Information and Memory

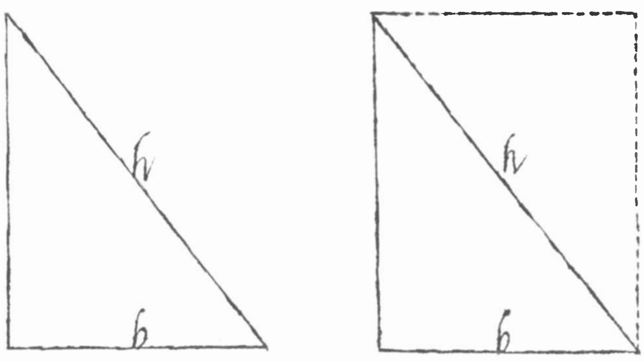
If a man sees six marbles, he can usually name their number without counting. With more marbles, he often makes mistakes. This indicates a limitation of perception that is overcome by resourceful stratagems

by George A. Miller

Some things are easy to remember. A short poem is easier to memorize than a long one; an interesting story is better recalled than a dull one. But brevity and wit are not all that is involved. Equally important is the way things fit together. If a new task meshes well with what we have previously learned, our earlier learning can be transferred with profit to the novel situation. If not, the task is much harder to master.

Imagine that you are teaching geometry to children. You have covered the business of calculating the length of the hypotenuse of a right-angled triangle when the base and the altitude are given. Now you are about to take up the problem of finding the *area* of a right triangle when the base and the hypotenuse are given. Suppose you were given your choice of the following two methods of teaching the children to solve the problem. In method A you would help them to discover that the area of a right triangle is half that of a rectangle with the same base and altitude, that the unknown altitude of the triangle in this case can be calculated from the given base and hypotenuse by use of the Pythagorean theorem, and that the area of the triangle can therefore be found by deriving the altitude, computing the area of the rectangle and then taking half of that. In method B you would simply tell the class to memorize six steps: (1) add the length of the base to the length of the hypotenuse; (2) subtract the length of the base from the length of the hypotenuse; (3) multiply the first result by the second; (4) extract the square root of this product; (5) multiply the positive root by the length of the base; (6) divide this product by 2.

Which method of teaching would you choose? Probably no one but an experimental psychologist would ever consider method B. Method A is productive and insightful; method B is stupid and ugly.



TWO METHODS may be used to teach children how to find the area of a right-angled triangle. The method associated with the triangle at left has the following algebraic steps: (1) $h + b = v$, (2) $h - b = w$, (3) $v \times w = x$, (4) $\sqrt{x} = \pm y$, (5) $+ y \times b = z$, (6) $z/2 =$ area. The method associated with the triangle at right proceeds: (1) find altitude by the Pythagorean theorem, (2) find area of rectangle from base and altitude, (3) area of triangle is half the area of rectangle. The first method is considered ugly and the second efficient. Why?

But just why do we find method B repulsive? What is repugnant about a procedure that is logically impeccable and that leads always to the correct answer? This question is raised by the psychologist Max Wertheimer in his provocative little book, *Productive Thinking*. An obvious answer is that a child taught by method A will understand better what he is doing. But until we can say what it means to understand what one is doing, or what profit there is in such understanding, we have not really answered Wertheimer's question.

It is helpful to consider the interesting fact that method B is the procedure we would use to instruct a computing machine of the present-day type. The machine is able to perform arithmetical operations such as addition, subtraction, multiplication, division and the extraction of roots. Instruction for the machine consists in writing a "program"—like the series of steps used in method B except that the computer's program must be even more explicit and detailed, with even less hint of the basic strategy. Computing machine engineers have their hearts set on some day designing machines which will construct programs for themselves: that is, given the strategy for handling a problem, the machine will understand the problem well enough to create all the appropriate operations or subroutines required to solve it. The desirability of such a development is obvious. In the first place, at present it takes many hours of drudgery to write the detailed instructions for all the steps a computer must take. Then, after the instructions have been written, they must be stored in the machine in some easily accessible form. In a large machine the number of subroutines may run into the thousands; it might actually be more economical to equip the machine with the ability to create them on demand rather than to build the necessary storage and access machinery. In other words, in a very elaborate computer it would be more efficient to store rules from which subroutines could be generated than to store the routines themselves.

