The Essential Message:
Claude Shannon and the Making of Information Theory

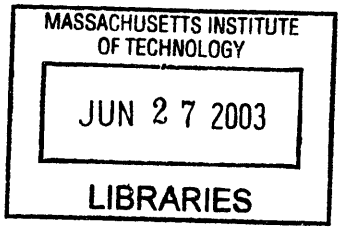
by

Erico Marui Guizzo
B.S., Electrical Engineering
University of Sao Paulo, Brazil, 1999

Submitted to the Program in Writing and Humanistic Studies
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Science Writing
at the
Massachusetts Institute of Technology

September 2003



© 2003 Erico Marui Guizzo. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly
paper and electronic copies of this thesis document in whole or in part.

Signature of Author:.....
Program in Writing and Humanistic Studies
June 16, 2003

Certified and Accepted by:.....
Robert Kanigel
Thesis Advisor
Professor of Science Writing
Director, Graduate Program in Science Writing

ARCHIVES

The Essential Message:
Claude Shannon and the Making of Information Theory

by

Erico Marui Guizzo

Submitted to the Program in Writing and Humanistic Studies
on June 16, 2003 in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Science Writing

ABSTRACT

In 1948, Claude Shannon, a young engineer and mathematician working at the Bell Telephone Laboratories, published "A Mathematical Theory of Communication," a seminal paper that marked the birth of information theory. In that paper, Shannon defined what the once fuzzy concept of "information" meant for communication engineers and proposed a precise way to quantify it—in his theory, the fundamental unit of information is the *bit*. He also showed how data could be "compressed" before transmission and how virtually error-free communication could be achieved. The concepts Shannon developed in his paper are at the heart of today's digital information technology. CDs, DVDs, cell phones, fax machines, modems, computer networks, hard drives, memory chips, encryption schemes, MP3 music, optical communication, high-definition television—all these things embody many of Shannon's ideas and others inspired by him.

But despite the importance of his work and its influence on everyday life, Claude Shannon is still unknown to most people. Many papers, theses, books, and articles on information theory have been published, but none have explored in detail and in accessible language aimed at a general audience what the theory is about, how it changed the world of communication, and—most importantly—what path led Shannon to his revolutionary ideas. "The Essential Message" presents an account of the making of information theory based on papers, letters, interviews with Shannon and his colleagues, and other sources. It describes the context in which Shannon was immersed, the main ideas in his 1948 paper—and the reaction to it—and how his theory shaped the technologies that changed one of the most fundamental activities in our lives: communication.

Thesis Supervisor: Robert Kanigel

Title: Professor of Science Writing
Director, Graduate Program in Science Writing

TH SS NT L M SS G

CL D SH NN N ND TH M K NG F NF RM T N TH RY

BY R C G ZZ

E(R)

"I only ask for information . . ."

Charles Dickens, *David Copperfield*[*]

[*] From a piece of paper with quotations kept by Shannon. *Claude Elwood Shannon Papers, Manuscript Division, Library of Congress, Washington, D.C.*

TWO MEN TALK about the past. The conversation is animated. They drink beer. It is an early evening in the summer of 1982. The two men sit in the living room of a large stuccoed house overlooking a lake in the suburbs of Boston. Bob Price is interested in things that happened more than thirty years ago. He wants to know about the origins of information theory. And Claude Shannon is the right person to ask: he invented the theory.[1] Shannon's information theory transformed one of the most fundamental activities in our lives: communication. His stunning new ideas made possible the information age that most of humanity lives in today.

"I'm looking at this 1945 cryptography report," Price says, "it's got the words 'information theory' in it. It says you're next going to get around to write up information theory. It sounds—"

"Oh, did it say that in there?" Shannon asks reluctantly.

"Yes, it sounds as though the cryptography gave you the mysterious link that made the whole... Well, the fan diagrams, for instance, if it hadn't been for cryptography would you have those fan diagrams?"

"What fan diagrams?"

Price points to diagrams in a book that resemble paper fans: "These, these."

"Oh, those are the fan diagrams," Shannon says laughing.

"Well, do you think that would have come out of cryptography? Or you had that already without the cryptography?"

Shannon pauses for a second. And then, lowering his voice, he says: "Well, I have no idea."

"OK, sure, well it's a long time ago."

"But not only that, Bob. You ask questions of where would things have come from... These complex hypothetical questions."

On that evening in 1982, Price tried hard to get onto the mind of the gray haired man sitting next to him. But the man resisted. Shannon wasn't rude or stubborn. Quite the contrary, he was known for his sharp sense of humor and even his self-deprecating nature. He never took himself too seriously.[2] How did he

come up with information theory? What about those fan diagrams? Well, it seemed Shannon couldn't care less.

* * *

MEMORABLE BREAKTHROUGHS IN science often have an origin myth, usually an epiphany that crystallizes in history that particular achievement destined to change the world. Archimedes in the bathtub. Newton under the apple tree. Einstein riding a beam of light.[3] Like a picture, static in time and limited in space, an origin myth doesn't tell the whole story—and the story it tells sometimes weaves fact and fiction. But an origin myth helps to popularize scientific achievements—and science heroes—that people otherwise wouldn't know about.

Information theory doesn't have an origin myth. And Shannon always repeated his creation was not the result of a single moment of clarity at a bathtub or under a tree. Didn't he have a "eureka moment"? "I would have," once he said jokingly, "but I didn't know how to spell the word."[4] Maybe that is why, despite the importance of his work and its influence on everyday life, Claude Shannon is still unknown to most people.[5] Today we use digital cell phones and send messages over the Internet but we know little about the ideas that contributed to make these things possible. Many of these ideas are the result of Shannon's theory, which sought the mathematical laws governing systems designed to transmit and manipulate information.

But if information theory doesn't have an origin myth, it has a very clear beginning. The field was founded in 1948 when Shannon published the paper considered his masterwork, "A Mathematical Theory of Communication."[6] The fundamental problem of communication, he wrote in the second paragraph, is that of reproducing at one point a message selected at another point. A message could be a letter, a word, a number, speech, music, images, video—anything we want to transmit to another place. To do that, we need a transmission system; we need to send the message over a communication channel. But how fast can we send these

messages? Can we transmit, say, a high-resolution picture over a telephone line? How long that will take? Is there a best way to do it?

