VOL. 77, NO. 6 NOVEMBER 1970

PSYCHOLOGICAL REVIEW

THEORY OF SERIAL PATTERN LEARNING: STRUCTURAL TREES[1]

FRANK RESTLE[2]

Indiana University

When college students learn patterned sequences, they divide them into subparts. Each subpart has the property that it can be generated unambiguously by simple rules. Such a rule system consists of E, the set of elements or events forthcoming, and I, the set of intervals leading from one event to the next. Parts or their generating rule systems can be the elements of higher order rule systems. This produces the "recursive E–I theory." One part is generated from another by any of a class of transitions such as repeating, transposing, or inverting. By applying such transitions as compound functions, one generates structural trees, which give a particularly simple account of certain regular patterns. Experimental results show that the difficulty of learning a transition within such a pattern depends on how high it is in the tree. The theoretical results are applied to the theory of music.

Serial pattern learning is the integration of a sequence of responses that are organized in a meaningful way (Lashley, 1951; Miller & Chomsky, 1963). It is distinct from the kind of serial learning often studied in the psychological laboratory, in which nonsense syllables or disconnected words are placed in a completely arbitrary order. Since there is no pattern to be learned in such material, ordinary serial rote learning cannot capture the nature of serial patterns —unless it be agreed that the apparent organization of serial tasks is really the product of a process of association whereby arbitrary elements are welded into a genuine structure. Recent experiments have shown clearly that (*a*) some patterns are easily learned, (*b*) other sequences of events hardly form a pattern at all, and (*c*) the pattern is a property inherent in the sequence of events (Restle & Brown, 1970a, 1970b). The present paper is concerned mainly with delineating the nature of a pattern and clarifying the major theoretical difficulties in this concept as it applies to the development of patterned and coordinated behavior.

Many forms of human behavior, from simple walking to the complexities of driving an automobile, speaking, playing the piano, or even playing chess, might reasonably be considered serial patterns. In these cases, the effect of the behavior lies in the smooth serial arrangement of serial elements, and the behavior is said to have a meaningful and organized pattern. Most behavior is inconvenient to study in the laboratory because (*a*) the "vocabulary" of elementary behaviors is too large, (*b*) the occasion for the behavior is difficult to control, or (*c*) the component responses have no separate environmental effects that make them convenient to record. Furthermore, many meaningful behaviors are so practiced that it is difficult to separate intrinsic structure (if there is any) from the

[1] This research was supported by National Science Foundation Grants GB-2848 and GB-5714 and by United States Public Health Service Grants MH-12541 and MH-16817. The theoretical and experimental work reviewed here has been developed jointly with Eric R. Brown, whose spirited and inventive interpretations are an inextricable part of this theory.

[2] Requests for reprints should be sent to Frank Restle, Department of Psychology, Indiana University, Bloomington, Indiana 47401.

effects of extended overtraining which might smooth and integrate an arbitrary sequence of movements.

In the experiments performed in the author's laboratory, college students each work with a response box having six buttons in a row, with an "event light" above each button. The lights are turned on in a fixed and regular pattern, and the subject (*S*) tries to anticipate each light correctly. Since the sequence of events is perfectly periodic, the problem can eventually be solved, but since *S* is required to anticipate each light in a short time (usually 2–3 seconds), the task is quite difficult and results in many wrong responses on early trials. The main experimental method has been to vary the sequential pattern of lights shown, to determine what locations in the sequence are most difficult to learn, and what errors are most frequently made.

A basic experimental study (Restle & Brown, 1970a) showed that none of five conventional stimulus–response theories of serial learning can handle even the simple data of serial pattern learning. The theories tested and found wanting were (*a*) that each item or element in a serial pattern is the stimulus to which the next prediction is associated, forming a simple S–R chain; (*b*) that *S* knows where he is in the list and associates his responses to the serial positions; (*c*) that *S*s learn by S–R associations, but that several adjacent past events constitute the stimulus term; (*d*) that *S*s bring particular past response patterns, like telephone numbers or familiar melodies, into the experimental situation and quickly learn patterns conforming to such prelearned units; (*e*) that *S*s enter the experiment with a set of fixed error tendencies which become eliminated by training.

Such theoreis take no proper account of the intrinsic organizing possibilities in the sequence, and it is the aim of this paper to develop the theoretical tools necessary to understand the serial patterns developed by *S*s in experiments.

An approach to the needed theory was given by Restle (1967) in a theory of two-choice behavior. The six-choice system provides a clearer look at the underlying process, and the present theory is concerned with this better experimental example of serial learning. The present stage of this theory is the result of a sequence of attacks on the problem, each solving one theoretical problem but opening another. The theory is most clearly presented by giving its simplest form first, then progressing through a series of refinements and improvements up to the present status of research. Basically, the theoretical problem is broken into two major questions: first, a study of the parts into which *S*s subdivide a long serial pattern, and an analysis of the inner structure that such a part has; second, the relationship between the parts and the way they are connected together to generate the whole sequence.

PARTS OF A SERIAL PATTERN

Human speech can be divided into passages, which in turn divide into sentences, clauses and phrases, words, and speech sounds. Music is divided into movements, sections, themes, and measures. It seems overwhelmingly obvious that long and complex serial patterns are divided into natural subparts, and that mastery is facilitated if the incoming sequence of events is somehow marked off into natural subparts. The first question to which this theory is devoted is, what is the nature of a "natural" or "useful" subpart.

Runs and Trills

The first experiments using serial patterns of six lights revealed the prominence of runs of events, like (2 3 4 5). If such a run appears anywhere within a periodic sequence, *S*s are likely to notice and use it both as a marker indicating the length of the periodic sequence and as the basis for correctly predicting the events later in the run, for example, predicting 4 and 5 of the sub-sequence (2 3 4 5). The data consistently found are that (*a*) the later events in a run are learned with very few errors and (*b*) runs are frequently overextended if that is possible, for example, after (2 3 4 5) the next prediction is likely to be 6, even if that is wrong.

Although runs are the most prominent landmarks, *S*s can also separate out trills, sub-sequences like (5 4 5 4). The tendency to use trills as sub-sequences can be detected in experiments (Restle & Brown, 1970b), but is noticeably weaker than the tendency to use runs. No other comparably universal organizing tendency at this level has been found by the author.

