Effects of Similarity, Familiarization, and Meaningfulness in Verbal Learning (1964)

with Edward A. Feigenbaum

Among the most commonly used paradigms in the study of verbal learning is the learning of nonsense syllables by the paired associate paradigm. The variables that have been shown to have important effects on the rate of learning include the levels of familiarity and meaningfulness of the syllables, the amount of prior learning, and the rate of presentation. In addition, the nature of the list, there are well-known serial position effects.

In previous papers (Feigenbaum, 1959, 1961; Feigenbaum and Simon, 1962), a theory has been set forth that undertakes to explain the performance and learning processes underlying the behaviour of Ss in verbal learning experiments. The theory, in its original version, makes correct quantitative predictions of the shape of the serial position curve (Feigenbaum and Simon, 1962), and the effect of rate of presentation on learning (Feigenbaum, 1959). Feigenbaum and Simon, 1962, as well as predictions of certain qualitative phenomena (for example, oscillation) (Feigenbaum and Simon, 1961).

In this chapter, a simplified and improved version of the theory is reported that retains these properties of the earlier theory while providing correct quantitative predictions of the effects of the other important variables: familiarization, meaningfulness, and similarity. The tests of the theory discussed here are based on comparisons of the performance of human Ss, as reported in published experiments on paired associate learning (Bruce, 1953; Chenoff, 1962; Underwood, 1953; Underwood and Schuld, 1960), with the performance predicted by the theory in the same experimental situations with the same, or equivalent, stimulus material.

The theory to be described is a theory of the information processing even underlying verbal learning. The precise statement of such a theory is most readily made in the information processing language of a digital computer, i.e., the language of computer programs.

This research has been supported by the specific kinds of processes that are involved in the performance of verbal learning tasks. These hypotheses take the form of subroutines that are component parts of the total program. Thus, there are performance subroutines which allow the program to produce responses that have previously been associated with stimuli, subroutines for learning to discriminate among different stimuli, and subroutines for acquiring familiarity with stimuli. Top-level executive routines, which organize these subroutines into a program, represent hypotheses about the Ss' understanding of the experimental instructions and the learning strategy he employs. The computer simulation of verbal learning behavior using the EPAM-III theory is, in essence, generation (by the computer) of the remote consequences of the information processing hypotheses of the theory in particular experimental situations.

In the first part of the chapter, a brief description of EPAM-III is provided. Since other descriptions of the program are available in the literature (Feigenbaum, 1959, 1961; Feigenbaum and Simon, 1962), only so much of the detail will be presented as is essential to an understanding of the experiments and the interpretation of their results. In the second part of the chapter, the results will be reported of comparisons of the behavior of EPAM-III with the behavior of human Ss in paired associate learning where similarity is the independent variable. In the third part of the chapter, the results will be reported of comparisons in which familiarization and meaningfulness are the independent variables.

A Brief Description of EPAM-III

The computer language in which EPAM-III is written incorporates one major performance system and two learning processes (Feigenbaum, 1959, 1961; Feigenbaum and Simon, 1962). When a stimulus (a symbol structure) is presented, EPAM seeks to recognize it by sorting it through a discrimination net. At each node of the net, some characteristic of the stimulus is noticed, and the branch corresponding to that characteristic is followed to the next node. With each terminal node of the net is associated an image that can be compared with any stimulus sorted to that node. If the two are similar, in the characteristics used for comparisons, the stimulus has been successfully recognized. We call such a stimulus familiar, i.e., it has a recognizable image in the discrimination net.