Before Shannon, engineers had no clear answers to these questions. At that time, a wild zoo of technologies was in operation, each with a life of its own—telephone, telegraph, radio, television, radar, and a number of other systems developed during the war.[7] Shannon came up with a unifying, general theory of communication. It didn't matter whether you transmitted signals using a copper wire, an optical fiber, or a parabolic dish. It didn't matter if you were transmitting text, voice, or images. Shannon envisioned communication in abstract, mathematical terms; he defined what the once fuzzy concept of "information" meant for communication engineers and proposed a precise way to quantify it. According to him, the information content of any kind of message could be measured in binary digits, or just bits—a name suggested by a colleague at Bell Labs. Shannon took the bit as the fundamental unit in information theory. It was the first time that the term appeared in print.[8]

In his paper, Shannon showed that every channel has a maximum rate for transmitting electronic data reliably, which he called the channel capacity. Try to send information at a rate greater than this threshold and you will always lose part of your message. This ultimate limit, measured in bits per second, became an essential benchmark for communication engineers. Before, they developed systems without knowing the physical limitations. Now they were not working in the dark anymore; with the channel capacity they knew where they could go—and where they couldn't.

But the paper contained still one more astounding revelation. Shannon demonstrated, contrary to what was commonly believed, that engineers could beat their worst enemy ever: transmission errors—or in their technical jargon, "noise." Noise is anything that disturbs communication. It can be an electric signal in a telephone wire that causes crosstalk in an adjacent wire, a thunderstorm static that perturbs TV signals distorting the image on the screen, or a failure in network equipment that corrupts Internet data. At that time, the usual way to overcome

noise was to increase the energy of the transmission signals or send the same message repeatedly—much as when, in a crowded pub, you have to shout for a beer several times. Shannon showed a better way to avoid errors without wasting so much energy and time: coding.

Coding is at the heart of information theory. All communication processes need some sort of coding. The telephone system transforms the spoken voice into electrical signals. In Morse code, letters are transmitted with combinations of dots and dashes. The DNA molecule specifies a protein's structure with four types of genetic bases. Digital communication systems use bits to represent—or encode—information. Each letter of the alphabet, for example, can be represented with a group of bits, a sequence of zeroes and ones. You can assign any number of bits to each letter and arrange the bits in any way you want. In other words, you can create as many codes as desired. But is there a *best* code we should use? Shannon showed that with specially designed codes engineers could do two things: first, they could squish the messages—thus saving transmission time; also, they could protect data from noise and achieve virtually error-free communication using the whole capacity of a channel—perfect communication at full speed, something no communication specialist had ever dreamed possible.

Measuring the information content of a message; channels with limited capacity; transmitting information with bits; compressing data; error-free communication. Shannon's concepts and results were not just surprising but counterintuitive. They went against some of the most fundamental beliefs in the communication field. Veteran engineers hardly welcomed—or believed—the theory.[9] Some thought the paper was confusing and badly written.[10] One mathematician was totally skeptical.[11] Even some of Shannon's colleagues didn't immediately understand his ideas; they found them interesting but not very useful.[12]

But at the same time, others at Bell Labs and places such as the Massachusetts Institute of Technology did recognize the significance of the work. And they were stunned. "I can't think of anybody that could ever have guessed

that such a theory existed," says Robert Fano, an emeritus professor of computer science at MIT and a pioneer in the field. "It's just an intellectual jump, it's very profound."[13] Fano and others showed that Shannon's information theory was not only correct—it was revolutionary.

And a revolution did come. Not in the 1940s. But today. CDs, DVDs, cell phones, fax machines, modems, computer networks, hard drives, memory chips, encryption schemes, MP3 music, optical communication, high-definition television. All these things embody many of Shannon's ideas and others inspired by him. Every time we save a file in our hard drives, play a disc on a CD player, send an email, or talk on our cell phones we are relying on the concepts of bits and codes and channels originated in Shannon's 1948 paper.[14]

But information theory, as Shannon has always said, didn't come from scratch in a single flash of inspiration. It was the result of several years of work in different places and contexts. He circulated among some of the brightest scientific minds of the twentieth century. As a student at the MIT, he worked under Vannevar Bush, the man who directed U.S. science efforts during World War II. Also at MIT, Shannon met and took a course with the brilliant—and eccentric—mathematician Norbert Wiener, the father of cybernetics. At the Institute for Advanced Study in Princeton, Shannon worked under another great mathematician, Herman Weyl, and met figures such as John von Neumann, Kurt Gödel, and Albert Einstein—Einstein once showed up for one of Shannon's lectures, but was apparently looking for the tea room and left.[15] During the war, at the Bell Telephone Laboratories, one of the top research organizations in the world, Shannon worked with the godfathers of control engineering, Hendrik Bode and Harry Nyquist. He also worked on cryptography and had lunch several times with a leading figure in the field, the British mathematician Alan Turing.

What path led Shannon to information theory? This was the question that Bob Price, a communication engineer from MIT, wanted to clarify. He left Shannon's house on that evening in 1982 with only one certainty: the answer was much more complex—and fascinating—than he expected.

* * *

"THE MOST MATHEMATICAL of the engineering sciences." That is how Shannon once defined the study of communication systems, for it combined both fields—mathematics and engineering—in a unique way.[16] Entering the University of Michigan in 1932 he still wasn't sure which he liked the best.[17] As a boy, he used to play with radios, remote-controlled models, and other electrical equipment. He once set up a telegraph line to a friend's house half a mile away using the barbed wires of a nearby pasture.[18] But math and abstract problems such as the cryptograms of Edgar Allan Poe's *The Gold Bug* also very much interested him. So while an undergraduate at Michigan he took courses in both electrical engineering and mathematics and graduated with two diplomas.[19] One day in 1936, wondering about a job, Shannon saw a notice for a position of research assistant at the Massachusetts Institute of Technology.[20] A group headed by the dean of engineering Vannevar Bush needed someone to operate an early type of mechanical computer, a machine known as the differential analyzer. The job seemed ideal for Shannon's skills and interests.[21] He applied and was accepted.

Shannon's work consisted in setting up differential equations into the machine for MIT's mathematicians, physicists, and visiting researchers. The differential analyzer was partly controlled by over a hundred relays, electromechanical switches largely used in the telephone system to route calls. Telephone engineers had to design and test intricate circuits containing thousands of relays, a complicated and tedious task that used to overwhelm them. In 1937, Shannon spent the summer working at Bell Labs, the research arm of the American Telephone and Telegraph Company, or just AT&T, the largest communication company in the country. Immersed in that environment, where the relay played a crucial role, and drawing on his experience with the differential analyzer at MIT, Shannon started thinking of a better way to study and design relay circuits.

The solution became his master's thesis, in which he showed how the algebra of logic invented by nineteenth century mathematician George Boole could facilitate enormously the design of relay circuits.[22] Shannon noticed that this kind of circuit is formed by switches that could be either on or off, while in Boolean algebra a complex statement—reasoning, Boole argued—was formed by simpler statements that could be either true or false. For Shannon, the connection was clear: on and off, true and false. The symbolic operations devised by Boole to manipulate a set of true-and-false statements could be used by engineers to design circuits with on-and-off relays—or any kind of switch, for that matter. To represent closed and open switches, he chose two symbols: `0` and `1`. Using Boolean algebra, Shannon transformed the design, analysis, and test of complicated relay circuits into an abstract, mathematical manipulation of zeroes and ones—something that engineers could easily do with pencil and paper.

