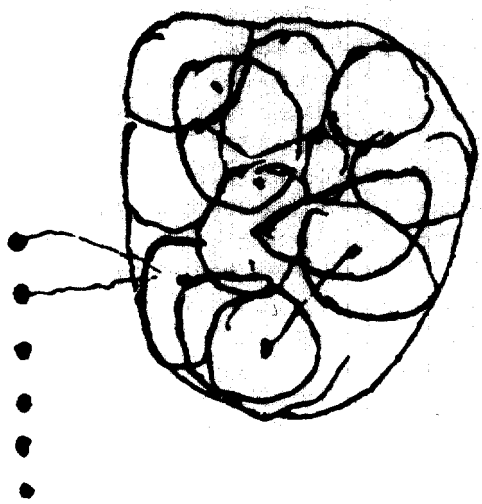
A Gamow Memorial Lecture delivered at the University of Colorado, Boulder, on October 5, 1982



Reflections on the Brain's Attempts to Understand Itself

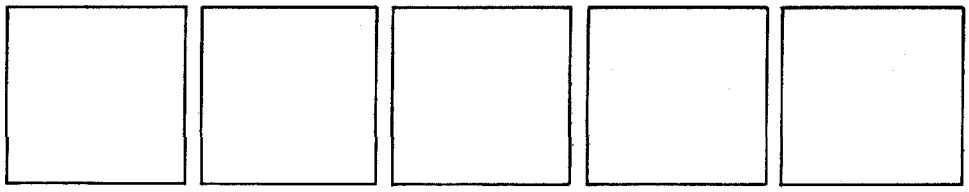
by S. M. Ulam

My choice of subject for this talk may seem strange, since I am not a psychologist, a physiologist, or a neurologist, merely a mathematician and an amateur, a dilettante, in the workings of the brain. However, it is fitting that I give such a talk in memory of the late George Gamow, a friend of mine. Though by training a physicist, he was able to make famous contributions in other sciences, such as astronomy and biology, that interested him toward the end of his life. He was, like me, an amateur, a dilettante, in biology. Nevertheless one of the most important discoveries of recent times in that field is due to him. It was Gamow who first pointed out that ordered arrangements of four chemical units—four "letters' '-along the DNA double helix, or chain, as he called it. might be codes for many biological processes, and that the codes for the manufacture of proteins might consist of three- or four-letter "words."

What I want to do today is talk about several of my own speculations, with some mathematical symbolism, concerning the operation of the brain. I believe that discoveries and breakthroughs within the next twenty years will lead to a better understanding of the mechanisms of the brain, of the processes of thought. It will not be a complete understanding—that would be too much to hope for—but it will give us some ideas of how the nervous system operates in lower animals and in humans.

Mathematicians may help in reaching this understanding, although for the time being I think that 99 percent of the progress will come from physiological and anatomical experiments. However, mathematics can be useful, for it is clear that the similarities between electronic computers and the nervous system are of great importance.

Another friend of mine, the late John von Neumann, was one of the pioneers in the planning and building of electronic computers. His book *The Computer and the Bruin,* which was published posthumously in 1957, is still one of the most elegant and understandable general introductions to the subject. 1 remember the discussions we had on how the advent of computers would enlarge the scope of experimentation in mathematical and physical sciences and about his specific interest in the partial analogies between computers, as they were planned in the early forties, and the processes of deductive thinking. We saw each other frequently at the time, either in Los Alamos or in



Princeton, and we would marvel at the few physiological facts then known about the brain, such as the number of neurons it contains. That number was of the order of ten billion, and their interconnections in the human cortex were known to be still more numerous. He would say: "Not only are there ten billion computing elements, but each is connected to many others, one hundred maybe! And maybe even to one thousand in the central part of the brain!" Well now, forty years later, the number of interconnections has been shown to be of the order of thousands, up to one hundred thousand in the central part of the brain. And the total number of connections, of axons and synapses, is of order 10^{14}. So you see, in the recent past the purely anatomical and physiological knowledge has vastly changed. The locations of certain centers in the brain and the differences between its right and left halves are also better known. And today more information is being gathered through studies of the electromagnetic signals being emitted constantly by the brain.

