Memory & Cognition
1973, Vol. 1, No. 3, 268-276

Attention and the measurement of perceptual learning*

DAVID LaBERGE
University of Minnesota, Minneapolis, Minnesota 55455

Novel and familiar letters were presented to Ss under conditions which controlled momentary attention states. The latencies of letter matching for the novel and familiar letters did not differ when Ss were expecting the particular letters which were presented. However, latencies to the two types of letters differed significantly when Ss were not expecting the particular letters which were presented. Additional exposures significantly reduced this difference, thereby generating a perceptual learning curve in terms of response latency. The main findings were interpreted in terms of a model of perceptual processing which involves mechanisms for hierarchical coding, selective attention, and automatic processing.

When comparing the perception of novel and familiar patterns in identification tasks, it seems reasonable to expect familiar patterns to show faster processing times. However, experiments which use artificial and familiar letters apparently fail to show consistent latency differences when number and familiarity of component features of these patterns are controlled. E. J. Gibson, who has done extensive work with artificial letters (Gibson, 1969), finds no differences in latencies between these patterns and familiar letters in same-different identification tasks (personal communication). The finding of no difference is not likely the result of very rapid learning of new patterns in the first few trials of the tasks. On the contrary, perceptual learning, viewed as feature selection and unitization, is considered to be a slow process as compared with the relatively rapid learning of associations (Estes, 1970).

On the basis of these considerations, it seems appropriate to examine the conditions under which perceptual processing is measured. A recently completed study of attention switching (LaBerge, 1973) showed that familiar stimuli such as tones and color patches may be perceptually analyzed while attention is directed elsewhere. In discussing these findings, it was suggested that perception of unfamiliar patterns might require the services of attention in some manner that simple tones and colors do not. The greater the familiarity of the pattern, the less the degree of attentional involvement during its perceptual analysis. As the patterns become more familiar to the S through many exposures, he might develop mechanisms which somehow automatically analyze the whole pattern before attention is involved.

The procedure which controlled the momentary attention of S during a trial in the color and tone experiment utilized a cueing technique developed by LaBerge, Van Gelder, and Yellott (1970). Each trial contained two stimuli, presented successively. The first stimulus, e.g., a color, served as a *cue* and informed the S as to the most probable stimulus which would appear next. When the second stimulus, the *target*, was the same as the *cue* stimulus, then the S did not need to switch attention to make his matching response. If the *target* was different from the *cue*, e.g., a tone, then the S had to switch his attention to the auditory modality before making a response. Sometimes the tone served as the *cue* and a color the *target*, so that the latency to a color could be obtained both under conditions when the S was expecting it and under conditions when he was not expecting it. It was concluded that tones and colors were processed without attention. This conclusion was based on the fit of two models to the latency data.

SERIAL MODEL

In this model, it is assumed that attention switching and perceptual analysis occur successively. In terms of the time components shown in Fig. 1, the latency of a

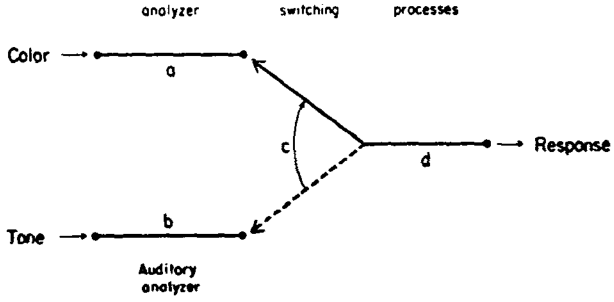

Fig. 1. Model of attention switching between tone and color processors.

---

*This research was supported by United States Public Health Service Grant MH-16270-04 and in part by Grant HD-06730-01 and the Center for Research in Human Learning through National Science Foundation Grant GB-17590. The author thanks Lee S. Brownston and Rohn J. Petersen, who ran the Ss, Steven M. Goldfarb and Anita L. Kaufmann for programming, and Teara Archwamety for statistical analyses. Special gratitude goes to S. Jay Samuels for discussions of all phases of this experiment.

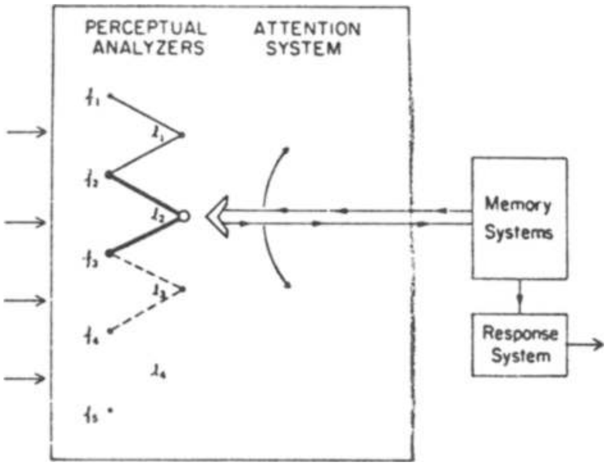

Fig. 2. Model of perceptual processing applied to letter features.

response to color given that the S expects a tone is expressed as: L(color/expect tone) = c + a + d.