In 1938, Shannon submitted this work to a conference of the American Institute of Electrical Engineers. His mentor, Vannevar Bush, soon received a letter from the organizers: the paper had been accepted and very much impressed the reviewers, one of which said: "To the best of my knowledge, this is the first application of the methods of symbolic logic to so practical an engineering problem. From the point of view of originality I rate the paper as outstanding."[23] Indeed, the paper surprised even Shannon's closest colleagues. "We used to talk about switching circuits, and I showed him some telephone diagrams I had," says Amos Joel, a friend of Shannon at MIT. "But all of a sudden—I really don't know how—he came with this whole idea of using Boolean algebra."[24] The importance of the work was immediately recognized, and Shannon was awarded the prestigious Alfred Noble Prize (an engineering award, not Sweden's Nobel Prize). His work, once called "one of the most important master's theses ever written,"[25] laid the basis of digital circuit design, an essential tool for the microelectronic industry—nineteenth-century logic made possible today's twenty-first century information technology.

At MIT, besides the recognition from the faculty and colleagues, Shannon's mathematical talent brought him also some unexpected situations. His enrollment in a flight training program raised concerns in the professor in charge of the course. The professor found Shannon "unusual" and went to talk to other members of the faculty who knew the young student. "From these conversations," the professor wrote in a letter to MIT President Karl Compton, "I am convinced that Shannon is not only unusual but is in fact a near-genious [*sic*] of most unusual promise." The professor asked Compton if Shannon should be withdrawn for any life risk however small wasn't justified in his case.[26] "Somehow I doubt the advisability of urging a young man to refrain from flying or arbitrarily to take the opportunity away from him, on the ground of his being intellectually superior," answered Compton. "I doubt whether it would be good for the development of his own character and personality."[27]

Once he had finished his master's thesis, Shannon began to look for a topic for a PhD dissertation. At about that time, Bush had been named president of the Carnegie Institution, a private, nonprofit research organization. The Institution had a department of genetics at Cold Spring Harbor in New York and Bush suggested that Shannon spend a summer there. Perhaps he could do for genetics what he had done for circuit switching. Shannon then worked on a PhD dissertation on theoretical genetics, which he completed in less than a year. He enjoyed being a "geneticist" for some time but didn't plan to stay in the field.[28] Shannon was driven by an endless curiosity and his interests were very broad. While studying genetics and learning how to fly, Shannon continued to work on Boolean algebra and relay circuits; his main project was to build a calculating machine to perform symbolic mathematical operations.[29]

But among all these things, Shannon still found time to work on a subject that had always interested him: the problems of communication—"the most mathematical of the engineering sciences," as he put it. "Off and on," he wrote in a letter to Bush in February 1939, "I have been working on an analysis of some of the fundamental properties of general systems for the transmission of intelligence,

including telephony, radio, television, telegraphy, etc." "Intelligence" was how engineers called the various electrical signals that flowed in the communication systems—a term they would soon replace for another: *information*. In his letter, Shannon included a series of equations and concluded, "There are several other theorems at the foundation of communication engineering which have not been thoroughly investigated."[30]

The place Shannon found to explore these ideas further was the Institute for Advanced Study at Princeton. Before graduating from MIT, he had applied for a one-year fellowship there. In the spring of 1940, he received the good news: he had been accepted.[31] Arriving at Institute later that year, Shannon went to his advisor, the German mathematician Herman Weyl, and said he wanted to work on problems related to the transmission of information.[32] Weyl, a disciple of David Hilbert—considered one of the greatest mathematicians of the twentieth century—had left Germany with the Nazis' rise to power in 1933. Weyl showed interest in Shannon's ideas, and soon they were discussing analogies between the transmission of information and the Heisenberg uncertainty principle.[33] At Princeton, at age 24, Shannon met several famed scientists. John von Neumann, in particular, impressed him a lot—"the smartest person I've ever met," he would say years later.[34] Einstein was also there. In the morning, driving to the Institute, Shannon used to see the German physicist walking in his bedroom sleepers; Shannon then used to waive at him—and Einstein waved back. "He didn't know really who I was," Shannon recalled later. "Probably he thought I was some kind of weirdo."[35]

Shannon attended many seminars at Princeton, not always related to the kind of thing that he was interested in. The mathematics department was oriented to pure mathematics or physics, not to engineering problems like communication.[36] Nevertheless, it was at Princeton, in 1940, that the first ideas of information theory began to consolidate in Shannon's mind. He seemed ready to work full time on the subject that interested him so much. Perhaps in a year or so he would work out and publish some significant results. But then the war came.

* * *

THE BELL TELEPHONE Laboratories at Murray Hill, in New Jersey, consist of a dozen new and old interconnected five-story buildings spread over an area of about a hundred acres. The redbrick style and the green areas make the place look like a college campus. Long corridors full of large pale-painted doors spread like a labyrinth all over the buildings. The complex was built during World War II to expand the research activities of the laboratories, then held almost entirely at its headquarters at 463 West Street in Manhattan. If New York focused on engineering, Murray Hill was created to focus on science[37], a successful initiative that would transform Bell Labs into one of the best industrial laboratories in the world. Following the end of the war, the organization experienced a very intense and lively period of activity, now seen as the golden years of the labs.

Shannon came to the Bell Labs by the summer of 1941. As he recalled later, he didn't fancy the idea of going to war and thought he could contribute a lot more working full time for the science and military joint effort.[38] The war was under way in Europe and the involvement of the U.S. was imminent. In fact, in June1940, President Franklin Roosevelt established the National Defense Research Committee.[39] Directed by Vannevar Bush, the committee's objective was to mobilize U.S. science to the war efforts, and one of the priorities was to address what Bush called "the antiaircraft problem." Airplanes were flying higher and faster, and traditional gunnery wouldn't be able to shoot them down. More agile and reliable gunfire control systems had to be developed. Bush was convinced that this technology wasn't receiving the necessary attention.[40]

To address this problem he called Warren Weaver, a mathematician and also a skillful administrator, who was working as a director for the Rockefeller Foundation, a major source of funding for innovative science. Weaver was put in charge of a NDRC division on gunfire control systems and quickly set dozens of groups to work on the problem, including industrial laboratories, companies, and

universities. Bell Labs received a major contract and became one of the largest groups. But why was a band of communication engineers suddenly working on antiaircraft technology?

