DISCUSSION

There are several techniques other than close scores that might be used to calibrate the complexity of the passage. Each has its advantages and its disadvantages.

One technique that has been used (Gray and Leary, 1955) is to let Ss read the passages and then ask them questions about it. The passage with the least wrong answers is called least complex. The disadvantage here is that the measure depends upon the questions as well as upon the passages. One can ask easy questions about difficult passages and vice versa. The questions one investigator would ask about a passage might bear little resemblance to those asked by another investigator.

Newman and Gernstam (1952) introduced a technique that has the advantage of giving an estimate of redundancy in bits. It is based upon the sum of the distributional constraint of the letters plus the simple unidirectional contingencies between the letters being predicted and the preceding letters. Since the words are scrambled before the analysis, this method not only ignores the interaction between letters, but also ignores constraints between words. The resulting disadvantage of this technique is that it reflects little or none of the increase in difficulty beyond the fifth-grade level. Cartarette and Jones (1963) used this method and found that the percentage redundancy of their fifth-grade reader was approximately the same as the average adult text. Therefore, if there is a need to calibrate these SS passages in bits of redundancy, it would probably be best to use Shannon's "guessing game technique" (1951) and a single, linguistically-simplified S.

Although close techniques do not lead to redundancy estimates, they are certainly more useful in calibrating a scale of complexity because their scores are related to the total constraint in the passages and they measure constraint which is ignored by the Newman and Gernstam technique. Thus, the close scores give a measure of readability over the whole range of difficulty, whereas the Newman and Gernstam technique becomes asymptotic at about the fifth-grade level.

REFERENCES


(Received March 10, 1969)

JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 6, 855-863 (1967)

Implicit Learning of Artificial Grammars1

ARTHUR S. REIDER2

Brown University, Providence, Rhode Island

Two experiments were carried out to investigate the process by which Ss respond to the statistical nature of the stimulus array, a process defined as "implicit learning." An artificial grammar was used to generate the stimuli. Experiment I showed that Ss learned to become increasingly sensitive to the grammatical structure of the stimuli, but little was revealed about the nature of such learning. Experiment II showed that information gathered about the grammar in a memorization task could be extended to a recognition task with new stimuli. Various analyses of the data strongly implied that Ss were learning to respond to the general grammatical nature of the stimuli, rather than learning to respond according to specific coding systems ingrouped upon the stimuli. It was argued that this "implicit" learning is similar in nature to the "differentiation" process of perceptual learning espoused by Gibson and Gibson (1955).

In recent years, the model of the verbal organism as an initiate and generalizing mechanism has been largely replaced by a model that characterizes him as a "sentence generating machine" who has learned a "generative grammar" in some implicit fashion (cf. Chomsky 1957, 1959; Miller and Chomsky 1963). The "implicit" is the problem, for the traditional learning paradigm—most notably the simple S-R approaches—do not seem adequate to explain the various behavioral phenomena which characterize a child learning to speak a natural language (cf. Brown and Fraser 1963 for a discussion).

The nearest thing to an investigation of implicit learning was carried out by Gibson and Gibson (1955) under the rubric "perceptual learning." The Gibson's argued that the phenomenon of perceptual learning, whereby an organism comes to perceive and respond to his environment in a reliable and efficient manner, should most parsimoniously be thought of as a 'differentiation' process, as opposed to an 'enrichment' process. By enrichment, they referred to the organism's adding of information to the stimulus display such as "forming a good Gestalt," and by differentiation they referred to a process whereby the organism becomes sensitive to the information already inherent in the stimulus display. Perceptual learning is assumed to be a primitive process certainly requiring no intellectual factors that would be beyond the scope of a two-year-old child, and is assumed to develop simply by constant exposure to the sources of variation in the environment. This similarity between the Gibsonsian differentiation process and what we are calling implicit learning has been noted elsewhere. Braine

1 This research was supported in part by grant G-18110 from the National Science Foundation directed by Richard B. Millward to whom the author is grateful for advice and guidance. The paper is based upon a thesis submitted to Brown University in partial fulfillment of the requirements for the Master of Arts degree.

2 Now at University of British Columbia, Vancouver, Canada.
(1963) pointed out that the children in his study on grammar learning who picked the correct linguistic symbol amply because it "ruined eight" were behaving like Sm in a perceptual learning experiment. If this similarity is more than illusory and if the mechanisms involved in the two types of learning are indeed synonymous, then the Gibsonian experiment of repeated exposures of the rat stimuli should be sufficient for a 5 to learn to respond to the grammatical nature of an artificial language. We ran the experiments on adult Ss rather than children and used artificial syntactic languages rather than natural languages. Our primary purpose was the study of the process by which Ss come to respond to the statistical nature of the stimulus array, and we felt that this technique provided better experimental control.

