Some Effects of Intermittent Silence

Author(s): George A. Miller

Source: *The American Journal of Psychology*, Jun., 1957, Vol. 70, No. 2 (Jun., 1957), pp. 311-314

Published by: University of Illinois Press

Stable URL: https://www.jstor.org/stable/1419346

a session. A relatively shorter period would be less likely to receive reinforcements on a given schedule, and might be expected to produce negative superstition more frequently. At the other extreme, an incidental stimulus which occupied half the experimental session would presumably share so nearly equally in the reinforcements that there would be no substantial separation of rates. The schedule and the performance generated are also relevant in determining the frequency of adventitious reinforcement. Finally, the nature and intensity of the incidental stimulus also may have their effect.

Pending an investigation of these parameters, it may at least be said that incidental stimuli adventitiously related to reinforcement may acquire marked discriminative functions.[3]

Harvard University W. H. MORSE
B. F. SKINNER

SOME EFFECTS OF INTERMITTENT SILENCE

Imagine that a monkey hits the keys of a typewriter at random, subject only to these constraints: (1) he must hit the space bar with a probability of $p(*)$ and all the other keys with a probability of $p(\mathrm{L}) = 1 - p(*)$, and (2) he must never hit the space bar twice in a row. I wish to examine the monkey's output, not because it is interesting, but because it will have some of the statistical properties considered interesting when humans, rather than monkeys, hit the keys.

In the monkey's output we will find runs of i letters in a row, where $i = 1, 2, 3, \ldots$, separated by single spaces. We will expect to find runs of length i with a probability

$$P_i = p(*)p(\mathrm{L})^{i-1}, \qquad i = 1, 2, \cdots, \qquad [1]$$

so that the probability of a word of length i will decrease exponentially as i increases.

Now suppose that there are A different keys on the typewriter, excluding the space bar. Then the number of different possible words of length i must be A^i. Since the probability of any particular word must be P_i divided by the A^i different words of the same length, it must be equal to $P_i/A^i = p(*)p(L)^{i-1}A^{-i}$, on the assumption that all sequences of letters are equally probable. If we represent this quantity, the probability of a word of length i, by the symbol, $p\ (w, i)$, we can write it as:

$$p(w, i) = [p(*)/p(L)]e^{-i[\log A - \log p(\mathrm{L})]}, \qquad [2]$$

[3] This work was done under a grant from the National Science Foundation.

where w is a word of length i and everything else other than i on the right side is a constant.

Since there are A different keys available, there must be A words one letter long, $A + A^2$ words of length no greater than two, $A + A^2 + A^3$ of length no greater than three, etc. Therefore, we can write the general expression that there must be

$$\sum_{i=1}^{k} A^i = A(1 - A^k)/(1 - A), \qquad (A \neq 1) \tag{3}$$

words of length less than $k+1$.

Now suppose that we rank order all the different words with respect to length. The one-letter words will all get ranks between 1 and A, the two-letter words will all get ranks between $A + 1$ and $A(1 - A^2)/(1 - A)$, the three-letter words will all get ranks between $[A(1 - A^2)/(1 - A)] + 1$ and $A(1 - A^3)/(1 - A)$, etc. Thus the particular word, w, of length i, will be assigned an average rank, $r(w, i)$, given by the expression

$$\begin{aligned} r(w, i) &= [[A(1 - A^{i-1})/(1 - A)] + 1 + [A(1 - A^i)/(1 - A)]]/2 \\ &= A^i[(A + 1)/2(A - 1)] - [(A + 1)/2(A - 1)]. \end{aligned} \tag{4}$$

It will prove convenient to rewrite Equation [4] as follows:

$$[2(A - 1)/(A + 1)] \cdot [r(w, i) + (A + 1)/2(A - 1)] = e^{i \log A}. \tag{5}$$

Now we can combine Equations [2] and [5]. This is seen most easily if we first rewrite Equation [2] as follows:

$$p(w, i) = \frac{p(*)}{p(L)} \left[e^{i \log A}\right]^{-[1-[\log p(L)/\log A]]}$$

so that it becomes obvious how to substitute the left side of Equation [5]:

$$p(w, i) = \frac{p(*)}{p(L)} \left[\frac{2(A - 1)}{A + 1} \left[r(w, i) + \frac{A + 1}{2(A - 1)}\right]\right]^{-1-[\log p(L)/\log A]]} \tag{6}$$

And this mess is what we wanted to derive. Actually, Equation [6] looks quite harmless after we define some new constants, substitute them in Equation [6], and write it as

$$p(w) = b[r(w) + c]^{-d}, \tag{7}$$

for now we have a simple relation between the probability of a word and its rank order. (Since probability decreases with length, once the words have been ordered with respect to increasing length, they are also ordered with respect to decreasing probability.)