However, I do not believe that now, or even in the near or distant future, it will be possible to gain what might be called a complete understanding of the brain's operation. My belief rests on very important and strange results in pure mathematics. These results, which date from 1930, are associated mainly with the name of Godel, a mathematician who worked at the Institute for Advanced Study in Princeton. Godel proved a theorem that says, roughly speaking, that in any mathematical system, any logical system, there exist statements that have sense but cannot be proved or disproved. So in every mathematical discipline one can conceive of at present, there are undecidable propositions, finite statements that, starting from axioms, one cannot demonstrate or show to be false.

Mathematics has a store of problems, some very old, whose solutions are not known. But it was assumed that, ultimately, yes or no solutions would be found. That was the belief of Hilbert, one of the greatest mathematicians of the last hundred years. Then Godel came along and showed that such a belief is no longer valid, that there are statements that are undecidable. This fact is of great philosophical significance. And beyond that, it could be a sort of consolation for our inability to attain a complete knowledge of various real phenomena.

So it is possible that some of the still unresolved mathematical problems are in *principle* undecidable on the basis of our present system of axioms. Many such problems are technically complicated, but let me give you one that is simple to state and understand.

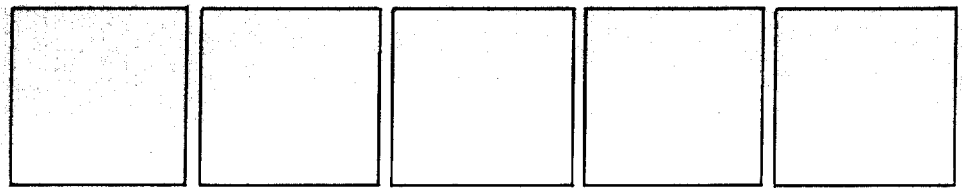
A prime number is an integer that is not divisible by any number except itself. The numbers 2, 3, 5, 7, ..., 41, 43, 47, et cetera, are all prime. The Greeks knew that there are infinitely many prime numbers. That is one of the oldest, greatest, and most beautiful discoveries in mathematics. Now certain pairs of prime numbers, such as 5 and 7, 11 and 13, 17 and 19, are called twins because they differ by only 2. The question is: How many twin primes are there, a fixed finite number or an infinity? Nobody knows the answer to this question, and it may be undecidable. I asked Professor Schmidt, a very famous number theorist, if he knew who first proposed this very old problem and whether he thought it might be undecidable. He did not know the answer to the former, and to the latter he answered, "One might not be able to decide whether it is undecidable!"

1 mention Godel's theorem to show the limitations of man's program to try to understand everything, even in a restricted domain. Perhaps the scope of the human brain is finite, or conversely, perhaps the growth of humanity, of its collection of brains, will, in terms of evolution, continue indefinitely and may reveal new points of view.

To continue the speculation on what the role of mathematics might be in the study of the brain, the time is not yet ripe to say its operation can be understood with abstract theories alone. But Gamow, who was perhaps the last great amateur in science, has shown us that it is possible to speculate—fruitfully, given some luck—on the great mysteries of nature. A Greek philosopher said that many are the wonders of the universe, but the greatest of all is the human mind. And Spinoza said that it is better to begin with small and modest truths. Starting from these premises, I want to give you now a few examples of biological questions that I think mathematics has already proved somewhat useful in answering, and how similar attempts and schematizations might possibly be of some use in partially understanding the nature of human perception.

One such question concerns the mechanism of recognition of external stimuli, say sights or sounds, and ultimately of ideas. Before recognition, there is perhaps discernment, discrimination. A priori it seems easier to see the difference between two objects than their similarity or analogy. We need to map the tremendous web of connections in the human brain into overlapping classes. But before we do this, here is an example of a mathematizable biological idea, one concerning the codes for the manufacture of proteins.

Gamow's suggestion about the existence of three- or four-letter codes for the constituent amino acids of proteins was almost correct. Many of the characteristics of living organisms are coded in very long sequences of four chemical units, which biologists call by the letters A, C, G, and T. Words are short strings of these letters. Finite sentences of several hundred words are codes for proteins, such as hemoglobins of various kinds. Today tens of thousands of these codes for proteins are known, and in some cases even the spatial forms of the proteins are known. A "reader" molecule goes along the DNA



"tape," reads the code, and deposits the information in other parts of the cell, in the ribosomes. This much is now understood. The functions of other parts of the long sequences, such as those called introns, are not yet understood, but they are not codes for proteins.