The idea of generation by rules. It is not enough to say that *S*s use runs and trills as sub-sequences, for a structural analysis of the pattern should explain why runs and trills are favored and should suggest how such a structural possibility would lead to advantages in the process of learning and performing the task. Why should certain sequences of events be noticed and used as subunits? The theory that comes most naturally to mind is that such sub-sequences are easily generated by rules and, therefore, have an inner simplicity that is of benefit to *S* in learning the sequence. The run (2 3 4 5) is generated merely by moving always one button to the right of the last light. The rule is nearly as simple as it could be, it generates the whole sequence, and it even generates the run-overextension errors found in great numbers. However, the rule-generation hypothesis is not as convincing as it seems at first glance, unless it is supplemented by a statement of the class of rules that may be used in generating sub-sequences and by some reasonable account of why those rules rather than others would be available to *S*.

The theoretical difficulties may be brought into focus by comparing the human learner with a computer. A computer easily solves a serial pattern by storing events in order as they occur, then reading out the results in the same order. Even if the computer chose to use subunits, there is no reason for it to act like human *S*s. For humans, (2 3 4 5) forms an easy subgroup, whereas (2 5 4 1) does not. A computer could handle one as easily as the other. For that reason, even if a computer chose certain preferred sub-sequences and then was skillful at learning sequences that could be divided into such sub-sequences, there is no reason why the sub-sequences should be runs and trills.

The reason computers do not care about the content of a sub-sequence is that the computer generates serial order by placing events in successive locations of memory. Given that capacity, a computer can learn any one serial arrangement of a given set of elements as easily as any other. Since human *S*s are not at all indifferent to content, it follows that when they learn a serial pattern they do not impose the order by sequential storage of elements, but instead generate the order from a system of rules. The rules, in turn, apparently do not have any sequential ordering built into them, such as would exist if the rules could refer to ordered arrays or sets of elements.

Events and intervals—the E–I theory. One system of rules that generates runs and trills, yet does not have any ordered sets within it, consists of a set E of events of elements, such as lights in a pattern-learning apparatus, and a set I of intervals (the term being used in the musical sense). As a sequential pattern unfolds, it is possible to think of the light moving from one place to another on the response box, and each such step has both a magnitude and a direction. Such steps are here called intervals. Let $E = \{e_1, e_2, \cdots\}$ be the set of events, and $I = \{i_1, i_2, \cdots\}$ be the set of intervals. Neither set is ordered, so there is no sequential information used in their definition. However, some combinations of a certain E and a certain I may combine to determine a particular sequence, at least if the starting point is provided. For example, let $E = \{2, 3, 4, 5\}$, $I = \{-1\}$, and provide the starting point as 5. The above pair (E, I) then uniquely determine the run (5 4 3 2). After 5 there is only one interval, -1, which produces Event 4, after 4 the interval produces 3, etc. In this case, after Event 2 the interval produces Event 1, which is outside E. This might stop the process, in which case the above structure uniquely produces the subunit.

The above system of rules generates an unambiguous sequence because I has only

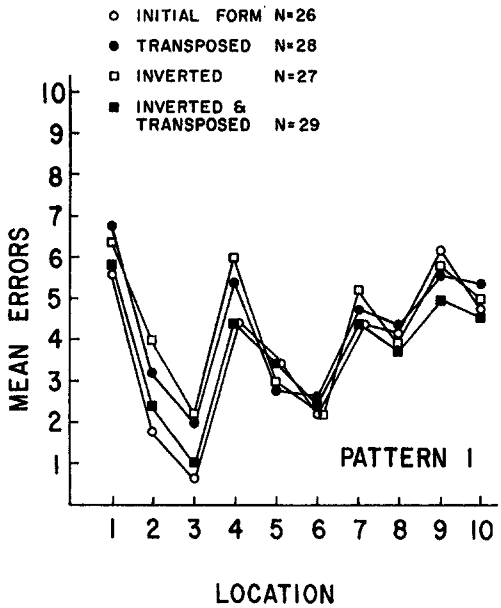

FIG. 1. Mean errors at each location (serial position) for four different forms of a given pattern of events.

one member. If it had two members—say, if I = {−1, −2} and E were as above—it could produce three different possible sub-sequences: (5 4 3 2), (5 3 2), and (5 4 2). Clearly, a set of rules of the form E, I will produce a unique sub-sequence if it has only a single interval. The important theoretical point is that the set of all possible runs is identical with the set of all sub-sequences using a single interval. The particular prominence of runs is supported, therefore, by the two theoretical facts that (*a*) the set of runs can be characterized as the products of a well-defined set of E, I rule systems, namely those having a single interval, and (*b*) all such rule systems have the property of producing unique and unambiguous sub-sequences, which in turn would be most useful in learning serial patterns.

A rule system can have more than one interval yet generate a relatively unambiguous sub-sequence. Consider the system having E = {3, 4} and I = {− 1, + 1} and starting at Event 3. There are two possible intervals, but Interval − 1 would lead to Event 2, which is not in the allowed set E. Therefore, *S* must choose the other interval, + 1, which leads to a prediction of Event 4. After Event 4, Interval + 1 leads to 5 which is not in the set E, so Interval − 1 must be used, leading to a prediction of 3. In this way the sub-sequence (3 4 3 4· · ·) is generated uniquely. Notice, however, that there is no way within this rule system for terminating the sub-sequence at the right time.

Application of the E–I theory. The two kinds of data that show how *S* organizes a sequence are his error profile and his high-frequency errors. Inspection of the error profiles for several versions of the same two patterns (Restle & Brown, 1970a) shows that the mean errors drop during a run. For example, Pattern 1 of that study consisted of (1 2 3 5 4 3 3 2 3 4), having runs in Locations 1–3, 4–6, and 8–10. There is a clear decline in errors during Locations 1–3 and during Locations 4–6, as shown in Figure 1. The last part of the sequence generates a more muddled profile, but this may be because it is (3 2 3 4) which starts like a trill (3 2 3 2). The other pattern in that experiment begins with a run of length 4 and ends with a run of length 3, and in both those blocks of locations show a generally decreasing profile of errors.

