 - Pr(8_n)$.

The multiple-choice case can be formulated in a comparable manner, where N_1 represents the trial on which the AHK group begins to drop out, and N_2 represents the start of discrimination between the *M* and the *N*:

$$Pr(M_n) = \begin{cases} 1/5 & \text{, for } n \leq N_1 \\ (1/2)\,[1-(1-\theta_1)^{n-N_1}] + (1/5)(1-\theta_1)^{n-N_1} & \text{, for } N_1 < n \leq N_2 \\ 1-(1/2)(1-\theta_2)^{n-N_2} & \text{, for } n > N_2 \end{cases}$$

$$Pr(N_n) = \begin{cases} 1/5 & \text{, for } n \leq N_1 \\ (1/2)\,[1-(1-\theta_1)^{n-N_1}] + (1/5)(1-\theta_1)^{n-N_1} & \text{, for } N_1 < n \leq N_2 \\ (1/2)(1-\theta_2)^{n-N_2} & \text{, for } n > N_2 \end{cases}$$

$$Pr(A_n) = \begin{cases} 3/5 & \text{, for } n \leq N_1 \\ (3/5)(1-\theta_1)^{n-N_1} & \text{, for } N_1 < n \leq N_2 \\ 0 & \text{, for } n > N_2\,. \end{cases}$$

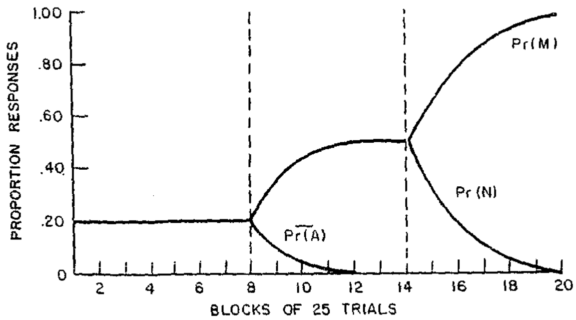

FIG. 4. Example of predicted proportion responses to each stimulus in multiple discrimination, 25-trial blocks. $Pr(A)$ is mean proportion responses to the A, H, and K. Parameter values used are the average values for the group: $N_1 = 205.7$, $N_2 = 353.6$, $\theta_1 = .047$, and $\theta_2 = .022$.

Fig. 4 shows the learning curve generated for the multiple discrimination by these equations using illustrative values of the parameters N_1, N_2, θ_1, and θ_2. Prior to N_1 the cues are chosen equally. At N_1 the AHK group begins to drop out with a rate given by θ_1, and $Pr(M_n)$ and $Pr(N_n)$ rise together to ½. At N_2 the stimulus N begins to drop out with rate θ_2, and $Pr(M_n)$ rises to 1.[4] The form of the learning curve for the 2-stimulus discrimination will be apparent from the preceding discussion.

Estimates of N', N_1, and N_2 for each S were obtained by setting them equal to the number of the last trial of the block preceding that block in which the

TABLE 2

PARAMETER ESTIMATES FOR EACH S ON TWO-STIMULUS AND MULTIPLE-STIMULUS DISCRIMINATIONS

S	N'	N_o-N'	$\hat{\theta}'$	N_1	N_2	N_o-N_2	$\hat{\theta}_1$	$\hat{\theta}_2$
160	230	150	.029	300	600	200	.023	.019
162	30	250	.011	125	225	175	.043	.026
165	280	100	.071	300	450	100	.025	.056
158	355	125	.045	100	200	200	.067	.022
159	130	150	.029	350	400	175	.035	.015
161	530	150	.024	275	400	350	.014	.008
166	80	150	.038	200	200	250	.120	.005
M	233.6	153.6	.035	205.7	353.6	207.1	.047	.022
σ	160.6	30.2	.018	88.5	140.4	71.6	.034	.016

Note.—Estimates are given for N', θ', N_1, N_2, θ_1, and θ_2, and for the values of $N_o - N'$ and $N_o - N_1$ resulting from these estimates. Means and standard deviations appear in the last two rows.