Some biologists are beginning to speculate on the importance of small differences that have been found to exist between the codes for a given protein in different species. For example, cytochrome c, which is important for the transmission of electrical impulses in nerves, differs slightly from one species to another but remains the same within a species. The biologist Emanuel Margoliash has tried to establish an evolutionary tree based on the quantitative differences in cytochrome *c* codes, on the gradations among them.

Mathematicians have studied in general the idea of comparing two elements *a* and b—two points in some space—by expressing the degree of their difference with a quantity called a distance. This distance, which is usually denoted by *p(a, b)*, should have the following properties. It should be positive definite: $\rho(a,b) > 0$ if $a \neq b$ and $\rho(a,a) = 0$. It should be symmetric: *p(a, b) = p(b, a).* And it should satisfy the triangle inequality: $\rho(a,c) \leq \rho(a,b) + \rho(b,c)$. This last property means that to go from *a* to c is no more difficult than to go from *a* to *b* and then from *b* to c. If such a distance exists for all pairs of points in a set S, then *S* is called a metric space.

I have said that the elements of the genetic code are sequences of symbols for four chemical units. For simplicity's sake and without changing any essentials, let us consider sequences of just two symbols, O and 1. For example, one such sequence *x* could be 0110101 and another sequence y could be 1000110. To get an idea of how much they differ, we want a distance *P(X, y)* between x and y. Let x_i be the *i*th symbol in *x* and y_i be the *ith* symbol in y, where $i = 1, 2, 3$, and so on. One distance we might consider is the sum of the absolute values of the differences between x_i and y_i:

$$\rho(x,y) = \sum |x_i - y_i|.$$

Suppose x and *y* are both of length N and *x* = 010101 ...0 and y = 101010...1. Then $\sum |x_i - y_i| = N$ since they differ in every place. This is one distance used by mathematicians. Another is the so-called Euclidean distance, $\sqrt{\sum (x_i - y_i)^2}$.

But our contention is that these distances are not suitable for biological objects. They are suitable for fixed objects, for sequences of symbols that are, so to say, rigid points of geometrical spaces, et cetera. But they are not well suited for flexible objects, such as strings of codes. To see this, consider the previous example of the two long sequences that differed in every place. They are in one sense almost identical since by erasing one symbol in each sequence they become the same. Two changes make the sequences identical! But according to the previous definitions of distance, the distance between them is N or $\sqrt{N}$ instead of just 2.

Let us try another definition of a distance. For example, we could define the new *p* as the minimum number of allowed changes that must be effected on one or the other sequence to make them identical. What could these allowed changes be? One might be the substitution of a O for a 1, or vice versa. Another could be the erasure, or the intercalation, of a O or a 1 at any place in the sequence. One can prove that this *p* has all the properties that a distance should.

A quantitative formulation of distance can be tried not only for the sequences of symbols in the genetic code but for a great variety of other objects. For example, one can try to define a distance between two sequences of musical notes, of acoustic signals, or between two drawings or sculptures, sets of points in two or three dimensions.

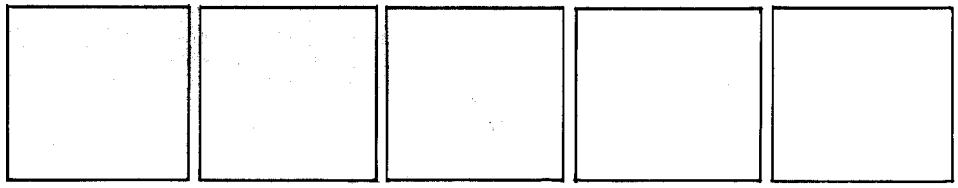
It is my speculation that in the brain, or more generally in the nervous system, there must be a mechanism that, perhaps in a qualitative way only, determines a distance between a perception stored in the memory and a newly presented perception. Recognition of the newly presented perception as known or unknown might mean that this distance is below or above a certain threshold. A perception insufficiently close to any of those already in the memory would be stored as a new perception.