An image is the internal informational representation of an external stimulus configuration that the learner has stored in memory. An image, thus, is comprised of the information the learner knows about, and has associated with, a particular stimulus configuration. An image may be elementary or compound. A compound image has, as components, one or more elementary or compound images which may themselves be familiar and which may possess their own terminal nodes in the discrimination net. For simplicity, in the current representation, letters of the Roman alphabet are treated as elementary stimuli whose characteristics may be different.
into more elementary familiar stimuli. On the other hand, syllables are compound stimuli, their components being, of course, letters, words, or phrases. A compound stimulus image, viewed from the bottom up, may be regarded as an association among the component stimuli. Thus, the net may contain stimulus images that represent pairs of syllables, these compounds being composed from other compound images, the individual syllables.

In performing the paired associate task, the program uses the stimulus, present in the window of the memory drum, to complete the symbol representing the pair comprised of the stimulus and its associated responses. We may designate this compound symbol by S',—for the second response member—and S—in the drum window. The compound symbol, S—that is, sorted through the net, and the image associated with the terminal is retrieved. We will designate this image by S'.—for the previous learning has been successful, it will be comprised of two components: an image of the stimulus syllable and an image of the associated response syllable. The response image, R', which has just been retrieved as the second component of the compound image, S'-R', identifies a net node where an image, say R', is stored. R' will have as its components symbols designating the constituent letters of the initials of the second term, T'. Each of these, in turn, identifies a terminal node. Associated with the terminal for a letter is not only an image of the usual kind (an afferent image) but also the information required to produce the letter in question, i.e., to print it. This information, which may call the efferent image, is produced to activate the response. Thus, the final step in the sequence is for the program to respond, say, A T.

It is a fundamental characteristic of this program that elementary symbols and compound symbols of all levels are stored in the discrimination net in exactly the same way. A syllable is simply a list of letters, and an R-S is simply a list of syllables. A single interpretive process can be of any sort, or any other symbol, elementary or compound. Moreover, the symbols discriminated by the net are not restricted to any specific sensory or motor mode. All modes can be accommodated by a single net and a single interpretive process. All interpretive symbols belonging to different sensory modes will possess different attributes: phonemes will have attributes like "voicing," "tongue position," and so on; printed letters will have attributes like "possession of closed loop," "possession of diagonal line," and so on. Because they possess entirely different attributes, the symbols will be sorted to different parts of the net

Finally, symbols may be of mixed mode. In a symbol, S-R, for example, S may be in the visual mode, R in the oral.

The Learning System

EPAM-III uses just two learning processes, one to construct and elaborate images at terminal nodes of the net (image building), and another to elaborate the net by adding new branches (discrimination learning). The first learning process also serves to guide the second.

When a stimulus S is in view and is sorted to a terminal, the stimulus can be compared with the image S' stored at the terminal. If there is no image at the terminal, the image-building process copies a part of S and stores the initial part of S as a visual image at the terminal. If there is already an image, S' at the terminal, one or more differences between S and S' are detected, and S' is corrected or augmented to agree more closely with the new S.

When a positive difference (not a mere lack of detail) is detected between a stimulus S and its image S', the discrimination learning process can use this difference to construct a new test that will discriminate between S and S'. The terminal node with which S was associated is then changed to a branch node, the test associated with the node S', is associated with a new terminal on one of the branches, and a new image of S is associated with a new terminal on another branch. Thus the discrimination learning process adds a new pair of branches to the discrimination net and attaches initial images to the branches.

Note that a stimulus S can be sorted to a terminal T only if S satisfies all the tests that point to the branches leading to T. But the image S' stored at T must also satisfy these tests. Hence, there can be a positive difference between S' and S only if S' contains more information than necessary for T. For instance, let S be the syllable KAW, and suppose that all the tests leading to the terminal T happen to be tests on the first letter, K. Then the image S' stored at T must have K as its first symbol, but it may differ from KAW in other characteristics. It might be, for example, the incomplete syllable K-B. The discrimination learning process could detect the difference between the W and B in the first two symbols, construct a test for this difference, and append the test to a new net node. The redundancy of information in the image, in this case the letter B, permits the further elaboration of the node.

