REBER, ARTHUR S., ***Implicit Learning of Artificial Grammars*** , Journal of Verbal Learning and Verbal Behavior, 6:6 (1967:Dec.) p.855

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

Implicit Learning of Artificial Grammars[1]

ARTHUR S. REBER[2]

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 imposed 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 imitative 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 paradigms—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 Gibsons 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 Gibsonian 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 who 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 simply because it "sounded right" were behaving like Ss 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 experimental technique of repeated exposures of the variation of the stimuli should be sufficient for a S 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 Ss: 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, 1958) was constructed with the five letters P,S,T,V,X as the vocabulary and a set of rules of sentence construction as the grammar. The grammar may be characterized as a Markovian process in which a transition from and state S_i to any state S_j 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 S_0, each transition produces a letter, and the sentence ends when state S_0' 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. Mandelbrot (1954) has shown that the number of sentences in such a language can be described by a "structure function." If $f(\lambda)$ denotes the number of possible sentences of exactly length λ, then the structure function is a recursion in λ which will specify $f(\lambda)$ for all $\lambda = 1,2,3,\ldots$. The structure function for the grammar illustrated in Fig. 1 can be shown to be $f(\lambda) = 2f(\lambda - 1) - f(\lambda - 2) + f(\lambda - 3) - f(\lambda - 4)$.

Given the first four values of $f(\lambda)$, 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 $f(4) = 3$. These values permit the computations: $f(5) = 4$, $f(6) = 7$, $f(7) = 11$, $f(8) = 16$. The latter three values of λ are of primary interest in this experiment because the length of sentences was restricted to 6, 7, and 8.

By means of a technique developed by Chomsky and Miller (1958), it can also be shown that each symbol in this language contains an average of .552 bits of information. The 6-, 7-, and 8-

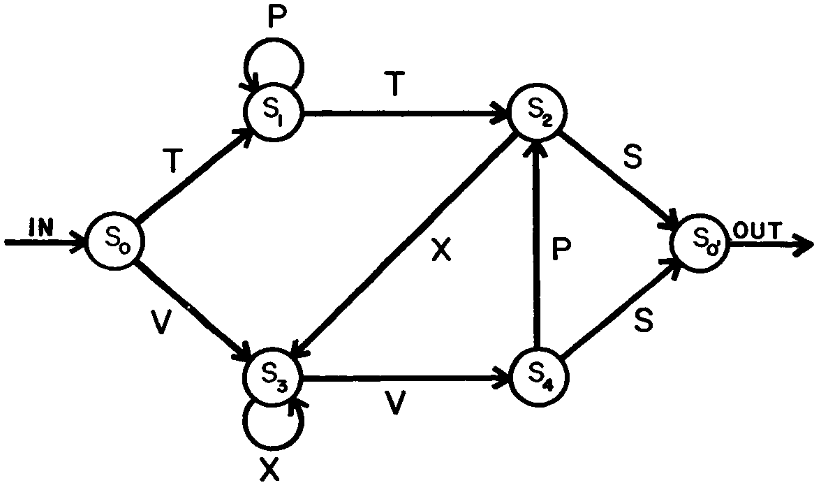

FIG. 1. Schematic state diagram of the grammar used to generate the grammatical stimulus items.

letter sentences, then, contain 3.312, 3.864, and 4.416 bits, respectively.

Random Items. As a control a set of random items was generated from the same 5 letters. The informational value of a symbol in a nonredundant system is given by $\log_2 n$, where $n =$ number of available symbols. Each letter in a random item contains $\log_2 5$, or 2.322 bits, as opposed to the .552 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 34 sentences of length 6–8. These were printed on 3×7-inch cards and constituted the grammatical stimulus items for the experimental Ss (S_g). For each S_g 28 items were selected at random and placed in a fixed order. Every S_g 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 S_g was then given a piece of paper and asked to reproduce the sentences. After each trial *E* removed the S_g's responses and informed him which items he had reproduced correctly and which ones he had not. No information was given about the nature of his errors. After the criterion of two consecutive correct reproductions of a given set was reached, 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 (S_r) learned the randomly formed items in identical fashion. There were 5 Ss in each group. They were undergraduates enrolled in an introductory psychology course.