PARALLEL MODEL

The assumption here is that some part of the perceptual analysis proceeds simultaneously with attention switching and some part of the perceptual analysis occurs after attention has been switched. Let k be the proportion of the perceptual analysis that occurs after attention has been switched and $1 - k$ be the proportion of perceptual analysis that occurs before attention has been switched. Thus, L(color/expect tone) = c + ka + d, $0 \leqslant k < 1$, $c > a$, where ka is the amount of attention time given to the visual analysis of the color input and $(1 - k)a$ is the time during which visual analysis proceeds automatically.

Data from eight switch and four nonswitch conditions in this study indicated that L(color/expect tone) = c + d, i.e., $k = 0$ and therefore the entire operation of visual analysis proceeded in parallel with the switching operation and was completed by the time attention reached the color analyzer. Thus, the serial model was rejected in favor of the parallel model.

To apply this model and procedure to the processing of letter-like patterns, the model is modified to represent the relationships of the component features of a pattern to the pattern taken as a unit and provision is made for a process by which unfamiliar patterns become familiar through exposure.

MODEL OF PERCEPTUAL PROCESSING

The main properties of the present model are represented in schematic form in Fig. 2. Like other coding models, this formulation rests on notions first described by Lawrence (1963). Incoming stimuli are first analyzed by feature detectors, f_i, and the ouputs of these detectors are organized into letter codes, l_j. Although not shown here, the letter codes could be further organized into codes representing letter clusters, syllables, words, and phrases. This hierarchical scheme assumes that the connections occur not between features, as earlier association theory would assume, but rather between features and higher order units. This part of the model codes features into letters much in the way memory items were coded into larger units in the recent work of Johnson (1972), Lesgold and Bower (1970), and more explicitly in the model for stimulus coding in short-term memory described by Estes (1972). The other main part of the model consists of an attention switching mechanism which is similar in some respects to one proposed by Kristofferson (1967).

The attention system is guided by information from memory and in Fig. 2 has selected the letter code l_2, e.g., the letter "a," and has activated the network into a state of heightened excitability, representing the momentary state of an S who expects to see that letter appear. Heightened expectancy reduces latencies for familiar stimuli, according to the findings of Bernstein and Reese (1965), LaBerge et al (1970), and Hinrichs and Craft (1971). Therefore, when l_2 is presented, the cluster of features is coded rapidly and read out through the attention system into memory stores for a matching operation or, in the case of highly practiced stimulus-response tasks, perhaps directly into a response system. If, on the other hand, l_1 is presented, e.g., the letter "b," then the features f_1 and f_2 are organized into the letter code l_1 automatically, i.e., while attention is in the process of being switched from Letter l_2 to l_1. Since the S has seen the letter "b" many times, the organized trace has become relatively permanent and can support coding of features into letters automatically. In fact, in this model it is assumed that the coding of familiar letters is obligatory much in the same sense as Shiffrin and Gardner (1972) and Eriksen and Spencer (1969) have concluded from experiments with tachistoscopically presented letters. Whether the attention selector chooses features or letters to read out remains optional, however.

We now turn to the case in which a new, unfamiliar letter pattern appears at the moment when S has focused his attention on Letter l_2. Suppose the new letter is represented by l_4. When l_4 appears, attention switches to Feature f_4 and then is switched to Feature f_5, and then additional time may be consumed in organizing these features for the match to be made. If a new letter has been experienced several times, as represented by the added lines at Letter l_3, then the organizing of features into a letter takes less time because some trace of previous coding operations has been laid down. Eventually, with enough exposure, l_3 and l_4 will have a code trace as strong and effective as that of l_1 and the features will process into letters automatically.

This model predicts, therefore, that under a switch task condition, the processing of new letters will take more time than the processing of old letters because, in the case of new letters, the sampling and organizing of feature detector outputs requires active attentive control, while for the old letters, it is accomplished automatically.

Attentional control of information flow can be regarded in two ways in the present model. Firstly, the "positioning" of the attention selector normally takes place prior to the onset of the stimulus and determines what analyzer outputs will be read out first. In addition, the selector may structure and raise the analyzers to a high degree of responsiveness, so that the momentary probability of detection is increased. This use of the term attention, with its emphasis on processes initiated prior to stimulus onset, conforms closely with that of Estes (1972).

A second way that information flow is altered by attention processes in this model is in the sequential sampling and organizing of features after the onset of the stimulus. For example, when the S is shown an unfamiliar pattern for which he has no preparation in the sense of attention first described, he cannot immediately read out from the analyzers that he has primed, but rather must move the selector to the appropriate analyzers, scanning the component features in some order that may or may not be automatically determined, and then organize the components for readout into memory stores for a match test. Since the selection and sequencing of these events occurs after stimulus onset rather than before, this may be regarded as a different type of attention (Shiffrin & Geisler, 1973).

A tacit assumption made for the present experiment is that the perceptual analyzing of Features f_4 and f_5 is as well learned as that of f_1 and f_2, and the only new learning required is the organizing or coding of their outputs into l_3 and l_4. Or, in terms of Gibson's theory (1969), the distinctive features have already been learned and now need only to be structured into larger units. On the other hand, it is not difficult to imagine letter-like patterns containing some unfamiliar features (e.g., Arabic graphemes), in which case perception of the new letter should take more time and perceptual learning should proceed at a relatively slower rate.

