31 The War Years

1940–1945

I came back to M.I.T. in the fall of 1939 and took stock of the situation. It was not hopeful. The passive alliance between Russia and Germany had dashed cold water on our wishful thinking that the Nazis were going to be held back in the east, and although the two countries finally fought each other, we did not at that time expect such luck. Moreover, the good will that had been building up for Russia, largely because we could scarcely see from what other direction a blow might come which could limit the Fascist aggression, was much dampened by Russia's aggressive policy in Finland.

In academic and technical circles, most of us were aware that a world war would ultimately engulf the United States as well as all other important countries. Therefore we began to turn our thoughts toward finding out in what sector of work we might make ourselves of use.

The possibilities of active military service had never been very real to me on account of my nearsightedness, and the passing of the years had not increased my physical fitness. I have never fancied myself as an administrator, nor, as a matter of fact, has anybody else ever fancied me in that capacity. It seemed obvious that I should have to turn to some sort of scientific research work

I had had an apprenticeship in ballistic computation in World War I. Ballistic computation consists in the making of tables for artillery and small-arms fire which give the range of the weapon and various other related constants in terms of the angle of elevation of the gun, the powder charge, the weight of the missile, and so on. This work had trained me in much more than purely ballistic matters, for it had left me with a pretty general sophistication in the ways of the computing room. I had also spent much time in recent years working with electrical engineers. Therefore I foresaw that my destined berth in the war would be some sort of job in which I should apply computational techniques to electrical engineering problems.

Furthermore, my work with Lee had given me a look-in on problems of engineering design.

All this was clear, but what was not clear was the direction from which the call would come. As the nerve-wracking wait of the "Sitzkrieg" began to give way to a more active and threatening military schedule, most of us came to see that the main problem before America would be to keep England in the war as an effective combatant until such time as we might enter it ourselves. This meant that the submarine and the bomber campaigns were the two chief menaces which we should help to conquer.

Luckily, England herself had given us the best possible lead for helping her in these fields, in the brilliant invention of radar. M.I.T. was pushing this sort of research from the very beginning, even before the start of the European war and long before we got into it. But for the time being this seemed to be a matter for specialists, and I was not a radar specialist.

The stream of refugees from Germany speeded up for the moment and then ceased altogether. These last driblets of immigration did not seem to me to consist altogether of persons of the same moral value as some of those who had come before. More than one of these last drippings of the wine press showed an eagerness to indoctrinate us in the irresistible momentum of the Nazi advance. Their zeal could scarcely have been exceeded if they had been paid propagandists. At last it became quite clear to us that in addition to the great cultural crop of fine, persecuted men and women who have enriched our intellectual life, there were those whose main objection to Nazism was that they were excluded from it. However, the new summer vacation came along in its good time, and we tried to make our life as cheerful as possible, notwithstanding the catastrophe about us. One cannot live in a perpetual atmosphere of gloom.

The Inghams, from the English Cambridge, were caught in America by the war and became our summer neighbors. They shared with us the pleasures of mountain walks and bathing in Bear Camp pond.

We had a curiously interesting visit that summer from the Hungarian mathematician Erdös, the Japanese mathematician Kakutani, and the English mathematician Stone. They had just got into trouble on Long Island by inadvertently approaching too close to a radio-beacon station. They were put in jail overnight as suspicious aliens but were released later when the authorities were able to reach their sponsor, Professor Veblen, of Princeton. It was just after this contretemps that they drove up to New Hampshire, and we had a most pleasant little session on our porch. At present, Kakutani is teaching in the United States, but Stone and Erdös have gone back to Europe.

At the end of the summer, Ingham returned to England as he had planned, but his wife and children and their maid stayed again in our valley the next year. Again we went for long walks together, the children now participating in them more easily. I have seen the family several times since their return to England, where I believe one boy is going to the university and the other is an officer in the Air Force. They still retain a real love for New Hampshire and our valley.

Wintner remained our summer neighbor. He and I had planned to work together during the year 1940–1941, and he came to Cambridge to do so. It was unfortunate that my attention was concentrated on war work at the time. I feel that I was to a certain extent unjust to Wintner in not living up to our informal contract, for he found himself able to ignore the pressures of the warlike atmosphere. I could not, and though I was willing to work with him with a part of my attention, I was unable to devote to it my full interest. Thus we went our two ways, which gradually drifted farther and farther apart.

By spring, the catastrophe of Norway had occurred and the catastrophe of France was threatening. The emotional solace of our country house, in which we had become used to take refuge from the buffets of the outer world, was of no avail to us when we were confronted by the imminent loss of European civilization.

In August 1940, the summer meeting of the American Mathematical Society was at Dartmouth. It was as pleasant as a meeting could be when nothing but the war could really command our attention.

The algebra of complex quantities is vital for telephone engineering, and the Bell Telephone Company instrument, a numerical computer, was made to fill a definite need in such work. Its importance derives from the fact that our Arabic notation for numbers gives to the number 10 an artificial position justified by custom alone and constituting no part of the real foundations of arithmetic. Instead of writing a number as so many units, so many tens, so many hundreds, and so on, we can just as easily write an integer as a sum of ones, twos, fours, eights, and so on. In this case, instead of the ten digits of our conventional scale of numeration, we need only the two digits, zero and one.

The Russian peasants use what amounts to a scale of this sort, known as the binary scale, for their addition, multiplication, subtraction, and division. It has the great advantage over the scale of ten that the multiplication table reduces to the statement that one times one is one.

For obvious reasons, it is easier to mechanize arithmetic on the binary scale than on the scale of ten, and the Bell System instrument accordingly employed the binary notation. The sole serious disadvantage in doing all arithmetic on the binary scale is simply that we have adopted the decimal system and the bulk of existing numerical results are given in this tradition. When we have a large amount of new computation to do, it is often worthwhile to ignore this fact and to translate all our initial data to the binary scale and all our final results back to the decimal scale.

One place where the binary system of numeration is used is in the employment of gauges for the measurement of the thickness of a mechanical part. Suppose that we have one accurate gauge of a thickness of one inch; one accurate gauge of a thickness of two inches; one accurate gauge of a thickness of four inches; and one of a thickness of eight inches. Then we can combine these to give accurate measurements by inches from one inch to fifteen inches. The code is the following: we combine our gauges on top of one another in these combinations:

1 inch	1 inch gauge
2 in.	2 in. gauge
3 in.	2 in. gauge and 1 in. gauge
4 in.	4 in. gauge
5 in.	4 in. gauge, and 1 in. gauge
6 in.	4 in. gauge and 2 in. gauge
7 in.	4 in. gauge, 2 in. gauge, and 1 in. gauge
8 in.	8 in. gauge
9 in.	8 in. gauge and 1 in. gauge
10 in.	8 in. gauge and 2 in. gauge
11 in.	8 in. gauge, 2 in. gauge, and 1 in. gauge
12 in.	8 in. gauge and 4 in. gauge
13 in.	8 in. gauge, 4 in. gauge, and 1 in. gauge
14 in.	8 in. gauge, 4 in. gauge, and 2 in. gauge
15 in.	8 in. gauge, 4 in. gauge, 2 in. gauge, and 1 in. gauge

This is equivalent to writing the numbers 1 to 15 in the following way: 1, 10, 11, 100, 101, 110, 111, 1000, 1001, 1010, 1011, 1100, 1101, and 1111.

I do not remember whether it was before or after the Dartmouth meeting that Vannevar Bush had sent around to the various members of the M.I.T. faculty a questionnaire asking for suggestions concerning the mobilization and use of scientists in case of war. This was a matter on which I had very definite opinions, and I was strongly in favor of a scientific collaboration which would cross the frontiers between one science and another, and which should at the same time be voluntary, thus preserving a large measure of the scientists' initiative and individual responsibility. I distrusted all plans that might depend on a high degree of subordination of individuals to a completely authoritative setup from above, which would assign each man the narrow frame within which he was to work. I suggested therefore the organization of small mobile teams of scientists from different fields, which would make joint attacks on their problems. When they had accomplished something, I planned that they should pass their work over to a development group and go on in a body to the next problem on the basis of the scientific experience and the experience in collaboration which they had already acquired.

But nothing resulted. Those who work almost exclusively with gadgets tend to develop a love for them, since they lack the unpredictable factors which affect the operation of the human being.

Gadgeteering very easily becomes a sort of religion. Luckily, the vicissitudes of the last twenty years of gadgeteering have somewhat shaken the faith of many men, including Bush, in the unlimited scope of the machine. However, there remain many people who have not been as directly confronted as Bush with the disadvantages as well as the advantages of machines, and these people follow the tendency of the day to favor the big laboratory and the big administrator.

On the way back from the meeting, I began to discuss with Levinson, who was now a full-fledged colleague, the general problem of computing machines, and to wonder whether this might not be the field in which I was destined to do my war work. For some time I had been considering for Bush the use of machines to solve systems of partial differential equations and I felt that a sort of television scanning would be the proper basis for the mechanization of partial differential equation problems. My new experience with the binary machine convinced me that electronic binary machines would be precisely the devices required for the high speed of computation needed in partial-differential-equation problems.