Now suppose that we take the values $A = 26$ and $p(*) = 0.18$ from the English language and compute the constants:

$$b = \frac{.18}{.82}\left(\frac{50}{27}\right)^{-1.06} = 0.11, \qquad c = \frac{27}{50} = 0.54,$$

$$d = 1 - \frac{\log 0.82}{\log 26} = 1.06$$

Then we have

$$p(w) = 0.11[r(w) + 0.54]^{-1.06} \tag{8}$$

If we wanted to round things off a bit, we might overlook the values 0.54 and 0.06 and so obtain the simpler expression,

$$p(w) = 0.11r(w)^{-1} \tag{9}$$

Equation [9] is very similar to a relation that is often called 'Zipf's Law.' Zipf found approximately $p(w)r(w) = 0.10$ in most natural languages.[1]

Research workers in statistical linguistics have sometimes expressed amazement that people can follow Zipf's Law so accurately without any deliberate effort to do so. We see, however, that it is not really very amazing, since monkeys typing at random manage to do it about as well as we do.

Equation [7] is due to Mandelbrot.[2] In his original derivation, Mandelbrot minimized the average cost per unit of information. Later, Simon advanced a different explanation of Zipf's rule, one based upon the model of a 'pure birth process.'[3] Simon's equation does not fit the verbal data as well as does Mandelbrot's, but it would seem to have the advantage of being derived from averaging rather than maximizing assumptions. The present argument should make it clear, however, as it was clear to Mandelbrot, that maximization is not essential to the derivation of Equation [7]. The assumption of maximization can be replaced by the assumption of random spacing.

In other words, Mandelbrot did more than derive Zipf's rule. He went on to show that the random placement of spaces which leads to Zipf's rule is actually the optimal solution. Our monkeys are doing the best possible

---

[1] G. K. Zipf, *The Psychobiology of Language,* 1935, 44-47; *Human Behavior and the Principle of Least Effort,* 1949, 19-55.

[2] B. Mandelbrot, Jeux de communication, *Publ. Inst. Stat. Univ. Paris,* 2, 1953, 1-124; Simple games of strategy occurring in communication through natural languages, *Trans. I. R. E.,* PGIT-3, 1954, 124-137. Also abstracted briefly in G. A. Miller, Communication, *Annu. Rev. Psychol.,* 5, 1954, 401-420.

[3] H. A. Simon, On a class of skew distribution functions, *Biometrika,* 42, 1955, 425-440.

job of encoding information word-by-word, subject to the constraints we imposed on them. If we were as smart as the monkeys, we, too, would generate all possible sequences of letters and so make better use of our alphabet. Instead, we use only a small fraction of the possible letter sequences. In our behalf, however, it should be added that the consequent redundancy serves as insurance against errors. Since it is possible to regard redundancy as reducing the effective size of the alphabet, it should be noted that variation in A has little effect in Equation [6] relative to the effects produced by changes in $p(*)$. Thus if redundancy reduced us to the equivalent of three letters, Equation [6] would change to

$$p(w) = \frac{p(*)}{p(L)} \left[r(w) + 1\right]^{-[1-2.2 \log p(L)]}, \qquad [10]$$

whereas, at the other extreme where $A \rightarrow \infty$,

$$p(w) = \frac{p(*)}{2p(L)} \left[r(w) + 0.5\right]^{-1}. \qquad [11]$$

Mandelbrot has found that the parameter d varies around 1.2, which corresponds rather well to Equation [10], when $p(L) = 0.82$, for then $d = 1 - 2.2 \log 0.82 = 1.187$. Similarly, an alphabet of $A = 3$ letters plus a space corresponds fairly well with Shannon's estimate of the redundancy of printed English.

It seems, therefore, that Zipf's rule can be derived from simple assumptions that do not strain one's credulity (unless the random placement of spaces seems incredible), without appeal to least effort, least cost, maximal information, or any branch of the calculus of variations. The rule is a simple consequence of those intermittent silences which we imagine to exist between successive words.[4]

Harvard University GEORGE A. MILLER

THE USE OF FOREIGN LANGUAGES BY PSYCHOLOGISTS, CHEMISTS, AND PHYSICISTS

A previous note demonstrated that psychologists do not draw from other national literatures in proportion to the distribution of materials by language and country.[1] It is interesting to test this same hypothesis concerning two sister sciences—chemistry and physics. The present data for

[4] This note was prepared under Contract AF33(038)-14343 and appears as report number AFCRC TN 56-60, ASTIA, Document No. AD 98822.

[1] C. M. Louttit, The use of foreign languages by psychologists, this JOURNAL, 68, 1955, 484-486.