The detailed data of that experiment, as well as a large number of preliminary and pilot studies, consistently show an excess of errors that can be identified as "run continuation" errors.

More direct evidence that trills and runs form functional sub-sequences comes from a study (Restle & Brown, 1970b) in which one group of *S*s was pretrained on a sub-sequence composed only of runs (6 5 4 3 5 4 3 2 4 3 2 1) and another group was pretrained on a sequence composed only of short trills (5 6 5 3 4 3 3 2 3 1 2 1). Both groups were then tested on the ambiguous sequence (2 1 2 3 4 3 4 5 6 5) which can be organized either as runs (2 1 2 3 4 3 4 5 6 5) or as trills (2 1 2 3 4 3 4 5 6 5). The experimental result was that *S*s pretrained on runs showed, in the ambiguous test sequence, an error profile and a pile-up of errors very characteristic of organization into run sub-sequences, whereas *S*s pretrained on trills (who had not learned their original sequence too well) gave corresponding evidence of having organized the ambiguous sequence into trill subunits.

One important characteristic of the E–I theory is that E–I rule systems cannot generate very many different unambiguous sub-sequences. Consider, for example, a sub-sequence candidate (1 2 4 3). Writing the sets of events and intervals then, $E = \{1, 2, 3, 4\}$ and $I = \{+1, +2, -1\}$. These rules can generate the sub-sequence, but they equally well generate (1 3 2 1 3 4 3 2 4), for example, and infinitely many other sub-sequences. The rule system that generates (1 2 4 3) is uniquely determined, for it is obtained merely by taking inventory of the events in the sub-sequence and the intervals. If that rule system also generates other sub-sequences, then it cannot generate any sub-sequence unambiguously, hence cannot itself be the basis for stable correct performance. Other sub-sequences that have ambiguous generating systems are (1 1 2 3), (6 5 4 2), (3 4 3 4 4).

The E–I theory makes a distinction between "good" subunits that can be generated unambiguously by rule systems, and other candidates that cannot be generated unambiguously. The ability to make such a distinction is the only thing that gives the theory real meaning, for a system that admitted all possible strings as possible subunits would be without value. However, having shown that (1 2 4 3) is not a well-organized subunit, the theory must attempt to do something with it, for *S*s can learn sequences that are not "good" subunits in the sense defined above.

Another way of describing the remaining theoretical question is to say that a theory of subunits is not itself a theory of serial pattern learning until it says how the subunits are integrated together. The simple E–I theory given above gives no basis for integrating subunits, and it is to this question that we now turn.

Recursive E–I Theory

The most economical approach to the integration of sub-sequences is to suppose that it works exactly like the organization of elementary events within a sub-sequence. Such a theory requires that a subunit, itself generated by an E–I system, should be able to serve as an element in a higher order rule system.

For example, consider the sequence (2 3 4 3 4 5) which cannot be generated unambiguously as a subunit. If this sequence is divided into two parts, (2 3 4) and (3 4 5), then each subpart is a run and has an unambiguous E–I generating system. In using the theory recursively, one now assigns names to the two subunits.

The first subunit, (2 3 4), has $E = \{2, 3, 4\}$ and $I = \{+1\}$. We use the standard notation of angle brackets to signify elements of an ordered pair, so that the generating rules for the first subunit are

$$A = \langle\{2, 3, 4\}, \{+1\}\rangle$$

giving E and then I. Let the second subunit, (3 4 5), be called B, where by the same definition

$$B = \langle\{3, 4, 5\}, \{+1\}\rangle.$$

Now the higher order generating system is developed by defining its set of elements, E, as consisting of the two rule-systems A and B. The interval between the two subunits appears to be -1, for that is the interval needed to get from the end of the first subunit to the beginning of the second. Therefore, the higher level generating system is

$$C = \langle\{A, B\}, \{-1\}\rangle$$

which also is an E–I structure. This system can generate (i.e., *S*s using such a structure could learn) patterns that cannot be generated unambiguously by first-order E–I patterns, but which can be analyzed as arrangements of such subunits.

Such a theory, then, has the theoretical advantage that it permits the definition of certain complex sequences that can be divided naturally into simple subunits. This, in turn, divides these more complex sequences into two subgroups, those that can and those that cannot be learned by recursive E–I systems.

For example, the sequences (2 3 4 3 4 5), (3 4 3 1 2 1), and (5 6 5 4 5 4 3 4 3 2 3 2) can all be learned as recursive E–I systems, for each can be divided into subunits consisting of either runs or trills (that can be generated by simple E–I systems) and then the subunits can be strung together in a simple "run" of subunits. That is, at the higher level, each of the above sub-sequences requires only a single interval between sub-sequences.

Other sub-sequences exist that cannot be generated by recursive E–I systems. Consider, for example, (2 3 4 3 4 4 5 6 1). It is complex and cannot be generated by a first-order E–I system and, furthermore, it has no subunits (other than pairs of successive elements) into which it can be divided. Such a sequence defies any complete patterning, though since some subunits can be extracted (2 3 4 and 4 5 6, for example), rapid partial solution of the problem might be expected.

This higher order, recursive form of the E–I theory, therefore, has the advantage that it specifies sequences that should be easy to learn even though they have more than one subpart and, therefore, cannot be generated from the simple E–I rules that do generate viable subparts.

The above system, operating at a higher and also at a lower level, seems to shed light on the general problem of serial organization, but has at least two serious theoretical weaknesses. First, in the definition of low-order systems, particularly trills, there is as yet no way of specifying within the rules how *S* should terminate a sub-sequence—how he would know that he has reached the end of a trill. The recursive E–I system does not aid in solving this problem. Second, the recursive E–I system contains ambiguities regarding the intervals between sub-sequences in higher order patterns. For example, consider the pattern 4 5 6 3 4 5 2 3 4. It is made up of three runs, 4 5 6, 3 4 5, and 2 3 4, each of which is a valid sub-sequence. The question is, What is the interval between such runs? One answer would be that it is the gap between the end of one run and the first element of the next, therefore, −3. However, in this case it seems more sensible to say that the run (3 4 5) arises from the run (4 5 6) by an interval −1, applied not to the last element of the run but to the first.