Results

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

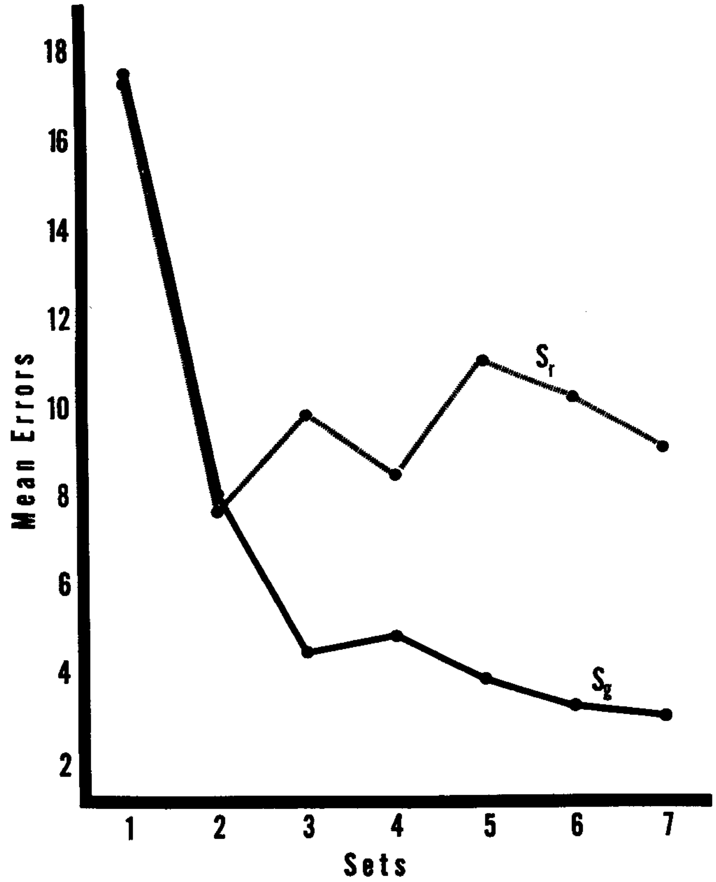

Fig. 2. Mean number of errors to criterion on each of the seven learning sets.

difference between the groups to be significant, $F(1, 8) = 14.87$, $p < .01$. Note however, that the two groups are indistinguishable from each other on Sets 1 and 2. It is not until Set 3 that the S_gs show a significant improvement in performance over the S_rs. 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 gramatical restraint) was not the cause of this change in the behavior of the S_gs. 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 S to adopt would be a stimulus-selection strategy. That is, learn one 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 S_rs 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 S_rs on Sets 3–7 only 29 (12%) were on previously correct items.

The S_gs, 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 S_rs had almost three times as many trials with 1 or 0 items learned as did the S_gs, and only one-half as many trials where more than one was learned. These differences are significant, $\chi^2(1) = 20.93$, $p < .001$. Note also that S_rs required almost twice as many trials as the S_gs. 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 S_rs hardly ever succeeding in correctly reproducing more than one item, the S_gs were just as likely to reproduce two or more as zero or one. A chi-square test showed a clear difference between the groups, $\chi^2(1) = 11.52$, $p < .001$. It should be recalled here that the S_gs are dealing with stimulus items which are 6-, 7-, and 8-letters long, while the S_rs 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 S_gs is amount of information processed. If

TABLE 1

FREQUENCY WITH WHICH VARIOUS NUMBERS OF ITEMS WERE LEARNED ON SINGLE TRIALS ON SETS 3–7

No. of trial	Group	Number of items learned 0	1	2	3	4	Total
All	Experimental *S*s (S_g)	10	38	23	4	1	76
All	Control *S*s (S_r)	63	67	15	1	0	146
First	Experimental *S*s (S_g)	2	10	10	2	1	25
First	Control *S*s (S_r)	7	16	2	0	0	25

Set 2 is compared with an average of Sets 5, 6, and 7 the S_gs show a 21% increase in information processed per trial (10.5–12.7 bits) while the S_rs show essentially no change (31.7–31.9). The difference between 10.5 and 12.7 represents an increase equivalent to processing an additional 4 letters on each trial.