**EXPERIMENT I**

**Method**

There were two groups of 16; one group worked only with "grammatical" stimulus items and the other only with randomly constructed items.

**Grammatical Items.** A finite-state language (cf. Chomsky and Miller, 1963) was constructed with the five letters P, S, V, T, and X in the vocabulary

and a set of rules of sentence construction as the grammar. The grammar may be characterized as a Markov process in which a transition from state S to any state S' produces a letter. A system of this sort can be illustrated by a state diagram (Fig. 1). A "sentence" begins when we enter at state S0; each transition produces a letter, and the sentence ends when state S1 is reached. The actual sentence produced depends upon the path taken. The language itself is simply the set of all possible sentences that the grammar can produce.

The number of possible sentences of a given length that can be generated was determined for each length up to 8. Macmillan (1954) has shown that the number of sentences in such a language can be described by a "structural function." If $f(n)$ denotes the number of possible sentences of exactly length $n$, then the structure function is a recursion in $n$ which will specify $f(n)$ for all $n > 1$. The structure function for the grammar illustrated in Fig. 1 can be shown to be $f(n) = f(n-1) + f(n-2) + f(n-3) - f(n-4)$.

Given the first four values of $f(n)$, all others may be found by using this recursion. It is easily seen from the state diagram that $f(1) = f(2) = 0$, $f(3) = 2$, and $f(4) = 5$. These values permit the computations: $f(5) = 4$, $f(6) = 7$, $f(7) = 11$, $f(8) = 16$. The latter three values of $n$ are of primary interest in this experiment because the length of sentences was restricted to 6, 7, and 8.

**Random Items.** A control a set of random items was generated from the same 5 letters. The informational value of a symbol in a nonrandom system is given by log $n$, where $n$ is the number of available symbols. Each letter in a random item contains log 5, or 2.322 bits, as opposed to the ABS bits for a letter in a grammatical item. Since it appeared that this would make random items more difficult to learn at the outset than grammatical items, the length of the random stimuli was held to 4, 5, and 6 letters.

**Procedure.** The structure function specifies 54 sentences of length 6-8. These were printed on 33×7-inch cards and constituted the grammatical stimulus items for the experimental Ss (S1). For each S1, 28 items were selected at random and placed in a fixed order. Every S1 was required to learn all 28 of the stimulus items presented in seven sets of four sentences each. Each sentence of a given set was presented through a viewing window for 5 sec. The S1 was then given a piece of paper and asked to reproduce the sentences. After each trial if removed the S1's responses and added the incorrect stimulus which it had reproduced correctly and which one he had not. No information was given about the nature of his errors. After the criteria of two consecutive correct reproductions of a given set was met, a new set of four sentences was learned in the same way until all 28 items were learned. The task was described as a memory experiment and no reference to rules of construction was made until after the learning task was completed. The control group Ss (S2) learned the randomly formed items in identical fashion. There were 5 Ss in each group. They were undergraduates confined in an interocular psychology course.

**Results**

The most obvious result was the consistent decline in the number of errors to criterion across the seven sets in the S1s (Fig. 2). An analysis of variance showed the

---

**Fig. 1. Schematic state diagram of the grammar used to generate the grammatical stimulus items.**

**Fig. 2. Mean number of errors to criterion on each of the seven learning sets.**
difference between the groups to be significant, P (L, 8) = .14.67, p < .01. Note however, that the two groups were indistinguishable from each other on Sets 1 and 2. It is not until Set 3 that the Ss show a significant improvement in performance over the Ss. This suggests that it should be fruitful to analyze the data for Sets 1 and 2 separately from the data for Sets 3-7.

Sets 1 and 2. On these sets, as was noted above, the two groups behaved identically, therefore the independent variable (the grammatical restriction) was not the cause of this change in the behavior of the Ss. It seems reasonable to assume that the learning which took place during Set 1 was related to some strategy to learn rather than to the stimulus material itself. Of the possible strategies the one that seems easiest for Ss to adopt would be a stimulus-selection strategy. That is, when a single item at a time, ignoring all others until that one is learned, then concentrate on the second item while ignoring the remaining two, and so on. On the assumption that Ss did use this strategy and used it efficiently, the data from Set 2 should display no cases (or at worst only a few) where an error was made on an item which had been reproduced correctly on one or more previous trials. In contrast the data from Set 1 should not exhibit this characteristic to any marked extent.