Further thought along this line suggests that the relationships possible between subunits are not the same as the relationships between elementary events. The shift from one event to the next can be described as an interval, but the relationship between one sub-sequence and another can be more complex. That is, in the above example it is apparent that (3 4 5) is a translation of the whole sub-sequence (4 5 6), hence one should not decide whether to apply the interval to the beginning or the end of the sequence—it is applied to the whole sub-sequence. Other transitions are possible besides simple transposition; for example, a sub-sequence can be repeated (2 3 4 2 3 4) or, in the symmetrical six-light apparatus, the mirror-image sequence can be developed, turning (2 3 4) into (5 4 3).

From such considerations there arises a theoretical approach to serial pattern learning which is more general and complete than those discussed above, and which is also more capable of exact and comprehensive analysis of the available data.

Structural Trees

The general idea of a truly hierarchical model for sequential learning is that the total sequence "concept" or system of rules serves to generate a sequence of certain elements. Each element is a rule system, which in turn can generate other elements. The elements of any rule system can be other rule systems, or (at the tips of the branches of this tree) may be specific events.

Certain patterns or sequences have the possibility of being generated by simple hierarchical trees, just as certain sub-sequences can be generated by a simple E–I rule system. However, there is a far greater richness of patterns that can be generated by trees and, in fact, the limitations on such a system are set mainly by the depth of the tree permitted (if any limitation can reasonably be set) and the variety of different transitions that *S* can use. We have already mentioned repeating, transposing, and taking the mirror image as transitions college students can use on the six-light apparatus.

The first theoretical task is to define these operations. Let X be any sub-sequence, and f(X) be another sub-sequence derived from X by a given operation of transition, f. For example, if X = (1 2 3) and f is transposing +1, then f(X) = (2 3 4). If X and Y are any sub-sequences, then let (X + Y) be the sequence generated by letting Y follow X. Then, in general, the way of building a sequence in serial pattern learning is by taking a sub-sequence X and concatenating to it another sub-sequence derived from X, thereby producing (X + f(X)). In the above example where X = (1 2 3) and f is the operation of transposing +1, the resulting sequence would be (X + f(X)) = (1 2 3 2 3 4).

Transitions like f are mainly used in this concatenating fashion, in building up serial patterns. Therefore, it is most convenient to define the operation of both transition and concatenating as a single operation. Capital letters will be used for such operations, and the mnemonics are as follows: T is transposition, and usually has a subscript and a number to indicate the distance of transposition intended; R means repeat; M means mirror image; and other operations may be introduced as needed.

Now, if X = (1 2 3), then R(X) = (1 2 3 1 2 3), which is, of course, (X + r(X)), where r is the operation of repeating. Similarly, M(X) = (1 2 3 6 5 4).

The notion of structural trees requires that these operations can be performed upon the products of other operations, following the mathematical concept of a compound operation. For example, starting with X = (1 2 3), one can produce T(X) = (1 2 3 2 3 4) and then M(T(X)) = (1 2 3 2 3 4 6 5 4 5 4 3).

The first characteristic of such a theory is that it provides a concise description of very long sequences, yet gives a kind of structural analysis at the same time. By giving an account of the inner structure, such a theory lays the groundwork for a theory of how *S* might generate such a sequence. The same theoretical idea, with somewhat different applications, has already been presented by Leeuwenberg (1969).

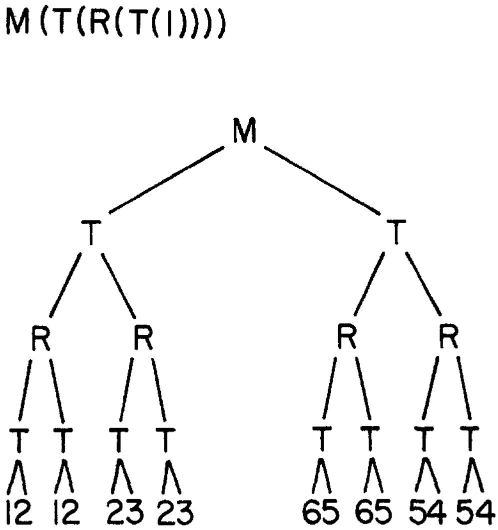

FIG. 2. Tree diagram of a long, regular binary pattern.

This theoretical approach gives a particularly full account of a class of regular binary trees and the sequences arising from them. Figure 2 shows a long sequence and the regular binary tree that would generate it. The second half of the sequence is the mirror image of the first; each half of the sequence is divided into two equal parts, the second being the transpose of the first. Each of these parts is the repetition of two equal parts, and each of those parts is an element and its transpose. In the algebraic notation of compound functions, this sequence would be described as M(T(R(T(1)))). The functions mentioned in the algebraic formula appear in the order from the top down in the structural tree.

To develop a correct description of a serial pattern, it is necessary to discover the subunits used, to discover the relationships between them which form second-order structures, and then to pursue this process, forming higher and higher order structures until the entire serial pattern is constructed. Such a process cannot give an interesting description of an amorphous or meaningless sequence, and, therefore, the theory as stated here is not easily used to assign "complexity" values to arbitrary sequences.

The use the author and his colleagues have

TABLE 1

FOUR SERIAL PATTERNS ARISING FROM REGULAR BINARY TREES

Functional formula	Sequence
M(R(T(R(T(1)))))	12122323121223236565545465655454
T(R(M(R(T(1)))))	12126565121265652323545423235454
T(M(R(T(R(1)))))	11221122665566552233223355445544
T(R(T(M(R(1)))))	11662255116622552255334422553344

made of this theory is more "constructive." Following the suggestions of the system, we have built very long sequences, which we previously had thought too long to be learned, but have concentrated on regular binary trees. In this way we have been able to deal with sequences which, according to this theory, should be highly organized at many different levels.

Learning of Long, Regular Sequences

In one recent unpublished experiment college students attempted to learn sequences of 32 events, presented 20 times. Data on four of the sequences are reported here, the results of 19, 26, 23, and 18 *S*s, respectively.