Thus, learning in EPAM-III involves cycles of the two learning processes. Through image building, the stimulus image is elaborated until it contains information sufficient to help to sort it to its terminal. Through discrimination, the information is used to distinguish between new stimuli and the stimulus that generated this terminal and grew its image. On the basis of such distinctions, the net is elaborated. The interaction of these two processes is fundamental to the whole working of EPAM. It is not easy to conjure up alternative schemes that will permit learning to proceed when the members of a pair of stimuli to be discriminated are not present simultaneously.

The stimuli that EPAM-III can make familiar and learn to discriminate are symbols, either elementary or compound, composed of elementary symbols, or compound symbols, of any kind.
The syllables employed in the EPAM simulation were the same as those used by Underwood. Row 2 in table 1 summarizes the data from the EPAM tests. Response similarity facilitated learning very slightly, while stimulus similarity impeded learning by as much as 40 percent. Since relative learning times are reported in both cases, there is one free parameter available for matching the two series. (In the normal course of events, the compound images $S^R$ are discriminated from each other on the basis of stimulus information, not response information. Hence all stimulus similarity makes difficulties for EPAM in discriminating and retrieving these images, and hence impedes learning. Response similarity, of course, has no such effect.)

The qualitative fit of the EPAM predictions to the Underwood data is better than the quantitative fit, although, considering the (a priori) plausible range of impact of the stimulus similarity variable on difficulty of learning, even the quantitative fit is not had. Nevertheless, we sought a much better quantitative prediction. This search led us into the following considerations. The prediction that is seriously out of line in table 2 is the prediction for the $M-L$ condition. The more carefully one scrutinizes the Underwood experiment and the Underwood materials, the more puzzling the experimental results become. Why do $S_s$ as the results indicate, respond in the $M-L$ condition so similarly to the way they respond in the $L-L$ condition, while their responses in the $H-L$ condition are so different from responses in either the $M-L$ or $L-L$ conditions? The answer is not to be found in the Underwood materials. We have analyzed the Underwood definition of "medium similarity" in terms of the information necessary to discriminate the items on a list of a given length (in EPAM-like fashion) and have found that Underwood's definition is quite careful and correct. By his definition, one should expect "medium similarity" lists to be midway in effect between his "low similarity" and his "high similarity" lists.

The answer, we believe, lies in the recoding, or "chunking," behavior of $S_s$, which would make the "medium similarity" stimulus list formally identical with the "low similarity" stimulus list under Underwood's definition. Suppose that many $S_s$ were engaged in the CVCs, i.e., recoding the items into the aural mode, instead of dealing with them directly in the visual-literal (presentation) mode. The recoded ("aural") syllables will be "chunked" into two parts: a consonant-vowel pair, and a consonant. In other words, the visual-literal stimulus objects of three parts (CVC) quite naturally recode into "aural" stimulus objects of two parts (C'C or CC'). Hence Underwood's "medium similarity" lists, none of the C chunks are duplicated, nor any of the C' chunks. The recoding, therefore, has transformed the "medium similarity" list into a "low similarity" list, by Underwood's definition.

To test this hypothesis for sufficiency from the point of view of the theory, we reran the EPAM (simulated) experiments using "aural" recordings of the original syllables. The modified predictions are given in row 3 of table 1. As the analysis above indicates, the $M-L$ condition is now different from the $L-L$ condition, but the prediction of difficulty for the $H-L$ condition is too low. Assuming that some $S_s$ are processing in the visual-literal mode and some in the "aural" mode, we have computed the average ($\frac{1}{2}$) of the two sets of EPAM predictions. This is given in row 4 of table 1. If we weight the average $\frac{1}{2}:1$ in favor of $S_s$ doing "aural" recoding, the result is as given in row 5 of table 1. Each of these averaging procedures gives a prediction which is closer than that for the Underwood lists nonrecoded.

It is clear that we still have much to learn about this low versus medium similarity problem. In this regard, we are currently attempting a direct experimental test of the "aural" recoding hypothesis.