The results of the analysis, while not in perfect agreement with a stimulus-selection strategy, are certainly compatible with it. The incidence of errors made on previously correct items dropped by 82.5% (from 40 to 7) from Set 1 to Set 2, while the total number of errors was reduced by only 55% (173-78), indicating that on Set 2 all Ss had a much greater tendency to learn an item the first time they reproduced it correctly than they did on Set 1. Such an analysis lends support to the interpretation that the superiority of performance on Set 2 over that on Set 1 reflects little more than a rather complicated warm-up effect.

Sets 3-7. Figure 2 shows that the behavior of the Ss does not depart from the level they achieved on Set 2, implying that they continued to respond strictly according to the strategy described. And in fact, this description does seem fairly adequate in accounting for their behavior. Of the 242 errors made by Ss on Sets 3-7 only 28 (12%) were on previously correct items. The Ss, while they also continued to learn items in an all-or-none fashion (only 9.4% of their 96 errors on Sets 3-7 were on previously correct items), departed noticeably from the simple stimulus-selection strategy. Rather than learning only one item at a time, they became capable of learning two or three items on a single trial. The first two lines of Table 1 show the frequency with which various numbers of items were learned on single trials during Sets 3-7. The Ss had almost three times as many trials with 1 or 2 items learned as did the Ss, and only one-half as many trials where more than one was learned. These differences are significant, x^2 (1) = 20.93, p < .001. Note also that Ss required almost twice as many trials as the Ss. The difference between the groups is even more striking if only the first trial of each set is looked at (last two lines of Table 1). While Ss hardly ever succeeding in correctly reproducing more than one item, Ss were just as likely to reproduce two or more as zero or one. A chi-square test showed a clear difference between the groups, x^2 (1) = 11.52, p < .001. It should be recalled here that the Ss are dealing with stimulus items which are 6-, 7-, and 8-letters long, while the Ss are learning strings of length 4, 5, and 6, and that for Sets 1 and 2 the items were clearly of equal difficulty.

One additional statistic which further indicates the increased learning capacity of the Ss is amount of information processed. It which is simple enough to discover yet efficient enough to facilitate learning even exists.

Another difficulty with such an explicit encoding explanation is that it implies that Ss have at least some verifiable knowledge of the rules of construction. This, however, was not the case. After the experiment was over, each S was in- formed that all of the 28 stimulus items were formed by the use of specific rules and was asked if he had any ideas about them. Since no S re- ported any patterns of letters or rules, the only possible interpretation is that while Ss were not able to recall any rules, they were able to make a certain number of random guesses because they were "recoding" individual symbols into larger chunks which decreased the absolute number of units. The coding system assumed was one based upon the "loops" or recurring symbols and sets of symbols in the grammar. By placing the loops within parentheses the grammar used in this experiment can be expressed as five basic sentence types: (a) T(V)X; (b) T(P/|PXX)X(VP/X)VS; (c) T(P/|PXX)X(VP/X)VS; (d) V(X)(VP/X)VS; and (e) V(X)(VP/X)VS. This shorthand method of writing the grammar illustrates the basic constraints of the language: T, V, F, X, VP, and S. It is not at all clear that an encoding system
otherwise, about the learning process. The conclusion indicated that 5s were
analogous to Gluskin’s perceptual learning. However, it is not impossible that the
coding systems of the type proposed by Miller were in use, the interactive verbal re-
ports of undergraduates are notoriously un-
reliable and usually incomplete. To correct
for this shortcoming, another study was
The procedure was modified so that
systematic coding schemes, if they exist, would be reflected in the data.

**EXPERIMENT II**

**Method**

The experiment was run in two parts: (a) a
learning phase similar to Exp. 1, in which the
rules were to be learned; and (b) a test phase,
in which performance was to be tested.

**LEARNING PHASE**

The finite-state grammar de-
scribed in Exp. 1 was again used to generate the
sentences used in this experiment. The
sentences were presented in a second
learning phase similar to Exp. 1, in which
the procedures were identical to those in
Exp. 1, except that the first two sentences
from each of the sets present to the
subjects were omitted.

