VERBAL CONTEXT AND THE RECALL OF MEANINGFUL MATERIAL

By GEORGE A. MILLER and JENNIFER A. SELFRIDGE, Harvard University

Communicative behavior, perhaps more than any of man's other activities, depends upon patterning for its significance and usefulness. An accidental inversion of words or letters or sounds can produce grotesque alterations of a sentence, and to scramble the elements at random is to turn a sensible message into gibberish. No attack upon the problems of verbal behavior will be satisfactory if it does not take quantitative account of the patterns of verbal elements.

We can dependably produce and distinguish only a small number of different letters or speech sounds. We must use these few elements to talk about millions of different things and situations. To stretch these few elements to cover these many needs, we are forced to combine the elements into patterns and to assign a different significance to each pattern. Since the number of possible patterns increases exponentially as the length of the pattern increases, this proves to be an efficient method of solving the problem.

Not all the possible patterns of elements are used in any particular language. In English, for example, the sequence of letters *qke* does not occur. It is reasonable to ask, therefore, why we do not exploit the available patterns more effectively. Is it not inefficient to ignore some patterns while others are greatly overworked?

The preference for some patterns at the expense of others forces us to produce more elements—letters, sounds, words, etc.—in order to make the same number of distinctions that we could make with the same elements if we used all possible patterns. To illustrate: imagine a language with 10 elementary symbols that is used to refer to 100 different things, events or situations. If we used all possible pairs of elements, we could refer to every one of the 100 things with one of the 100 pairs of 10 symbols. If, however, we refuse to use some of the pairs and so rule them out of the language, it is necessary to make up the difference by using triads. Thus the language uses patterns of three elements to make distinctions that could be made with patterns of two elements.

On further consideration, however, this kind of inefficiency does not appear a complete waste of time. By favoring some patterns rather than others the language is protected against error.[1] More specifically, in Engish we recognize immediately

* Accepted for publication August 30, 1949.

[1] C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.*, 27, 1948, 379-423, 623-656.

that an error has occurred if we read in our newspaper, "Man bites dxg." The pattern *dxg* is not admitted in English, and so we catch the error. If, however, all patterns of elements were admissable, *dxg* would have some semantic rule and we would not be able to catch the mistake. The number system is an example of the efficient use of ten symbols, but it is highly susceptible to mechanical errors. If a man says that his telephone number is 9236 we have no way of recognizing that he has or has not made an error.

Patterns are unavoidable, and a preference for some patterns provides insurance against errors. Thus it seems reasonable that the statistical studies of different languages all show that some patterns of elements are greatly overworked while others occur rarely or not at all.[2] The present interest in verbal patterning, however, is in the light these observations can throw upon the psychological problem of verbal context.

Verbal context. Psychologists use the word context to refer to the totality of conditions influencing a behavioral event. For the present discussion we want to restrict this broad definition and to consider only the antecedent verbal conditions. When a man talks, his choice of words depends upon his training, his needs and intentions, the situation and audience. These factors comprise the total context in which his words must be studied. By verbal context, as opposed to total context, we mean only the extent to which the prior occurrence of certain verbal elements influences the talker's present choice. If the talker has said "children like to," his choice for the next word in this pattern is considerably limited—*elephant, punished, loud, Bill,* and many other words are highly unlikely continuations.

By verbal context, therefore, we mean the extent to which the choice of a particular word depends upon the words that precede it. In the statistical sense, this definition of verbal context is given in terms of dependent probabilities.[3] The probability that event *C* will occur is not the same after *A* as it is after *B*. The statistical dependencies between successive units form the basis for a study of verbal context.