What does the model predict when old and new patterns are presented under nonswitch task conditions, i.e., when the S receives the pattern that he expects? In Fig. 2, momentary attention to the old letter l_2, e.g., the letter "a," is represented by a heavy line which is meant to indicate that the analyzer network is raised to a state of high excitability. When, on other trials, the cue directs attention to a new letter code l_4, it is assumed that the S may temporarily organize Features f_4 and f_5 into l_4 prior to the onset of the letter and activate this network to an excitability state as high as that illustrated for the letter l_2. This temporary coding may take place no matter how unfamiliar the letter is, so long as the component features are familiar and there are not too many of them to organize at one time.

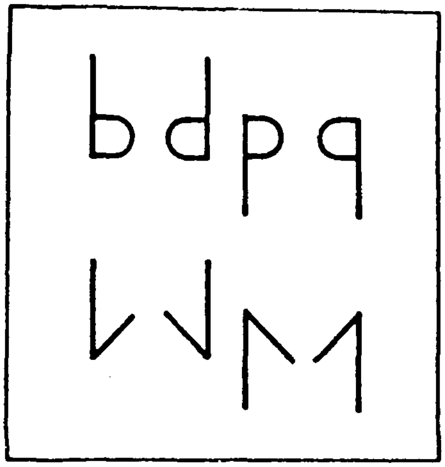

Fig. 3. Old and new letter patterns.

On the basis of these assumptions, then, the processing times for new and old letters are predicted to be equal under nonswitch task conditions. Also, the latencies should be considerably less than the latencies obtained under switch task conditions, because there will be no time component involved for switching attention.

The general procedure used here to control attention can be regarded as a combination of Donders's c method and the cueing technique (LaBerge, Van Gelder, & Yellott, 1970). On each trial, the S was first given a letter, e.g., "a," as a cue which induced him to expect the same letter to reappear after a short blank interval. If that letter reappeared, the S was to press the button; if another single letter appeared, he was not to press the button. On a low percentage of trials, two unexpected letters were presented simultaneously. The S was instructed to press the button when these two letters were the same, regardless of the cue given at the beginning of the trial. Thus, the S pressed the button on two types of trials: one on which he was expecting that particular letter, henceforth to be called a *primary* trial, and one on which he was not expecting a particular pair of matching letters, henceforth termed a *secondary* trial. The important comparisons will be between latencies to new and old letters under primary (nonswitch) and secondary (switch) conditions.

METHOD

Stimuli and Apparatus

Lowercase letters were presented on a 27-cm diagonal television screen by means of a Tektronix scan converter under the control of a NOVA computer. The letters of main interest are shown in Fig. 3. The arrow-like patterns will henceforth be termed "new letters." These were constructed so that component features would be familiar to college Ss and so that the organization of the features would have some formal similarity to that of the four control letters, b d p q. The letters

Table 1
Cue-Stimulus Contingencies for Practice Blocks*

Cue	Stimulus	Response	Frequency
a	a	+	8
a	c	–	1
a	e	–	1
a	o	–	1
a	h h	+	1
a	k k	+	1
a	h k	–	1
a	k h	–	1
n	n	+	8
n	m	–	1
n	u	–	1
n	r	–	1
n	f f	+	1
n	t t	+	1
n	f t	–	1
n	t f	–	1

**Block size = 30*

were formed by placing dots in a 7 by 15 matrix. An old letter was made up of 24 dots and a new letter of 19 dots. In an attempt to compensate for the larger number of dots required for old letters, the height of the new letters was increased by 1 dot. Therefore, the size of the old letters was 1.2 x .8 cm and that of the new letters was 1.4 x .8 cm.

The letters appeared in either the upper half or the lower half of the screen. On a primary trial, a letter first appeared in the upper half of the screen and served as a *cue* informing the S as to what letter to expect to appear next on the lower half of the screen, termed the *target*. On a secondary trial, the *cue* was followed by a pair of *target* letters, both different from the *cue* letter. The first member of the pair appeared in the location where the primary trial letter normally appeared. The second member of the pair simultaneously appeared seven spaces (70 mm) to the right of that letter.

The S sat at a table and viewed the screen at eye level 120 cm from the end of the table. The button was 2.5 cm in diam and was mounted on an inclined plane in front of the S. The S kept the fingers of his preferred hand on the button at all times.

Subjects

There were two groups of Ss, 16 in each group. All were volunteers of college age and had no previous experience with experiments of this type. They were paid according to their performance in the experiment in a manner to be described later.

Design

Practice Blocks

A practice block contained both primary and secondary trials with the letters a and n as primary tasks and the letters f h k t as secondary tasks. The catch trials for the letter a were e, o, and c. The catch trials for the letter n were m, r, and u. The cue-stimulus contingencies within a block are given in Table 1. The proportion of positive responses was always greater than 50% in order to induce the S to focus his attention on the cued stimulus on virtually 100% of the trials (LaBerge et al. 1970). The order of cue-stimulus trials was randomized by the computer.