It seems, therefore, that even the computing machine realizes that method B is ugly. Each subroutine is an isolated operation that must be stored in its proper place, and no attempt is made to tie these steps to other information available to the machine. So we can see that one superiority of method A lies in the fact that it makes more efficient use of the capacity for storing information. In the teaching of geometry to a child, method A highlights the relations of the new problem to things that the child has already learned, and thus it provides the rules by which the child can write his own subroutines for computation. In essence the ugly method is less efficient because it requires the child to master more new information.

The intimate relation between memory and the ability to reason is demonstrated every time we fail to solve a problem because we fail to recall the necessary information. Since our capacity to remember limits our intelligence, we should try to organize material to make the most efficient use of the memory available to us. We cannot think simultaneously about everything we know. When we attempt to pursue a long argument, it is difficult to hold each step in mind as we proceed to the next, and we are apt to lose our way in the sheer mass of detail. Three hundred years ago René Descartes, in an unfinished treatise called *Rules for the Direction of the Mind*, wrote:

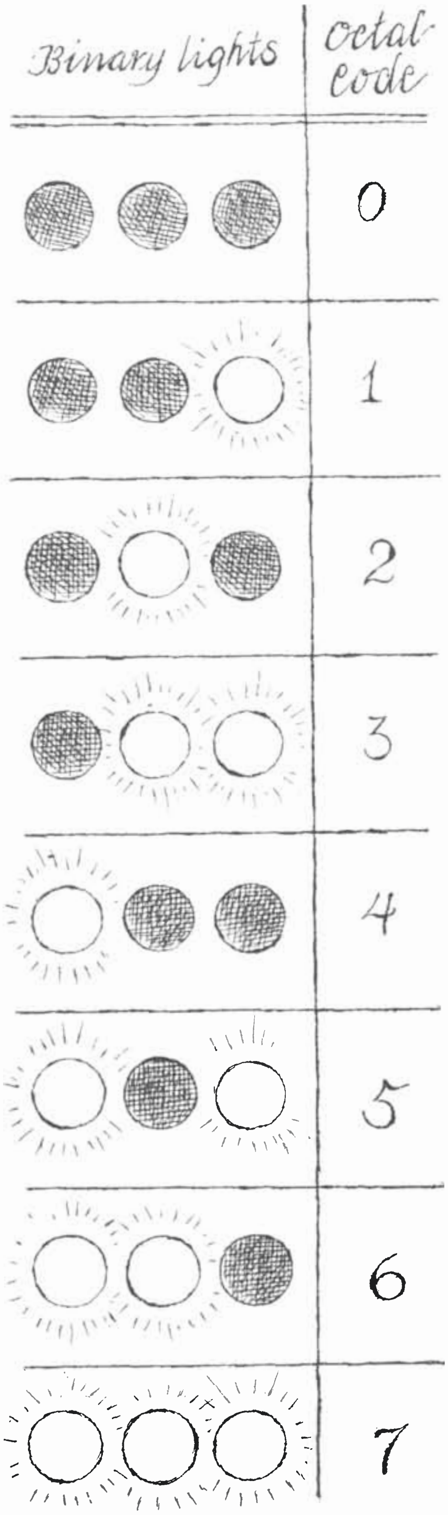
"If I have first found out by separate mental operations what the relation is between the magnitudes A and B, then that between B and C, between C and D, and finally between D and E, that does not entail my seeing what the relation is between A and E, nor can the truths previously learned give me a precise knowledge of it unless I recall them all. To remedy this I would run them over from time to time, keeping the imagination moving continuously in such a way that while it is intuitively perceiving each fact it simultaneously passes on to the next; and this I would do until I had learned to pass from the first to the last so quickly, that no stage in the process was left to the care of memory, but I seemed to have the whole in intuition before me at the same time. This method will relieve the memory, diminish the sluggishness of our thinking, and definitely enlarge our mental capacity."

Descartes's observation is familiar to anyone who has ever memorized a poem or a speech, or mastered a mathematical proof. Rehearsal or repetition has the very important effect of organizing many separate items into a single unit, thus reducing the load our memory must carry and leaving us free for further thinking. In terms of logic, the process is like the substitution of a single symbol for a longer expression which would be clumsy to write each time we wanted to use it.