I saw that for a machine to work properly on partial differential equations, it would have to go through an almost incredible bulk of work in an almost incredibly short time. This suggested to me that the future of high-speed computing machines for these particular purposes could not lie in Bush's models, which represented physical quantities by electrical or mechanical quantities, but rather in some enormous extension of the ordinary desk computer, working as I have said, on a scale of two rather than on a scale of ten.

Now that I had come to take a serious interest in problems of the speed of computation, I was forced to consider the relative merits of two grand strategies in computational methods. One of these, which Bush followed, was called analogy computation, in which the numerical digits of computation are represented as measurable physical quantities. The other, the digital mode of computation—which is that of the desk computing machine—represents a number by the sequence of its digits.

The important point in the distinction between the analogy computers and digital computers is that the latter do essentially what we ourselves do with arithmetical operations on paper. When we represent a number as 56, we mean that it is a combination of five tens and six units. When we multiply this by 38, which is three tens and eight units, we go through the operations indicated in the following table:

$$\begin{array}{r} 56 \\ \underline{38} \\ 48 \\ 40 \\ 18 \\ \underline{15} \\ 2128 \end{array}$$

We never have to go beyond the multiplication table and the simple rules of addition, nor do we represent our 56 and our 38 as quantities such as 56 degrees or 38 inches.

There are digital multiplying machines in which ten plays no role, which operate on the binary scale. This is how they work: Let me consider the operation in which $7 \times 5 = 35\cdot$

$7 = 4 + 2 + 1$

$5 = 4 + 1$

In the scale of 2, these statements are equivalent to writing 7 in the form 111, and 5 in the form 101—meaning that 7 is one 4, one 2, and one 1; while 5 is one 4, no 2's, and one 1. When we multiply these we get the following schedule of operations:

$$\begin{array}{r} 111 \\ \underline{101} \\ 111 \\ \underline{111} \\ 11211 \end{array}$$

If we remember that, in our scale of notations, 2 = 10, the number 11211 may be written 12011 or 20011 or 100011. The last representation is the truly binary one, which uses no digits other than 0 and 1. Here this means 32 + 0(16) + 0(8) + 0(4) + 2 + 1 = 35. This method is called the method of multiplying on the scale of two. I repeat, it is quite as digital as ordinary multiplication on the scale of ten.

In contrast, a particular analogy-computation instrument proceeds as follows: In an electrodynamometer, two coils are attracted to one another in proportion to the product of the current carried by these two coils, and this attraction can be measured by an appropriate sort of scale instrument. If, then, one coil is carrying seven units of current and the other five units, the scale will read something proportional to thirty-five. This sort of instrument for multiplying is known as an analogy instrument, because we are replacing our original situation, in which certain quantities are to be multiplied, by a new situation, in which we set up two currents in analogy to the original quantities and read off the product by a physical situation analogous to the original one.

Digital computing machines thus differ from analogical computing machines in that they can theoretically be read to the complete degree of precision of the numbers introduced, while analogical machines are restricted to the degree of accuracy with which the original situation is truly analogous to the corresponding situation, which we use to replace it in our computation. The Bush machines for solving differential equations are strictly analogy machines.

As to the relative merits of these machines, the great flexibility of electrical and other measuring apparatus makes the construction of a fairly good analogy machine easier than that of a fairly good digital machine. However, when it comes to high speed or high accuracy, the advantage is all with the digital machine. There are very few physical measurements which can be made with an accuracy greater than one part in ten thousand, and this corresponds to a determination of four decimal digits and not quite fourteen binary digits.

Moreover, to take a measurement with this degree of accuracy can scarcely be a truly instantaneous process. Analogy machines are intrinsically slower than is needed to fulfill the demands of the very fastest and bulkiest computation, so that I felt that they had already reached their apogee.

When it came to the matter of digital machines, I was forced to consider the real essence of the action of such a machine. The ordinary desk computer determines the position of certain wheels on the basis of the position of certain other wheels. Each such position is a choice between ten alternatives. It is not difficult to represent these ten alternatives by ten projections on a metal wheel, but the use of metal wheels involves very disagreeable and restrictive problems of inertia and friction.

It seemed to be preferable in every way to replace the mechanical choice characteristic of existing digital machines by an electronic choice of digits. The two advantages to be expected here were the vastly reduced inertia of a stream of electrons as compared with a sequence of mechanical parts, and the greater technical ease of canceling quasi-frictional losses—that is, resistance losses—by amplification. As a result of both of these, I was quite certain that the coming high-speed computing machines would be electronic digital machines. I may say that the idea had begun to come up in various places in the literature and that, in accepting this approach to computing machines, I was simply representing the spirit of the times.

As I have said, a decimal digital machine uses a choice among ten possibilities as its fundamental decision while a binary machine uses a choice between two alternatives. I suppose that the general use of the scale of ten came from the ten digits of our two hands. Certain races such as the Mayas have apparently counted on fingers and toes together, and use the scale of twenty. It is an interesting reflection that if the human race had been constructed after the pattern of the Walt Disney cartoons, with four digits on each hand, we should probably have adopted the scale of eight, which is merely a slight variation of the scale of two, since $2 \times 2 \times 2$ is 8.

However, there is a fortuitous advantage that, while it does not recommend the scale of ten, at least makes it easier to use it than, let us say, the scale of thirteen. Computing machines of the decimal type depend on the use of wheels with ten equally spaced teeth. To construct them involves laying out a decagon—a polygon with ten angles. This is a simple problem in plane geometry, while the construction of a regular figure with thirteen sides is not.

In the case of an electronic circuit, however, the parts which are equivalent to wheels are not dependent on ordinary plane geometry, nor is ten a particularly easy number to represent. The natural choices to be made in an electric circuit will be between pairs of alternatives.

There are well-known circuits already existing with two alternative positions of equilibrium, and these are known as flip-flop circuits. About the only easy way of constructing a circuit with ten choices seemed to be by a combination of flip-flop circuits. The logic of flip-flop circuits would lead to a combination of choices among a number of alternatives—a combination which was a power of two. Thus, it seems that the only natural way to construct a set of ten alternatives is to construct sixteen and throw away six.

In the design of a machine, you pay in effort and in cost not only for everything that the machine does but also for everything that the machine might be made to do; and to use sixteen alternatives for the purpose of selecting ten represents a wastefulness of 37 1/2%. I came to the conclusion that the high-speed computing machine for the solution of partial differential equations would be a digital electronic machine on the scale of two.

For work on a scale of two, we need to use machines which have only two possible alternatives, such as the presence or absence of a punched hole in a card. This device was already familiar in the Hollerith machines made by the International Business Machines Corporation. This particular way of writing a number in the scale of two is, however, unsuitable for a thoroughly high-speed machine. Punching a hole in a piece of cardboard is a slow process when one considers speed on the scale of millionths of a second per operation. Some such scale is needed before we can consider a computation as really high-speed, and, moreover, the problem of disposing of the used cards and of keeping a sufficiently large stock of new cards soon becomes almost astronomical.

The speed of punching the paper might be greatly increased by using an electric spark rather than a mechanical punch, but this would leave the problem of the bulk of material just as bad as it ever was. Thus, I was naturally driven to the idea of steel tape on which a magnetic mark would be made by an electromagnet. One can read such a mark at high speed and erase it at high speed, leaving the tape blank and ready for another use.

One of the main problems with such a tape is to make the marks so small that as many as possible can be kept clearly distinguishable on a given area. This requires extremely small pole pieces for the marking and reading magnets. It seemed to me that the smallness of these pole pieces might be vitiated by the spread of the magnetic field within the tape, unless the tape itself or at least its effective magnetic layer were extremely thin.

I therefore had the idea, partly on my own and partly through conversation with those of my colleagues who were better acquainted with technical developments in this subject than I was, that the best thing to do would be to treat separately the two requirements of the tape: to carry magnetization and to hold together. We might do this by imposing a thin magnetic layer on a non-magnetic material which would have all the strength. I had thought of a thin iron layer superimposed on a tape of brass or non-magnetic metal, but I had also thought—I believe through the suggestion of a colleague—of the device which now dominates the field: of a paper tape carrying a thin layer of magnetic oxide of iron.

I have recently talked with a friend at the IBM company concerning present practices in high-speed computing machines, and in particular in those which work according to what is now known as the Monte Carlo method, solving partial differential equations by an extremely often repeated process of averaging. Apparently the devices I suggested in 1940 are substantially those which are now employed.

The odds in a gambling house are extremely regular and predictable, and the Monte Carlo method consists in setting up a mathematical problem as an ideal game, in playing it out a large number of times, and in determining the theoretical winnings. The computational device which I suggested in 1940 had the same non-static character as the Monte Carlo method of the present day and also depended on the playing out of an ideal game.