The group at Bell Labs had realized that the aim of a gun and a telephone call had a lot in common. Not long before the war, engineers began to consider text, speech, and images as a single entity—electrical signals—that could flow in the telephone network and had to be transformed, switched, and amplified. Now the same approach applied to fire control: the coordinates of an enemy airplane had to be transformed into electrical signals that could flow inside the gunfire control systems and had to be manipulated.[41]

The Bell Labs team was working on a kind of electronic computer that could track a plane, estimate its future position, and aim a gun, a fraction of a mile ahead, so that the shell had time to get there—and shoot down the target. The coordinates of the plane were supplied by optical equipment devices similar to telescopes or by radar. The problem was that these devices were not perfectly accurate. The coordinates they provided had errors and deviations—"noise." If plotted in a graph, the coordinates wouldn't form a smooth line, but a wildly zigzagging curve. If these coordinates were used to aim a gun, the shell would hardly destroy the target.

The problem of predicting the future position of a plane with noisy coordinates was solved by Norbert Wiener, at MIT. A child prodigy, Wiener received his B.A. in mathematics from Tufts College in Boston in 1909 at age 15.[42] In the fall of the same year he entered Harvard as a graduate student and four years later received his PhD degree with a dissertation on mathematical logic. Wiener continued his studies in Europe, where he worked under Bertrand Russell, G. H. Hardy, and David Hilbert. After World War I, he came back to the United States and joined the mathematics department at MIT.[43] Wiener realized that the prediction problem required a statistical treatment because the coordinates varied in an unpredictable manner—the trajectory was unknown and the errors were random. Much as in the case of weather forecast, it was necessary to evaluate past

and present conditions to predict a future situation. Wiener then developed mathematical tools to analyze the statistical behavior of the noisy coordinates, "filter out" the errors—or "smooth" the data—and estimate the future trajectory of the target.[44] He worked out a complete mathematical solution that for the first time clearly combined the fields of statistics and communication engineering.[45] His novel results were published in a report of restricted circulation dubbed the "Yellow Peril"—because of its yellow cover and its frightening mathematics.[46]

The Yellow Peril was an influential publication during wartime,[47] and Shannon read the document with interest.[48] As a researcher in Bell Labs' mathematics department under the legendary Hendrik Bode (every electrical engineering student today knows "Bode plots" from control textbooks), he was also working on the problem of the trajectory prediction. While engineers at the labs developed the machine itself—the hardware—Bode's group worked on the "software." They realized that Wiener's solution was mathematically perfect, but not the best one to be implemented in practice; it assumed, for example, a signal varying infinitely in time, while real trajectory measurements lasted just a few seconds.[49] To attack the problem and devise a more practical solution, the group drew on their knowledge of communication systems. In a classified report, the authors, Shannon among them, noted that "there is an obvious analogy between the problem of smoothing the data to eliminate or reduce the effect of tracking errors and the problem of separating a signal from interfering noise in communications systems."[50] Electronic filters and other concepts and devices developed for the telephone network could now be used in gunfire control systems. In that report, they proposed a better way to remove the errors from the noisy coordinates, including details on how to implement this solution with an eletromechanical computer.

The solution came in the end of the war and was not used in the Bell Labs' gunfire control systems that were sent to the battle field—and which played an important role in the war, shooting down thousands of V-1s, the rocket-powered bombs fired by the Nazis against targets in England.[51] But the results obtained by

the mathematics group and also by Wiener at MIT and others in the NDRC's division on gun fire control proved important to advance the understanding of the means we use to represent the world—be it speech or airplane coordinates—in the realm of electronic machines.[52] When electrical signals flow in gunfire computers, in the telephone network, or in any other system, we are dealing essentially with the very same process: transmission and manipulation of information.

* * *

NOT LONG AFTER the beginning of the war, Shannon was working also with cryptography, another of his assignments. Bell Labs had several projects on secrecy systems, especially on speech scrambling, techniques to protect a telephone conversation from eavesdroppers. One of these projects was possibly the most secretive of the whole war: the development of an encrypted radiotelephone system to connect Washington and London—known as the "X System."

Engineers at Bell Labs had been experimenting with various methods to scramble speech. But when they found one they thought was fairly good, someone always figured out a way to break it. They then turned to the only system they knew could create virtually unbreakable messages: the telegraph.[53] In telegraphy, messages are sent by the opening and closing of contacts in an electric circuit. When the contacts are closed, an electric pulse is sent—a "dot" is a short electric pulse and a "dash" is a longer one. When the contacts are open, no electric current flows in the line—and a "blank space" is sent. The telegraph system, therefore, used the presence and absence of current, or sequences of on-off pulses, to represent all messages. Engineers knew that by combining a message with a random sequence of on-off pulses known just to the sender and the receiver—which they called a "key"—they could obtain a perfectly secure message; this encrypted message would also be a random sequence of on-off pulses and there

was no way to attack this cryptography scheme—it was virtually unbreakable.[54] Could the same idea be applied to speech?

In the telephone system, speech is transformed into an electric signal that varies proportionally to the air vibrations of a person's words and sounds. The signal is analogous to the air vibrations it represents—and we call this signal "analog." Bell Labs researchers realized it was difficult to scramble analog signals; sometimes just by listening carefully to an analog-encrypted conversation it was possible to understand what was being said. They needed, therefore, something similar to the on-off pulses of telegraphy.

The solution came with two techniques now known as "sampling" and "quantization." The idea was to approximate a continuous signal by a series of "steps"—as if we superimpose the continuous signal by what seems a stairway that goes up and down following the shape of the signal. When we chose the number of steps used in the stairway, we are "sampling" the signal (each step is a sample). Also, we can imagine that each step in the stairway has a different height from the "ground;" when we determine these heights, we are "quantizing" the signal. In this way, a continuous signal is transformed into a discrete sequence of numbers. Now, this sequence of numbers could be combined to a random numeric key using special computing operations to create an encrypted conversation.

Throughout the war, the X System was used by the high commands in the United States and in England to work the war strategy, sure that the enemy couldn't eavesdrop the conversation. It was one of the first *digital* communication systems[55], for it transmitted information with digits, while analog systems use continuous, "analogous" signals.

The development of the X System was very secretive and Shannon had no access to all the details—"They were the most secret bunch of people in the world," he recalled years later.[56] He worked on a small part of the puzzle without seeing the whole picture. Nevertheless, he played an important role in the project; Shannon was asked to inspect the "heart" of the system: the encryption scheme.

His job was to verify that nothing had been overlooked and that the method used was really unbreakable.[57]

During that time, Shannon's interactions with the other researchers were very restricted, and many of his interlocutors couldn't discuss what they were doing. One of these interlocutors was Alan Turing. An eminent British mathematician, one of the world experts in cryptography and secrecy systems,[58] Turing was a leading member of the team that broke the Nazi secret code "Enigma." His computing machines at Bletchley Park, fifty miles northwest of London, deciphered the encrypted messages and delivered Hitler's plans straight to Prime Minister Winston Churchill's desk.