The practical advantages of this unitizing process were vividly illustrated for me the first time I saw one of those digital computing machines that have small neon lights to show which relays are closed. There were 20 lights in a row, and I did not see how the men who ran the machine could grasp and remember a pattern involving so many elements. I quickly discovered that they did not try to deal with each light as an individual item of information. Instead, they translated the light pattern into a code. That is to say, they grouped the lights into successive triplets and gave each possible triplet pattern a number as its name, or symbol. The pattern all three lights off (000) was called 0; the pattern off-off-on (001) was called 1; off-on-off (010) was called 2, and so forth. Having memorized this simple translation, the engineers were able to look at a long string of lights such as 011000101001111 and break it down into triplets (011 000 101 001 111) which they immediately translated into 30517. It was much easier to remember these five digits than the string of 15 lit and unlit lights.

Reorganization enabled the engineers to reduce the original complexity to something easily apprehended and remembered without changing or discarding any of the original data. There is an analogy between this simple trick and the process described by Descartes. Each step in a complex argument is like a single light in the binary sequence. Rehearsal organizes the steps into larger units similar to the engineers' triplets. Repeated rehearsal patterns the long argument into larger and larger units which are then replaced in thought by simpler symbols.

The first person to propose an experimental test of the span of a man's instantaneous grasp seems to have been



COMPUTER LIGHTS are quickly read by engineers using the code illustrated here.

Sir William Hamilton, a 19th-century Scottish metaphysician. He wrote: "If you throw a handful of marbles on the floor, you will find it difficult to view at once more than six, or seven at most, without confusion." It is not clear whether Hamilton himself actually threw marbles on the floor, for he remarked that the experiment could be performed also by an act of imagination, but at least one reader took him literally. In 1871 the English economist and logician William Stanley Jevons reported that when he threw beans into a box, he never made a mistake when there were three or four, was sometimes wrong if the number was five, was right about half the time if the beans numbered 10 and was usually wrong when the number reached 15. Hamilton's experiment has been repeated many times with better instrumentation and control, but refined techniques serve only to confirm his original intuition. We are able to perceive up to about six dots accurately without counting; beyond this errors become frequent.

But estimating the number of beans or dots is a perceptual task, not necessarily related to concepts or thinking. Each step in the development of an argument is a particular thing with its own structure, different from the other steps and quite different from one anonymous bean in Jevons's box. A better test of "apprehension" would be the ability to remember various symbols in a given sequence. Another Englishman, Joseph Jacobs, first performed this experiment with digits in 1887. He would read aloud a haphazard sequence of numbers and ask his listeners to write down the sequence from memory after he finished. The maximum number of digits a normal adult could repeat without error was about seven or eight.

From the first it was obvious that this span of immediate memory was intimately related to general intelligence. Jacobs reported that the span increased between the ages of 8 and 19, and his test was later incorporated by Alfred Binet, and is still used, in the Binet intelligence test. It is valuable principally because an unusually short span is a reliable indicator of mental deficiency; a long span does not necessarily mean high intelligence.

A person who can grasp eight decimal digits can usually manage about seven letters of the alphabet or six monosyllabic words (taken at random, of course). Now the interesting point about this is that six words contain much more information, as defined by information theory, than do seven letters or eight digits. We are therefore in a position analogous to carrying a purse which will hold no more than seven coins—whether pennies or dollars. Obviously we will carry more wealth if we fill the purse with silver dollars rather than pennies. Similarly we can use our memory span most efficiently by stocking it with informationally rich symbols such as words, or perhaps images, rather than with poor coin such as digits.

The mathematical theory of communication developed by Norbert Wiener and Claude Shannon provides a precise measure of the amount of information carried. In the situation we are considering, the amount of information per item is simply the logarithm (to the base two) of the number of possible choices. Thus the information carried by a binary digit, where there are two alternatives, is $\log_2 2 = 1$ bit. In the case of decimal digits the amount of information per digit is $\log_2 10 = 3.32$ bits. Each letter of the alphabet carries $\log_2 26 = 4.70$ bits of information. When we come to make the calculation for words, we must take into account the size of the dictionary from which the words were drawn. There are perhaps 1,000 common monosyllables in English, so a rough estimate of the informational value of a monosyllabic word selected at random might be about 10 bits.