[4] Note that $Pr(A_n)$ for $N_2 < n$ is arbitrarily set equal to 0. This is not a necessary restriction in the model but is done on the assumption that its actual value by Trial N_2 will be small.

appropriate probabilities appeared to separate (Table 2).[5] More sophisticated methods of estimation could be used, but it was felt that these estimates obtained by visual inspection were sufficient for the purposes of this paper. The reader may determine whether or not significant biases seem to have been introduced by the values chosen. The means of the obtained estimates are: $\overline{N'} = 233.6$, $\overline{N_1} = 205.7$, and $\overline{N_2} = 353.6$. Thus, in the 3-8 discrimination it took an average of approximately 5 days for *S*s to reach N' and about 3 more days to reach criterion. In the multiple discrimination, N_1 was reached in about 4 days, N_2 in 3 additional days, and criterion after another 4 days.

These estimates permit the construction of meaningful summaries of the data for the whole group in the form of modified Vincent curves. The learning curves for each *S* were aligned at Trial N' and the 2 parts thus formed were each Vincentized in fifths (Atkinson, Bower, & Crothers, 1965). For example, Subject 160 has 9 blocks of trials prior to N' and 6 blocks from N' to N_c. The first 9 blocks were collapsed into 5 parts by multiplying $Pr(3)$ in each block by an appropriate fraction. Thus, $Pr(3)$ in the first fifth is given by $(5/9)(.55) + (4/9)(.60) = .57$. This was done for both parts of the 3-8 discrimination and the resulting probabilities were averaged over *S*s. In the case of the multiple discrimination the same procedure was followed, this time aligning *S*s at N_1 and N_2 and Vincentizing in fifths between these points. The results—a two-segment Vincent curve for the 3-8 discrimination and a three-segment Vincent curve for the multiple discrimination—are shown in Figs. 5a and 6a. Each segment of the Vincent curve is drawn so that its length is proportionate to the mean number of trials spent by the group on that segment (e.g., in the ratio of 205.7:147.9:207.1 for the multiple discrimination).

For purposes of comparison, Vincent curves were constructed in the normal manner, that is, by dividing the *total* number of learning trials into 10 parts (3-8 discrimination) in the same manner as above, but without regard for Trial N'. The data for the multiple discrimination was divided into 15 such parts. These Vincent curves appear in Figs. 5b and 6b. As would be expected, this averaging procedure tends to obscure the significant features of the data, although the initial period of random choice and the increase in $Pr(N)$ as the AHK group is eliminated are still perceptible. And even though the normal (or unsegmented) Vincent curves are not as informative as those constructed using the parameter estimates, they still preserve more information than do the mean learning curves of Figs. 2a and 2b.

Consideration of Table 2 reveals a very small variability among *S*s in the number of trials between N' and criterion, N_c. The standard deviation of the

[5]This analysis was originally undertaken to find out if there were *any* differences between the normal and operate *S*s. They do not differ significantly in total trials to criterion, in the trials spent in each segment of their learning curves, i.e., the values of N', N_1, and N_2, or in the estimated learning-rate parameters (Mann-Whitney *U* test, Siegel, 1956).

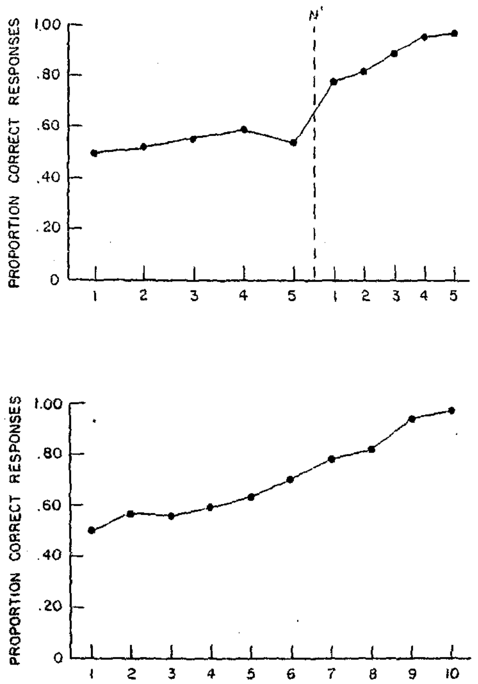

FIG. 5. (a) Two-segment Vincent curve for 2-stimulus discrimination. Each point within the segments before and after trial N' is the mean proportion correct responses for one-fifth of that segment (see text) ($N = 7$). (b) Regular Vincent curve of proportion correct responses in 2-stimulus discrimination. Each point is mean proportion for one-tenth of the total trials ($N = 7$).