**Test Phase**

The instructions read to the Ss were the
same as those used in the first experiment; the
task was described as a memory experiment, and
the experimenter was male of grammar rules or sub-

**Results**

The first thing to note is that the two
subject pools were apparently very different.
The college undergraduates performed 73.5% correct responses on the 88 test
items, whereas the Ss who had completed the
course in grammar had a mean of only 68.2%. The difference is significant (t(8) = 7.53, p <
.001). Since each group considered inde-

**TABLE 1**

<table>
<thead>
<tr>
<th>Item</th>
<th>G</th>
<th>NG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>94</td>
<td>92</td>
</tr>
<tr>
<td>NG</td>
<td>91</td>
<td>94</td>
</tr>
<tr>
<td>Total</td>
<td>185</td>
<td>186</td>
</tr>
</tbody>
</table>

The similarity between the total number of G and NG responses is a consistent result, for it
eliminates the possibility of response bias.

To investigate the possibility of coding systems the test trial data were analyzed in
a number of ways: (a) analysis of the
errors committed on NG items, (b) analy-

**REFERENCES**

[1] The NC items were broken down
into five basic types depending upon the position
in the text where the error occurred. The
random item (R), all of which contained multiple
errors, had a significantly higher detection rate than any of the others, x²(1) = 6.04, p <
.05. A single error occurring in the last (L) or next-to-last
(L-1) position also proved to be a more effective
cue than an error occurring in the first
(S) or in the internal (I) position. These
results indicate that Ss focused most of their
attention on the end of the item. In fact, the com-
detected error rates on the L and L-1 items was
significantly higher, x²(1) = 8.63, p < .01, than the
combined rate for L and I items. So far as
NC items are concerned, there is no significant
difference between the error rate of the end of a
string and the next to last position. This may be
explained by the fact that multiple errors were
more likely to be detected than a single error in any position.

**C Items**

The study analysis, however, treated
only half of the data. It failed to explain why a S
would prefer to omit a C item. Unfortunately, there
is no simple way to classify G and NG responses and there
is no way to analyze the responses to them. The
best that can be done is to examine the
five basic sentence types the grammarians generate. The
proportion of times that each type of incorrect
classified G was made in the group is not significantly
different. Even Type V, which is the
simplest form of sentence, has a probability of
incorrect rejection as high as that of the other
more complex types. Since Type V has only one
loop and the others have three, this result
calls into question the hypothesis that 5s used
computer systems based upon the loops of the

**conclusions**

[2] The NC items were broken down
into five basic types depending upon the position
in the text where the error occurred. The
random item (R), all of which contained multiple
errors, had a significantly higher detection rate than any of the others, x²(1) = 6.04, p <
.05. A single error occurring in the last (L) or next-to-last
(L-1) position also proved to be a more effective
cue than an error occurring in the first
(S) or in the internal (I) position. These
results indicate that Ss focused most of their
attention on the end of the item. In fact, the com-
detected error rates on the L and L-1 items was
significantly higher, x²(1) = 8.63, p < .01, than the
combined rate for L and I items. So far as
NC items are concerned, there is no significant
difference between the error rate of the end of a
string and the next to last position. This may be
explained by the fact that multiple errors were
more likely to be detected than a single error in any position.

**C Items**

The study analysis, however, treated
only half of the data. It failed to explain why a S
would prefer to omit a C item. Unfortunately, there
is no simple way to classify G and NG responses and there
is no way to analyze the responses to them. The
best that can be done is to examine the
five basic sentence types the grammarians generate. The
proportion of times that each type of incorrect
classified G was made in the group is not significantly
different. Even Type V, which is the
simplest form of sentence, has a probability of
incorrect rejection as high as that of the other
more complex types. Since Type V has only one
loop and the others have three, this result
calls into question the hypothesis that 5s used
computer systems based upon the loops of the

**conclusions**

[2] The NC items were broken down
into five basic types depending upon the position
in the text where the error occurred. The
random item (R), all of which contained multiple
errors, had a significantly higher detection rate than any of the others, x²(1) = 6.04, p <
.05. A single error occurring in the last (L) or next-to-last
(L-1) position also proved to be a more effective
cue than an error occurring in the first
(S) or in the internal (I) position. These
results indicate that Ss focused most of their
attention on the end of the item. In fact, the com-
detected error rates on the L and L-1 items was
significantly higher, x²(1) = 8.63, p < .01, than the
combined rate for L and I items. So far as
NC items are concerned, there is no significant
difference between the error rate of the end of a
string and the next to last position. This may be
explained by the fact that multiple errors were
more likely to be detected than a single error in any position.
range, 16–27. The other two have values which are suspiciously low (.07 and .09). However, in item-by-item analysis according to errors committed on each of the five basic types of G and NG sentences revealed many inconsistencies. The only pattern discovered was in one S who rejected any item with the same letter repeated more than three times (e.g., VVVVXX, TTPPFTFT), a criterion which does not necessarily reflect grammaticality.