A person who can repeat nine binary digits can usually repeat five words. The informational value of the nine binary digits is nine bits; of the five words, about 50 bits. Thus the Wiener-Shannon measure gives us a quantitative indication of how much we can improve the efficiency of memory by using informationally rich units. The computer engineers who group the relay lights by threes and translate the triplets into a code can remember almost three times as much information as they would otherwise.

It is impressive to watch a trained person look at 40 consecutive binary digits, presented at the rate of one each second, and then immediately repeat the sequence without error. Such feats are called "mnemonic tricks"—a name that reveals the suspicious nature of psychologists. The idea that trickery is involved, that there is something bogus about it, has discouraged serious study of the psychological principles underlying such phenomena. Actually some of the best "memory crutches" we have are called laws of nature. As for the common criticism that artificial memory crutches are quickly forgotten, it seems to be largely a question of whether we have used a stupid crutch or a smart one.

When I was a boy I had a teacher who told us that memory crutches were only

SIR WILLIAM HAMILTON, a 19th-century Scottish philosopher (not to be confused with Sir William Rowan Hamilton, the mathematician), observed: "If you throw . . . marbles on the floor, you will find it difficult to view at once more than six . . . without confusion."

Binary (1 bit)	Decimal (3.3 bits)	Alphabetic (4.7 bits)	Syllabic (10 bits)
110100	4972	XJR	for, line
0100110	86515	AYCZ	nice, it, act
10010011	021942	EDLYG	time, who, to, air
101100010	3776380	QJPEVJ	by, west, cent, or, law
0010101110	28201394	DLXBAHC	bay, sea, ten, red, ask, mob
11010001011	918374512	HOKOMSFB	go, how, ice, save, hat, sue, way
101001110110	1038204665	FQGUJRZVM	odd, gas, call, at, ant, pay, get, was
0001010111011	57048621937	DNKSNWJUWT	by, game, log, free, so, you, car, big, why

SPAN OF IMMEDIATE MEMORY depends mainly on the number of items to be memorized and is relatively independent of the amount of information per item. In this table the amount of information is measured in "bits," or binary digits. A binary digit can be 1 or 0, and hence conveys a minimum amount of information. A person who can repeat nine binary digits can usually memorize seven decimal digits, six letters or five words (*row above broken line*). The other rows compare the span for other groups of items.

one grade better than cheating, and that we would never understand anything properly if we resorted to such underhanded tricks. She didn't stop us, of course, but she did make us conceal our method of learning. Our teacher, if her conscience had permitted it, no doubt could have shown us far more efficient systems than we were able to devise for ourselves. Another teacher who told me that the ordinate was vertical because my mouth went that way when I said it and that the abscissa was horizontal for the same reason saved me endless confusion, as did one who taught me to remember the number of days in each month by counting on my knuckles.

The course of our argument seems to lead to the conclusion that method A is superior to the ugly method B because it uses better mnemonic devices to represent exactly the same information. In method A the six apparently arbitrary steps of method B are organized around three aspects of the total problem so that each aspect can be represented by symbols which the student has already learned. The process is not essentially different from the engineers' method for recoding a sequence of binary lights.