Secondary Test Blocks

These blocks always used a, g, n, or s as cues and tested either old or new letter pairs. A given block had either old or new letters as secondary tasks. The cue-stimulus contingencies for one of two types of blocks utilizing an old letter secondary task are given in Table 2. The other type of old letter secondary block substituted the letters d and p for b and q under the a cue conditions and made similar types of substitutions under the g, n, and s cue conditions. Two blocks which tested pairs of new letters under secondary test conditions were constructed in *exactly* the same manner as those for the old letter pairs, substituting the four arrow-like letters for the b d p q letters.

Primary Test Blocks

These were constructed with either old letters or new letters as cues. The blocks used only single letter tests. Therefore, S was to respond only to an expected letter. The cue-stimulus contingencies for an old letter block are shown in Table 3. A primary test block with new letters was constructed in a parallel manner.

Procedure

In order to compare latency differences between new and old letters under secondary and primary tests, two groups were tested. One group received their first exposures to the new and old letters under secondary test conditions. The other group received their first exposures to these letters under primary test conditions.

Group 1

The 16 Ss of this group were given six blocks of practice trials on the first day and began each subsequent day with a practice

Table 2
Cue-Stimulus Contingencies for Secondary Blocks*

Cue	Stimulus	Response	Frequency
a	a	+	8
a	c	–	1
a	e	–	1
a	o	–	1
a	b b	+	1
a	q q	+	1
a	b q	–	1
g	g	+	8
g	y	–	1
g	j	–	1
g	a	–	1
g	d d	+	1
g	b b	+	1
g	d b	–	1
n	n	+	8
n	m	–	1
n	u	–	1
n	v	–	1
n	p p	+	1
n	d d	+	1
n	p d	–	1
s	s	+	8
s	c	–	1
s	x	–	1
s	z	–	1
s	q q	+	1
s	p p	+	1
s	q p	–	1

**Block size = 56*

block. Two secondary test blocks followed the practice block on Days 2-6. On Day 2 only, before each secondary block began, the Ss were shown a card for 15 sec. on which the secondary test letters for that block (new or old) were printed. The order in which the new and old blocks were given each day and the type of block given were balanced as much as possible over days and Ss. On Days 4 and 5, two primary blocks were given following the secondary test blocks in an attempt to increase the learning of the new letters.

Group 2

The 16 Ss of this group were treated exactly like the Ss of Group 1 on Day 1. On Day 2, following the practice block, they were given primary test blocks using new and old letters. Then they were given secondary test blocks using these letters. Before each primary block began, the Ss were shown a card for 15 sec on which the letters for that block (new or old) were printed. The orders in which old and new letter blocks were given were balanced across Ss. This group was tested only 2 days.

At the beginning of the first practice trials of Day 1, the Ss were instructed to watch the cue letter in the upper part of the screen and to respond with a buttonpress if the letter which subsequently appeared in the lower part of the screen was the same as the first letter. The S was told that occasionally a pair of letters would appear in the lower part of the screen and that he was to press the button if these letters matched, regardless of the cue letter given on that trial. He was also told that when he responded correctly and fast to the letter which matched the cue letter, he would receive a short burst of low-level noise in the earphones, and that he would be paid 2c for every burst he received but would lose 4c for every error he made. He was also told that he could not receive feedback bursts for fast correct responses to the pairs of letters. These bursts were omitted to discourage Ss from forming expectancies of secondary test trials. Following each block, the S was informed of the number of fast corrects and errors.

The computer determined whether or not a response on a trial was faster than the criterion latency which the E typed into the computer before each block of trials. The criterion latency value for a block was the mean latency of the primary tasks of the previous block for Blocks 2 and 3 on Day 1. Thereafter, 20 msec was added to the mean to calculate the criterion so that Ss received feedbacks on a majority of the primary trials. This was done for both groups on all subsequent blocks.

The duration of the event sequence on a trial was as follows: The cue letter was shown for 1,000 msec, followed by 1,000 msec of blank, followed by the letter or pairs of letters. The letter or letter pair remained on the screen for 1,000 msec unless a correct response occurred and terminated the image.

Table 3
Cue-Stimulus Contingencies for Primary Blocks*

Cue	Stimulus	Response	Frequency
b	b	+	6
b	d	–	1
b	p	–	1
d	d	+	6
d	b	–	1
d	q	–	1
p	p	+	6
p	q	–	1
p	b	–	1
q	q	+	6
q	p	–	1
q	d	–	1