I made a report of my suggestion to Vannevar Bush, but I did not get a very favorable reception. Bush recognized that there were possibilities in my idea, but he considered them too far in the future to have any relevance to World War II. He encouraged me to think of these ideas after the war, and meanwhile to devote my attention to things of more immediate practical use.

Later on I found that he had no very high opinion of the apparatus I had suggested, especially because I was not an engineer and had never put any two parts of it together. His estimate of any work which did not reach the level of actual construction was extremely low. The only satisfaction I can now get is that I was right something like ten years before the techniques to prove my ideas were developed.

Having discarded this as my task in the threatening war, I began to look around for other places where I might be more useful. At one time I had ideas concerning a mathematical and mechanical method for encoding and decoding messages. My ideas would certainly have worked, but in this field it is not enough for an idea to work. In fact, for it to have any merit at all, it must work considerably better than existing devices, or any devices easily invented.

The question of what it means for one decoding-coding device to work better than another is not simple. It can be taken for granted that any ciphered text which is sufficiently long can be decoded by a possible enemy if you give him enough time, and it also must be considered that the problem of decoding a cipher message is not necessarily trivial even when the cipher is known. A good cipher must combine a certain ease of decoding by machine, or by a recipient in the know, with a large measure of difficulty for an enemy to decode it without the help of a knowledge of the cipher.

As is usually the case when there are two different requirements for a system or an apparatus, this does not lead one to a single apparatus but to a number that is large according to the weight one puts on each of the two requirements. Thus, there are easy ciphers which will be useful in the field for messages which need only an hour's secrecy, and there will be difficult ciphers for messages whose secrecy must be maintained for months. There will be a whole spectrum of ciphers in between. For this reason, the design of ciphers is not a matter into which one can go cold, without knowledge of the existing tradition and of the practical demands for each particular use. Again I had to look around for another possible field of usefulness. I found this in the design of fire-control apparatus for anti-aircraft guns.

When I was a boy, fire control was conceived primarily from the point of view of coastal batteries and of battleships—that is, of gun platforms whose motion relative to the target was so slow that there was a chance for a considerable amount of computation by very crude manual devices before the target should have passed out of effective range.

Even in the First World War, the airplane had changed all this. The problem of shooting down an airplane is not, of course, like that of lobbing a mortar shell into a fortress but like that of shooting ducks on the wing. The duck will not stay still while you are shooting; and if you aim your gun where you see the duck, the bird will be considerably ahead of that point by the time the shot arrives. You must shoot ahead of your target, and you must estimate that amount you shoot ahead quickly and accurately. If your estimate is not correct you will probably not have another chance at that bird.

The result is that from the very beginning it was necessary to build into the control system of the anti-aircraft gun some mechanical equivalent of a range table which would automatically allow the gun the necessary lead over the plane to make the shell and the plane come to the same place at the same time. To some extent this a purely geometrical problem, but in its finer developments it involves an improvement of our estimate of the future position of the plane itself. This must be estimated from the past positions, or at any rate from the observed past positions. The problem of predicting the future position of the plane is what the mathematicians call a problem of extrapolation.

My previous electrical engineering work had made me familiar with the theory of *operators*. An operator in this sense represents a device which will change a certain electric input into a corresponding electric output.

Mathematically speaking, the operator may be represented by a transformation formula, but not all such transformation formulas lead to operators which are physically realizable. The main condition for physical reality which we must impose on an operator is that the output should involve only the past and present of the input. It will be seen that the problem of shooting ahead of an airplane demands that the realizable operator approximate the future position of the plane which could, in fact, be ascertained only by a non-realizable operator. Only a prophet with the knowledge of the mind of the aviator could predict the future position of an airplane with absolute certainty, but there are often enough, in fact, means which will allow one to accomplish the minor task of a quite correct prediction.

The mathematical processes which suggested themselves to me in the first instance for prediction were, in fact, impossible of execution, for they assumed an already existing knowledge of the future. However, I was able to show there was a certain sense in which these processes might be approximated by processes free from this objection.

I do not wish to lose myself here in technicalities which will be understood only by scientists or engineers. However, I did in fact consider certain possibilities of approximating to non-realizable operators by realizable operators. I suggested these notions to Professor Caldwell, who was normally in charge of M.I.T. work on Bush computing machines, and who was now engaged in applying these machines to war problems. After the custom of those times, Caldwell immediately put a classification on my ideas, so that thereafter I could no longer speak freely of them to anyone with whom I wished to talk.

For a trial setup for my problem, Caldwell and I were tempted to make use of Bush's differential analyzer because of the ease of assembling its parts for the simulation of a large range of different problems. In this the differential analyzer resembled a Meccano set; and in fact, when the English tried to follow in Bush's footsteps and construct a differential analyzer, they used ordinary Meccano parts with very creditable success.

We made several experimental runs with different settings of our apparatus, and we found that those which we had considered in advance to be better actually were. Our instrument was an assembly of a number of adding devices, multiplying devices, and integrating discs.

At this stage, prediction theory was made a government project, and a young engineer, Julian Bigelow, who had worked some time with International Business Machines, was assigned to the problem. This was the beginning of a long collaboration between us. Bigelow is a quiet, thorough New Englander, whose only scientific vice is an excess of scientific virtue.

He is a perfectionist, and no work that he has ever done is complete enough in his eyes to satisfy him.

He used to be an enthusiastic aviator, but this sport became impossible in war time and is, at any rate, too expensive for the average man. Most accidents of private flying are not serious, in the sense that the aviator can walk away from them unhurt, but there are no minor repairs to an airplane. They must be done by certified mechanics and have the O.K. of the Civil Aeronautical authorities. Since they generally occur at remote points, they constitute a real burden to the pocket. For many years, Bigelow nursed a series of old and decrepit cars. For the ordinary automobilist a car is an instrument to get somewhere, but for the enthusiastic gadget man it represents a challenge to his ability to overcome difficulties. Such an engineer will never be content with a car that functions normally. He is either trying to construct a super car or he is exercising his ingenuity in making run a car which, by all the canons of the motorist, should have been consigned to the junk heap years ago. If you ride with such a motorist, you will be safe from accidents of any consequence, but you will never, never travel without adventure. I remember one occasion when Von Neumann was interested in consulting with Bigelow as a possible engineer for a computing machine project. We telephoned from Princeton to New York, and Bigelow agreed to come down in his car. We waited till the appointed hour and no Bigelow was there. He hadn't come an hour later. Just as we were about to give up hope, we heard the puffing of a very decrepit vehicle. It was on the last possible explosion of a cylinder that he finally turned up with a car that would have died months ago in the hands of anything but so competent an engineer.

Bigelow and I began to try to ascertain the limitations of our prediction method, for it was almost certain that we should find serious limitations. This time, instead of trying out our predictor on a smooth curve, we tried it on graphs made of two straight lines joining each other at an angle.

It must be understood that the predictor consisted of one member which was made to follow a given curve and of another member which, on the basis of these past data, was expected to indicate the curve a little further in the future. This second member we will term the follower. When we put our apparatus to work on a curve which was not smooth, but in which one straight segment of a line was succeeded by another segment at an angle with the first, the predictor still worked, but in a very peculiar fashion.

What was interesting and exciting, and in fact not unexpected, was that the pieces of apparatus designed for best following a smooth curve were oversensitive and were driven into violent oscillation by a corner. We tried to test this repeatedly, and we always got the same result. Then the idea suggested itself to me: perhaps this difficulty is in the order of things, and there is no way in which I can overcome it. Perhaps it belongs to the nature of prediction that an accurate apparatus for smooth curves is an excessively sensitive apparatus for rough curves. Perhaps we have here the example of the same sort of malice of nature which appears in Heisenberg's principle, which forbids us to say precisely and simultaneously both where a particle is and how fast it is going.

The more we studied the problem, the more we became convinced that we were right and that the difficulty was fundamental. If then we could not do what we had wanted but had not really hoped to do (that is, to develop a perfect universal predictor), we should have to cut our clothes to fit our cloth and develop the best predictor that mathematics allowed us to. The only question was: What did we mean by the best predictor? If errors of inaccuracy and errors of hypersensitivity always seemed to be in opposite directions, on what basis could we make a compromise between those two errors?

The answer was that we could make such a compromise only on a statistical basis. For the actual distribution of curves which we wanted to predict, or let us say for the actual distribution of airplanes that we wanted to shoot down, we might seek a prediction making some quantity a minimum; and the most natural quantity to choose at the start, if we should be guided by easy computation, if not military significance, was the mean square error of prediction.

This means that we took the square of the error of prediction at each time, or in other words, the square of the difference between the predicted value and the true value. We then took the average of this over the whole time of the running of the apparatus. This average of the square error was what we were trying to minimize.