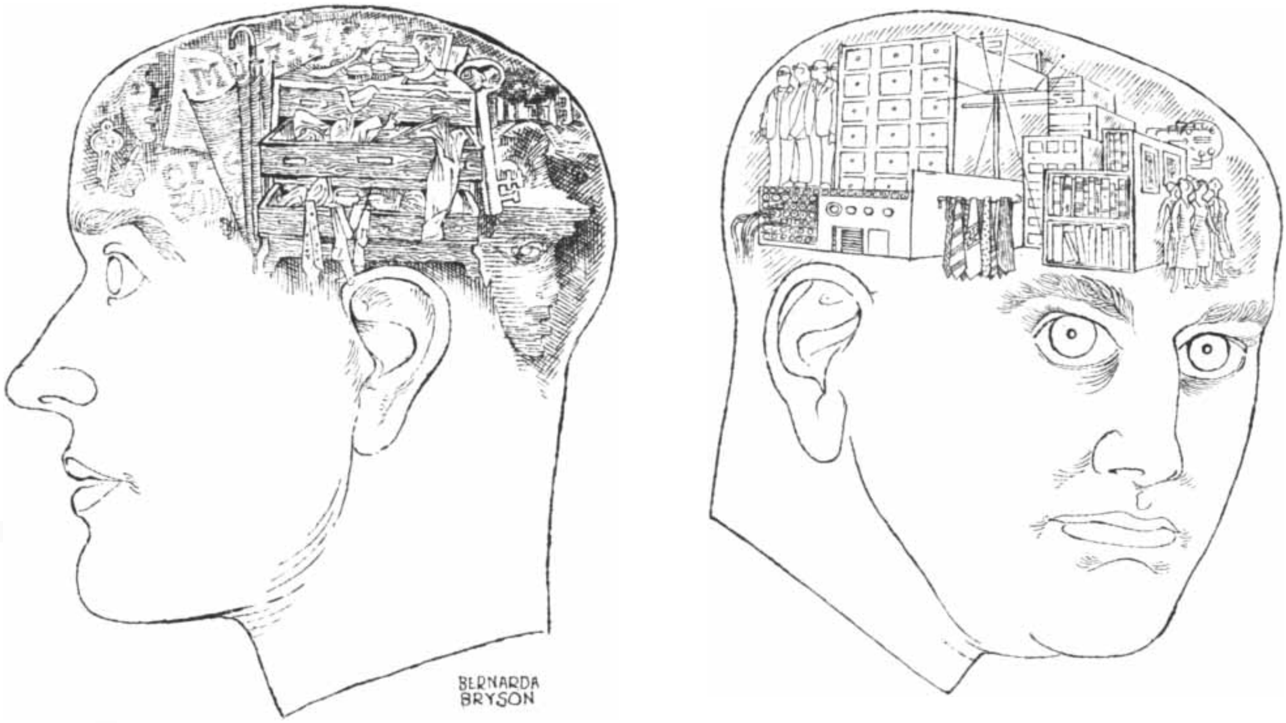
It is conceivable that all complex, symbolic learning proceeds in this way. The material is first organized into parts which, once they cohere, can be replaced by other symbols—abbreviations, initial letters, schematic images, names, or what have you—and eventually the whole scope of the argument is translated into a few symbols which can all be grasped at one time. In order to test this hypothesis we must look beyond experiments on the span of immediate memory.

Our question is: Does the amount of information per item (*i.e.*, the number of possible alternative choices per item) affect the number of items we can remember when there is a large amount of material to be mastered? For example, is it more difficult to memorize a random sequence of 100 monosyllabic words than 100 digits or 100 letters of the alphabet? The question is important because it has a bearing on how we can organize material most efficiently for learning.

In an exploratory study that S. L. Smith and I devised at the Harvard University Psychological Laboratories, the subjects were required to memorize three different kinds of lists of randomly chosen items. One list was constructed from a set of 32 alternatives (all the alphabet except Q plus the numerals 3, 4, 5, 6, 7, 8 and 9), another from a set of eight alternatives, and the third from just two alternatives. The subject read a test list at the rate of one item every second and then had to write down as much of the list as he could remember in the correct order. The lists ran to 10, 20, 30 or 50 items. If the subject failed to reproduce the list exactly, it was presented again. The number of presentations required before the first perfect reproduction measured the difficulty of the task.

We were not greatly surprised to find that the subjects did somewhat better (*i.e.*, needed about 20 per cent fewer trials) on the binary-choice lists than on the other types. After all, a run of, say, six zeros or six ones is easy to remember and therefore in effect shortens the list. But on the other two types of lists (eight alternatives and 32 alternatives) the subjects' performances were practically indistinguishable. In other words, it was just as easy to memorize a list containing a lot of information as one of the same length containing less information.