In January 1943, Turing came to Bell Labs in New York to consult on speech scrambling. He stayed in the country for two months and had occasional conversations with Shannon.[59] They couldn't discuss Turing's work on the Enigma nor their work at the laboratories. So during lunch or at teatime in the cafeteria, they talked about things that interested both of them and that they could discuss freely, things such as computers and the possibility of a machine simulating the human brain.[60] And what about systems capable of manipulating and transmitting information? Working on the breaking of the Enigma, Turing had developed a kind of measure of information very similar to the one Shannon would develop in his 1948 paper. Shannon's unit of information was the "bit." Turing's was the "ban."[61] Did the British mathematician contribute any insight to information theory? "He somehow didn't always believe my ideas," Shannon told Bob Price. "He didn't believe they were in the right direction. I got a fair amount of negative feedback."

Despite their disagreement, Shannon and Turing were working on fields that shared many concepts and methods—but with different goals. While in cryptography you want to protect a message from eavesdroppers, in information theory you want to protect a message from transmission errors. In both fields you need a measure of information and you deal with coding and decoding methods. So during the war, while studying cryptographic techniques and consulting on

projects like the X System, Shannon could carry on a parallel work, the one he started at Princeton. He could work on the problems of transmission of information.

In 1945, he wrote a classified report titled, "A Mathematical Theory of Cryptography," in which he used probability theory to study the subject in an unprecedented mathematically rigorous way. To some, the paper transformed cryptography from an art to a science.[62] In this work, Shannon introduced several concepts that would appear later in his "A Mathematical Theory of Communication," terms such as choice, information, and uncertainty, revealing the close connection between the two fields. In fact, in a footnote in the beginning of that report he wrote: "It is intended to develop these results in a coherent fashion in a forthcoming memorandum on the transmission of information."[63] Shannon was literally announcing what would be his 1948 seminal paper. Towards the end of the text, when discussing some problems in the encryption of English text, he used yet another term, one that would become always associated with his own name: "information theory."[64]

* * *

ONE DAY IN 1947, Shannon went to his colleague Brockway McMillan, a mathematician from MIT who had recently joined Bell Labs, and asked for help with a problem. Shannon said he needed a proof for a theorem related to the reliability of communication and sketched a diagram on a piece of paper. On the left side, he drew a bunch of points, each below the other. He did the same thing on the right side. He then connected some of the points: from one point on the left departed several lines to points on the right. The resulting figure resembled a paper fan. Shannon's explanation, however, was not very clear; he didn't express the problem in mathematical terms so that McMillan could really understand what had to be proved. McMillan found those ideas somewhat obscure. He couldn't help his colleague.[65]

Shannon worked on information theory almost entirely alone. He eventually talked about it with some colleagues, but never officially and never in detail as he did for antiaircraft control systems or cryptography. As in most of his works, Shannon waited until he could grasp the problem clearly in his mind to finally write up his results. It was not different with his 1948 paper. "There were no drafts or partial manuscripts," wrote Robert Gallager, a professor of electrical engineering at MIT, in a recent paper on Shannon. "Remarkably, he was able to keep the entire creation in his head."[66] For Shannon, information theory was almost a hobby, which he kept in the scarce spare time he had. During wartime, many researchers worked ten hours a day, seven days a week.[67] He worked on information theory sometimes at night, at home, not during office hours.[68]

Shannon worked first at the laboratories' headquarters in Manhattan, during the beginning of the war, and later moved to Murray Hill when the first buildings were erected around 1941. Not long after, he got his own office on the fourth floor of Building 2 on the east part of the campus, a sizable space with two generous windows overlooking a green area. His door was closed most of the time. While his colleagues got together to fly kites and play word games during lunch, Shannon preferred to stay alone, working on his own.[69]

The mathematics department functioned as a consulting group, providing expert advice to other researchers from Bell Labs or other organizations such as the military.[70] But its members also worked on their own projects, they could pursue "exploratory" research that wasn't dictated by an engineering department.[71] Shannon very much appreciated this freedom to follow his own interests. He liked to lock himself in his office and just think.

Shannon did interact with others. He liked to laugh and play jokes. Sometimes he invited a colleague for a chess match in his office. The matches attracted others, who watched by the door while Shannon, most of the time, beat his opponent. "He wasn't a great champion, but he was quite good," says McMillan, who occupied the office next door to Shannon's. "Most of us didn't play more than once against him."[72]

But it was alone that Shannon created information theory. In an interview in the late 1970s, he said his ideas were most developed around 1943 to 1945. Why it wasn't published until 1948? "I guess laziness," Shannon said, adding that those were busy times because of the war and information theory "was not considered first priority work." It was cryptography that allowed him to work on information theory. "That's a funny thing that cryptography report," he told Price about the 1945 confidential memorandum, "because it contains a lot of information theory which I had worked out before, during those five years between 1940 and 1945."[73] In fact, the connection between the two fields is so straight that many believe that cryptography originated information theory. This is not true, as Shannon was thinking about the problems of communication much before coming to Bell Labs. But the question thus remains: what led Shannon to his novel approach to communication, to his 1948 paper, to the ideas that stunned engineers? If it was not cryptography, what did? Price asked this question to Shannon on that evening in 1982 and his answer was: "The real thing for me was Hartley's paper."[74]

* * *

RALPH HARTLEY JOINED the research laboratory of Western Electric, the manufacturing arm of AT&T, in 1913. He started working on transatlantic wireless communication and later with telephone and telegraph systems. In 1925, AT&T and Western Electric combined their engineering departments to form Bell Labs, and Hartley became a member of the new organization. Three years later, he published an important paper titled "Transmission of Information" in the *Bell System Technical Journal*, an in-house publication.[75]

In this work, Hartley proposed to find a quantitative measure for the transmission of information that could be used to compare various systems, such as telegraph, telephone, and television. At that time, the very notion of information was fuzzy and he began the paper clarifying what the concept meant

for engineers. Hartley considered it necessary to eliminate what he called the "psychological factors" in communication.[76] He imagined an operator sending a message in a telegraph cable. Opening and closing the contacts of the telegraph, the operator sends a sequence of symbols down the line. What messages should be considered "valid"?

We can imagine, for example, operators who speak different languages. A message that is not intelligible to one operator could be meaningful to another. In other words, from the engineering standpoint, the meaning of a message is not important. Given a set of symbols such as an alphabet, any combination of characters should be considered "valid." In fact, Hartley imagined that instead of a human operator, the message could be generated by an automatic mechanism, a random event such as a ball rolling into pockets that determine the opening and closing of the telegraph contacts. This randomly generated message should be considered as "valid" as any other; the telegraph system should be able to handle and transmit it as well as any other message.