All four patterns arise from regular binary trees, the only differences being the order of the operations, that is, which operations are to be found at which level of the tree. These are the four subsequences which did not, in their development, produce any strong conflicting organization at any point. The four sequences are shown in Table 1, along with their formal descriptions.

If the hierarchical theory is correct, then these sequences should be quite learnable and, in fact, performance was approximately 75% correct over the first 20 trials, despite the usual scattering of ineffective *S*s. Much more important is the profile of errors. A hierarchical tree structure leads to the hypothesis that *S* should have the most difficulty in learning and correctly predicting the highest order transformations in the sequence, and should find the lowest order transformations easiest. There are several possible reasons, the most obvious of which is that there are many more examples of the lower than of the higher order transitions within the sequence; there is only one transition from the first half to the second half, but there are (in sequences of length 32) a total of 16 examples of the lowest order transition.

To a rough approximation, the difficulty of a given transition may be discerned in the errors made right after the transition must be made. The transition from first half to second half will be found most clearly in response to the first element of the second half, though failure to make the transition may produce errors later, and it is even possible that an erroneous system of predictions may produce correct responses for part of the sequence and then manifest itself elsewhere. However, the particular sequences under consideration do not show such side effects in a prominent way, so it is reasonable to suppose that the difficulty of any location might be predicted from the level of transition immediately preceding it.

If so, then these binary patterns should show a very complicated "serial position effect" or profile. Worst performance should be at Location 17, the first location of the second half. Next worst should be Locations 9 and 25, starting the second and fourth quarters. Next worst would be Locations 5, 13, 21, and 29, followed by the next level which has Locations 3, 7, 11, 15, 19, 23, 27, and 31. The easiest locations, at the lowest level of all, should be the even-numbered ones. Rather than showing this saw-toothed profile, it is simpler to segregate the types of locations, so as to see if performance is really better at the lowest level (even numbered) locations, and progressively worse the higher in the tree. The results of such an arrangement of the data are shown in Figure 3 from which it can be seen that all four patterns show the same monotonic trend.

Not only is the profile of errors in good agreement with the hierarchical tree, but it should also be noticed that the level of performance on these sequences is quite high, averaging near 80% correct responses of 20 trials, that is, about four errors per location total.

This theoretical approach makes other

predictions, which are now being tested. One is that under suitable conditions, *S* may make errors by mistaking what part of the total pattern his is in. Errors will not only consist of overrunning low-order structures (as in the run overextension error predicted from the E–I theory), but may also arise from overextending higher order structures. If this is so, then errors may sometimes be attributed to a tendency to "generalize" from one part of the pattern to another, remote location which has the same relative position within the tree. Another further development of the theory is to put imperfections within a sequence; for example, to take a perfect binary pattern and change the element at one location. This should result in many errors at that particular location as *S* makes the response that would naturally arise from the general tree structure. Eventually, *S*s would no doubt learn the specific inhomogeneity, but then might make certain *corresponding* errors, making the correction in the wrong part of the pattern. Thus, the effects of the inhomogeneity might be expected to "resonate" through the tree structure, and analysis of specific errors would then permit a detailed analysis of the structure learned.

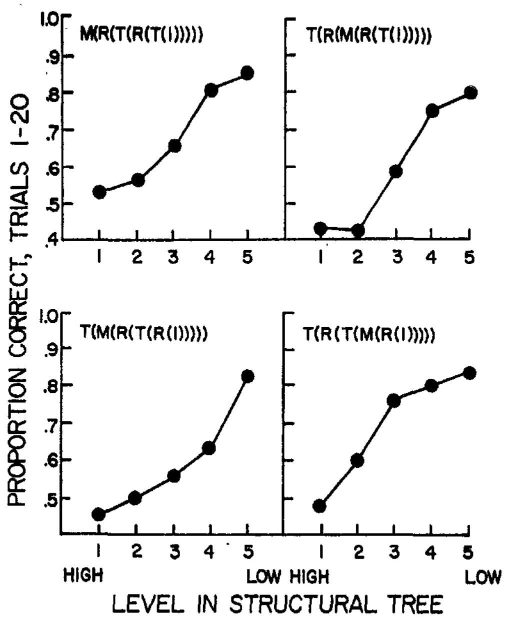

FIG. 3. Mean errors of various locations arranged by their level in the structural tree, for four different sequences.

The Problem of Ending and Connecting Subparts

The value of the theory of structural trees is shown not only in its ability to analyze long, symmetrical sequences, but also in its ability to solve some of the problems left unsolved by the recursive E–I theory.

Ending subparts. The E–I theory could not specify how trill sub-sequences could be stopped on time. For example, consider the trill (3 4 3 4). In the E–I theory, E = {3, 4} and I = {−1, +1}. This system would also generate (3 4 3) or (3 4 3 4 3), etc. It was suggested that *runs* were stopped whenever the system first generated an event that is outside the range of E, but that was a slightly artificial approach and had the further disadvantage that it did not explain *S*s' strong tendency to overextend runs (except by saying that for some reason E was very difficult to learn).

The theory of structural trees would generate (3 4 3 4) as R(T(3)). That is, it would say that one starts with Event 3, transposes to obtain (3 4), then repeats that to obtain (3 4 3 4). This system permits a definite end to the trill. It should be very simple to end trills at lengths 2, 4, or 8, since these would merely require symmetrical subtrees, T(3), R(T(3)), and R(R(T(3))). Trills of length 3 present a conceptual difficulty that may be overcome by supposing that *S*s can adopt an operation of elision, leaving off the last of a trill of length 4. This might explain why there is a tendency to overextend trills from length 3 to length 4. According to the structural trees theory, a trill is merely a transposition, and the number of repetitions is controlled by the depth of the tree of repetitions.

Connecting subparts. A second question, not well solved by the E–I theory, was how to connect subparts. Does the "interval" lead from the *end* of one subpart to the beginning of the next, as seems natural in connecting (2 3 4) with (4 3 2), or from the *beginning* of one subpart to the beginning of the next, as seems natural in connecting (1 2 3 4) with (2 3 4 5)? According to the theory of structural trees, the answer is that the notion of "interval" is replaced with the more general idea of "transition." Thus, (4 3 2) is an inversion of (2 3 4), whereas (2 3 4 5) is a transposition of (1 2 3 4).