Very similar results have been obtained at the University of Wisconsin by W. J. Brogden and E. R. Schmidt, who did their experiments for other reasons and without knowledge of the hypothesis Smith and I were trying to test. They used verbal mazes with either 16 or 24 choice points and they varied the number of alternatives per choice point from two to 12. Here again the length of the list of points that had to be learned, and not the number of alternatives offered at each choice point, determined the difficulty of the test—with the same excep-



FANCIFUL HEADS were drawn by Bernarda Bryson to depict René Descartes's *Rules for the Direction of the Mind*, described in the text of the article. The individual at left has presumably not had the benefit of the rules, whereas the man at right has.

tion that we found, namely, that it was slightly easier to remember where only two choices were offered.

Tentatively, therefore, we are justified in assuming that our memories are limited by the number of units or symbols we must master, and not by the amount of information that these symbols represent. Thus it is helpful to organize material intelligently before we try to memorize it. The process of organization enables us to package the same total amount of information into far fewer symbols, and so eases the task of remembering.

How much unitizing and symbolizing must we do, and how can we decide what the units are? The science of linguistics may come to our aid here. Language has a hierarchical structure of units—sounds, words, phrases, sentences, narratives—and it is there that one should seek evidence for a similar hierarchy of cognitive units.

It has been estimated that English sentences are about 75 per cent redundant: that is, about four times as long as they would need to be if we used our alphabet with maximum efficiency. At first glance this fact seems paradoxical. If length is our major source of difficulty, why do we deliberately make our sentences longer than necessary? The paradox arises from a confusion about the definition of sentence length. Is a sentence 100 letters, or 25 words, or 6 phrases, or one proposition long? The fact that all our books contain 75 per cent more letters than necessary does not mean that 75 per cent of the ideas could be deleted. And it is those larger subjective units, loosely called ideas, that we must count to determine the psychological length of any text.

A sequence of 25 words in a sentence is easier to recall than a sequence of 25 words taken haphazardly from the dictionary. The sentence is easier because the words group themselves easily into familiar units. In terms of psychological units, a 25-word sentence is shorter than a sequence of 25 unrelated words. This means that the word is not the appropriate unit for measuring the psychological length of a sentence. Perhaps linguistic techniques for isolating larger units of verbal behavior will provide an objective basis for settling the question.

When we memorize a sentence, all our previous familiarity with the lexicon and grammar of the language comes to our aid. It is one of the clearest possible examples of the transfer of previous learning to a new task. And the transfer is profitable because it serves to reduce the effective length of the material to be remembered. By learning the language, we have already acquired automatic habits for unitizing those sequences that obey the rules of the language.

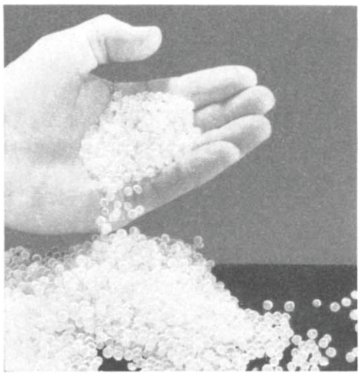
There are three stages in the unitizing process. All three were described in the 17th century by John Locke in his famous *Essay Concerning Human Understanding:* "Wherein the mind does these three things: first, it chooses a certain number [of specific ideas]; secondly, it gives them connexion, and makes them into one idea; thirdly, it ties them together by a name." Men form such complex ideas, Locke said, "for the convenience of communication," but the combination of ideas sometimes leads to confusion because it is "the workmanship of the mind, and not referred to the real existence of things." The development in the 20th century of a mathematical theory of communication enables us to see more clearly how this process serves the convenience of communication and, coupled with the fact that it is the length, not the variety of the material that limits our memories, gives us an important insight into the economics of cognitive organization.

Organizing and symbolizing are pervasive human activities. If we can learn to perform them more efficiently, perhaps we shall indeed be able, as Descartes promised, to "relieve the memory, diminish the sluggishness of our thinking, and definitely enlarge our mental capacity."