Thus we could set up the prediction problem as a minimization problem and give it a definite mathematical form once we made certain assumptions concerning the statistics of curves to be predicted. The branch of mathematics dealing with the minimization of quantities associated with curves is known as the calculus of variations, and it has a very well-known and well-recognized technique. In many cases it leads to setting up a certain differential equation for the function or curve which is to fulfill the minimization condition, but there are cases (and this case of ours was one of them) where it leads to the related sort of equation known as an integral equation.

This was lucky for me, for integral equations were well within my field of interest; but the even luckier thing was that the particular integral equation to which the problem leads is a slight extension of the one which had been considered by Eberhard Hopf and myself. The result was that not only was I able to formulate the prediction problem but also to solve it; and what was even luckier was the fact that the solution came out in a simple form. It was not hard to devise apparatus to realize in the metal what we had figured out on paper. All that we had to do was make a quite simple assembly of electric inductances, voltage resistances, and capacities, acting on a small electric motor of the sort which you can buy from any instrument company.

We made an apparatus which translated the height of a point above a given base line into an electrical voltage. We passed this voltage, which varied with the time, through an electrical combination of resistance wires, condensers, and magnetic coils. At another point in the system, we took off the voltage and measured it continuously by a voltmeter. The actual type of voltmeter which we used gave us a continuous graph of the output voltage. It was this output which was to serve as a prediction of the voltage a certain length of time in the future.

The next problem which I had to take up concerned prediction in the case in which the data from which we were to predict were not precisely given. This also led to a minimization problem, in which we had to specify not only the statistics of the data which we were supplied but also the statistics of the errors at the same time. This minimization problem led to another Hopf-Wiener equation. This was solvable by the same methods, and we obtained a very satisfactory theory.

In scientific work it is not enough to be able to solve one's problems. One must also turn these problems around and find out what problems one has solved. It is frequently the case that, in solving a problem, one has automatically given the answer to another, which one has not even considered in the same connection.

This proved to be true in the new prediction theory. The concept of predicting the future of a message with a disturbing noise on the basis of the simultaneous statistics of the noise and message turned out to contain in itself the whole idea of a new method for separating noises and messages in what would be in some sense the best possible way.

This happened at a very opportune time, for the new technique of radar had come to meet serious difficulties. In radar too it was important to pick out a confused and faint message from the background of noise. Noise to the electrical engineer means not only noise that you hear but any unwanted electrical disturbance. For example, the flutter and flicker which you see in a badly regulated television set is noise. The messages which come in through a piece of radar apparatus and serve to confuse the image for which we are looking instead of to define it are known as noise.

The separation of noise from messages is the function of a wave filter. Wave filters go back many years in the history of telephone engineering and are pieces of apparatus which manage to free a message from part of the accompanying noise. Originally they were designed to pass all messages over a certain range of frequencies (or pitches) with as little change in their intensity as possible and to weaken other immediately adjacent frequencies by as much as possible.

When, with the precedent of telephone filters, such filters were built for television, and it was found that, after a certain point, the sharper the design of the filter, the worse it would function. Why was this the case? The answer is that the telephone filter is tailored to the specific characteristics of the human ear. The human ear is a very accurate instrument for perceiving pitch and a fairly accurate one for perceiving loudness, but it is a very poor instrument for perceiving what is known as phase; or, in other words, the precise time that the oscillation of the air passes through zero. An alternating current is represented, as I have said, not by one quantity which gives it intensity, but by two, which give it intensity and phase. The picture of an alternating current somewhat resembles the teeth of a comb. As I move the comb forward and backward along its edge, I change a certain quantity known as phase. In sounds, this phase change is not completely imperceptible, but it is not particularly important, and the earlier filters for telephone work and for other sound work did not pay much attention to differences in phase.

Radar, like television, appeals to the eye; and in the sort of message that radar or television transmits, the eye is quite as sensitive to phase errors as to amplitude errors. Thus, the phase distortion which the old telephone-type filter generated in radar and television was too high a price to pay for the excellent transmission of amplitude over a large range of frequencies. To minimize the total error in television and radar, it was necessary to cut down the phase distortion, at the cost of permitting a little more amplitude distortion than would by itself be best. In seeking for this balance between the two distortions, the method which I had suggested—although not ideal—would at least work, and was far better than any which had previously been used.

I do not mean to say that others had not become aware of the failure of the earlier forms of filter design for radar nor that these other people had not understood what was in essence the reason for this failure, but simply that my method gave for the first time a simple, compact, and reasonable way of attacking the problem on the level of fundamentals.

Bigelow and I started up a little laboratory to explore the possibilities of predictors. We had a staff of two men working with us. One was an excellent machinist and electrician, who put our ideas into the metal almost as fast as we could conceive them. The other was a computer who had been an accountant. Reader, if you ever have to start a computing laboratory, be warned by me and do not take as a computer an accountant, no matter how honest and efficient. Your computer must work to so and so many places of accuracy. This means so and so many significant figures, whether the significance of the digits begins six places before or six places after the decimal point. Your accountant works to cents, and he will work to cents until hell freezes over. Whatever numbers our accountant computed he kept at all stages to exactly two places after the decimal point, whether they were numbers in the millions, where even the first place to the left of the decimal point was of no possible significance, or numbers which begin only five places after the decimal point.

This was his conscience, that he should be accurate to the last cent; and he simply could not understand that physical quantities are not measured in cents but on a sliding scale of values in which the cents of one problem might be the dollars of another. In particular, when he had to obtain a small number as the difference between two nearly equal large numbers, he could never realize that these large numbers would have to be measured to a much greater degree of accuracy than that which was to be demanded for their difference.

I took my responsibility in this project very seriously. I tried to work against time, and that is a thing for which I am completely unsuited. More than once I computed all through the night to meet some imaginary deadline which wasn't there. I was not fully aware of the dangers of Benzedrine, and I am afraid that I used it to the serious detriment of my health.

Be that as it may, I found one very disagreeable fact, that the burden of secrecy in my project weighed heavily on me and that Benzedrine plays hob with one's ability to keep a secret properly. It superimposed on my not very secretive nature a garrulity which was completely unfitting at the time. I had to give it up and look for a more rational way of strengthening myself to bear the burdens of war work.

My work was under the supervision of Dr. Warren Weaver, of the Rockefeller Institute. We made several trips, Bigelow and I, to consult with him, and to compare our ideas with those of other people working on prediction theory and on the smoothing of anti-aircraft data. We traveled south two or three times to Fort Monroe, in Virginia, and to an army camp on the North Carolina coast. There we found workers from the Bell Telephone Laboratory who were more than eager to exchange ideas with us, and we pooled our resources with them and with other workers in the same field. At these meetings, after our travel and our hard work, I am afraid that I often went to sleep.

When we got back home, we made an experimental setup for generating the sort of irregular functions which arise in the aircraft-prediction problem, and then we designed a prediction apparatus on the basis of the statistical observations which we made on this setup. We actually were able to construct a predictor which would show the shape of a pattern of voltage in time, let us say, half a second before it occurred. This allowed us to check on our theory and find the criterion for a piece of apparatus that would give us good prediction.

The problem of generating an irregular curve with a statistically controllable degree of irregularity was quite interesting. We had reflected on the ceiling a spot of light, moving in a more or less periodic course. We tried to follow this spot with another, reflected in a mirror which was controlled by a certain apparatus. In this apparatus, the actual motion of the spot was not proportional to the turning of the crank which regulated it but to a rather complicated mixture of derivatives and integrals of this motion. Moreover, the crank was attached to a system of weights and springs, which was very far from giving the kinesthetic sensations which one would naturally associate with an apparatus of the sort. In other words, the spot had to be moved by a control which was complicated to begin with and, furthermore, felt completely wrong. Naturally, each person would respond to the apparatus in a somewhat different way; and we based our set of the predictor not merely on the general behavior of the apparatus but on the specific ability of an individual person to control the apparatus at a certain specific epoch in his training.

We were gratified by our clear and consistent results. On one hand, we had made a mechanical setup which threw a great deal of light on the way in which we act when we are confronted with an artificial problem and on the nature of humanly caused irregular action. On the other hand, we had found a way to duplicate in some degree the properties of the type of irregular motion of an airplane in flight. Thus we had some hope of the theory which could be used for the design of a practical apparatus for bringing down airplanes.

The importance of our ideas in connection with the control of anti-aircraft fire was double. There are two human elements which must be considered in this control. On the one hand, when the airplane pilot is flying and taking evasive action, his pattern of flight has a great deal to do with not only the limitations of his plane but those of his nervous system, so that his action is not too different from that of the hypothetical human action we had designed. On the other hand, the anti-aircraft gunner uses a technique in which he cannot follow his target perfectly but in which he introduces certain random errors because of the limitations of his sense organs and muscles. These two sorts of human elements are combined as part of the semimechanical processes by which the anti-aircraft gunner brings down his target.