Identifying subparts in ambiguous patterns. In some experiments (i.e., Restle & Brown, 1970b), ambiguous sequences were used which permitted two very different, though incomplete, structural trees. The pattern (2 1 2 3 4 3 4 5 6 5) is an example. Looked at as trills, its subparts are (2 1 2), (3 4 3 4), (5 6 5). Looked at as runs, this pattern has two main parts with two left-over elements: (2) (1 2 3 4) (3 4 5 6) (5). It was shown experimentally that either organization might be favored depending on pretraining (Restle & Brown, 1970b).

This result sets a problem for any structural theory. Many sequences may be ambiguous, especially in one or another part. What, then, is to control the set of subparts used?

The structural-tree theory states that any firmly established set of subparts should be related so that one part can be generated from another, and so that the whole pattern can be generated from a single tree. This principle does not determine which of two possible organizations will prevail. However, it argues that certain mixed structures, like, for example, (2 1 2) (3 4 3) (4 5 6) (5), may be unlikely. At any rate, *S*'s choice of one subpart is not independent of his choice of others. Instead of *S*'s choosing among many subparts, we may think of his choosing among a relatively few tree organizations, each of which establishes its own set of subparts.

Inhomogeneous Trees

In the trees described above, every part could be generated from the previous part (at some level) by a single transformation. Such structures are too simple for communication, and even for aesthetic purposes. The sonata allegro, in music, traditionally uses two contrasting themes. This has several advantages, one of which is that the two contrasting themes are highly distinguishable, so that *S* is not easily confused as to where he is in the sequence.

Inhomogeneous patterns may be either easier or more difficult than homogeneous ones. It is more difficult to generate an inhomogeneous pattern, but easier to discriminate the parts. The effects on serial pattern learning remain to be analyzed.

Right-Branching Trees

Early studies by the author showed the importance of runs, but a recent study showed how this tendency can overcome severe distortions and interruptions. The *S*s learned the sequence 1122665522335544, a regular binary sequence that can be written T(M(T(R(1)))). A common and persistent error was, at Location 9, to make Response "3," making the initial segment 112266553. It is difficult to see where this 3 comes from, for it is not an association to 5 or to 55, nor the transfer of any simple interval. It appears to be a continuation of the earlier segment 1122, which was interrupted by 6655 but is continued later. In agreement with this hypothesis, it was found that Response "4" appeared as a strong and persistent error at Location 13, as an apparent continuation of 6655 as well as of 2233.

With a regular binary tree it is easy to express the sub-sequence 11222233 as T(T(R(1))), but it is more difficult to express 11223344, which is easier to learn. The trouble is in expressing a relatively long run, which cannot arise simply from binary trees.

A "right-branching tree" has the form shown in Figure 4 and is characterized by the following property: at each level the tree breaks down into two parts—to the left, a single element or small tree built on a particular element; to the right, a subtree with the same property as the parent tree. A binary tree, by contrast, breaks down into two equal parts.

Chomsky (1965) characterizes trees by the rewrite rules that show how to go from higher to lower levels of the tree. In a binary tree, the two parts are comparable and would be characterized by $S \rightarrow S_1 + S_1'$, where, in the present applications, S_1' would be some transition from S_1. A right-branching tree, by contrast, would be $S \rightarrow x + S'$, where x is either an element or a small subtree.

In Figure 4 the nodes of the right-branching tree are marked with the transition used. A superscript shows how many times the transition is still to be used. Thus, one may write the tree generating (12345) as $T^4(1)$. Then,

$$T^4(1) \rightarrow 1 + T^3(t(1))$$

where t(1) is the transpose of 1, namely, 2. Thus,

$$T^4(1) \rightarrow 1 + T^3(2).$$

Suppose that in general, f is any function taking one element to the next. Then the tree $F^n(e)$ is a right-branching homogeneous tree starting with Element *e*. The general rewriting rule is

$$F^n(e) \rightarrow e + F^{n-1}(f(e)).$$

Even more generally, if S is any sequence to which f can be applied, then

$$F^n(S) \rightarrow S + F^{n-1}(f(S)).$$

Two examples will be given to show how this operates. Consider, first, how to produce the sequence $R^2(T^3(2))$. Starting at the outside,

$$R^2(T^3(2)) \rightarrow T^3(2) + R^1(r(T^3(2)))$$

and the second term rewrites as

$$R^1(T^3(2)) \rightarrow T^3(2) + T^3(2)$$

so that the original sequence becomes

$$R^2(T^3(2)) \rightarrow T^3(2) + T^3(2) + T^3(2).$$

The next stage of rewriting gives

$$\begin{aligned} T^3(2) &\rightarrow 2 + T^2(t(2)) \\ &= 2 + T^2(3). \end{aligned}$$

Then,

$$\begin{aligned} T^2(3) &\rightarrow 3 + T^1(t(3)) \\ &= 3 + T^1(4) \end{aligned}$$

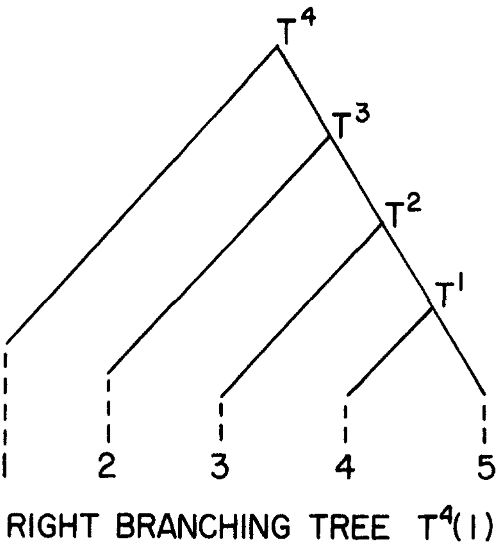

FIG. 4. Diagram of the right-branching tree of a run.

and

$$\begin{aligned} T^1(4) &\rightarrow 4 + t(4) \\ &= 4 + 5. \end{aligned}$$

Reassembling all of the parts from the successive rewritings,

$$T^3(2) \rightarrow 2 + 3 + 4 + 5,$$

whence

$$R^2(T^3(2)) \rightarrow (2\ 3\ 4\ 5\ 2\ 3\ 4\ 5\ 2\ 3\ 4\ 5).$$

This is a right-branching tree of right-branching trees.