Aside from this minor example, the data do not reveal any consistent pattern of responding on the part of any S. It is concluded then that if 5S were using strategies they were not consistent in their use of them; there is no simple set of rules for characterizing the behavior of any one S.

Analysis of Response Probabilities. Consider the following assumptions: (1) a S knows the grammaticality of the ith item with probability p, and does not know it with probability 1–p; (2) when he does not know, he guesses correctly with probability g and incorrectly with probability 1–g. Since each item was presented twice and there was no feedback, these are four possible outcomes for each item: (a) correct on both presentations; (b) correct on the first, as wrong on the second; (c) as wrong on the first, correct on the second; and (d) errors on both. The following equations describe these four possibilities:

\[ P_i(C) = p \times (1-p^2) \]

\[ P_i(E) = P_i(C) = (1-p)(1-g) \]

where \( p \) is the average of all \( p_s \) and \( g \) is \( S \). These equations describe what is probably the simplest detection model possible. On each trial the S either knows the grammaticality of the item or he does not, and when he does not, he guesses. The actual value of any one of the equations can be read from the data and used to estimate \( p \). The values listed in Table 3 were obtained by using the empirical values of \( P_i(C) \) and \( P_i(E) \) for the values of \( P_i(C) \), \( P_i(E) \), and \( P_i(E) \). These calculations were first carried out for the G and NG items independently and then since the estimated \( p_s \) were essentially the same for all items pooled. The fit is fairly good; in some cases a* fall short of significance. With N = 6, the power of the test is quite high and the lack of clear statistical significance lends support to the model.

Two other implications of the model are of interest: (a) a S never innocently thinks he knows the grammaticality of an item, and (b) a S knows the grammaticality of the ith item with a probability constant across all \( i \), regardless of the characteristics of the item and regardless of whether or not he has seen the item before. The first of these implications is related to the prediction that the proportion of items with two errors should be equal to the proportion of items with only one error. It is of interest because it logically precludes the use of a strategy which was not inconsistent with the general, since any such strategy would force items into the (CE) category and inflate it beyond (CE) and (EC). The model then demands further support from the fact that \( P_i(C) \) is not significantly greater than either \( P_i(C) \) or \( P_i(C) \).

The latter is related to the prediction that \( P_i(C) = P_i(C) \). When confronted with an item a S will either know or not know its grammaticality. If he does know it, then he will automatically always know it, and if he does not know it, then he will automatically never know it, and all should be forced to guess. This is a rather strong statement and may perhaps not yield a strictly accurate picture of the data, for there may be a tendency for \( P(C) \) to be higher than \( P(C) \), \( x^2 = 1.48, .05 < p < .06 \), indicating that there was some learning during the testing phase.

Discussion

The four analyses of the data suggest the following: (a) that the memorisation task provided the Ss with information about the grammaticality of the stimulus sequences; (b) that such information is abstracted out of the environment by S without recourse to explicit strategies for responding or systems for recoding the stimuli; and (c) that such implicitly learned information can be applied efficiently in a transfer-recognition task.

However, while it is reasonable to assume that Ss can become sensitive to the statistical nature of their environment without using explicit or verbalizable strategies, it is not reasonable to assume that such learning takes place irrespective of the stimulus array. When ordered stimuli are presented to a S who has appropriate coding schemes available then he will use such schemes to organize the stimuli (G6mer, 1966). G6mer further implied that when the stimuli are not patterned or S cannot discern the pattern, then he will impose his own organizational schemes. In our case the stimuli were patterned but in a fashion which biased against the possibility of there being appropriate coding schemes available to S. Nevertheless Ss did come to respond efficiently, they developed a strong sensitivity to the lawful-ness that existed in the stimulus array. The process by which such an efficient responding is developed is what is meant by implicit learning. It is closely akin to Gibson and Gibson's (1955) perceptual learning and is, we believe, a rudimentary inductive process which is intrinsic in such phenomena as language learning and pattern perception.

References


(Received March 18, 1969)

*The values of \( P_i(C) \) were used to estimate the \( p \) parameter; the estimates were .88, .60, and .81 for the grammatical, nongrammatical, and pooled data, respectively. The \( x^2 \) tests were run with frequency data.