At the beginning of the war the only known method of tracking an airplane with an anti-aircraft gun was for the gunner to hold it in his sights by a humanly regulated process. Later on in the war, as radar became perfected, this process was mechanized. It became possible to couple directly to the gun the radar apparatus by which the plane is localized, and thus to eliminate the human element in gun pointing.

However, it does not seem even remotely possible to eliminate the human element as far as it shows itself in enemy behavior. Therefore, in order to obtain as complete a mathematical treatment as possible of the over-all control problem, it is necessary to assimilate the different parts of the system to a single basis, either human or mechanical. Since our understanding of the mechanical elements of gun pointing appeared to us to be far ahead of our psychological understanding, we chose to try to find a mechanical analogue of the gun pointer and the airplane pilot.

In both these cases, the operators seemed to regulate their conduct by observing the errors committed in a certain pattern of behavior and by opposing these errors by actions deliberately tending to reduce them. This method of control appeared to us not unlike a method already known in electric circuits and now being applied in servomechanisms, or systems by which we switch in an outside source of power for control purposes, such as occurs in the power steering of a truck. We call this negative feedback.

We use negative feedback in controlling the power input to the gun turret of a ship. When the direction in which the gun is pointing and the direction in which our computing apparatus says that the gun ought to point are different, this difference is used to regulate a power input to the turret, which will be of such a nature as to bring the turret more nearly into the desired position.

It is a maxim of the physiologist that the pathology of an organ throws a very great light on its normal behavior. We asked ourselves the question:

Does a negative-feedback apparatus have a specific recognizable pathology? Here we were on firm ground.

To see the general purpose of negative feedback apparatus, let us take the case of a gun turret controlled by a handle. If this handle works the gun turret directly, then the same pressure on the handle will produce very different results if the gun turret is cold and the grease is sticky from what it will produce if the gun turret is warm and the grease flows easily. It will produce different results if the gun is depressed, thereby increasing the moment of inertia of the gun turret, from those which we shall find if the gun is elevated and the turret has a small moment of inertia about a vertical axis. The primary purpose of feedback control on the gun turret is to make the response of the turret more nearly proportional to the push on the lever and thus less dependent on the variable friction, the inertia, and other external circumstances.

Not only is a feedback system less dependent on changes of load than a system without feedback, but this dependence becomes increasingly less as more and more of the motion is fed back—or, in other words, if the feedback is put through larger and larger amplification. However, this improvement in behavior does not continue indefinitely, for after a certain stage, with a large measure of amplification in the feedback, the apparatus will go into spontaneous oscillation and behave in such a wild way that we have decreased rather than increased the load independence of the apparatus. We expected that if human control also were to depend on feedback, there would be certain pathological conditions of very great feedback, under which the human system, instead of acting effectively as a control system, would go into wilder and wilder oscillations until it should break down or at least until its fundamental method of behavior should be greatly changed.

This suggestion, which emanated equally from Bigelow and from me, I brought to the attention of my neurophysiologist friend, Dr. Rosenblueth. He had not yet left for Mexico and was still Dr. Cannon's colleague at the Harvard Medical School. The specific question we put was: Are there any known nervous disorders in which the patient shows no tremor at rest, but in which the attempt to perform such an act as picking up a glass of water makes him swing wider and wider until the performance is frustrated, and (for example) the water is spilled?

Dr. Rosenblueth's answer was that such pathological conditions are well known, and are termed intention tremors; and that very often the seat of the disorder lies in the cerebellum, which controls our organized muscular activity and the level on which it takes place. Thus, our suspicions that feedback plays a large role in human control were confirmed by the well-established fact that the pathology of feedback bears a close resemblance to a recognized form of the pathology of orderly and organized human behavior.

Within the last two years I have had an experience which may be regarded as a commentary on the ideas I am putting forward here. Suddenly my little granddaughter, who was staying with us, developed a purpose tremor of exactly the nature of the one discussed here. We took her at once to the hospital and found that she was suffering from some form of encephalitis involving the cerebellum. This was a condition with very grim possibilities, but she had the splendid luck to make a perfect recovery without any aftereffects. If I were a superstitious man, this experience and many others reported by medical. men might make me suppose that a disease has a vicious personality and wished nothing more than to revenge itself on the scientist who has pursued it.

To go back to the work of our team of three: we wrote up these ideas in an article, but what was even more important was that Bigelow and I felt that we could safely go ahead with the treatment of the human links in the control chain as if they were pieces of feedback apparatus. Accordingly, we felt justified in proceeding from our crude experimental setup in the direction of the design of a complete apparatus for anti-aircraft control and prediction.

Since it was clear that the anti-aircraft control apparatus was in essence a feedback loop and contained in its construction many subsidiary feedback loops, we had to find out something of the characteristics of these loops. These characteristics were not available, so that the over-all apparatus which we designed was essentially crude and of unverifiable behavior. Under the circumstances, it was not considered advisable to proceed much further with it, partly as such tentative behavior as we could compute mathematically did not suggest an excellent performance.

Our ideas were eagerly taken up by other workers in the field and did lead to a very definite improvement in practice, in particular of the part that consisted in filtering out experimental errors of observation. We were not finally commissioned to perfect our own design, but I was in fact asked to write a book on time series, extrapolation, and interpolation. This book was reproduced in a photo-lithographed form which, because of the yellow paper in which it was bound, came to be known as the "yellow peril," a name previously confined to the yellow books of Springer's mathematical series. My textbook was very freely used, not only during the war by the designers of systems of control for the aiming and firing of anti-aircraft guns, but also by servo-engineers and electrical-communication men, both then and later. It was reprinted after the war in an amplified and improved edition with an appendix by Professor Norman Levinson, who made the use of my methods considerably clearer.

The work I did on the statistical treatment of anti-aircraft-fire control has led eventually to a general statistical point of view in communication engineering. The years that have gone by between then and now have secured the general acceptance of this point of view in communication engineering, but they have also led to much more. I might almost say that the whole of engineering is rapidly assuming a statistical aspect, and that this is passing over to less orthodox fields such as meteorology, sociology, and economics.

Let me go back to my earlier remarks concerning Willard Gibbs and the revolution he and his contemporaries created in physics. The orthodox Newtonian view of physical dynamics provides for certain equations involving rates, known as differential equations. With the aid of these equations and with a knowledge of the initial values, or values at time zero, of the variables whose rates are being determined in relation to their numerical values, we can inch our way along in developing the history of these phenomena. At each time we know our values. From these we determine their own rates of change, and this gives us an approximation to the same values and rates an instant later.

If we choose an instant of time that is sufficiently short, we may progress along the history of our phenomena to any point in time which we wish to attain. This is the method of the astronomers in calculating the orbits of the planets and the method of the ballistics expert for the study of the flight of projectiles in determining their paths.

In astronomy, as I have said before, the computation of these orbits is a very precise mathematics, and our initial data are very accurately known. This is not the case in most ballistic or engineering problems. In firing a shell, for example, the angle of elevation is known only to a very limited degree of accuracy, as are the weight of the projectile, the charge of powder, and the various atmospheric conditions. The result is that we start with none of this precisely given but with each set of data given only within a certain range. The traditional way of solving the ballistic equation is to assume the initial data as given precisely. Then we find the range, the angle of impact, the velocity of impact, and other significant quantities, and we immediately start to revise these with the aid of methods of interpolation or correction, reckoned by procedure which is entirely distinct from the first.

In this process we waste a good deal of effort, first in making our data unrealistically accurate, and second in correcting our imperfectly realistic results. There is, however, another method which is now coming into use, and which finds its spiritual father in Willard Gibbs.

Gibbs pointed out that when a dynamical system develops according to its own laws as, for example, when a top spins freely, something occurs very much like the flow of a fluid. To characterize a top, we need a point in a certain space, but this space is not the same as the familiar three-dimensional one of solid geometry. The position of a top requires six co-ordinates—or measurable quantities—to give its position and six to give its momenta; and these together form a twelve-dimensional array, which we may call by analogy a twelve-dimensional space. In this space there is a certain measure of volume, so that a set of tops which fill a given volume at one time will fill an equal volume at any other. This type of invariance of volume is to be found in all dynamical systems in which there is no power input or output.

This flow may be conceived as a flow of probability, and so it was conceived by Gibbs. The probability that a particle will be at one time in a given region of this peculiar space is the same as that at a later time it will be in the corresponding region into which the first has flowed.

Thus the typical equation of flow is no longer a general system of what is known as ordinary differential equations but a set of integral equations. These integral equations relate past distributions to future distributions, in such a way that if we superimpose different past distributions we superimpose the corresponding future distributions. A system like this, in which sums in the output answer to sums in the input, is known as linear, and the integral equations of the flow method of treating dynamics may be taken as linear.

The method is quite practicable computationally; and if the problem to be solved is of any high degree of complexity, this method may well be easier than the purely Newtonian method. Certain simplifications of methods of this type are now being used extensively by members of the mechanical-engineering department at M.I.T.

