Free Recall of Self-Embedded English Sentences*

GEORGE A. MILLER AND STEPHEN ISARD†

Harvard University, Cambridge, Massachusetts

Subjects were given five trials to memorize 22-word sentences that varied in degree of self-embedding. Sentences with self-embedded phrase structure proved more difficult to learn, which is interpreted to mean that our capacity to deal with recursive interruptions may be extremely limited.

The problem posed for our subjects in the present experiment will be familiar to anyone who has tried to discuss algebra over the telephone. It is not too difficult to indicate by appropriate emphasis and intonation that "A plus B times C" refers to $A + BC$, rather than to $(A + B)C$, but when the expression becomes only a little more complicated both talker and listener begin to flounder rather badly. Spoken communication is not the best medium for indicating parenthetical constructions.

It is interesting, therefore, that in spite of our inability to cope with complicated parenthetical constructions in vocal forms, all natural languages, including English, make provision for just such constructions in the sentences we speak. For example, the sentence, The man who said that a cat killed the rat is a liar, is perfectly grammatical and has one sentence (a cat killed the rat) nested inside of another (the man who said that is a liar). But now carry the process another step and put the dog chased the cat inside a cat killed the rat, in the form of a relative clause: a cat that the dog chased killed the rat. When all three are put together

* This research was supported in part by a grant from the National Science Foundation No. NSF-G-16486 and in part by a PHS Research Grant No. MH 05120-02 from the National Institutes of Health, United States Public Health Service, to Harvard University, Center for Cognitive Studies. The preparation of this report was supported by the Department of Defense, Advanced Research Projects Agency, Contract SD-187 with Harvard University, Center for Cognitive Studies. We are indebted to Kathryn Ojemann McKean and to Dr. Lee E. McMahon for assistance in preliminary experiments.

† Present address: Department of Mathematics, University of California, Berkeley, California.
RECALL OF EMBEDDED SENTENCES

into a single sentence, The man who said that a cat that the dog chased killed the rat is a liar, the result begins to be a bit confusing. Add another relative clause for the sentence, the boy owns the dog, and we get a really difficult, but still perfectly grammatical sentence: The man who said that a cat that the dog that the boy owns chased killed the rat is a liar. Or we can work in the other direction, and wrap another sentence around that one: It is more likely that the man who said that a cat that the dog that the boy owns chased killed the rat is a liar than not. Unless special and rather arbitrary rules are introduced to prevent it, this sort of grammatical onion could grow indefinitely.

Obviously, people do not talk this way. There are in English alternative constructions that enable us to say all this in a much simpler way: It is more likely than not that the man is a liar who said that the rat was killed by a cat that was chased by the dog that is owned by the boy. Since both are equally acceptable according to the rules of grammar, any preference for the latter must have some psychological, rather than linguistic, explanation. This fact seems to have been clearly stated first by Yngve (1960), although he used it as the basis for certain generalizations about linguistic structure and evolution that we would not endorse (Miller and Chomsky, 1963). Yngve points out, quite correctly, that the discontinuous constituents of the nested sentence impose a severe load on our short-term memory, whereas the alternative form does not. In order to deal with nested constructions, the language user must hold in memory the still unresolved portion of one constituent while he is processing another. When two or three initial portions must be remembered, all in proper order, the task becomes quite difficult. Nested constructions, therefore, pose a problem of some psychological interest.

Nesting is also of considerable interest for the general theory of grammar. Two types are generally distinguished: nesting refers to the insertion of one construction into the middle of another; when the inserted construction is of the same grammatical form as the construction into which it is inserted, it is called self-embedding. (These definitions can easily be made precise; for a general review of this and related matters, see (Chomsky and Miller, 1963).) The admissibility in natural languages of an indefinite number of self-embedded elements was used by Chomsky (1956) to demonstrate that a formal characterization of natural grammars in terms of the finite-state, Markovian models used by information theorists is impossible in principle; later (Chomsky, 1959) he was able to prove that the additional power of the phrase-
structure grammars that linguists had been using was attributable entirely to their capacity for generating self-embedded constructions.