Discussion

It seems rather clear from the above analyses that, while the general learning strategy is the same for all Ss, those in the experimental group were learning to be more efficient, presumably due to the grammatical structure of their items. In Miller's (1958) study of free recall it was hypothesized that Ss were able to recall more redundant (grammatical) items than random items 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(P)TS; (b) T(P)TX(X)(VPX(X))VS; (c) T(P)TX(X)(VPX(X))VPS; (d) V(X)(VPX(X))VS) and (e) V(X)(VPX(X))VPS.

This shorthand method of writing the grammar illustrates the basic constituents of the language: T, V, P, X, VPX, and S. It is not at all clear that an encoding system 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 verbalizable knowledge of the rules of construction. This however, was not the case. After the experiment was over, each S_g was informed that all of the 28 stimulus items were formed by the use of rigorous rules and was asked if he had any ideas about them. Since no S replied, they were all asked four leading questions: (a) What letter or letters may sentences begin or end with? (b) Can sentences begin with a P?, an S?, a T?, a V?, an X? (c) Can sentences end with a P?, an S?, etc. (d) Were there any recurring themes or sequences of letters which seemed to reflect any rules? Not one S answered the first question correctly. By giving successive hints, all Ss were eventually prodded into answering the second and third correctly, and the only concrete response to the fourth was from one S who felt that the sequence VPS (which occurred in 9 of his items and which he had seen a total of 48 times) was significant.

Such a peculiar combination of highly efficient behavior with complex stimuli and almost complete lack of verbalizable knowledge about them is similar to that observed by Braine (1963). The children in his study were shown examples of sentences constructed of nonsense words. His Ss were able to create new, grammatically correct sentences but were not able to verbalize any of the rules. He reported that their typical answer was "I don't know, that just seems (or sounds) right."

The study described here however, does not permit any statements, speculative or otherwise, about the learning processes. The conclusion indicated is that Ss were learning the grammar in a manner analogous to Gibsonian perceptual learning. However, it is not impossible that coding systems of the type proposed by Miller were in use; the introspective verbal reports of undergraduates are notoriously unreliable and usually incomplete. To correct for this shortcoming, a second study was run. 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. I, and (b) a testing phase during which Ss were required to make use of what they had learned about the grammar during the first part. It was hoped that the performance of Ss in the test phase would reflect what they had learned in the memorization task and thereby permit a more rigorous discussion of the learning process.

Learning Phase. The finite-state grammar described in Exp. I was again used to generate the stimuli. Twenty sentences 3–8-letters long were selected so that each type of grammatical construction was included, e.g., for each length sentences beginning with both T and V were used, and for all lengths, where it was possible, at least one example of each of the three loops of the grammar (P, X, and VPX) was employed. These 20 sentences were presented to Ss in four sets of five each; and as in the first experiment, the criterion for learning was two consecutive correct reproductions of a set.[3]

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 no mention was made of grammars or rules of sentence construction. The Ss were not informed that there was a test phase in the experiment until after they had completed the learning task.

Testing Phase. Upon completion of the learning phase each S was told that the 20 stimuli he had just learned were formed by a complex but rigorous set of "grammatical" rules. No information was given about the rules themselves. The Ss were merely told to consider each of the 20 strings of letters to be examples of grammatically permissible sentences. To clarify this point, examples of grammatical and ungrammatical English sentences were given. Each S was then run on a series of 88 test trials. On each trial he was presented with a new string of letters and required to make a decision about its grammaticalness. He was instructed to make his decisions on the basis of what he had learned about the grammar from the 20 sentences he had previously memorized. No feedback was given until all 88 trials were completed.

The list of 88 test stimuli actually contained only 44 different items, 22 grammatical (G) and 22 nongrammatical (NG), each of which was presented twice. Later questioning, however, revealed that no S suspected this repetition, and so the statistical analyses discussed below were carried out on the basis of 88 independent trials. Of the 22 nongrammatical items a few (4) were formed randomly and quite obviously "looked wrong." The others contained rather subtle errors and were not so obviously incorrect. It was hoped that the errors which Ss made in rejecting G items and in accepting NG items would reflect what they had learned about the grammar during the memorization task. Moreover, it was assumed that if any systematic coding schemes were employed by Ss they would be reflected as patterns of responding.

Ten Ss were run; five were high-school seniors who were recipients of NSF Summer Fellowships and five were undergraduates. They were each paid $2.50 for the single two-hour session.