I want to talk about this sort of approach to the recognition of visual perceptions. Let us take, for example, the case of recognition of two-dimensional pictures. My conjecture is that the brain uses several different distances to compare such pictures after they are registered on the retina, recorded or recoded on several layers behind the retina, and deposited in the brain.

What distance might be appropriate for comparing two two-dimensional pictures, that is, two sets of points in a plane? Let the two sets be *A* and *B*. We are interested in some possible *p(A, B).* Distances between sets have been studied by mathematicians. One of these, the Hausdorff distance $p_H(A, B)$ *is* defined as follows. Let ρ_E be the ordinary distance between two points. Given a point *x* in *A,* find the point y in *B* for which $\rho_E(x,y)$ is a minimum; that is, find $\min_{y \in B} \rho_E(x,y)$. Do this for all *x* in *A* and then find the maximum of these minima, $\max_{x \in A} \min_{y \in B} \rho_E(x,y)$. Now find $\min_{x \in A} \rho_E(x,y)$ for a given *y* in *B* and $\max_{y \in B} \min_{x \in A} \rho_E(x,y)$. Then

$$\rho_H(A,B) = \max_{x \in A} \min_{y \in B} \rho_E(x,y) + \max_{y \in B} \min_{x \in A} \rho_E(x,y).$$

But this Hausdorff distance, like some of the distances mentioned in connection with one-dimensional sequences, can be objected to in biological applications. Obviously, *pH,* as defined, depends on aspects of *A* and *B* that are of little con-



sequence to recognition. For example, *B* may be just a magnified version of *A* or congruent to *A* but rotated or translated. In these cases the meaningful distance should be very small.

By repeating, or iterating, the idea of Hausdorff as follows one can arrive at a more satisfactory distance. For a given set, or picture, *A*, let us consider the class of *A's* that "look like" *A*, that, for example, are replications of *A* in various sizes or are obtained from *A* by some rotation or translation. Call this class of sets an impression of *A* and denote it by *A*. We proceed analogously for *B* and obtain a class of sets *B*. Now we may define a distance between the impressions of *A* and *B* as follows:

$$\rho(\mathcal{A},\mathcal{B}) = \max_{A\in\mathcal{A}} \min_{B\in\mathcal{B}} \rho_{\mathrm{H}}(A,B) + \max_{B\in\mathcal{B}} \min_{A\in\mathcal{A}} \rho_{\mathrm{H}}(A,B).$$

This is a more satisfactory measure of the difference between *A* and B. Needless to say, distances between three-dimensional objects can be defined analogously.

One can define still other distances between sets of points. or signals, in two or mm-e dimensions. It is possible, for example, to express such a measure of similarity or dissimilarity as a distance between encodings of the set points in terms of orthogonal functions, such as those used in Fourier expansions. [See "An Ulam Distance."]

I shall now describe a computer experiment Robert Schrandt and I did in the early sixties at Los Alamos. The experiment concerned the use of distances in the recognition of handwritten letters and involved the second conjecture that I want to present in this talk, one about the role of impressions, or examples, in the process of recognition.

The idea of the experiment was to provide the computer with a great many handwritten examples of the letters *a* and h—actually with a great many sets of coordinates of points outlining the letters—and then make the computer decide if a new example was an *a* or a *b*. It would have been prohibitively tedious to provide, say, 512 examples of each letter. (Powers of 2 are convenient when dealing with computers; hence the number 5 12.) Instead we used a stratagem by which the computer itself generated the examples. I remembered a proof of mine that there exist on the interval, and analogously on spaces of higher dimension, two functions *f* and *g* such that any continuous function can be approximated by one of their compositions—fg, ffg, fgf, $fggf$, $fgfg$, et cetera. So we gave the computer only one example each of *a* and *b* and also two transformations of each, which served as *f* and *g*. By programming the computer to produce compositions up to the order of 10 of the transformations, we obtained 512 examples each of a and *b*. When displayed on a screen, these looked indeed like various handwritten versions of the original ***a*** and *b*. Some were slightly tilted, others appeared to have been written by a shaky hand, and so on. Then the computer was asked to decide whether a new handwritten sample was an ***a*** or a *b* by computing the Hausdorff distances between the sample and the examples it had created. The computer's decisions were correct in more than 80 percent of the cases! Of course, the same method works in the case of more than two letters or other standardized figures.