A second example illustrates the complexity of organization that can be attained, for it is a binary tree having two right-branching parts, each of which has a small binary tree. The sequence is

$$S = M(T^3(R(1))).$$

The tree diagram of this sequence is shown in Figure 5. The derivation is straightforward:

$$S \rightarrow (T^3(R(1))) + m(T^3(R(1))).$$

Combining a few steps,

$$\begin{aligned} T^3(R(1)) \rightarrow R(1) &+ t(R(1)) \\ &+ t(t(R(1))) + t(t(t(R(1)))), \end{aligned}$$

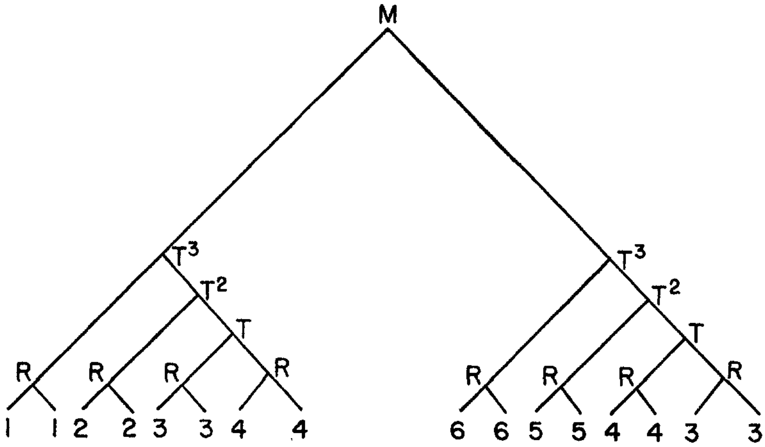

FIG. 5. Diagram of a mixed binary and right-branching tree, $M(T^3(R(1)))$.

and since $R(1) \rightarrow (11)$,

$$T^3(R(1)) \rightarrow 1\ 1\ 2\ 2\ 3\ 3\ 4\ 4,$$

and

$$M(T^3(R(1))) \rightarrow 1\ 1\ 2\ 2\ 3\ 3\ 4\ 4\ 6\ 6\ 5\ 5\ 4\ 4\ 3\ 3.$$

Notice that a binary tree is merely a special case of a right-branching tree in which there is only one right branch; that is, $T(X) = T^1(X)$.

Formal Recapitulation

The structural theory of serial pattern learning can be summarized briefly. Any sequence can be rewritten as follows:

$$S \rightarrow A + B \qquad [1]$$

where "+" signifies concatenation.

If F is any transition function and S any sequence to which it applies, then

$$F^n(S) \rightarrow S + F^{n-1}(f(S)). \qquad [2]$$

Furthermore,

$$F^0(S) \rightarrow S. \qquad [3]$$

With this system, the structural tree of any regular sequence, using any mixture of binary and right-branching trees, can be constructed. Furthermore, using the first rule (Equation 1), it is possible to give an account of nonhomogeneous trees, namely those requiring use of such an arbitrary rewriting rule.

APPLICATIONS

The main experimental application of the present ideas has been to experiments in serial pattern learning. The above theoretical development is the current state of several years of continuous interplay between theory and experiment, attempting to account for details of experimental outcomes.

The first major finding (Restle & Brown, 1970a) was that different versions of the same pattern, transposed or given a mirror-image transformation, produce the same difficulties for *S*s. In particular, the number of errors at each location and the most frequent errors were found to be the same in various versions of the same pattern. Patterns tend to be divided into parts by *S*s, and a common pattern of data includes the following properties: (*a*) more errors at the beginning than at the end of a subunit; (*b*) a large number of errors right after a subunit, in which the organizing principle of the subunit is extended too far. If, for example, (2 3 4) is a subunit, many *S*s will overextend it and respond 5 in the next location even if it is wrong.

In recent experiments, relatively long sequences of up to 32 locations have been used, mainly sequences generated from regular binary trees. The detailed data are then combed for specific errors, at specific locations, that are unusually frequent. When such a high-frequency error is lo-

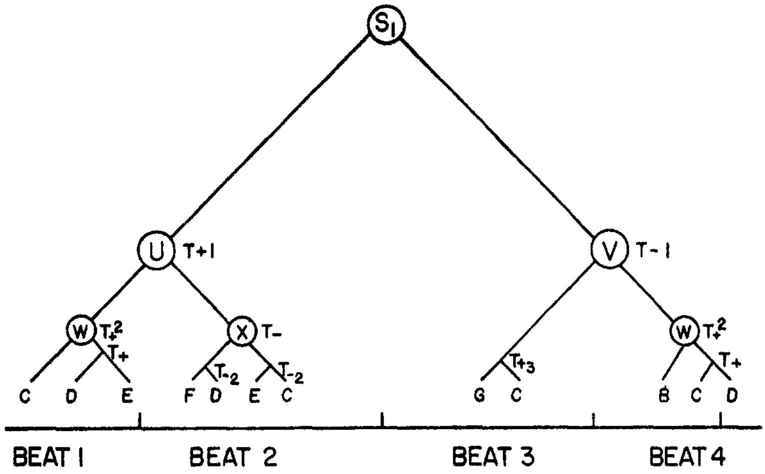

FIG. 6. Tree diagram of the first measure of Bach's Two-part Invention No. 1.

cated, the attempt is made to find a reason for it, either as an association learned in a previous sequence or elsewhere in this sequence, as a generalization, as the overextension of a principle, or as the use of too simple a structural tree or the wrong tree.

From such studies it has become apparent that the above principles regarding subunits have wide applicability if one realizes that subunits may have subunits, that structural principles of many types can be overgeneralized, and that *S*s, before they learn a given sequence, may use cognitive structures different from those intended by the experimenter.