The fact that an indefinite number of self-embeddings is grammatically acceptable, yet at the same time psychologically unacceptable, would seem to imply that a clear distinction is necessary between our theory of the language and our theory of the language user. This simple conclusion is resisted, however, by many workers. Some linguists resist it because they believe that the description of a language would be useless if it did not describe the verbal behavior of people who use the language. And some psychologists resist it because they believe that the description of behavior exhausts the entire range of topics that a scientist is licensed to study. Without entering into this somewhat confusing argument, let us say simply that the distinction between knowing a rule and obeying it seems to us both valuable and necessary, and on this basis it is possible to find a reasonable division of labor between linguists and psychologists, one that does not annex to either the territory that rightfully belongs to the other.

The present study, therefore, should be viewed as a contribution to our psychological understanding of the language user, and not as an attempt to revise or limit our statement of the rules of grammar in any way.

MATERIALS AND PROCEDURE

Sentences with varying degrees of self-embedding were recorded and played to listeners who attempted to learn them one at a time by the method of free recall.

Six sentences, all 22 words long, were used. In their unembedded forms, these sentences were: They saw the old lady that kicked the dog that followed the boy that delivered the newspaper that told about the fire; Jack kissed the girl that met the man that wrote the book that told the story that was about a nuclear war; We cheered the football squad that played the team that brought the mascot that chased the girls that were in the park; She liked the man that visited the jeweler that made the ring that won the prize that was given at the fair; He teased the little boy that joined the club that stole the green apples that made the pie that tasted very good; and She thanked the producer that discovered the plot that became the script that made the movie that was applauded by the critics. All relative clauses were introduced by that, in order to avoid stylistic variations among that, who, and which; in preliminary experiments it had been found that these
variations were quite difficult to remember, and since they tended to obscure the main effects of self-embedding, they were simply eliminated.

Each subject memorized all six sentences, but each in a different syntactic form. He learned one sentence presented with no self-embedding, another with one degree of self-embedding, another with two, three, or four, and one sentence with the words scrambled in haphazard order. The different degrees of self-embedding were all obtained in the manner illustrated for the following sentences: (0) She liked the man that visited the jeweler that made the ring that won the prize that was given at the fair; (1) The man that she liked visited the jeweler that made the ring that won the prize that was given at the fair; (2) The jeweler that the man that she liked visited made the ring that won the prize that was given at the fair; (3) The ring that the jeweler that the man that she liked visited made won the prize that was given at the fair; (4) The prize that the ring that the jeweler that the man that she liked visited made won was given at the fair; and (random) Won given liked that that the fair man made visited prize the at the the she that jeweler was the ring that. Six sentences, each presented in these six forms, gave 36 items in all. Note that exactly the
The 36 items were read aloud by one of the authors (GAM) and recorded on magnetic tape. This recording was then duplicated and spliced to produce the test materials actually used in the experiment. Considerable pains were taken to obtain satisfactory renditions of each sentence, since in exploratory experiments with unrecorded presenta-
tions it had been impossible to decide whether the learners' poor performance on the self-embedded strings should be attributed to their inability to understand and remember such constructions, or to the speaker's inability to read them properly, or both. Practice and re-recording were continued, therefore, until in the judgment of the talker a spoken version of each sentence was recorded that contained all of the auditory clues to grouping that he could express vocally. The random strings of words were read with simulated sentence intonation, as similar to the sentences as possible.

In the learning situation, the recorded sentences were played over earphones to the subject. As soon as the sentence finished he attempted to repeat it verbatim. His repetition was recorded on an office dictation machine and scored later. A single sentence was presented and repeated five times in this way. Then another sentence was presented five times, etc., until the subject had memorized all six sentences, each in a different syntactic form. Six different orders of presentation of the sentences were prepared, in counterbalanced design, and four subjects learned them in each order. Before the test began, however, each subject memorized one practice sentence with two self-embeddings in order to eliminate the "warm-up" effect and to ensure that he understood the task and the use of the dictation equipment; data obtained from this practice sentence were discarded.

RESULTS

The responses by the subjects were scored in terms of the number of words recalled in the same order as the words in the original sentence. (The total number of words correct, regardless of order, was also examined; the scores were slightly higher, particularly for the random strings, but the relative differences as a function of embedding were not affected.) Omissions were counted as errors; novel words added by the subject during recall were simply ignored.