For Hartley, therefore, communication could be thought of as the successive selection of symbols from a finite set of symbols. A person, for example, mentally selects words from a vocabulary to talk to another. "By successive selection a sequence of symbols is brought to the listener's attention," he wrote. "At each selection there are eliminated all of the other symbols which might have been chosen."[77] Hartley observed that a symbol conveys information because there are other possibilities to it. You can't communicate if your vocabulary has just a single word. The more words you have, the more *choices* you can make—and the more information you can communicate. Hartley expected that by measuring this "freedom of choice" he could obtain a measure of information.

But it couldn't be a direct measure. It couldn't simply be that as choices rise, information rises because these two variables—the number of choices and the amount of information—increase in different rates. Suppose a telegraph operator has an alphabet with 26 letters. When sending a message with, say, one

single letter, there are 26 possible choices. For two letters, the number of choices increases to 676 (there are 26 times 26 possible combinations for two characters). For three letters, the number is much higher: 17,576. If we continued in this way, the number increases extremely fast—a so-called exponential growth. Hartley realized that, while the number of possible choices increases exponentially, there is no such exponential increase in the information conveyed. For a communication engineer, a telegram with 20 characters should convey twice the information of one with 10 characters.

Is there a way to make the measure of information proportional to the number of selections? In other words, is there a way to make each further selection add a fixed amount of information to the total so that, instead of growing exponentially, this total amount grows linearly? Hartley found a way to accomplish that using the mathematical function that is the "inverse" of exponentiation. This function is the logarithm. If you "apply" the logarithm to an exponential curve, the curve becomes a straight line. Using the logarithm, Hartley derived a formula that transformed the exponential growth of choices into a linear growth of information.[78] With this formula, a twenty-character telegram would contain twice the information of a ten-character one—just as expected.

But so far Hartley had considered just discrete processes, such as a telegraph operator who chooses one symbol from a set of symbols. What about other forms of communication, such as telephone, radio, or television, in which information is sent as continuous signals? Hartley demonstrated that the problem is essentially the same: a continuous signal could be approximated by a series of successive steps that follow the shape of the signal. This is exactly the idea of "quantization" used by Bell Labs engineers in the X System. So communication in the continuous case could also be considered a succession of choices—the choices being the "heights" of the steps—and the amount of information could be quantified in the same way.

Hartley had a measure of information that could in principle be used for discrete or continuous signals. Now he could analyze what limits the transmission

speed of a system. Bell Labs engineers knew that they could send images over the "air," from one antenna to another, as in the television system. But why was it so hard, as experiments showed, to send images over a telephone line?[79] What makes one channel so different from the other and how much information each could transmit?

In most communication systems, information is transmitted by an electrical signal. Suppose you are sending a signal that goes up and down continuously, in a smoothly varying way. Imagine you are holding one end of a string, the other end fixed, and you move your hand up and down with a constant speed. In this case, the undulation in the string varies regularly and endlessly, always in the same way—it is predictable. A signal like this has always the same frequency and it conveys no information. But suppose you start to move your hand wildly, with varied speeds. Now you are generating an undulation much more complicated, with an arbitrary and unpredictable shape. This signal contains not only one but several frequencies—it is made up of several components. The more frequencies a signal contains, the more rapidly it can change—and the more information it can convey. This range of frequencies is called bandwidth. If a signal contains frequencies from 200 to 3,200 hertz, for example, its bandwidth is 3,000 hertz. This is approximately the bandwidth required to transmit a telephone conversation. A television transmission requires roughly two thousand times more bandwidth, or 6 million hertz.[80] That is why it was so hard to send an image over a telephone line: the bandwidth of this channel is too "narrow." The art of communication, therefore, is to match the message to the medium.

Very close to Hartley, another Bell Labs researcher explored similar ideas. Born in Sweden, Harry Nyquist came to the United States at age eighteen and studied electrical engineering. In 1917, he received a PhD degree in physics from Yale University and joined AT&T.[81] Nyquist's seminal work on control engineering would transform him into a legendary figure in the field.

Nyquist made important contributions to communication as well. In the 1920s, he studied the transmission of signals in the telegraph system with a

mathematical tool known as Fourier analysis, which decomposes a complicated signal into a sum of simpler components.[82] This allowed him to understand deeply how these components affect the transmission speed. In accordance to Hartley, he also concluded that the bandwidth defines how much information you can send over a channel.[83] But Nyquist went further in understanding the boundaries between continuous and discrete representations. He showed that once you had approximated a continuous signal with a series of discrete steps, you could use these steps to reconstruct not just a similar signal, but an *identical* one. The finite bandwidth of a signal limits the amount of information it can carry. So if you had enough "steps," the reconstructed signal would be a perfect copy of the original. "The crucial point is that a finite amount of information implies an essentially discrete message variable," wrote James Massey, a former professor of electrical engineering at ETH, Zurich's Swiss Federal Institute of Technology, in 1984. "The world of technical communications is essentially discrete or 'digital'."[84]

Hartley and Nyquist sought a way to measure the maximum transmission speed of a given system. They attacked the problem not only by inspecting electric currents with oscilloscopes and voltmeters; they used mathematics and abstract tools in a novel approach to understand communication. Especially important was the connection they established between discrete and continuous signals. "Nyquist's and Hartley's notions began to resemble digital representations, as their techniques analyzed the subtleties of converting discrete pulses to and from the continuous world," writes David Mindell, a historian of technology at MIT, in the book *Between Human and Machine*. "These men laid the groundwork for the theory of information that Claude Shannon would articulate in 1948."[85]

Indeed, Shannon cites both in the very first paragraph of "A Mathematical Theory of Communication." Hartley's paper, in particular, which Shannon read while a student at Michigan, impressed him very much.[86] He mentioned Hartley's work in his February 1939 letter to Vannevar Bush.[87] And he mentioned it also to Herman Weyl when he arrived at Princeton.[88] Shannon thought the paper was

very good, but it left a lot more to explore. His thinking on information theory, as he recalled later, began with Hartley's paper.[89]

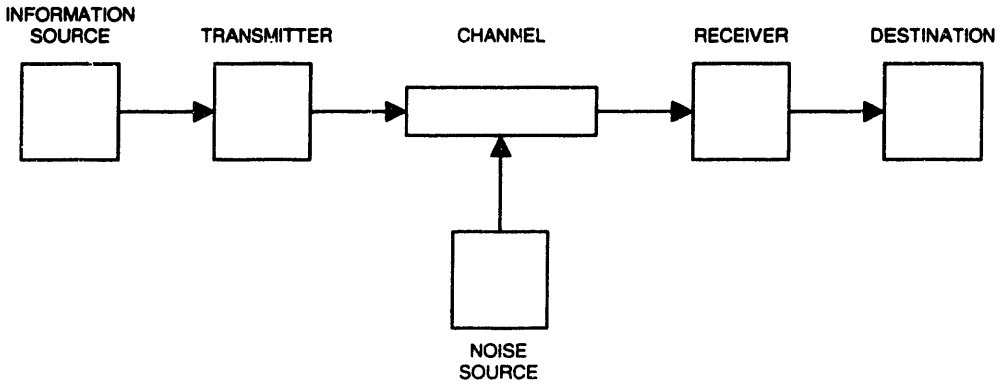
After Hartley's and Nyquist's work in the 1920s, communication theory "appears to have taken a prolonged and comfortable rest," in the words of John Pierce, a Bell Labs and an eminent communication engineer.[90] World War II brought important advances, but the field remained a kind of technological archipelago, with several islands of knowledge that didn't talk to each other. Radar, developed during the war, was still a secretive subject. The telephone system was operated almost entirely by AT&T. At universities, communication engineering textbooks amounted to the impressive quantity of two; students learned radio transmissions and technologies such as AM and FM.[91] Communication had advanced significantly during wartime, but it was far from a unified science.