length of this segment is 30.2 trials as compared to 160.6 trials for the standard deviation of N'. This suggests that the total number of trials which S requires to learn the discrimination is primarily determined by the length of the initial segment, rather than by the rate at which the discrimination is attained after it once begins. If this is correct, the variability of N_1 and of $N_2 - N_1$ in the multiple discrimination would be expected to be greater than the variability in $N_c - N_2$. The observed standard deviations are 88.5 trials and 140.4 trials, compared to 71.6 trials in the last segment. The second segment, $N_2 - N_1$, would be expected to have the greatest variability because it includes any variability which does exist in the number of trials taken to eliminate the AHK group. Although one would certainly hesitate to attach too much significance to these values, they seem consistent with the idea that the most variable process is the initial elimination of responses to irrelevant cues before overt discrimination begins. To be very clear on the matter of relative variability would require that one develop a model which made predictions for the Ns on the basis of some parameter(s), whose value(s) could then be estimated. The variability among *Ss* of the values of these parameters as compared with those of θ', θ_1, and θ_2 would provide a sharper answer to this question than does comparison of variability in the number of trials in different segments. It should be noted that one cannot conclude from these data that this variability depends on differences in the abilities of *Ss*, and the idea discussed above does not re-

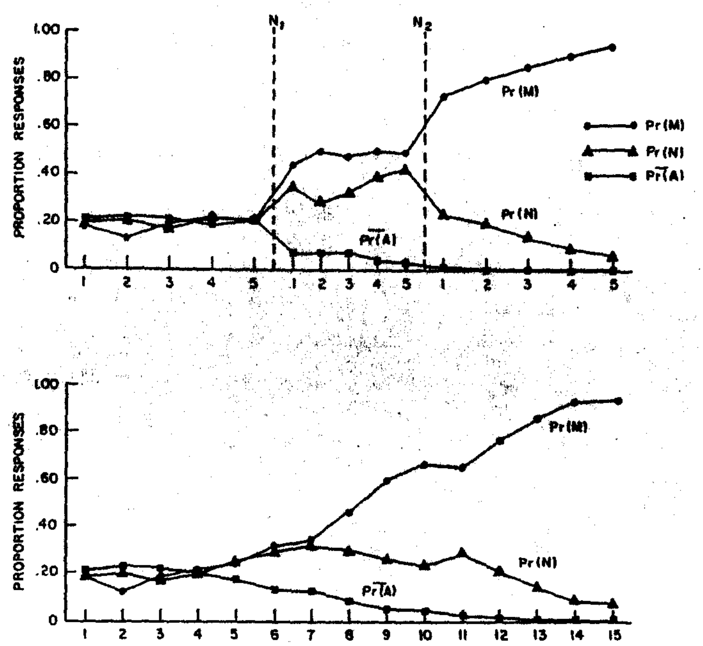

FIG. 6. (a) Three-segment Vincent curve for multiple-stimulus discrimination. Each point within the segments marked off by trials N_1 and N_2 is the mean proportion of responses for one-fifth of that segment ($N = 7$). (b) Regular Vincent curve of proportion responses to each stimulus in multiple-stimulus discrimination. Each point is mean proportion for one-fifteenth of the total trials ($N = 7$).

quire that it do so. In fact, there is a lack of correlation between the lengths of the initial segments, N' and N_1, on the two discriminations ($\rho = .13$). Zeaman and House (1961), however, did find that the initial plateaus of their *Ss* were related to individual differences in mental age. They had no evidence that the rate of divergence was dependent on mental age.

The estimated values of N', N_1, and N_2 can also be used to obtain estimates of θ', θ_1, and θ_2 for each *S*. For θ' this was done by setting the observed number of incorrect responses after N' equal to the theoretical value and solving for $\hat{\theta}'$:

PRIBRAM, K. H., BLEHERT, S. R., & SPINELLI, D. N. The effects of crosshatching and undercutting the inferotemporal cortex on discrimination learning and retention. *J. comp. physiol. Psychol.*, 1966, in press.

PRIBRAM, K. H., GARDNER, K. W., PRESSMAN, G. L., & BAGSHAW, M. An automated discrimination apparatus for discrete trial analysis (DADTA). *Psychol. Rep.*, 1962, 11, 247-250.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

ZEAMAN, D., & HOUSE, B. *An attention theory of retardate discrimination learning.* From the Psychological Laboratories of Mansfield State Training School and the Department of Psychology of the University of Connecticut, 1961. (Progress Report No. 3, Vol. 2)

Accepted June 6, 1966.