To illustrate the operation of conditional probabilities in our verbal behavior, consider the set of all possible sequences 10 letters long. We could construct a table listing them. The first row of the table would be the pattern *aaaaaaaaaa,* 10 consecutive *a*'s. The second would be *aaaaaaaaab,* then *aaaaaaaaba, aaaaaaaabb, aaaaaaabaa,* etc., until all possible arrangements of letters, spaces, commas, periods, hyphens, quotes, colons, numbers, etc., were exhausted. Altogether there would be about 50 different symbols, and the table would contain 50^{10}, or about 100,000,000,000,000,000 different patterns. Then we would examine some English writing and try to determine the relative frequencies of occurrence of the patterns. Only a small fraction of the 50^{10} alternatives actually occur in English. The table would show

[2] G. K. Zipf, *Human Behavior and the Principle of Least Effort,* 1949, 1-343.

[3] G. A. Miller and F. C. Frick, Statistical behavioristics and sequences of responses, *Psychol. Rev.,* 56, 1949, 311-324.

strong dependencies. For example, the letter *q* is always followed in English by the letter *u*, and so all those entries in the table that contained a *q* followed by anything but *u* would not occur in English. It is not possible to predict the relative frequency of *qe*, for instance, by multiplying the relative frequencies of *q* and of *e*.

If such a table existed, along with the relative frequencies of occurrence, it would be possible to construct sequences of letters that reflected the statistical dependencies of English verbal patterns. We can imagine similar tables constructed for shorter or longer sequences of letters. A table for all patterns of 2 symbols would represent the relative frequencies of pairs; for 3 symbols, triads, etc. The longer the sequence, the more information the table contains about the pattern of dependencies in our molar verbalizations.

To illustrate the use of such information we shall borrow a device used by Shannon. Suppose we have no knowledge at all of the relative frequencies of occurrence, but only a list of the 50 different symbols. Then, for all we know, any sequence of symbols might be permissible. If we tried to construct a message in the language, the best we could do would be to draw at random from the 50 different symbols. We have no reason to think that one sequence of symbols is more likely than another. Proceeding in ignorance to construct a message, we might produce something like this: *cplp'rzw(p".:k!)"ntegznqO?i6vlaur :8h,* etc.

Now suppose that we have a reliable tabulation of the relative frequencies of 'patterns' of one symbol. We know, therefore, that *e* and the space between words are more likely to occur than are *?* and *z*. With this information we can increase the chance of constructing a meaningful message, although our chances are still very small. If we draw successive symbols according to their relative frequencies of occurrence in English, we might produce something like this: *wli hnrooye lricocnri mae c zg 2eaya,* etc.

The next step is to imagine that we have a tabulation of relative frequencies of occurrence of patterns of two symbols. Now it is possible to improve the statistical approximation to English by drawing in the following way. Begin by drawing any likely pair. Suppose the pair is *au*. Now look at all the pairs starting with *u* and draw from them according to their relative frequencies of occurrence. Suppose the result is *ud*. Now look at all pairs starting with *d* and draw one of those, and so proceed to build up the message. Notice that each draw depends upon the preceding draw—the preceding draw determines from which set the present draw is to be made. Drawing in this way reflects the conditional probabilities of successive symbols. A message constructed in this way might read *aud ren stiofivo omerk. thed thes bllale,* etc.

If we have a tabulation of sequences of three letters, we can construct a message that reflects the conditional probabilities of English triads. First we draw a likely triplet, say *ann,* then draw next from the triplets starting with *nn* and obtain *nna,* then from the triplets beginning *na,* etc. The preceding two symbols determine from which set the next triplet is drawn. In this way a message might be produced that would read: *annation ef to the acticas. Oth rested,* etc.

With a tabulation of sequences of four letters we might produce: *influst intradio be decay, the condive,* etc. By tabulating the relative frequencies of longer sequences and drawing successive items so as to reflect these frequencies, we can construct messages that reflect the statistical dependencies of English as extensively as we please.

For convenience, we shall refer to these different ways of constructing a statistical English as orders of approximation, and shall number them from *0* to *n*. At the zero order we have no knowledge of relative frequencies, at the first order we know the relative frequencies of individual symbols, at the second order we know the relative frequencies of pairs, at the *n*th order we know the relative frequencies of *n*.