The averaged results for the 24 subjects who served in this study are displayed graphically in Fig. 3; since there were 22 words per sentence, each point is based on a sample of 528 observations.

The first thing to note is that learning occurred; for all types of sentences the average percentage of the words recalled and in correct order increased on successive repetitions. The second point to note is that all of the grammatical sentences were much easier than the haphazard strings of words; obviously, the presence of grammatical structure, even
in the self-embedded form, was a considerable aid and permitted the subjects to transfer their normal linguistic skills to this learning situation. Finally, and of most interest, self-embedding made the sentences more difficult to learn. With one exception, on every trial the number of errors increased as the degree of self-embedding increased. The exception occurred on trials four and five, where there is a minor reversal of the general trend in the case of three and four degrees of self-embedding.

An alternative way to analyze the data is to consider the trial on which a subject achieved his first perfect recall of each type of sentence. These distributions are summarized for the 24 subjects in Table I for zero through four self-embeddings; no subjects achieved a perfect recall of the scrambled strings within the limit set by the five trials allowed. From this table we see, for example, that two of the 24 subjects repeated the unembedded sentence perfectly the first time they heard it, that seven more got it on the second trial, etc. This way of presenting the data makes the differences in difficulty appear somewhat more dramatic than do the learning curves of Fig. 1.

**Fig. 3.** The percentage of words recalled correctly and in the correct order is plotted as a function of the number of trials given, with the syntactic complexity of the sentences as the parameter. Points are averages for 24 subjects.
RECALL OF EMBEDDED SENTENCES

TABLE I

<table>
<thead>
<tr>
<th>Trial</th>
<th>Number of embeddings</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>15</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>16</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>18</td>
<td>18</td>
<td>10</td>
</tr>
</tbody>
</table>

DISCUSSION

If we take the difference between the unembedded sentences and the random strings of words as a measure of the maximum effect on performance in this situation that can be produced by rearranging the order of the words, then it is obvious that the shift in performance attributable to self-embedding is only a fraction of the maximum possible shift. If we consider only the average performances on the first trial (since the unembedded sentences are so well recalled on subsequent trials that the comparison is difficult), 16.4 words were recalled in the correct order for unembedded sentences, 12.7 for sentences with four self-embeddings, and only 4.4 for the random strings of words.

Although precise calculations are probably unwarranted, a rough estimate based on the following argument serves to indicate that this is approximately what might have been expected. According to the argument (Miller, 1956a, b), these sentences differ in difficulty because, from a psychological point of view, they differ in length.1 If we think of the unembedded sentences as five psychological units (five simple sentences)

1Behavioral evidence in support of this hypothesis can be obtained by observing eye movements. Drs. Norman Mackworth and J. S. Bruner (personal communication) have recorded the eye movements of subjects reading these sentences and have found that the number of fixations and the number of regressive movements of the eyes both increase markedly as the amount of self-embedding increases; if the number of fixations can be used as an index of the number of psychological units that a reader must assimilate, then self-embedding clearly operates to increase the effective number of units per sentence. It should be noted, however, that some readers make a first pass through the entire sentence before they return to fit the pieces together—a multiple-pass strategy that was not available to the listeners in the present experiment.
long, then the average unit is about $22/5 = 4.4$ words long. Four of those units are broken in two by the four degrees of embedding, giving the sentences with four embeddings an effective psychological length of nine units, with the average unit about $22/9 = 2.4$ words long. The random strings would be 22 units long, where each unit is one word long. To obtain the number of these hypothetical units that subjects could retain in immediate memory, therefore, we simply divide the total number of words correctly recalled by the number of words per unit. The result is 3.7 units for unembedded text, 5.2 units for sentences with four self-embeddings, and 4.4 units for the random strings. These are reasonably equal, although the figures for four self-embeddings is too high, which may mean that the subjects could, to a certain extent, organize the discontinuous constituents as single units, i.e., that there were fewer than nine units involved.