**Block size = 32*

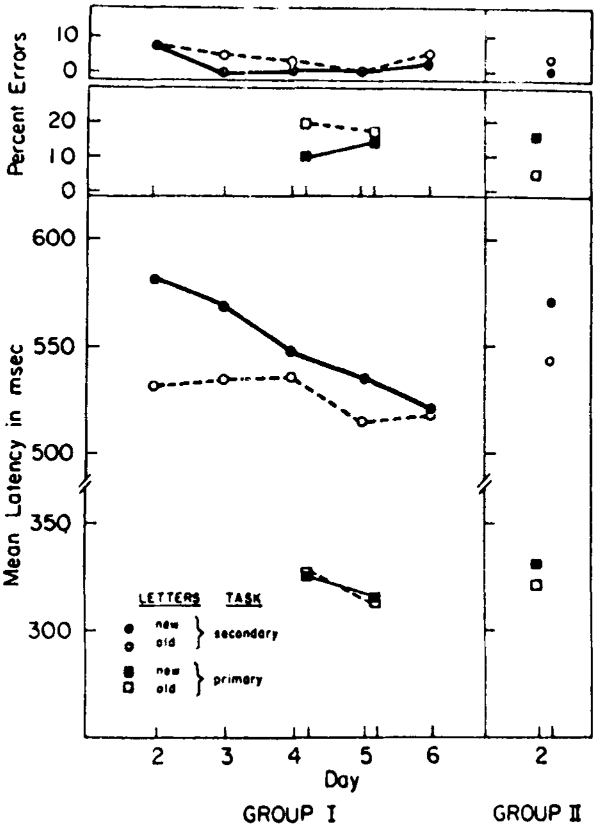

Fig. 4. Mean latencies and percent errors to new and old letters under primary and secondary test conditions. Each group contained 16 Ss.

The intertrial interval was 1,000 msec. About 2 min of rest was given between blocks.

All trial events were controlled by the computer, which recorded latency and frequency data and, at the end of each block, exhibited these data by a line printer.

RESULTS

The comparisons of processing times for correct responses to new and old letters under primary and secondary conditions are shown in Fig. 4. Each point of the secondary latency conditions is based on 8 responses per S or 128 responses. Each point of the primary latency conditions is based on 24 responses per S or 384 responses. For Group 1, the Day 2 difference in mean latency for old and new letters under secondary conditions was 48 msec and is significant [$F(1,112) = 16.19$, $p < .001$]. For Group 2, the Day 2 difference in mean latency for old and new letters under primary conditions was 9 msec and is not significant [$F(1,48) = 3.08$, $p > .05$]. In fact, the mean latency for new letters was less than that for old letters for 9 out of 16 Ss. When Group 2 subsequently received secondary tests of old and new letters, the difference in mean latency was 29 msec and is significant [$F(1,112) = 5.34$, $p < .05$].

Functions showing changes in latency to new and old letters under secondary test conditions over Days 2-6 are given by the data of Group 1. While both curves show a

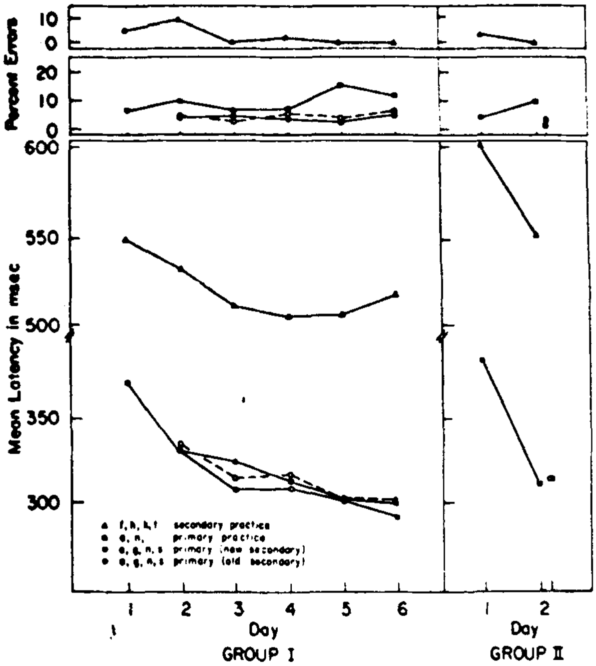

Fig. 5. Mean latencies and percent errors to Letters a and n on primary practice trials, to Letters f, h, k, t on secondary practice trials, and to a, n, g, s on primary trials when either new or old letters were given as secondary tests.

downward trend, the curve for the new letters drops much more rapidly. Specifically, the new letter curve decreases 59 msec from Day 2 to Day 6, while the old letter curve decreases 14 msec over this range, which would indicate that the new letter curve is falling over four times as fast as the old letter curve. The difference in the slopes of the two curves was tested by analysis of variance using the data from all 5 days, and the slope interaction of letter type and day is significant [$F(4,448) = 3.29$, $p < .05$]. In addition to the finding that the two secondary curves of Fig. 4 have different rates of decline, it appears that they converge.

The error data of Fig. 4 indicate that secondary tasks do not produce a high rate of errors; in fact, they produce fewer errors than do primary tasks. Further, the initially longer latency for new letters under secondary conditions is not due to a higher rate of errors for new letters. Rather, it appears that the error rate for new letters is, if anything, slightly lower than that for old letters. As for the errors on primary tasks, the data of the two groups do not indicate a consistent difference.

Mean latencies of errors are lower than latencies of corrects under all conditions. For the primary conditions, errors are 46 msec lower than are corrects over both groups. For secondary conditions, errors are approximately 140 msec lower than are corrects, though there are very few errors available on which to base these estimates. There are no indications that type of letter has any effect at all on error latency under either primary or secondary conditions.