Bruce's $S$'s (1953) learned two successive lists of paired associative nonsense syllables. On the second list, response syllables, or stimulus syllables, or neither, could be the same as the corresponding syllables on the first list. Thus, using current designations, Bruce's three conditions were (A-B, C-D), (A-B, A-C), and (A-B, C-B, respectively. In summary, he found that learning of the second list was somewhat easier than learning of the first when all syllables were different (A-B, C-D), much easier when the response syllables were the same (A-B, C-B), and a little harder when the stimulus syllables were the same (A-B, A-C) (see table 2). The relative difficulties are compared using the A-B, C-D group as the nonrecoded.

Nonsense syllable lists of low Glazve value and low intralist similarity were used when the experiment was replicated with EPAM.

Table 1. Comparison of EPAM with Underwood's (1953) Data on Intralist Similarity (Relative Number of Trials to Criterion, LL = 100).

<table>
<thead>
<tr>
<th>Condition of Stimulus and Response Similarity</th>
<th>Data</th>
<th>L-L</th>
<th>L-M</th>
<th>M-L</th>
<th>L-H</th>
<th>H-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underwood</td>
<td>100</td>
<td>96</td>
<td>109</td>
<td>104</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>EPAM-III</td>
<td>100</td>
<td>88</td>
<td>141</td>
<td>91</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>&quot;visual only&quot;</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>114</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPAM-III</td>
<td>100</td>
<td>91</td>
<td>121</td>
<td>92</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>&quot;visual&quot; and &quot;aural&quot; mixed: 1:1</td>
<td>100</td>
<td>96</td>
<td>114</td>
<td>97</td>
<td>125</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Comparison of EPAM with Bruce's Data on Interlist Similarity (Relative Number of Trials to Criterion).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>130</td>
<td>96</td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>
The normalized results are shown in the second line of table 2. The effects in the simulated experiment were qualitatively the same as in the actual data. If we compare the EPAM, A-B, A-C and A-B, C-B with A-B, C-D, we find that identity of stimulus syllables impeded learning less, and identity of response syllables facilitated learning to the same degree in the simulation as the human Ss. The ratio of difficulty for the A-B, A-C compared with the A-B, C-D conditions, where total number of different syllables discriminated and learned was the same, was 1.73 for the human Ss and 1.49 for EPAM.

From our analysis of the data of the Underwood and Bruce experiments, we conclude that EPAM provides satisfactory explanations for the main observed effects of intralist and interlist stimulus and response similarity upon the learning of paired associate nonsense syllables. These effects predicted by EPAM-III are in the right direction and are of the right order of magnitude, although there is room for improvement in the quantitative agreement.

FA\MIL\\R\D\ AND MEANINGFULNESS

Among the other independent variables that have been shown to have major significance for the learning of two syllable pairs, the nonsyllables are familiarity and meaningfulness. A thorough discussion of the definition of these two variables can be found in Underwood and Schulz (1960).

The degree of familiarity of a syllable is usually not measured directly; instead, it is measured by the amount of Familiarization training to which S has been exposed with that syllable. In the following discussion, they are not synonymous. "Familiarization" will be used when reference is made to specific experimental conditions and operations, "Familiarity," on the other hand, will refer to a condition internal to an S: the state of information about a syllable in the categories of an S who has gone through some kind of familiarization training. Thus, familiarity is an intervening variable not directly observable. The use of this intervening variable hardly needs to be defended, since it is the rule rather than the exception in theory building in the natural sciences as well as the behavioral sciences.

Familiarization training is accomplished by causing S to attend to the syllable in question in the context of some task other than the paired-associate learning task before given him subse-

quently. It should be noted that there is no way of discovering, with this definition, how familiar a syllable may be for an S due to his experience prior to the experiment. Although the syllables are not meaningful words, the consonant-vowel combinations contained in them occur with varying frequency in English. The meaningfulness of a syllable, on the other hand, is generally determined by measuring the number of associations that Ss make to it in a specified period of time. For these learning experiments are generally selected from available lists that have been graded in this way for meaningfulness.