Results

The first thing to note is that the two subject pools were apparently different. The college undergraduates had a mean of 73.5 correct responses on the 88 test trials while the NSF high-school students had a mean of only 65.2. The difference is significant ($t(8) = 5.75$, $p < .01$). However, since each group considered inde-

---

[3] Only one order of presentation was used and it was such that it tended to obscure the learning across sets reported in Exp. I, i.e., the four sets contained 29, 28, 35, and 31 letters, respectively. Nevertheless there was still a hint that Ss improved over sets, the mean trials to criterion were 9.2, 5.4, 7.8, and 5.2 for Sets 1, 2, 3, and 4, respectively.

pendently performed at a level far above chance, and since the emphasis of the study is on the types of errors committed and not on their number, it seemed reasonable to pool all 10 Ss and treat them as a single group. Additional analyses were also carried out in which each S was treated independently in case there were systematic differences in types of errors.

Table 2 gives a breakdown of the test-trial data according to G and NG items. The two most salient features are the high proportion of correct responses and the relative equality of the number of G and NG responses. Out of the 88 items presented to each S an average of 69.4 were correctly identified as to their grammaticalness. This indicates that not only was simple memorization of grammatical strings sufficient to impart information about their construction, but that this information could be used in a transfer task. The similarity between the total number of G and NG responses is a convenient result, for it eliminates the possibility of response biases.

TABLE 2

THE FREQUENCY WITH WHICH GRAMMATICAL (G) AND NONGRAMMATICAL (NG) RESPONSES WERE MADE TO GRAMMATICAL AND NONGRAMMATICAL ITEMS

Item	Response G	Response NG
G	345	95
NG	91	349
Total	436	444

To investigate the possibility of coding systems the test trial data were analyzed in a variety of ways: (a) analysis of the errors committed on NG items, (b) analysis of errors committed on G items, (c) item-by-item analysis of individual S's responses, and (d) analysis according to response probabilities without regard to item type.

NG Items. The NG items were broken down into five basic types depending upon the position in the item where the error occurred. The random items (R), all of which contained multiple errors, had a significantly higher detection rate than any of the others, $\chi^2(1) = 6.46$, $p < .01$. 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 which occurred in either the first (F) or in an internal (I) position. These results indicate that Ss focused most of their attention on the end of the items. In fact, the combined detection rates on the L and L-1 items was significantly higher, $\chi^2(1) = 8.63$, $p < .01$, than the combined rate on F and I items. So far as NG items are concerned then, an error at the end of a string is a more salient cue than one at the beginning or in the middle of a string, and multiple errors render a string more likely to be detected than a single error in any position.

G Items. The above analysis, however, treats only half of the data. It fails to explain why a S would ever reject a G item. Unfortunately, there is no simple way to classify G sentences and there is no simple way to analyze the responses to them. The best that can be done is to examine the five basic sentence types the grammar generates. The proportion of times that each type was incorrectly called NG was calculated. There were no significant differences. Even Type a, 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 a has only one loop and the others have three, this result calls into question the explanation that Ss used coding systems based upon the loops of the grammar.

Intrasubject Analysis. The above two analyses suggest little that might account for the behavior of the Ss as a group. Individually however each S could very likely have developed his own "template" to which he attempted to fit each item, calling it G when it matched and NG when it failed to match. The existence of such a process was checked for by forming the ratio between the number of items on which a S responded inconsistently (i.e., G on one presentation and NG on the other) and the total number of items. Any marked departure of this ratio from zero must be interpreted to mean that the S was not using a consistent matching-to-template strategy. On the basis of this statistic, eight of the Ss may be eliminated, their scores falling in the range .16–.27. The other two have ratios which are suspiciously low (.07 and .09). However, an 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., VXXXXVS, TPPPPPTS); a criterion which does not necessarily reflect grammaticalness.