Consider this statistical English now in terms of verbal context. With a zero-order approximation to English there are no contextual influences whatsoever on the choice of successive symbols. At the *n*th-order approximation, however, each symbol is selected in the context of the preceding *n*-1 symbols. As the order of approximation is increased, the amount of context for each symbol is increased, and the contextual constraints (dependent probabilities) have a chance to operate. As the order of approximation is increased, the messages we can construct become more and more familiar, reasonable, meaningful. The more we permit contextual restraints to operate, the better are our chances of producing a message that might actually occur in English.

We have, therefore, a scale for what can be loosely called 'meaningfulness.' At one end are the random jumbles of symbols we customarily call nonsense, and at the other end are patterns of symbols that could easily appear in our daily discourse. Equipped with this quantitative estimate of 'the degree of nonsense' or 'amount of contextual constraint,' we can proceed to study certain psychological problems that have been phrased in terms of meaningfulness.

An experimental illustration. Briefly stated, the problem to which this concept of verbal context has been applied is, How well can people remember sequences of symbols that have various degrees of contextual constraint in their composition? The experimental literature contains considerable evidence to support the reasonable belief that nonsense is harder to remember than sense. This evidence has suffered, however, from a necessarily subjective interpretation of what was sensible.

In the present experiment, the learning materials were constructed at several orders of approximation to English. These materials were presented to *S*s whose recall scores were then plotted as a function of the order of approximation.[4]

Learning materials. In the preceding examples we have used patterns of letters to illustrate the effects of contextual constraints. There is, of course, no necessity to limit the argument to letters. It is possible to use words or even sentences as the component elements that are arranged according to the statistical structure of English. In the present experimental illustration the materials were constructed with words as the units of analysis.

In theory, the construction of materials to incorporate the statistical structure

[4] The experiment was carried out by the junior author and is presented in detail in her honor's thesis, *Investigations into the Structure of Verbal Context,* 1949. The thesis is on file in the Library of the Psychological Laboratory, Memorial Hall, Harvard University.

of English over sequences of several words requires a tabulation of the relative frequencies of such sequences. Such a tabulation would be exceedingly long and tedious to compile. An alternative method of construction is available, however, which makes the procedure practicable. Instead of drawing each successive word from a different statistical distribution indicated by the preceding words, we draw the word from a different person who has seen the preceding words.

At the second order, for example, a common word, such as *he, it* or *the,* is presented to a person who is instructed to use the word in a sentence. The word he uses directly after the one given him is then noted and later presented to another person who has not heard the sentence given by the first person, and he, in turn, is asked to use that word in a sentence. The word he uses directly after the one given him is then noted and later given to yet another person. This procedure is repeated until the total sequence of words is of the desired length. Each successive pair of words could go together in a sentence. Each word is determined in the context of only one preceding word.

For higher orders of approximation the person would see a sequence of words and would use the sequence in a sentence. Then the word he used directly after the sequence would be added, the first word of the sequence would be dropped, and the new (but overlapping) sequence would be presented to the next person. By this procedure we constructed sequences of words at the second, third, fourth, fifth and seventh orders of approximation.

For the first order approximation to English a scrambling of the words in the higher orders was used. By drawing words at random from the contextually determined lists, we obtained as good an approximation to the relative frequencies of individual words in English as these higher order lists provided. The alternative method of selecting words at random from a newspaper might have given a sample quite different in difficulty (familiarity).

A zero order approximation to English could be obtained by drawing at random from a dictionary. Most dictionaries contain too many rare words, however, so we drew from the 30,000 commonest words listed by Thorndike and Lorge.[5] This source had the additional advantage that it listed separately all forms of the word, whereas the dictionary lists only the lexical units. Words drawn at random from this list of 30,000 words are selected independently and without any constraints due to adjacent words or the relative frequencies of appearance of the words in English.

A final set of words was taken directly from current fiction or biography. These lists represent a full contextual determination.