Kodak reports to laboratories on:

the plasticizerless plastic for the kitchen middens of tomorrow . . . how to draw a crowd into your tent . . . non-hygroscopic acetylcholine iodide

Beyond the squeeze bottle



Since we are stepping up the pace at Longview, Texas, to make 40,000,000 pounds annually of these *Tenite Polyethylene* pellets, it is desirable to look beyond squeeze bottles, sheet, pipe, wire covering, and housewares, much as our vast learning in dye chemistry is winning favor from a color-responsive populace who have known polyethylene only as the white, waxy one. No stone must be left unturned. From under one there might crawl out some triviality, colorful or otherwise, that within a twelvemonth *has* to go on every single ****** ******* in service. Likely as not, the gizmo's daddy turns out to be a guy better wooed with numbers than with warm, reassuring words. All right:

Average Properties of Tenite Polyethylene

Thermal coefficient of expansion	11×10^{-5} in/in/°F
Thermal conductivity	8×10^{-4} cal/cm²/sec/°C/cm
Specific heat	
solid (70-105F)	0.55
liquid (250-285F)	0.70
Brittleness temperature	<-70C
Tensile strength at fracture, 73F	2150 to 1100 psi (depending on formulation selected)
Dielectric strength (.030″ specimen at 60 cycles)	1000 v/mil (short time)
Volume resistivity	10^{19} ohms/cm
Surface resistivity	$>4 \times 10^{14}$ ohms
Dielectric constant	2.3
Dissipation factor	<0.0005

For more such, write to Eastman Chemical Products, Inc., Kingsport, Tenn. (Subsidiary of Eastman Kodak Company), requesting the 16-page specification booklet on Tenite Polyethylene, *the plasticizerless plastic by which future antiquaries may date our kitchen middens. During the discussion of how the beads might get fabricated into the objects you need, we have every intention of making a friend out of you.*

Slowdown in color

In the heat of debate we once heard an advertising man cry out, "What's a product? Anybody can make a product. The real art is selling a product."

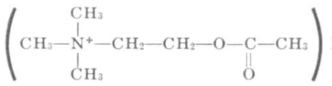
Though since moved on to fields where his artistry could more lushly flower, he wasn't entirely wrong, just too sweeping in his value judgements. In the market place—particularly in the industrial market place—many a wonderfully ingenious and efficient product of the engineering mind and hand fails to ring the bell as loud and clear as expected, simply because too few potential customers know how the thing works. One way to draw a crowd into the tent for educational purposes is to show them movies. Showmanship isn't all; some mechanisms can be seen at work in no other way than through movies which slow down the action fifty times or more. Sometimes recognition of this is all the showmanship needed.

There was a time when these high speed movies were used only for development and trouble-shooting. Long miles of high speed film still quite justify themselves in the form of black-and-white rush negatives shown once to taut little engineering groups, but more and more high speed shooting is done on *Kodachrome Film* and even on *Commercial Kodachrome Film*, which is chosen only with advance knowledge that numbers of full-color copies will be required for circulation.

"High Speed Motion Pictures," a new booklet obtainable from Eastman Kodak Company, Sensitized Goods Division, Rochester 4, N. Y., tells about the Kodak High Speed Camera *and about the films spooled for this kind of movie making.*

A delightful business

You sit one fine morning reading your mail and lo, there is a letter from a researcher at a medical school in the green hills of Vermont who wants to know why you don't put up acetylcholine as the



quaternary iodide. Somehow, possibly by reading it in a book, he seems to have learned that acetylcholine iodide is not hygroscopic at all, whereas that well known vasodilator acetylcholine itself and its well known bromide (Eastman 2117) absorb so much water from the air that to try to weigh them out accurately is a nuisance. There doesn't seem to be a blessed reason in the world why you can't put up acetylcholine iodide. You even think of a way to make it from starting material less costly than acetylcholine. You try it and it works and the iodide *is* non-hygroscopic. That particular researcher is pleased at what an agreeable fellow you are, and you are pleased at the prospect of all the acetylcholine iodide that all the other researchers are going to buy from you.

So now all you have to do is to advertise that Acetylcholine Iodide (*Eastman 7209*) *costs $2.60 for 10 g. and is one of some 3500 highly purified organic chemicals sold by Distillation Products Industries, Eastman Organic Chemicals Department, Rochester 3, N. Y.* (*Division of Eastman Kodak Company*).

Price quoted is subject to change without notice.

This is one of a series of reports on the many products and services with which the Eastman Kodak Company and its divisions are . . . *serving laboratories everywhere*

Kodak
TRADEMARK