In addition to purely computational advantages in the more complicated cases, this method is also essentially superior to the Newtonian method of computation from the logical point of view. The reason is this: what we put into our problem not only consists of precise data which we later have to ease off in accordance with the inaccuracy of the equations and the initial conditions but contains intrinsically the very inaccuracy which hinders our work. We are thus not overcomputing and relieving the effect of this overcomputing by an *ad hoc* study of its errors but putting all our cards on the table at the beginning. What we finally get is what we want, neither more nor less. This cuts down a lot of unnecessary effort, but it also increases the real precision of what we are doing.

No scientific measurement can be expected to be completely accurate, nor can the results of any computation with inaccurate data be taken as precise. The traditional Newtonian physics takes inaccurate observations, gives them an accuracy which does not exist, computes the results to which they should lead, and then eases off the precision of these results on the basis of the inaccuracy of the original data. The modern attitude in physics departs from that of Newton in that it works with inaccurate data at the exact level of precision with which they will be observed and tries to compute the imperfectly accurate results without going through any stage at which the data are assumed to be perfectly known.

If we follow in these unprecise problems the sort of computation which the astronomer uses in determining the orbits of the planets, we may happen to choose initial conditions which lead to final results not typical of the wider range of initial conditions with which we have operated, and this instability of our orbit may drive us to a false reckoning of our ultimate error.

As I have said earlier in the description of my work on prediction, the more sensitive our instruments become, the more unstable they will be. These cause an error of a different sort from that of imprecision, but an error equally serious. What I have said of mechanical instruments is true of methods of computation. The balance between errors of imprecision and errors of instability is something which we can compute only on a statistical basis. Why not, then, assume the statistical basis at the very beginning and obtain both mean result and error by a unified method of computation?

If this recognition of the statistical nature of all science is already proving to be valuable in the most Newtonian type of mechanical-engineering computation, how much more must it then be the natural method of computation in those fields in which our errors of observation are naturally very large!

Let us consider meteorology as one example. We know a good deal about the dynamics of the atmosphere; and if our observations of the initial conditions were extremely good, we might expect that we could compute the future in a purely Newtonian way, even though this way is very likely to involve a great deal of overcomputation. What we actually know about the atmosphere, though, is a sampling taken in no more than three or four observations per day per hundred thousand cubic miles of the atmosphere.

Recently, under the influence of John von Neumann, there has been an attempt to solve the problem of predicting the weather by treating it as something like an astronomical-orbit problem of great complexity. The idea is to put all the initial data on a super computing machine and, by a use of the laws of motion and the equations of hydrodynamics, to grind out the weather for a considerable time in the future.

The catch is that all the observations of the weather bureau give only limited information at a very few points, with colossal gaps between them. These one can fill up only by some sort of statistical reasoning. Thus, an adequate meterological method must partake both of dynamics and statistics. There are clear signs that the statistical element in meterology cannot be minimized except at the peril of the entire investigation.

I do not mean to deny the importance of dynamics, but I do mean to assert the virtues of that Gibbsian approach in which this dynamics is treated as a statistical flow.

Meterology is typical of most of those numerical sciences which have come to the fore late in the history of science. In economics, the so-called econometric science of economic dynamics suffers under the radical difficulty that the numerical quantities which are put into the dynamics are not well defined and must be treated as gross statistical estimates. Who knows precisely how to define a demand, and how to measure it in terms which will satisfy most other economists? Can any two economists check on the amount of unemployment in the United States at a given time?

Econometrics will never get very far until two steps are taken: One of these steps is that the observation of the quantities—demands, inventories, and the like—with which econometrics operates must be subject to the same criteria of precision and rigor as the dynamics by which they are combined. The other is that we should recognize from the beginning the statistical and imperfectly precise nature of the quantities with which we operate and that we should go over to a Gibbsian treatment of them.

What I have just said about meteorology and econometrics applies equally to sociological dynamics, to biometrics and in particular to the very complicated study of the nervous system which is itself a sort of cerebral meteorology. It belongs to the very grammar of the use of mathematical methods in semiprecise sciences. It is the heart of the engineering of the future.

This new technique was foreshadowed in my war work on anti-aircraft fire-control predictors and was carried further in my development of communication theory. As yet it has penetrated to only a few initiates in the appropriate fields of scientific work, but it is philosophically right, and it bids fair to change the entire face of all the precise and semiprecise sciences.

When I first wrote about prediction theory, I was not aware that some of the main mathematical ideas had already been introduced in the literature. It was not long before I found out that just before the Second World War an important little paper on the same subject had been published by the Russian mathematician Kolmogoroff in the *Comptes Rendus* of the French Academy of Sciences. In this, Kolmogoroff confined himself to discrete prediction, while I worked on prediction in a continuous time; Kolmogoroff did not discuss filters, or indeed anything concerning electrical engineering technique; and he had not given any way of realizing his predictors in the metal, or of applying them to anti-aircraft-fire control.

Nevertheless, all my really deep ideas were in Kolmogoroff's work before they were in my own, although it took me some time to become aware of this. A series of papers by Kolmogoroff and such pupils of his as Krein continued to appear in the *Doklady (Reports of the Russian Academy of Sciences),* and although these papers still stuck for the most part to the concept of prediction previously developed by Kolmogoroff, somewhat narrower than my own, I am by no means convinced that Kolmogoroff was not independently aware of the possibility of some of the applications I had made. If that was so, he must have had to keep them out of general publication because of their importance for the military-scientific work of the Soviets. A recent paper by Krein, in which he makes an explicit allusion to my own work in the field of applications, convinces me of this.

I have never met Kolmogoroff, and indeed I have never been in Russia, nor have I been in correspondence with him or with any of his school. Thus what I say about him is largely surmise. At an early stage of my work for the United States military authorities, before I had seen Kolmogoroff's paper, the question came up whether anybody abroad was likely to be in possession of ideas similar to mine. I said that they would unquestionably receive no particularly ready reception in Germany; that my own friends Cramer, in Sweden, and Lévy, in France, might well have been thinking along similar lines; but that if anyone in the world were working on these ideas it would most likely be Kolmogoroff in Russia. This I said because of my knowledge that for twenty or thirty years hardly had either of us ever published a paper on any subject but the other was ready to publish a closely related paper on the same theme.

Within the last two or three years I have seen a Russian book on prediction theory, communication theory, and similar topics which makes extensive references to both Kolmogoroff's work and to my own. It gives

Kolmogoroff the priority, and although this priority is only partial, I have just said that there is good reason for considering him not only as an independent discoverer of large parts of the subject but as the first man to write on it. The book takes my own work very seriously and treats me much more fairly than I would expect in a Soviet book, international relations being what they are.

The "yellow peril" book is still playing an important role in American research work, both for military purposes and for more general uses. It was with the permission of the government that it has been reprinted, and it must have been a copy of this book which filtered into Russia and served as a basis for the Russian comments of which I have just written.

From this point on, my work, or rather the work of my group, has spread out to cover a very wide field of communication theory and practice. In the first place, the "yellow peril" is most definitely a statistical treatment of problems of communication. When the book was written, almost nobody had thought of communication in these terms. I think I am to be pardoned for a certain pride in saying that the statistical approach to communication theory is now accepted almost everywhere.

I approached information theory from the point of departure of the electric circuit carrying a continuous current, or at least something which could be interpreted as a continuous current. At the same time, Claude Shannon, of the Bell Telephone Laboratories, was developing a parallel and largely equivalent theory from the point of view of electrical switching systems. This represented a direct development of his previous work on the use of the algebra of logic in switching problems.

As I have said before, Shannon loves the discrete and eschews the continuum. He considered discrete messages as something like a sequence of yeses and noes distributed in time, and he regarded single decisions between yes and no as the element of information. In the continuous theory of filtering, I had been led to a very similar definition of the unit of information, from what was at the beginning a considerably different point of view.

In introducing the Shannon-Wiener definition of quantity of information (for it belongs to the two of us equally), we made a radical departure from the existing state of the subject. For many years it was believed that the carrying power of a communication line per unit time was to be measured by the band width it could carry.

A band width of 200 cycles was supposed to be able to carry twice as much information per second as a band width of 100 cycles. This supposition ignored the fact that, in the absence of noise, any band width would be enough to carry any amount of information in one second. One single voltage measured to an accuracy of one part in ten trillion could convey all the information in the Encyclopedia Britannica, if the circuit noise did not hold us to an accuracy of measurement of perhaps one part in ten thousand.

In the early days of the telephone art, very few lines were burdened with messages to the ultimate limit of their message-carrying capacity. As the art developed and the new modes of communication like radio and television demanded a more complete exploitation of the message space available, it became clear that the noisiness of the line or air channel is another important factor which we must take into consideration. The ether is full of disturbances which the radio man terms static, and no conductor, be it metallic or gaseous, can carry electricity in smaller lumps than the single electron. The irregularity of the stream of electrons is known as the shot effect, and it is an important consideration in all modern communication design.