By these devices we constructed sequences of words with eight different degrees of contextual constraint. In the following discussion we shall refer to these lists as 0, 1, 2, 3, 4, 5, 7 and text-orders of approximation. At each order four lists of different length—10, 20, 30 and 50 words—were constructed. Thus the experimental design called for 32 different lists. Two such sets of 32 lists were constructed. Since the lists require considerable time to compile and since they may be some general interest, one of the sets of 32 is reproduced in full in the appendix to this paper.

Experimental procedure. Each set of 32 lists was read aloud and recorded on a wire recorder. A man's voice was used. The words were read slowly and distinctly

[5] E. L. Thorndike and I. Lorge, *The Teacher's Wordbook of 30,000 Words,* 1944.

in a near monotone, with a short pause between words. At the beginning of the recording the instructions were given and a single practice list was presented to make sure the *Ss* understood their task. They were to listen until a list was finished, at which time a bell sounded signalling them to begin writing what they had just heard. The *Ss* were instructed to write the words they remembered as nearly in their correct order as possible. Order was not used, however, as a criterion for scoring their responses. All eight of the 10-word lists were given first, proceeding from least to greatest contextual determination, then the 20-word lists in the same order, then the 30-word lists, and finally the 50-word lists. Short rest

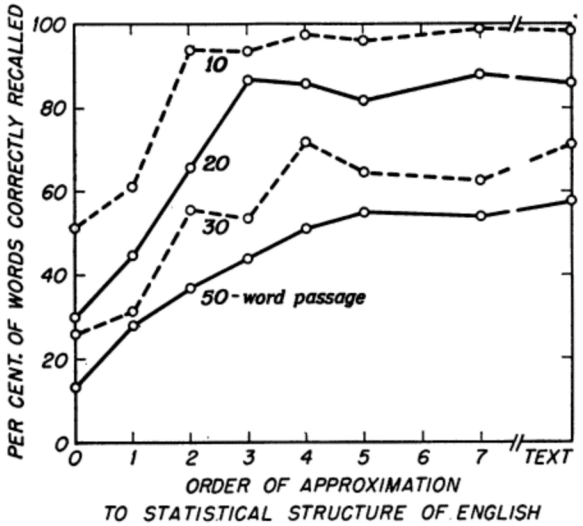

FIG. 1. Percentage of Words of the Lists of Different Lengths that were Correctly Recalled at the Various Orders of Approximation to the Statistical Structure of English.

periods (5 min.) were given between the 20- and 30-word lists and between the 30- and 50-word lists.

Two groups of 10 *Ss* were used. One group heard and recalled one of the sets of 32 lists, the other group heard and recalled the second set. The *Ss* were principally students at Harvard and Radcliffe. It was *E*'s impression that a larger number of *Ss* would not have reduced the irregularities in the results, for most of the variability seemed attributable to sampling peculiarities in the lists themselves. Several more sets of 32 lists would be needed before an accurate estimate of the functional relations could be made. The results are adequate, however, to indicate the approximate magnitudes and general trends of the functions.

Ss' answers were scored for the number of words that they had written that had occurred in the test material. The number recalled, regardless of order, was expressed as a percentage of the total number presented.

Results and discussion. The experimental data are summarized in Figs. 1 and 2. In Fig. 1 the recall-score, expressed as a percentage, is plotted as

a function of the order of approximation to the statistical structure of English, with the length of the lists as the parameter. In Fig. 2 the same data are replotted to show the relation of the recall-score to the length of the list, with the order of approximation as the parameter. In both figures the functions represent the mean scores for all 20 *S*s. It is clear from the results that percentage recalled increases as the order of approximation is increased and decreases as the length of the list is increased. Inspection of Fig. 1 leads to a reasonable suspicion that the two variables, length and