Since high meaningfulness and high meaningfulness both facilitate nonsense-syllable learning, there has been much speculation that the two phenomena might be the 'same thing.' This, in fact, is the central hypothesis examined in the Underwood and Schulz monograph. In one sense, meaningfulness and familiarity are de-

meanably not the same, for the amount of familiarization training can be given with low-meaningful syllables without significantly increasing their meaningfulness. However, Underwood and Schulz (1960) adduce a large body of evidence to show that there is a strong relation running the other way i.e., that meaningfulness of words is correlated with their frequency in use in English, and that ease of learning nonsense syllables is correlated with the frequency, in English, of the letters that compose them (for low pronunciability, or with their pronunciability.

The data are of course greatly complicated by the fact that Ss may handle the material in either the usual or natural mode, and that most Ss probably encode into the latter, at least part of the time. Hence, for relatively easily pronounceable syllables, frequency of phoneme pairs in the English language would be a more relevant measure of frequency than frequency of the printedigrams or trigrams. Thus, the finding by Underwood and Schulz that pronunciation is a better predictor than trigram frequency of ease of learning does not damage the hypothesis that familiarity of the components is a relevant factor. The critical variable is meaningfulness; familiarity, in turn, is a function of previous exposure.

We conclude that high meaningfulness implies high familiarity, although not the converse.

If this is so, then the correlation of meaningfulness with the number of possible trials to criterion for the least and most meaningful conditions, as represented by the neighborhood of 2.5. That is to say, syllables of very low meaningfulness take about two and one-half times as long to learn as syllables of very high meaningfulness (and about one-half times as many errors are made during learning).

Before the significance of this 2.51 ratio is considered further, it is necessary to discuss one difficulty with the hypothesis that familiarization and meaningfulness (via similarity) facil-
ate learning primarily because of the two being integrated prior to the associative trials. The effects reported in the literature with meaningfulness as the independent variable are generally much larger than expected for familiarization. No one has produced anything like a 2.51 gain in learning speed by familiarization training.

There is now some evidence, primarily in a doctoral dissertation by Chernoff (1962), that the main reason for this discrepancy is that the familiarization training in experiments has been too weak, has stopped too soon. It appears that no one has carried out familiarization training with his Ss to the point where the syllable integrat-
ion achieved is comparable to the integration of syllables of high meaningfulness.

Chernoff's experiment can be summed up as follows. First, in his experiments, both meaningfulness and familiarization of both stimuli and responses. Thus, he had 16 condi-
tions: all possible combinations of H-H, L-H, L-L, and H-L, in terms of meaning-
fulness with F-F, U-F, L-F, and U-U for familiarity. Second, he employed a more thor-
ough familiarization training technique for the F condition than had any previous investigator. The syllables were presented one at a time to S at about a 2.5-sec rate. The S was required to pronounce each syllable immediately, and was asked to recall the syllables in any order. If an incorrect syllable was given, S was told that it was not a member of the list. If, when yes or no, S could not be made to choose, five more familiarization trials were adminis-

4. The two levels of the meaningfulness variable were constructed as follows (using CVG's): 4.5-100 Glaze, 85-100 Kruger, 67-94 Archer, 85-100 Noble's (m), 3:8-3:7 Noble (a), 0:5-53 Glaze, 30-72 Kruger, 48-48 Archer, 3:4-1:8 Noble's (m), 1:38-1:94 Noble (a).
Effects of Similarity, Familiarization, and Meaningfulness in Verbal Learning

In the present interpretation of the mechanism of familiarity, correct responses were obtained with a combination of three trials of response familiarization and one trial of stimulus unfamiliarization: additional familiarization did not facilitate performance. Such an asymptote, however, was not attainable with any amount of stimulus familiarization, or with any amount of stimulus unfamiliarization without response familiarization. The asymptotes in the latter two cases were 27 errors and 98 errors, respectively, and were reached with three trials and two trials, respectively, of familiarization.