The latency and error data for practice trials and for the letters a g n s under old and new letter secondary conditions are shown in Fig. 5. The mean latency to the primary practice letters a and n decreases rapidly from Day 1 to Day 2 for both groups. The mean latency to the secondary practice letters f h k t shows a similar decrease. Data from a g n s, the primary task letters used when new and old letters were given secondary tests, were separated according to whether a new or old letter had secondary tests during a block. The comparison is given by the two curves of Group 1 plotted from Days 2 to 6. Apparently, there is no difference in latencies for these letters due to type of letter being tested on secondary trials.

The most notable feature of these data taken as a whole is that latency drops most rapidly from Day 1 to Day 2 and that there is further reduction in latency over Days 2-6, though the rate is only about 8 msec per day for the letters a g n s. The reduction in latency over this period for secondary tasks f h k t appears to be only about 4 msec per day.

Error latencies for letters a g n s and f h k t showed the same general pattern as that of old and new letters already reported.

DISCUSSION

The perception model proposed here generates three important predictions about the processing of letter patterns, and all three predictions are clearly confirmed by the data. Firstly, the model predicts that processing time for unfamiliar (new) and familiar (old) letters should not differ when the S is *attending to* the letter pattern at the moment it is presented. Secondly, the processing time for new letters should be greater than that for old letters when the S is *not attending to* the letter pattern at the moment it is present. Finally, the difference in processing time under this nonattention condition should gradually disappear as the S receives additional exposures of the new letter pattern.

If pairs of new and old letters had been presented under a procedure which did not control S's attention on each trial, but which allowed it to fluctuate with the influence of previous trial events and S's idiosyncratic expectations, then the mean latency to a given pair of letters would have resulted from a mixture of latencies from different processing events: some from trials when S was expecting the pair presented, others from trials when S was expecting some other pair. To see what effect this mixture could have on the obtained latency difference between new and old letters, let Fig. 1 be the model and let a be the time to process a new letter (instead of a color) and b the time to process an old letter (instead of a tone). We will omit d from consideration here, because it will drop out when we subtract the equations for new and old letters. Further, we will assume that $c > a$ and $c > b$.

Referring to Fig. 1, we note that the latency of a new letter given that the S expects that new letter is $L(\text{new/new}) = a$, and similarly, $L(\text{old/old}) = b$.

$L(\text{new/old}) = ka + c$. $L(\text{old/another old}) = c$. When we subtract new and old letter latencies under secondary tests, we obtain $L(\text{new/old}) - L(\text{old/another old}) = (ka + c) - c = ka$, which is the indicator used in the present experiment to measure the amount of attention needed to process a new letter.

Consider now a procedure which mixes primary and secondary trials within a block, e.g., giving a pair of new and a pair of old letters without cues but with appropriate catch trials. Suppose that on a given trial the S expects the new letter with probability p and the old letter with probability $(1 - p)$. Then $L(\text{new}) = pL(\text{new/new}) + (1 - p)L(\text{new/old})$ and $L(\text{old}) = pL(\text{old/new}) + (1 - p)L(\text{old/old})$. Subtracting we obtain $L(\text{new}) - L(\text{old}) = [pa + (1 - p)(ka + c)] - [(1 - p)b + pc] = p(a - c) + (1 - p)(ka - b + c)$. This expression does not allow us to estimate differences in perceptual processing of new and old letters free of the particular c, b, and p values involved in a given experiment. However, in the special case that $b = a$ and $p = \frac{1}{2}$, $L(\text{new}) - L(\text{old}) = ka/2$, but this yields *half* the difference obtained by the experimental procedure which allows a direct subtraction of the secondary task latencies. An additional disadvantage of the mixture procedure is that the variance of the latencies will be quite large owing to the fact that the overall mean latency from primary and secondary tasks combined consists of two substantially disparate groups of latencies, depending on the presence or absence of the attention switching time component.

In view of these considerations, it is not surprising that procedures in which momentary attention states are not controlled could fail to show significant differences in the identification time of familiar and unfamiliar letter patterns.

It would seem that the repeated exposures of the novel patterns in this experiment increase the familiarity of the patterns. The indicator of this perceptual learning is latency of pattern identification on trials when S is not expecting to see it. However, this indicator requires a correction factor, owing to the fact that latency of asymptotically learned patterns continues to decrease slightly over trials. Therefore, in order to trace the course of perceptual learning of a new pattern, one must select an appropriate familiar pattern to test concurrently in order to provide a baseline against which perceptual learning of the new pattern can be compared. In the present experiment, the letters b d p q were selected because they are presumed to be at the asymptote of familiarity for college age Ss, and any further decrease in latency in identification tasks should be due to task-related factors such as improvements in orienting to the proper location on the display screen at the right moment, and possibly improvements in time to activate a response.

If one can safely assume that these task-related processes are learned at the same rate regardless of whether one is learning a new pattern or merely experiencing repeated exposures of a pattern already very familiar, then a straightforward subtraction of the two curves should yield a curve which represents perceptual learning of the new patterns.