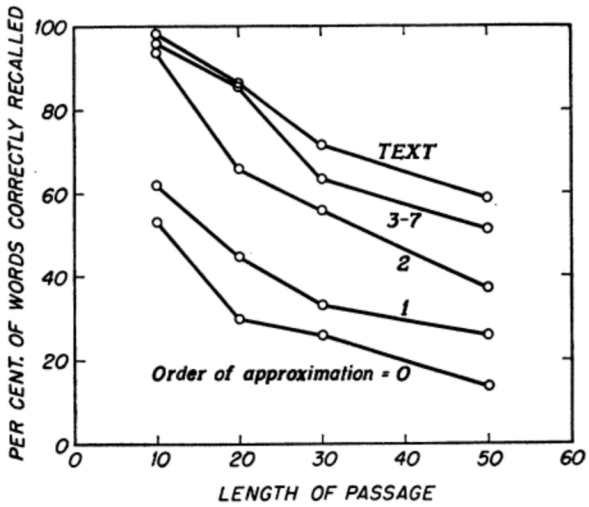

FIG. 2. Percentage of Words of the Various Orders of Approximation to the Statistical Structure of English that were Correctly Recalled at the Different Lengths of Passages Learned.

order of approximation, interact. With the short, 10-word lists there is little to be gained from contextual bonds extending over more than two words. With the 20-word lists the *S*s remembered as well at the third order of approximation as they did for the textual material. With the 50-word lists, however, only orders 5 and 7 are comparable to the textual material in terms of percentage recalled. It would seem, therefore, that the longer the passage the greater is the usefulness of contextual associations extending over long sequences of items.

By a strict interpretation of the word 'nonsense,' one is forced to conclude that all orders of approximation less than the full text are nonsense. Consider an example from Order 5:

house to ask for is to earn our living by working towards a goal for his team in

old New-York was a wonderful place wasn't it even pleasant to talk about and laugh hard when he tells lies he should not tell me the reason why you are is evident.

The experimental results show that this kind of gibberish is as easily recalled as a passage lifted from a novel. Thus there are kinds of nonsense that are as easy to recall as are meaningful passages. The significant distinction is not to be drawn between meaning and nonsense, but between materials that utilize previous learning and permit positive transfer and materials that do not. If the nonsense preserves the short range associations of the English language that are so familiar to us, the nonsense is easy to learn.

The experiment shows, therefore, that the problem of meaning vs. nonsense in verbal learning need not be approached in terms of a qualitative dichotomy, but can be studied as a functional relation between quantitative variables. The results indicate that meaningful material is easy to learn, not because it is meaningful *per se,* but because it preserves the short range associations that are familiar to the *S*s. Nonsense materials that retain these short range associations are also easy to learn. By shifting the problem from 'meaning' to 'degree of contextual constraint' the whole area is reopened to experimental investigation.

Psychologists familiar with the problems of verbal learning will recognize the usefulness of the kind of material employed in this illustrative experiment. For example, is retroactive inhibition affected by interpolating different orders of approximation to English between the original learning and the recall? What is the effect of using original and interpolated materials of the same or of different orders of approximation to English? Do the higher approximations to English show the same differences between recall after sleep and recall after waking activity that the lower approximations show? Is it possible to show a continuum from the short-term reminiscence that can be demonstrated with syllables to the long-term reminiscence that can be shown with poetry? How does the span of immediate memory vary with the order of approximation? Is the superiority of distributed over massed practice a function of the order of approximation of the materials to the statistical structure of English? Can differences in learning and recalling different orders of approximation be demonstrated as a function of age?

The operational analysis of meaningfulness makes it possible to ask such questions and to see how one would proceed to answer them. The problem now is to collect the experimental data.

Summary

A quantitative definition for verbal context is given in terms of dependent probabilities. The definition is used to construct lists of words with varying degrees of contextual determination. When short range contextual dependencies are preserved in nonsense material, the nonsense is as readily recalled as is meaningful material. From this result it is argued that contextual dependencies extending over five or six words permit positive transfer, and that it is these familiar dependencies, rather than the meaning *per se,* that facilitate learning.