* * *

SEVENTY-SEVEN PAGES, twenty-three theorems, seven appendixes with mathematical proofs. A lengthy paper for today's standards, "A Mathematical Theory of Communication" was published in two parts in the July and October 1948 issues of the *Bell System Technical Journal*, then widely circulated among electrical engineers and those interested in the communication field. The paper presented a unifying theory of communication, a mathematical theory that didn't depend on a particular system or technology.

Shannon imagined communication as the flow of messages along a series of stages, which he represented schematically with a block diagram (*see figure below*). The diagram showed a series of blocks, one connected after the other, through which a message is transmitted. The first stage is the *source*, where the message is originated. The second stage is the *transmitter*, which transforms—or encodes—the original message into a form suitable for transmission. The encoded message is then sent over the communication *channel*. During its way through the

channel, the message may be affected by errors. In other words, the channel is plagued by *noise*. And noise is everywhere: it is in space, where magnetic storms can disturb a signal; inside electronic equipment, where spurious currents can corrupt data; within an optical fiber, where energy losses degrade the light transmitted. It is impossible to eliminate all the noise from a channel—you have to live with it. When the encoded message leaves the channel, it reaches the *receiver*, which performs the inverse operation of the transmitter; that is, it decodes the message and delivers it to the final stage: the *destination*.



COMMUNICATION BLOCKS Shannon's classic schematic diagram of a general communication system. All practical systems can be broken down into parts that perform the same functions as these boxes.

Shannon's block diagram became a classic concept in the communication field. Its generality reflected the fundamental character of the theory. "All the many practical different schemes—radio, TV, carrier pigeon—can be broken down into parts that perform the same functions as these boxes," says David Slepian, a colleague of Shannon in Bell Labs' mathematics department.[92] With a simple figure showing a bunch of interconnected boxes, now the main questions of communication could be clearly formulated. How much information does a source produce? What is the best way to encode a message in the transmitter? Is

there a limit to the amount of information we can send over a channel? How badly will the noise affect the transmission? What characteristics of the receiver and the destination are important for the communication process?

Shannon began his paper by noting that frequently the messages produced by an information source have *meaning*. That is, they refer to things or concepts that "make sense" to people. But from the engineering standpoint, Shannon observed, these "semantic aspects" were not important—agreeing with Hartley's attempt to eliminate the "psychological factors" in communication. For Shannon, too, any message selected should be considered "valid." What is meaningful to a person—a certain kind of music, a text in a foreign language—can be meaningless to another. And a system should be able to transmit *any* message. So what matters for communication engineering, Shannon said, is that a message is selected from a set of possible messages. As he recalled later, he wondered about various kinds of sources. What is the simplest one? Is there a fundamental information source to which we can compare all the others? Shannon then thought about the toss of a coin.[93]

Consider, for example, the case when you toss a coin and want to tell the result to a friend. This is the simplest source of information: there are just two outcomes—heads and tails—that are equally likely. (One might think about a coin with heads on both sides. But you don't need to toss this coin to know the outcome. So this coin produces no information.) So we can regard the toss of a fair coin as having unitary information. And there are many ways you can tell the result to your friend. You can simply shout the outcome. Or you can jump once for heads and twice for tails. Or you can agree on more complicated schemes, using a flashlight, smoke signals, or flags. No matter how you decide to communicate the result, the amount of information conveyed is always the same.

We can denote each possible outcome by the digits `0` and `1`—called binary digits. When we use binary digits, we are counting in the binary system, or what is called "base 2" (as opposed to the "base 10," or the decimal system, which we normally use); in this case the unit of information is the *bit*. The story

goes that, one day during lunch, some Bell Labs researchers were thinking of a better term for "binary digit." What about binit? Or maybe bigit? John Tukey, one of the men at the table put an end to the discussion: the best and obvious choice, he said, was "bit."[94] To communicate the outcome of the toss of a coin—or any fifty-fifty probability selection for that matter—we need to send just one bit of information. We send either the digit 0 or the digit 1. The fact that there are *two* digits and only *one* has to be sent is at the very basis of the concept of "information": information can be conveyed by a digit because there exists an *alternative* to it. If there isn't another possible choice—as in the case of the coin with heads on both sides—the amount of information is zero.

For Shannon, information was a measure of uncertainty. But *not* in a negative way. Uncertainty in a sense of something newsy. You don't want to read last week's newspaper. You want today's paper because it brings things you don't know yet, are uncertain about. Communication, therefore, was the resolving of uncertainty. The more uncertainty, the more information needed to resolve it.

Hartley had used this same idea of a message chosen from a set of possible messages to derive his measure of information. His formula worked fine for the cases when the messages have all the same chance of being selected. Shannon, however, noticed that usually choices occur in a much more complex way. Communication can't be just like the toss of a coin. In written English, for instance, some letters are used more frequently than others. The letter E appears much more than Z, Q, or X. And more than that, we also form words, so a particular selection depends also on the previous ones. Take a common word like THE. So in English there is a great chance of an E to be selected after a T and an H. But Hartley's formula couldn't be used to reflect such situations. Actually, it couldn't be used even for the simple toss of an unfair coin, for it couldn't deal with the different probabilities for each face. Shannon realized he needed to generalize Hartley's formula, make it mathematically more powerful so that *any* information source could be "measured." To derive a new formula, he sought a

new mathematical model for the source, one that could generate complex things like English.

Shannon realized he needed a model that could produce messages based on complex rules of probabilities. One that could produce a selection depending on the previous selections made. And one that should be able to generate not some, but *all* possible messages of a certain source. The mathematical model that can do all that is known as a stochastic process. The interaction of molecules of boiling water in a pan, the fluctuations in the price of stocks, the "random walk" of a drunk in a sidewalk—these are all examples of phenomena that can be modeled as a stochastic process. They are essentially random events. At a certain instant, you can't predict the precise position of each water molecule, or the exact price of the stock, or where the drunk will be. But stochastic processes have a *statistical* behavior. And we can analyze statistics to draw certain conclusions. Thus using a stochastic model, physicists can deduce the temperature of the water; stock analysts can have an idea of the variation of prices; and mathematicians can estimate how far the drunk will go.