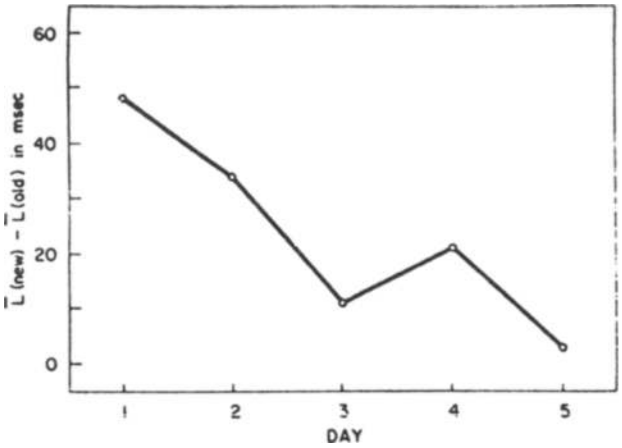

Fig. 6. Perceptual learning curve of new letters based on subtraction of secondary tests of new and old letters shown in Fig. 4.

Figure 6 shows a learning curve obtained by subtraction of the two secondary latency curves of Group 1 of Fig. 4. In terms of the model shown in Fig. 2, this latency learning curve is represented by the formation of organizing links between features and a letter code, a process which could be termed unitization (Estes, 1970) or chunk learning (Miller, 1956). The curve in Fig. 6 presumably is an indicator of the gradual formation of a network at l_4 in Fig. 2. As training proceeds, the links between Features f_3 and f_4 and l_4 become more permanent until the l_4 network looks like the l_1 network, where features are unitized into a letter automatically. Evidently this learning is of the "slow" variety (Estes, 1970), because the secondary latency tests of Group 2 in Fig. 4 show a substantial difference between old and new letters despite the 32 primary trials with new letters given just prior to the tests.

In considering the unitization or chunking process in perception, the present emphasis is not placed on aggregates of feature detectors, but rather on the way the feature detection outputs are organized into a new unit or Gestalt. This viewpoint implies that the particular way lower level outputs are organized into higher level units may be influenced and guided by active attentional processes, an assumption which was made explicitly when the model predicted that, under primary test conditions, one would process new letters as fast as old letters. Preparation for receiving a new pattern should involve not only an activation of feature detectors, but also the temporary activation of their organization into a unit. This activation may be sustained over a period of time by closely allied memory systems acting through attention, so that when the pattern is presented to the receptors, it is analyzed quickly by the unit network and the result is matched in the memory location which has been sustaining the analyzer network in its state of heightened excitability. Organizing at the perceptual level may speed a perceptual match in memory because the match may be made in one operation, as opposed to matching by comparing a list of discrete features. Patterns which are constructed of too many features to be organized as a whole would presumably require serial readouts by moving the positions of the attention selector of Fig. 2 across several features, and the matching operation in memory would then occur sequentially, all requiring more time to accomplish than was the case when the input was first organized at the analyzer level.

The learning curve shown in Fig. 6 presumably reflects one of perhaps several stages of perceptual learning, namely, the organizing or unitizing of the component features of a pattern. Another earlier learning stage is the selection of appropriate features to observe, termed distinctive-feature learning by Gibson (1969), a process related to selection of relevant cues in concept identification tasks. Which particular features of a pattern are selected depends upon the set of patterns that the given pattern is to be discriminated from. For example, variations in height of a vertical line is a distinctive feature in distinguishing the letters h and n, but variation in line thickness is typically not one of the features selected. The control of selection in reaction time experiments by type of catch trial has been investigated elsewhere (Posner & Mitchell, 1967; LaBerge, 1971a, b).

In tasks of the kind used in the present experiment, selective learning of distinctive features typically proceeds by a trial and error search (cf. Zeaman & House, 1963) and the representation of learning is usually given in terms of percent correct response over trials. The unitization stage of perceptual learning, on the other hand, makes considerable gains after error reduction has stabilized, as exemplified in Fig. 4, and this stage of perceptual learning therefore must be represented by some measure other than percent corrects or errors.

The perception model of Fig. 2 is intended to generalize to higher levels, e.g., to letter clusters and words (LaBerge & Samuels, 1973) as well as to visual figures of a nonsymbolic nature, and to patterns in other sensory modalities such as speech and musical patterns.

Modes of organizing may vary in their dependency on attentional control. It is likely that phonemes are organized into syllable units by young children without attention because these children have trouble segmenting syllables into phonemes. They simply are not aware of the way phonemes may be organized into syllables (Savin, 1972). Yet some higher-order organizing of words into sentences may need to be under tight control of attention for some Ss if they are to generate clear and meaningful speech or prose; but for other Ss, this organization apparently comes automatically and therefore "easily."

One kind of situation needed to explore modes of organization is one in which the same features are used to produce two or more different units. An example is the reversible figure, such as the Necker cube. By contextural effects, real or imagined, we seem to be able to move the attention selector from one unit to the other at will and "reverse" the figure without changing the set of features being selected.