The details of table 4 show some exceedingly complicated relations. For example, if syllables have received no prior familiarization, one trial of stimulus familiarization reduces errors more than one trial of response familiarization (reductions of eight and four errors, respectively) from 39 to 14 in the non-familiarization case. On the other hand, for syllables that had already received one trial of stimulus and response familiarization, an additional trial of stimulus familiarization reduced errors only by three, while an additional trial of response familiarization reduced errors by 11, from a level of 35. Other similar results may be read from table 4. Many of the numerous small anomalies in the literature on familiarization training may be attributable to the lack of control over the amount of prior familiarity that 8s had with the syllables used in the experiments.

In table 3, column 6, we show the predicted effect, estimated from the EPAM data of table 4, of familiarization training with syllables that were already somewhat familiar before the experiment began (i.e., that had previously received one simulated trial each of stimulus and response familiarization). Under the F-F condition, we would have 21 errors to criterion; under the U-F condition (one stimulus familiarization trial, three response trials), 21; under the F-U condition (one stimulus and response familiarization trials, respectively), 32; and under the U-U condition (one S and one R familiarization trial), 35. The resulting index of relative difficulty for the four conditions are 1.0, 1.0, 1.5, and 1.7, respectively, as shown in column 6. These may be compared with the values 1.0, 1.2, 1.6, 1.8, for the actual data in column 5. In other words, the fact that the effects in column 2 and even in column 4 are somewhat smaller
due simply to the fact that the syllables were already slightly familiar to the Ss at the beginning of the experiment.

Conclusion
In this chapter we have compared the predictions of EPAM-III, a theory of human verbal learning, with data from the experiments of Bruce, Chenoff, Underwood, and others, on the effects of intralist and interlist similarity, of familiarization, and of meaningfulness upon difficulty of learning. We find that there is good quantitative, as well as qualitative, agreement between the published data and the predictions of the theory. Finally, we have used our findings to discuss the relation between familiarity and meaningfulness, and have shown that most of the known facts can be explained by supposing that a symbolic structure necessarily becomes familiar in the process of becoming meaningful, but that the converse is not necessarily the case.

References

3.3
One-trial and Incremental Learning (1967)

with Lee W. Gregg

Determining to what extent rote verbal learning is incremental, and to what extent a task takes place in a single trial has been a central problem of learning theory during the past decade. In the light of the accumulated experimental evidence, the answer to "Is rote learning incremental?" may only be "Sometimes." The problem, as Atkinson, Bower, and Coothers (1965, p. 118) observe, is to define the "sometimes." "At present writing," they say, "a challenging task confronting the theorist is to identity those experimental conditions under which the one-element (i.e., all-or-none) model works, and those where it does not, and (2) to construct a more structural model which essentially reduces to the one-element model in the former cases but accounts for the discrepancies in the latter cases."

It is the aim of this chapter to show that an information processing theory, EPAM (Elementary Perceiver and Memorizer), makes quantitative predictions about one-trial learning and departures from it that are consistent with the main body of empirical evidence. The first version of EPAM was constructed by Feigenbaum and Newell, 1962, pp. 307-320.

This investigation was supported by public Health Service Research grant MH-07272 from the National Institute of Mental Health. The authors are greatly indebted to Edward A. Feigenbaum, with whom they have had numerous discussions of the implications of the EPAM theory for one-trial learning. They are grateful also to Andrew P. Chenoff and Kenneth R. Laughery, who participated in conducting the experiment, and whose data are reanalyzed here.

1. The earliest public description of EPAM is in Edward A. Feigenbaum (1959). The pre-