Although this argument may have some general validity, it is not adequate to explain the finer details of the experimental results. In particular, it does not account for the fact, obvious on inspection of Fig. 2, that the performance curves for zero and one degree of self-embedding are quite close, as are the curves for three and four embeddings, with performance on the doubly embedded sentences falling somewhere in between. In order to explain this spacing, which was also characteristic of the results obtained in preliminary experiments, a more explicit argument is required.

Suppose that for the moment we think in terms of an analogy with information processing by computer systems. Let us imagine that anyone who knows English has something corresponding to a relative clause subroutine that can be called to process sentences containing such structures. When this subroutine is called, the main sentence-analyzing routine is interrupted and the point at which it must be resumed is stored temporarily until the subroutine has been executed. As long as the sentence contains only one such construction, little difficulty will result. But now suppose that, while the subroutine is being executed, a second such construction is encountered, so the subroutine is required to call itself. If this recursive feature were not available, confusion would result; the temporary memory for the point of re-entry into the main routine might be erased, for example, so that when it resumed, the main routine would have to treat subsequent words as if they began a new constituent of the sentence. This analogy can help us to understand why perform-
ance with zero and one embedding were approximately equivalent, but two, three, and four embeddings were more difficult.

Several related observations can be cited to lend credibility to this analogy. In several nonsystematic observations it has been noted that nested constructions, which would involve different types of subroutines, are less difficult to understand and remember than are self-embedded constructions of the same degree. That is, if, in the middle of executing the relative clause routine, some other type of subroutine is called, there does not seem to be as much interference in remembering the point at which the main program must be resumed after the relative clause subroutine is completed. The analogy in terms of grammatical subroutines seems to offer a possible way to explain this difference between nesting and embedding (which, of course, should be validated in further experiments). We need assume merely that entering a different subroutine does not erase the temporary storage of the point of re-entry on completion of the original subroutine.

Moreover, Mackworth and Bruner's recordings of eye movements while reading such sentences confirms the introspective impression that the difficulty does not begin until the long string of apparently unrelated verbs is encountered toward the end of the self-embedded sentence. At this point the recursive eye movements begin and one feels that all grasp of the sentence structure has suddenly crumbled away. Difficulty at this point is, of course, exactly what the subroutine analogy would lead us to expect, since not until then would the loss of re-entry addresses have any serious effect on the processing of the sentence.

As Table I indicates, however, some of the subjects handled two self-embeddings as well as they handled one, which suggests that the subroutine analogy needs to be modified slightly to admit the possibility that some subjects can handle two self-embeddings. Examination of the data for individual subjects seems to support this suggestion.

Of the 24 subjects, 12 made as few errors (summed over all 5 trials) in memorizing a sentence with one relative clause as they did with a sentence having none; these two types of sentence are, for most people, about equally easy. If we take as a subject's optimal performance on sentences with simple syntax the better of his two scores on sentences with zero and one embedding, we find that six subjects actually exceeded this score (made fewer total errors) in learning a sentence with two embeddings. However, no one exceeded this optimal performance for sen-
tences with three or four self-embeddings. In short, most subjects could handle one relative clause, and some managed two, but no one learned sentences with three or four self-embedded relative clauses without encountering difficulty. Apparently, some subjects were able to retain the point at which the main program was to be resumed and simultaneously to retain the point at which the subroutine was to be resumed, even when the second subroutine called was the same as the first. Retaining three or four such points of resumption, however, seems to have been impossible.

It is difficult to do more than speculate about these matters, for we are here faced by a type of cognitive task—the handling of interruptions—about which very little is known. Drs. Allen Newell and H. A. Simon (personal communication) have pointed out that our ability to resume interrupted tasks at the precise point we left them off is a matter of considerable importance for the simulation of cognitive processes by computers, and that a series of studies aimed at this type of temporary memory would be particularly valuable and not too difficult to design.

In any case, if we are willing to regard our ability to process self-embedded sentences as a measure of our ability to deal with recursive interruptions in general, then the present experiment demonstrates that, in the absence of environmental reminders and memory aids, our capacities in this direction are very limited indeed. Exactly how significant this fact may be for our information processing theories of cognition in general, and for our theory of the language user in particular, cannot be determined, however, until we understand more clearly what the limit is and what factors affect it.

**Received:** September 5, 1963

**References**