It was only shortly before World War II that the load on communication channels became heavy enough for this intrinsic noise to become a serious practical barrier to the use of the lines for even more communication. Thus, the statistical point of view in communication theory, which I had anticipated so many years before with my generalized harmonic analysis, and which Shannon and I had jointly made fundamental at the beginning of World War II, became inevitable and basic shortly after the war had begun.

The work we were doing on feedback in connection with the fire-control machine and the nervous system introduced another revolution, which, like the first, has received universal acceptance in the course of the last few years. When I first came to Tech, electrical engineering was divided into two fundamental parts, which were known in Germany as the technique of weak currents and the technique of strong currents, and in the United States as communication engineering and power engineering.

The distinction between these two fields is valid, but the nature of this distinction and the place to make it was not understood for a considerable period. The generating station for a television sender or a trans-Atlantic radio sender may use relatively large quantities of power, but it is directed primarily toward communication; while the fractional-horsepower motor used in a dentist's drill may employ relatively small quantities of power. Nevertheless, the first piece of apparatus is primarily oriented with respect to the message and the second one with respect to the energy consumed.

At a period at which this distinction was not fully appreciated, servomechanisms for the control of gun turrets and other pieces of heavy apparatus were naturally assumed to belong to power technique rather than to communication technique. The whole tradition of power technique was to consider electric currents and voltages as varying in time, while the whole tradition of communication technique, particularly under the influence of Heaviside, had led to consideration of a message as a sum of a large or infinite number of different frequencies. It was not easy to see that the frequency treatment, rather than the time treatment, was just as appropriate for the servomechanism as for the telephone, the telegraph, and television.

I think that I can claim credit for pointing this fact out and for transferring the whole theory of the servomechanism bodily to communication engineering. My whole point of view in these matters made me regard the computing machine as another form of communication apparatus, concerned more with messages than with power. Its nature, as I saw it, was that of a series of switching devices, so enchained together that the information coming out of a number of stages of these was introduced into a subsequent stage as ingoing and regulating information.

It was clear that while these switching devices might be gear wheels and the like, they could equally well be mechanical relays or the electrical relays which depend on vacuum tubes and other electronic phenomena. I was much more disposed, as I have said, to use switching devices which made a choice between two alternatives than those which made a choice between ten alternatives, and I tried to bring this concept of computing machines to the attention of the engineering public.

It was at Harvard, under the supervision of Howard Aiken, that I found the first of the newer switching computers dependent on relays. Aiken was developing them under a government grant. I was much struck with Aiken's work, which I greatly admired, and which Aiken himself considered as the modern carrying-out of the crude computers developed by Babbage in England a hundred years ago. Babbage had formed an excellent conception of their mathematical possibilities but had almost no understanding of the mechanical problems to which they gave rise.

I was surprised to find that Aiken was completely committed to the relatively slow mechanical relay as the mechanical computer's first tool and that he did not put any enormous value on the speed which could be derived by the use of electronic relays. This limitation of point of view has now been discarded by Aiken himself, who has become one of the most active and original inventors and designers of electronic computers. But at that time he labored under a curious moralistic quirk in accordance with which he considered work with mechanical relays as essentially sound and right and work with electronic relays as unnecessary and ethically sloppy.

Here I wish to emphasize again a weakness of attitude joined with a great strength in those men who show practical ingenuity in the devising of gadgets. It is the desire to fix the technique of a subject forever at the precise point to which their ingenuity has carried them and then to offer a profound intellectual and moral resistance—a block, in fact, to later work which departs from their principles. We mathematicians who operate with nothing more expensive than paper and possibly printer's ink are quite reconciled to the fact that, if we are working in a very active field, our discoveries will commence to be obsolete at the moment they are written down or even at the moment they are conceived. We know that for a long time everything we do will be nothing more than the jumping-off point for those who have the advantage of already being aware of our ultimate results. This is the meaning of the famous apothegm of Newton, when he said, "If I have seen further than other men, it is because I have stood on the shoulders of giants."

Yet the commercial possibilities of the invention in the metal tend to blind the industrial worker to this fundamental fact and to make him hope that he can hold back the stream of progress at the precise stage where he had made his own contributions. The patent system and the commercial value of an inventor's idea as something salable tend to push him in this direction. This is not realistic. As a practical man, the inventor should have the very practical consciousness that for many years his greatest contribution will not be a single gadget but the furthering of the whole stream of thought and ideas concerning an enormous class of gadgets past, present, and future. He should come to terms with this streaming of thought and realize that, just as he has gone beyond those who were born before him, he himself and his work will have to serve rather as a stepping stone to the future than as the end to which science and technique must finally arrive.

However, my interest in the development of computing machines carried me far beyond those machines past, present, or to come, which are made of brass and copper, glass and steel. The brain and the nervous system also share in the main characteristics of computing machines. Parallel to the yes and no of a relay is the fact that a nervous fiber can exist in what are fundamentally only two states: the state of carrying a message and the state of not carrying a message. This is the so-called all-or-none law of the nervous system; and, although it may not be as precisely true as its crude, cold formulation would suggest, it is sufficiently true to represent a fundamental fact of nervous conduction.

A nerve fiber, it is true, may be stimulated by messages of varying intensity, but the ultimate fate of each of these messages is either to die out and fail to reach the end of the fiber or to continue as what the chemists would call a self-catalyzing process and start an impulse which will go from one end of the fiber to the other. When it has reached the end of the fiber, its subsequent history is so nearly independent of the original strength of stimulation that this strength may be entirely neglected. Thus, there is a certain analogy between a nerve fiber and a flip-flop electric circuit, an electric circuit with two, and only two, states of equilibrium. This analogy is so close that, long before the message reaches the end of the fiber, it carries its information in the form of a number of impulses rather than in the form of the strength of the impulses.

Not only are nerve fibers switching devices, but they are devices which lead into other switching devices. The nerve fibers communicate with one another by junction points or junction systems known as synapses, and in these the question whether a new message is established in an outgoing fiber depends on the precise set of incoming messages received from various fibers. In the simplest cases, the synaptic system has a threshold, which means that if more than a certain number of incoming fibers receive messages within a certain critical interval of time, the outgoing fiber fires, and otherwise not.

We are so used to feedback phenomena in our daily life that we often forget the feedback nature of the simplest processes. When we stand erect, it is not in the manner in which a statue stands erect, because even the most stable statue needs to be fastened to some sort of pedestal or it would fall over. Human beings stand erect, however, because they are continually resisting the tendency to fall down, either forward or backward, and manage to offset either tendency by a contraction of muscles pulling them in the opposite direction. The equilibrium of the human body, like most equilibria which we find in life processes, is not static but results from a continuous interplay of processes which resist in an active way any tendency for them to lead to a breakdown. Our standing and our walking are thus a continual jujitsu against gravity, as life is a perpetual wrestling match with death.

In view of this, I was compelled to regard the nervous system in much the same light as a computing machine, and I communicated this idea to my friend Rosenblueth and to other neurophysiologists. I managed to get a group of neurophysiologists, communication engineers, and computing-machine men together at Princeton for an informal session, and I found on the part of each group a great willingness to learn what the other groups were doing and to make use of their terminology. The result was that very shortly we found that people working in all these fields were beginning to talk the same language, with a vocabulary containing expressions from the communication engineer, the servomechanism man, the computing-machine man, and the neurophysiologist.

For example, all of them were interested in the storage of information to be used later, and all of them found that the word *memory* (as used by the neurophysiologist and the psychologist) was a convenient term to cover the whole scope of these different fields. All of them found that the term *feedback,* which had come from the electronics engineer and was extending itself to the servomechanism man, was an appropriate way of describing phenomena in the living organism as well as in the machine. All of them found that it was convenient to measure information in terms of numbers of yeses or noes, and sooner or later they decided to term this unit of information the *bit.* This meeting I may consider the birthplace of the new science of cybernetics, or the theory of communication and control in the machine and in the living organism.

I had hopes that this new science was going to pass through a rapid development over a broad front. The subject has developed greatly, and I have participated in its later phases. However, the times were not favorable for the normal growth of new ideas, and I have had to watch very carefully through a period where what I intended as a serious contribution to science was interpreted by a considerable public as science fiction and as sensationalism.

Science fiction is in vogue, and it is the fashion even among certain serious scientists to see merit in its writings. I myself as a child was a devotee of Jules Verne and H. G. Wells, to whom the present literature of science fiction owes its origin, but it is an infinitely slicker and more pernicious article. On the one hand, it leads to fantasies of power and of brutality quite as devastating as anything in the thud-and-blunder type of gangster story or the most uncomic comics. On the other hand, it is helping to create a generation of youngsters who believe that they are thinking in scientific terms because they are using the language of science fiction. It is a real difficulty in our schools of science and engineering to have to try to educate young men who believe that they have a calling toward science merely because they are accustomed to playing with the ideas of destructive forces, other planets, and rocket travel.