Since these experimental applications must be discussed in the detailed reports of the experiments, it is useful to look beyond the laboratory for other applications. A recent theoretical paper on music recognition (Deutsch, 1969) notes that a specific melody is described as a list of intervals, namely, (*a*) a descending second, (*b*) a descending fourth, (*c*) an ascending fourth, and (*d*) an ascending second, and then the sequence is abstracted so as to be transposable. This idea is then used as the basis for a neurophysiological model. Using the ideas of serial pattern learning, Deutsch's account of a melody can be extended to provide a basis for subunits and higher order structure.

Simon and Sumner (1968) have applied to music the concepts of the information-processing approach to letter-series completion (Simon & Kotovsky, 1963). Their approach is to characterize a program which will extract patterns from musical scores, using a variety of notions from music theory. They introduce a considerable richness of concepts at will, and then reconstruct the conventional analysis of music by a strictly formal procedure.

The present theory of structural trees can be applied to the analysis of musical melodies, and the result is surprisingly informative. The first Two-part Invention of J. S. Bach has been analyzed, using a very limited and strict interpretation of the theory of structural trees.

The first measure, right hand, consists of two parts, half made up of sixteenth notes (called U in Figure 6) and half of eighth notes (called V). Melody U divides into two parts; X, a run (C–D–E) leading into Y, two descending thirds (F–D, E–C). X and Y are two contrasting subparts. Each makes the beginning of Y be the natural overextension of X, perhaps so that overextension errors can be avoided.

The second half of the measure, V, extends to the second measure. It consists of G, the harmonically important dominant with C, the tonic, forming a short

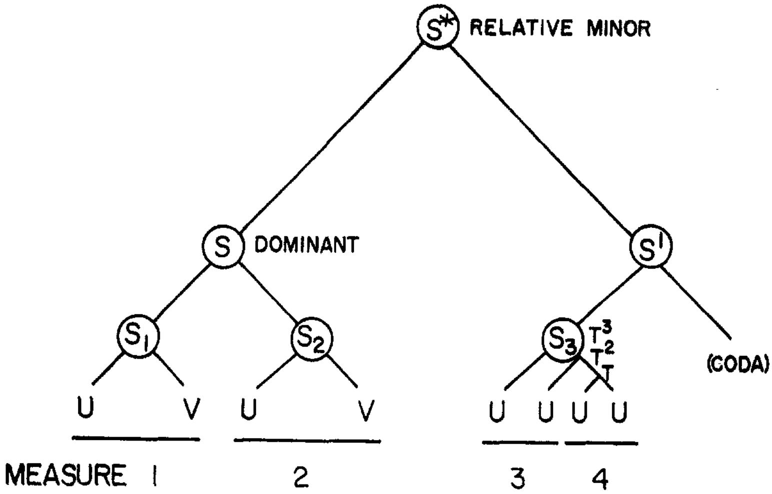

FIG. 7. Tree diagram of the first four measures of Bach's Two-part Invention No. 1.

harmonic cadence. This is followed by the run B–C–D, which is a transposition of the first submelody, W. The above analysis is somewhat ambiguous, and is chosen partly on the assumption that the decorative mordent is attached to note B, and may help divide the phrase V into its two subparts.

The entire melody S_1 is repeated in the right hand but transposed to the key of G (a harmonically simple transposition, if not one easily described in these purely melodic terms). Then Submelody U is inverted and transposed four times downward. This passage takes most of Measures 3 and 4 and leads into Measure 5. Thus, the first four measures of the first voice can be described by the tree shown in Figure 7, in which the music is also shown.

The above analysis deals with the right hand or first voice. The second voice, having echoed Submelody U, then begins a sequence of ascending runs (each represented by a right-branching tree), with the runs themselves being assembled into another right-branching tree. These ascending runs may be thought to be the Subunit W of U, seen in Figure 6.

The full development of even this elementary piece is far too complicated for complete exposition here, and the system proposed is, by a wide margin, too weak in musical resources for the full job. The relationship between the two voices cannot be expressed, rhythm is not emphasized (except to the degree that notes within a submelody are played close together, and pauses between submelodies are longer), and the various key relationships are not clearly stated.

Actually, the above analysis resembles conventional theoretical analysis of a musical piece. Its inclusion in this paper is justified not by its originality, nor its value to musicians, but because it arises from nonmusical studies of serial pattern learning yet applies to highly patterned music.

It is possible to study the learning of a simple musical piece from the point of view of serial pattern learning. When the tree structure of the piece is known, then the various difficult locations (at the highest nodes of the tree) and the results of oversimplification of the tree can be predicted. If such predictions agree with the errors students are observed to make when learning the piece, then some validation of the method within music and music education would be attained. On this question the writer can report his own memories of struggles with such pieces, and informal observations of musical practicing within the household. It appears that certain errors are much more probable than others, and that such errors can be derived from a structural analysis, though no detailed theoretical or empirical investigations have been carried out.

REFERENCES

CHOMSKY, N. *Aspects of the theory of syntax.* Cambridge, Mass.: M.I.T. Press, 1965.

DEUTSCH, D. Music recognition. *Psychological Review,* 1969, **76**, 300–307.

LASHLEY, K. S. The problem of serial order in behavior. In L. A. Jeffres (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951.

LEEUWENBERG, E. L. L. Quantitative specification of information in sequential patterns. *Psychological Review,* 1969, **76**, 216–220.

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

RESTLE, F. Grammatical analysis of the prediction of binary events. *Journal of Verbal Learning and Verbal Behavior,* 1967, 6, 17–25.

RESTLE, F., & Brown, E. R. Serial pattern learning. *Journal of Experimental Psychology,* 1970, **83**, 120–125. (a)

RESTLE, F., & BROWN, E. R. Serial pattern learning: Pretraining of runs and trills. *Psychonomic Science,* 1970, in press. (b)

SIMON, H. A., & KOTOVSKY, K. Human acquisition of concepts for sequential patterns. *Psychological Review,* 1963, **70**, 534–546.

SIMON, H. A., & SUMNER, R. K. Pattern in music. In B. Kleinmuntz (Ed.), *Formal representation of human judgment.* New York: Wiley, 1968.

(Received January 12, 1970)