The conjecture is that in the brain, in the visual system and in the memory, perhaps only a few visual perceptions are permanently stored, and, when presented with another, the brain produces. for comparison, many deformations either of what is in the memory or of what is presented. If this is so, the storage capacity of the memory would be enormously enhanced.

At present one can only speculate about the mechanisms by which the brain might produce the deformations. Some are obvious, such as a tilt of the head or a change in size. One can also only speculate about what distances or how many are used in the decision. One may also speculate that a similar mechanism directs the recognition of objects within the body. Could it be that the antibodies produced by the immune system have an analogous way of recognizing antigens? Again, deformations might be used to produce a large number of examples for such discrimination and recognition.

The next higher stage in the operation of the brain might be a more complicated analysis of impressions. Instead of considering impressions of single objects, the brain might study a succession of two or three, even a "movie" of ten or more. Combined with recognition of the passage of time, this could lead to development of primitive logic or elementary reasoning, perhaps in the form of the statement *post hoc ergo propter hoc* (after, therefore because) or its reverse *ante* ***hoc ergo qua hoc*** (before, therefore as a reason for).

Our comprehension of less elementary learning should involve the mathematical idea of measuring complexity. In recent years quite a number of mathematicians, including Jan Mycielski and Andre Ehrenfeucht, both professors at this university, have done some very interesting work on this subject. With proper changes some of their results could be applied to investigating the operation of the nervous system.

It is clear that one of the most important mysteries about the brain is the organization of the memory, including the means of access. As I surmised earlier, some form of memory must exist in the visual, auditory, olfactory, and immune systems-and even in the system for differentiation itself. A mechanism for producing many examples from one would certainly seem a very efficient way of using the storage capacity of the visual and auditory memories. In the course of evolution, special devices, or tricks, must have developed to increase the scope of recognition and of the complementary process of registering perceptions as new.

Let me give an example of a trick for efficient use of a computer. Suppose we have stored in its memory a great many, say 10^6, eight-digit numbers arranged sequentially and want the computer to decide whether a given number is among those stored. The computer can do this extremely fast by comparing in succession the digits from first to last. Suppose now that we want the computer to decide whether the given number differs from any of the stored numbers by, say, 1 in any of the eight positions. We might program the computer to do this by deciding whether any of the 106 numbers in its memory is that close. That would be a very lengthy operation. There is a much better way to proceed, a way that requires only sixteen times the effort required for the computer to decide whether a single number is among those stored. We first program the computer to produce from the given number the sixteen numbers that do differ by 1 in any of the eight positions and then to decide whether any of the sixteen is among those in its memory.

This example illustrates that a mechanism for producing auxiliary perceptions for comparison with perceptions stored in the memory would be an advantageous acquisition of the nervous system. So also would a mechanism for producing variations of what is stored in the memory for comparison with external stimuli. Perhaps a physiological or anatomical arrangement might serve such functions. Clearly these are merely guesses as to special characteristics the nervous system may have acquired in the course of evolution. ■

An Ulam Distance

by Willium A. Beyer

Stan had often referred, as he did in this lecture, to a distance between sets based on an encoding of the set points in terms of orthogonal functions. However, he had never explicitly defined such a distance. I do so now to honor the originator of so many seminal ideas.

Let A and B be two-dimensional finite sets enclosed in a square. Let n_A and n_B be the number of points in A and B, respectively. Let $\{f_{i,j}\}$ be a complete set of orthogonal functions on the square, such as two-dimensional Fourier trigonometric functions. Define $\mu^A_{i,j}$ and $\mu^B_{i,j}$, the encodings of A and B mentioned above, as follows:

$$\mu^A_{i,j} = \frac{1}{n_A} \sum_{x \in A} f_{i,j}(x)$$

and

$$\mu^B_{i,j} = \frac{1}{n_B} \sum_{x \in B} f_{i,j}(x).$$

Then $\mu^A_{i,j}$ and $\mu^B_{i,j}$ are functions on the nonnegative lattice points of the plane. Finally, let $\rho(f_1, f_2)$ be some selected distance between such functions. Then $\rho(\mu^A_{i,j}, \mu^B_{i,j})$ is a distance between the sets A and B—*an* alternative to the Hausdorff distance defined in the lecture. ■