This vicious daydreaming is largely a product of World War II, which has done so much to demoralize the whole generation of science. The period of the war was one in which the status of science and that of mathematics were changing rapidly. In the first place, leisure was vanishing in every sector of life. Before the war, I used to find the M.I.T. boys playing a game or two of bridge after lunch in one of the lounges of Walker Memorial. I often participated in these games.

I did not regard the time as wasted either by myself or by the students, for between games we used to have an occasion for wide-ranging discussions which might be pure bull session or might involve a real play of ideas. From the beginning of the war, everyone was in deadly earnest, and all chance of intellectual play was restricted. To the present day, it is hard to find young men who dare to take enough time off from work to consider what their work is about. The hours spent in the fantasy of space books are no replacement for a good bull session.

Before the war, and particularly during the depression, positions in science were not easy to get. The requirements for these positions had become exceedingly high. During the war, this situation had changed in two respects. First, there were not enough men to carry out all the scientific projects which the war involved. Secondly, in order to carry out these projects at all, it became necessary to organize the work so as to use those with a minimum amount of training, ability, and devotion.

The result was that young men who should have been thinking of preparing themselves in a long-time way for their careers lived in a lighthearted way from hand to mouth, confident that the existing boom in scientists had come to stay. Such men were in no state to accept the discipline or hard work, and they evaluated whatever intellectual promise they might have as if it had been already realized in performance. With the older men crying out for assistance and manpower, these boys would shop around for those masters who would demand least and grant them the most in indulgence and flattery.

This was a part of a general breakdown of the decencies in science which continues to the present day. In most previous times, the personnel of science had been seeded by the austerity of the work and the scantiness of the pickings. There is a passage in Tennyson's "Northern Farmer: New Style" which says: "Doänt thou marry for munny, but goä wheer munny is!"

Thus, an ambitious man with slightly anti-social tendencies or, to put it more politely, indifferent to spending other people's money, would formerly have avoided a scientific career as if it were the plague itself. From the time of the war on, these adventurers, who would have started out as stock promoters or lights of the insurance business, have been invading science.

The old assumption which we used to make must be discarded. We all knew that the scientist had his vices. There were those among us who were pedants; there were those who drank; there were those who were overambitious for their reputations; but in the normal course of events we did not expect to meet in our world men who lied or men who intrigued.

When I began to emerge from my sheltered life into the scientific confusion of wartime, I found that among those I was trusting were some who could not be held to any trust. I was badly disillusioned more than once, and it hurt.

The meeting to consolidate communication theory took place well after Pearl Harbor. It may surprise the reader that in all this talk about war work, I have not mentioned Pearl Harbor and the actual entry of America into the war. The fact is that all of us had long been convinced that the war was coming to America in some form or other, and the actual opening of hostilities did not change my work on means of defense.

In the fall of 1941, the tension of the successive defeats experienced by the Allies in Scandinavia, Holland, and France, the Battle of Britain, and the ambiguous and to-and-fro situation in North Africa, had grown as great as one thought one could bear, yet it became complicated by the fairly general feeling that something was about to blow off in Japan. While none of us was exactly prepared for Pearl Harbor, I do not think that we were convinced that a military dictatorship like Japan would play the game of war and diplomacy according to the standard rules, particularly when these rules were manifestly advantageous to us. Thus, Pearl Harbor came, to me at least, as much more of a shame and a humiliation than a surprise.

Pearl Harbor and our subsequent entry into World War II on both sides of the globe had a number of direct effects upon me. It is true it could not involve me in war research any more than I was already involved, because I was fully occupied in this direction and had been for more than a year. However, the war wiped out plans which Manuel Sandoval Vallarta and I had made to go to South America in the interest of international good will, on funds primarily emanating from the State Department (or, in his case, from the Mexican Government).

What I felt to be much more important to me was what was going to happen to my dear friends the Lees. We had just managed to secure for them passage on a liner from Hong Kong. Then Pearl Harbor came and prevented the boat from sailing, or at least from sailing with our friends on board. Thus, Lee, who had already gone for five years through the Chinese-Japanese War without any adequate contact with his profession, was sentenced to wait five years more, until after V-J Day, before we could bring him over to the United States. During all that time, an offer which we had secured him from the electrical-engineering department of M.I.T. stood open, or at least ajar; and when he finally arrived he was able to step into an instructorship. This has now been succeeded by an assistant professorship and an associate professorship.

The problem of what professional man is to do when he comes back as a sort of Rip Van Winkle who has slept for a decade and wakes to find a changed world is very difficult. The obvious thing might appear to be to spend a year or two in studying the various developments of the intermediate period. This must be done to some degree or other, but it is not a completely adequate way of treating the situation. The very bulk of new material tends to produce an intellectual indigestion in the student. He must come into competition with the younger generation who have learned the field the easy way while it has been developing and who are at home in it. Our Rip Van Winkle cannot expect to compete with them.

What made Lee's situation easier was that I had recently developed a considerable part of the statistical theory of communication engineering in the "yellow peril." I pointed out to Lee that the one way to avoid being disastrously behind the game was to move deliberately ahead of the game and thus secure an advantage of some years while the other people were catching up. Lee saw the point.

The situation was made considerably easier by the fact that for years we had been so much in the habit of working together that my mental processes and ways of writing were quite familiar to him. Thus, he took over the problem of working out in detail the communication and engineering consequences of ideas which I had only sketched in general terms and of making himself interpreter to the engineering public of the field which I was later to call cybernetics.

Lee has established himself in this program and has been busy for some years in carrying it through to a most successful conclusion. He is now writing a book on communication engineering from the new standpoint, and he is showing in it a great patience, thoroughness, and consideration of the reader. For me, close as I am to the origin of the subject, such a detached treatment would be impossible.

Lee has presented the new ideas to quite a number of government and industrial laboratories. He has brought up a whole generation of young electrical engineers to do research along statistical lines and to employ my point of view as a habitual approach to communication problems. He has also organized very successful summer meetings, so that engineers already actively engaged in the communication industry have been able to come to M.I.T. for a refresher course on the cybernetic viewpoint.

In these ways the difficulties of ten years' isolation have been bypassed successfully. The head start which Lee has had in the new methods has allowed him time to catch up with what has happened between 1936 and 1946, and, what is more, the work in which he has been engaged has given him specific problems to use as a touchstone for his understanding of and familiarity with the research of the intermediate period. In other words, we have seen in the pay-off of the policy which the two of us began to adopt the moment the Lees arrived in Boston's South Station after the war.

With Lee back at M.I.T., I was greatly encouraged to go on with further investigations of servomechanisms and the entire class of topics which I was later to give the name of cybernetics. As I have said, Lee is himself now finishing a book on the matter. However, not all that the two of us could do together on that, nor all that any hundred people could do, would suffice to cover more than a small part of the literature on servomechanisms and on the automatic factory, to which our early work has led. The automatic factory bids fair to become the norm rather than the exception, even within the lifetime of those who are now in college. It is giving rise to a new profession of experts who are able not only to design these factories but to set up on them problems of the most varied sort. The modern technique of automatic-factory design is well beyond the ambit of a theoretical man like me. As I shall show in a later chapter, I have conceived it to be my primary function not so much to develop the automatic factory further as to explain its nature and its consequences, and to alert both labor and management to the need of facing these intelligently.

This is a section of doi:10.7551/mitpress/11597.001.0001

Norbert Wiener—A Life in Cybernetics

Ex-Prodigy: My Childhood and Youth and I Am a Mathematician: The Later Life of a Prodigy

By: Norbert Wiener

Citation:

Norbert Wiener—A Life in Cybernetics: Ex-Prodigy: My Childhood and Youth and I Am a Mathematician: The Later Life of a Prodigy
By: Norbert Wiener
DOI: 10.7551/mitpress/11597.001.0001
ISBN (electronic): 9780262347051
Publisher: The MIT Press
Published: 2018

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.

The MIT Press

This book was set in ITC Stone Sans Std and ITC Stone Serif Std by Toppan Best-set Premedia Limited. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Wiener, Norbert, 1894-1964, author. | Kline, Ronald R., writer of foreword. | Container of (work): Wiener, Norbert, 1894-1964. Ex-prodigy. | Container of (work): Wiener, Norbert, 1894-1964. I am a mathematician.
Title: Norbert Wiener--a life in cybernetics : Ex-prodigy : my childhood and youth, and I am a mathematician : the later life of a prodigy / Norbert Wiener ; with a new foreword by Ronald R. Kline.
Description: Cambridge, MA : The MIT Press, 2018. | Includes bibliographical references and index.
Identifiers: LCCN 2017042823 | ISBN 9780262535441 (pbk. : alk. paper)
Subjects: LCSH: Wiener, Norbert, 1894-1964. | Mathematicians--United States--Biography.
Classification: LCC QA29.W497 A25 2018 | DDC 510.92 [B] --dc23 LC record available at https://lccn.loc.gov/2017042823

10 9 8 7 6 5 4 3 2 1