Returning to the particular findings of the present experiments, it appears that perceptual learning of a visual pattern may be revealed in terms of processing time if the S is given the pattern at moments when he does not expect it, i.e., under conditions which here are termed secondary tasks. Taken together, these data support a model in which a hierarchical analysis of stimuli can occur automatically if the stimuli have been coded sufficiently often in the past. In fact, perceptual learning of a pattern may then make its analysis into a unit obligatory, while the level of analysis at which attention selects the output remains optional. For example, Ss may be trained to press a button when a letter having a vertical line is presented. When the letter b appears, the S will process the features into the letter code automatically, but his attention system selects for further processing only outputs from the vertical line analyzer.

It is tempting to speculate how association of these patterns with responses might vary depending upon when the association training is initiated during the course of perceptual learning. For example, one could attempt to associate some arbitrary letter name, N(l_4), with the new letter, l_4, during the feature selection stage of perceptual learning, during the unitization stage, or after the unitization stage has become automatic. If associative learning is attempted before the relevant features are selected, then the response may be linked with irrelevant features, resulting in confusions and response errors. If the associative learning is attempted at the end of the selection stage but before unitization has begun, the associative links may be formed between separate features and the responses, resulting in confusions when other letters which share these common features are presented. But if associative learning is held off until unitization has progressed to some level, then perhaps only one associative link need be formed between the pattern and the response. In this way, confusions should be minimized and associative learning should, in fact, proceed quite quickly. With the simple new letters of the present study, which presumably may be temporarily unified by attention, single associative links may perhaps be laid down before unitization is completely automatic. But if the patterns resist such temporary unitization by attention, owing to a large number of component features, or if some features themselves are not automatically processed, then single-link associative learning should be unlikely at this stage and additional unitization training should be necessary before associations to the pattern as a whole can be accomplished.

At the risk of oversimplification, one might propose the following rule of thumb for treatment of perceptual learning: When an S is learning to identify a pattern, he should focus attention on it in order to select the appropriate features and to organize them into a unit. When he is being tested on the degree of learning accomplished, he should *not* focus attention on the pattern, but rather focus attention elsewhere at the moment the pattern is presented. If the momentary focus of attention is critical in the acquisition and testing of perceptual learning, then it would seem highly desirable to design our presentation procedures to put attention under the control of the E rather than leaving it to fluctuate at the whim of the S.

REFERENCES

Bernstein, I. H., & Reese, C. Behavioral hypotheses and choice reaction time. Psychonomic Science, 1965, 3, 259-260.

Eriksen, C. W., & Spencer, T. Rate of information processing in visual perception: Some results and methodological considerations. Journal of Experimental Psychology Monographs, 1969, 72(2, Pt. 2), 1-16.

Estes, W. K. *Learning theory and mental development.* New York: Academic Press, 1970.

Estes, W. K. An associative basis for coding and organization in memory. In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory.* New York: Wiley, 1972.

Gibson, E. J. *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts, 1969.

Hinrichs, J. V., & Craft, J. L. Verbal expectancy and probability in two-choice reaction time. Journal of Experimental Psychology, 1971, 88, 367-371.

Johnson, N. F. Organization and the concept of a memory code. In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory.* New York: Wiley, 1972.

Kristofferson, A. Attention and psychophysical time. In A. F. Sanders (Ed.), *Attention and performance.* Amsterdam: North-Holland, 1967.

LaBerge, D. On the processing of simple visual and auditory stimuli at distinct levels. Perception & Psychophysics, 1971a, 9, 331-334.

LaBerge, D. Effect of type of catch trial on generalization gradients of reaction time. Journal of Experimental Psychology, 1971b, 87, 225-228.

LaBerge, D. Identification of two components of the time to switch attention: A test of a serial and a parallel model of attention. In S. Kornblum (Ed.), *Attention and performance.* New York: Academic Press, 1973.

LaBerge, D., & Samuels, S. J. Automatic processing in reading subskills. Technical Report No. 3, 1973, Minnesota Reading Research Project, University of Minnesota.

LaBerge, D., Van Gelder, P., & Yellott, J. I. A cueing technique in choice reaction time. Perception & Psychophysics, 1970, 7, 57-62.

Lawrence, D. H. The nature of a stimulus: Some relations between learning and perception. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 5. New York: McGraw-Hill, 1963. Pp. 179-212.

Lesgold, A. M., & Bower, G. H. Inefficiency of serial knowledge for associative responding. Journal of Verbal Learning & Verbal Behavior, 1970, 9, 456-466.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Posner, M. I., & Mitchell, R. F. Chronometric analysis of classification. Psychological Review, 1967, 74, 392-409.

Savin, H. B. What the child knows about speech when he starts to read. In J. F. Kavanagh and I. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading.* Cambridge, Mass: M.I.T. Press, 1972.

Shiffrin, R. M., & Gardner, G. T. Visual processing capacity and attentional control. Journal of Experimental Psychology, 1972, 93, 72-82.

Shiffrin, R. M., & Geisler, W. S. Visual recognition in a theory of information processing. In R. Solso (Ed.), *The Loyola symposium: Contemporary viewpoints in cognitive psychology.* Washington, D.C: Winston, 1973.

Zeaman, D., & House, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency.* New York: McGraw-Hill, 1963. Pp. 159-223.

(Received for publication December 22, 1972;
revision received February 6, 1973.)