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 Ss 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 grammaticalness of the *i*th item with probability p_i, and does not know it with probability $1 - p_i$, (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, there are four possible outcomes for each item: (a) correct on both presentations; (b) correct on the first, an error on the second; (c) an error on the first, correct on the second; and (d) errors on both. The following equations describe these four possibilities:

$$P_r(\mathrm{CC}) = p + (1-p)g^2$$
$$P_r(\mathrm{CE}) = P_r(\mathrm{EC}) = (1-p)g(1-g)$$
$$P_r(\mathrm{EE}) = (1-p)(1-g)^2,$$

where p is the average of all p_is and $g = .5$. These equations describe what is probably the simplest detection model possible. On each trial the S either knows the grammaticalness 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 $P_r(\mathrm{CC})$ to "predict" the values of $P_r(\mathrm{CE})$, $P_r(\mathrm{EC})$, and $P_r(\mathrm{EE})$. 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 each case χ^2 fell short of significance. With $N = 440$ 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 incorrectly thinks he knows the grammaticalness of an item, and (b) a S knows the grammaticalness of the *i*th 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 any strategy which was not isomorphic with the grammar, since any such strategy would force items into the (EE) category and inflate it beyond (CE) and (EC). The model then, derives further support from the fact that $P_r(\mathrm{EE})$ is not significantly greater than either $P_r(\mathrm{CE})$ or $P_r(\mathrm{EC})$.

The latter is related to the prediction that $P_r(\mathrm{CE}) = P_r(\mathrm{EC})$. When confronted with an item a *S* will either know or not know its grammaticalness. If he does know it, then he will assumedly always know it, and if he does not know it, then he will assumedly never know it, and

TABLE 3

TEST OF THE SIMPLE DETECTION MODEL

	Grammatical items		Nongrammatical items		All items	
	Pred.	Obs.	Pred.	Obs.	Pred.	Obs.
$P_r(\mathrm{CC})$[a]	.686	.686	.700	.700	.693	.693
$P_r(\mathrm{CE})$	.105	.068	.100	.077	.102	.073
$P_r(\mathrm{EC})$	.105	.128	.100	.109	.102	.118
$P_r(\mathrm{EE})$	.105	.118	.100	.114	.102	.116
	$\chi^2(2) = 4.26$		$\chi^2(2) = 1.73$		$\chi^2(2) = 5.64$	
	$p < .20$		$p < .30$		$p < .10$	

[a] The values of $P_r(\mathrm{CC})$ were used to estimate the p parameter; the estimates were .581, .600, and .591 for the grammatical, nongrammatical, and pooled data, respectively. The χ^2 tests were run with frequency data.

will always be forced to guess. This is a rather strong statement and may perhaps not yield a strictly accurate picture of the data, for there appears to be a tendency for P_r(CE) to be smaller than P_r(EC), $\chi^2(1) = 4.88$, $p < .05$, indicating that there was some learning during the testing phase.

Discussion

The four analyses of the data suggest the following: (a) that the memorization task provided the Ss with information about the lawfulness 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 (Garner 1966). Garner 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 Ss. Nevertheless Ss did come to respond efficiently; they developed a strong sensitivity to the lawfulness that existed in the stimulus array. The process by which this 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

BRAINE, M. D. S. On learning the grammatical order of words. *Psychol. Rev.*, 1963, **70**, 323–348.

BROWN, R. AND FRASER, C. The acquisition of syntax. In C. Cofer and B. Musgrave (Eds.), *Verbal behavior and learning.* New York: McGraw-Hill, 1963, Pp. 158–197.

CHOMSKY, N. *Syntactic structures.* The Hague: Mouton, 1957.

CHOMSKY, N. Review of B. F. Skinner, *Verbal Behavior., Language,* 1959, **35**, 26–58.

CHOMSKY, N. AND MILLER, G. A. Finite state languages. *Inform. Cont.*, 1958, **1**, 91–112.

GARNER, W. R. To perceive is to know. *Amer. Psychologist,* 1966, **21**, 11–19.

GIBSON, E. AND GIBSON, J. J. Perceptual learning: differentation or enrichment? *Psychol. Rev.*, 1955, **62**, 32–41.

MANDELBROT, B. On recurrent noise limiting coding. *Proc. Symp. on Info. Networks,* Polytech. Inst. of Brooklyn, 1954, 205–221.

MILLER, G. A. Free recall of redundant strings of letters. *J. exp. Psychol.*, 1958, **56**, 485–491.

MILLER, G. A., AND CHOMSKY, N. Finitary models of language users. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology.* Vol. 2., New York: Wiley, 1963, Pp. 419–491.

(Received March 18,1966)