APPENDIX

Lists Used in Recall Experiment

0-order approximation

10: byway consequence handsomely financier bent flux cavalry swiftness weather-beaten extent

20: betwixt trumpeter pebbly complication vigorous tipple careen obscure attractive consequence expedition pane unpunished prominence chest sweetly basin awoke photographer ungrateful

30: crane therewith egg journey applied crept burnish pound precipice king eat sinister descend cab Idaho baron alcohol inequality Illinois benefactor forget lethargy fluted watchtower attendance obeisance cordiality dip prolong bedraggle

50: hammer neatly unearned ill-treat earldom turkey that valve outpost broaden isolation solemnity lurk far-sighted Britain latitude task pub excessively chafe competence doubtless tether backward query exponent prose resourcefulness intermittently auburn Hawaii unhabit topsail nestle raisin liner communist Canada debauchery engulf appraise mirage loop referendum dowager absolutely towering aqueous lunatic problem

1-order approximation

10: abilities with that beside I for waltz you the sewing

20: tea realizing most so the together home and for were wanted to concert I posted he her it the walked

30: house reins women brought screaming especially much was said cake love that school to a they in is the home think with are his before want square of the wants

50: especially is eat objections are covering seemed the family I that substance dinner raining into black the see for will passionately and so I after is window to down hold to boy appearance think with again room the beat go in there beside some is was after women dinner chorus

2-order approximation

10: was he went to the newspaper is in deep and

20: sun was nice dormitory is I like chocolate cake but I think that book is he wants to school there

30: the book was going home life is on the wall of you are ready to the waltz is I know much ado about it was a dog when it was

50: you come through my appetite is that game since he lives in school is jumping and wanted help call him well and substance was a piano is a mistake on this is warm glow in and girl went to write four turtledoves in my book is fine appearance of the

3-order approximation

10: tall and thin boy is a biped is the beat
20: family was large dark animal came roaring down the middle of my friends love books passionately every kiss is fine
30: happened to see Europe again is that trip to the end is coming here tomorrow after the packages arrived yesterday brought good cheer at Christmas it is raining outside as
50: came from the beginning and end this here is the top spins in a house by the library is full of happiness and love is very nice of her that fell from the window she went home from work to pass the cigarettes down to earth he picked an apple

4-order approximation

10: saw the football game will end at midnight on January
20: went to the movies with a man I used to go toward Harvard Square in Cambridge is mad fun for
30: the first list was posted on the bulletin he brought home a turkey will die on my rug is deep with snow and sleet are destructive and playful students always
50: the next room to mine silver in Pennsylvania is late in getting home on time my date was tremendous fun going there skiing this day would end and have no more objections to his speech on the radio last night played the viola in the orchestra and chorus performed the

5-order approximation

10: they saw the play Saturday and sat down beside him
20: road in the country was insane especially in dreary rooms where they have some books to buy for studying Greek
30: go it will be pleasant to you when I am near the table in the dining room was crowded with people it crashed into were screaming that they had been
50: house to ask for is to earn our living by working towards a goal for his team in old New-York was a wonderful place wasn't it even pleasant to talk about and laugh hard when he tells lies he should not tell me the reason why you are is evident

7-order approximation

10: recognize her abilities in music after he scolded him before
20: easy if you know how to crochet you can make a simple scarf if they knew the color that it
30: won't do for the members what they most wanted in the course an interesting professor gave I went to at one o'clock stopped at his front door and rang the
50: then go ahead and do it if possible while I make an appointment I want to skip very much around the tree and back home again to eat dinner after the movie early so that we could get lunch because we liked her method for sewing blouses and skirts is

Text

10: the history of California is largely that of a railroad
20: more attention has been paid to diet but mostly in relation to disease and to the growth of young children
30: Archimedes was a lonely sort of eagle as a young man he had studied for a short time at Alexandria Egypt where he made a life-long friend a gifted mathematician
50: the old professor's seventieth birthday was made a great occasion for public honors and a gathering of his disciples and former pupils from all over Europe thereafter he lectured publicly less and less often and for ten years received a few of his students at his house near the university