In information theory, a stochastic process is used as a model to generate the messages. It is a mathematical machine that runs endlessly spilling out the messages according to probability rules. And we can define the rules. You can start with simple rules and then introduce more and more rules—more constraints to how the messages are generated. Shannon gave some examples of how rules can change a stochastic process that generates English text.

You can start with the simplest model possible. Write each letter of the alphabet in small pieces of paper and put them in a hat. Now take one piece, write down the letter, and put the piece back in the hat. Shannon did this experiment with 26 letters and a space and included a typical sequence in his paper:[95]

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD
QPAAMKBZAACIBZLHJQD.

Keep doing this and, in principle, you will come up with all Shakespeare's plays. But the way things are written, be it a grocery list or *Hamlet*, is not like a bunch of monkeys typing randomly on a typewriter. Language has a statistical structure, and you can incorporate this structure into a stochastic model. You can create a model that takes into account that some letters appear more frequently than others. You can also include in the model the probability for a certain letter to be selected after, say, two letters were previously chosen. Shannon included an example of this case:

> IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE.

Rather than continue with letters, Shannon then jumped to word units. He considered the probabilities of two words appearing together and generated a sequence of this kind using words from a book:

> THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

Shannon then concluded that a stochastic process—in particular, one special kind of stochastic process known as a Markov process—could be a satisfactory model of English text. In more general terms, he noted, a sufficiently complex stochastic process could represent satisfactorily *any* discrete source.[96]

Once he had the model, Shannon proceeded to find a measure of information. "Can we define a quantity which will measure, in some sense, how much information is 'produced' by such a process, or better, at what rate information is produced?" he wrote in the paper. Shannon deduced a formula that

was a generalization of Hartley's—a kind of weighted sum that takes into account the probability of each possible selection. The formula can be applied to the simple throw of a coin, to a message written in any language, to a ball taken from a box with five red balls and three blue ones, or to any other probability process. In the simplest case of the toss of a coin, Shannon's formula assumes the following form:

$$H = -p \log p - q \log q$$

In this expression, H is the amount of information, measured in bits; p is the probability for heads and q the probability for tails; *log* means the logarithm, as used by Hartley. (Since we are working with the binary system, we use the "logarithm to the base 2," which means that the "log" in the formula is actually "$\log_2$". The logarithm of a number to the base 2 is how many times you need to multiply 2 to get the number. For example, the logarithm of 8 to the base 2 is 3, because you need to multiply 2 three times to get 8. Others bases can be chosen, but in information theory the base used is 2, and the unit of information is the bit.)

For a fair coin, heads and tails have the same probability, thus p and q have the same value: 50 percent. Plug these values in the formula and you get H equals to 1 bit, just as expected. Now imagine you have an unfair coin, one that has one side heavier than the other, weighted away from fifty-fifty probability. Lets say it gives heads 70 percent of the time and tails just 30 percent. Use the formula and H works out to be approximately 0.9 bit. The amount of information decreased. Does that make sense? Yes, because the unfair coin is more predictable; when it lands as heads, we are less surprised, we have learned less. So the unfair coin produces less information on average than the fair one. Toss ten fair coins and you get 10 bits. Toss ten unfair coins and you get 9 bits.

But of course, communication is not only about tossing coins. Information sources are things like a telephone conversation or a television broadcast. And in most systems, information is represented with electric signals. Shannon knew

from Hartley and Nyquist that a continuous signal could be converted into a discrete form and the theoretical results such as the measure of information would still hold. He adopted the "conversion" method described by Nyquist in his papers on telegraphy and developed it further into what is known now as the "sampling theorem."[97] And perhaps more importantly, Shannon knew that a system that transmitted discrete information could actually *work*. During the war, Bell Labs engineers had built things such as the X System and had explored communication schemes that mixed continuous and discrete signals.

The team working on the X System had found that the technique they developed to transform a signal into digits was very similar to one devised and patented years earlier by a British engineer. This method also consisted in approximating a continuously variable signal—like the analog signal of a telephone call—by a series of steps with different "heights" ("samples" of the signal). But instead of sending the numeric values directly, as in the X System, they would be converted to the base 2, that is, to numbers with just two digits: `0`s and `1`s. A "height" of 23, for example, would be transformed into the binary number, or code, `10111`. If a `1` means the presence of an electrical current and a `0` the absence of current, you can send any message using just on-off signals—uniform and unequivocal electric pulses. This method of transforming a signal into codes and then into pulses was called "pulse code modulation," or just PCM. It was widely studied at Bell Labs after the war and later implemented in various telephone equipment.[98] What we call now a digital communication system has its roots in PCM and other methods to manipulate and transmit information in discrete sequences of digits.

The terms "analog" and "digital" appeared nearly simultaneously during the war. Both technologies evolved together as computer and communication engineers decided which was better for the various problems they had to solve.[99] Shannon realized the potential of the discrete representation—that with digital information one could do things it was impossible with analog information. "Before Shannon, the continuous, or analog case was considered the basic one for

communications with the discrete, or digital case a kind of extreme special case," says Massey, from ETH. "Shannon reversed this viewpoint."[100]

In his 1948 paper, Shannon showed that one of the advantages of a digital system was that you could choose how you represent your message with bits—or how you *encode* your information. We can, for example, analyze statistical patterns of the messages produced by a source and use this knowledge to compress the messages before transmission. Consider the case of English text. We could use the same number of bits to encode each letter. We could assign, say, `00001` for `A`, `00010` for `B`, and so on. But if `E` appears more frequently than `Z`, why use the same number of bits? We could, for example, assign just one bit for `E` and a code with more bits for `Z`. This same idea could be applied to all letters of the alphabet and by using this more efficient code we could save a good deal of transmission time. The idea of using efficient codes was not new. Samuel Morse used short combinations of dots and dashes for the most common letters and longer combinations for uncommon ones. While the letter `E` was represented by a single dot, an `X` was a dash-dot-dot-dash sequence.[101]

But Shannon's insight was to take this idea even further. We could assign codes with different lengths not just for the letters individually, but for pairs or groups of letters (`ED`, for instance, is very common, while `QZ` very rare). And we could do the same thing for words. Frequent words like `THE` would have shorter codes while infrequent words would have longer ones. It is possible to obtain such efficient codes because of the statistical nature of language—certain patterns repeat themselves. Shannon called these repetitive and predictable patterns "redundancy." Eliminating redundancy it was possible, on average, to compress information and save transmission time.

"Two extremes of redundancy in English prose," he wrote in his 1948 paper, "are represented by Basic English and by James Joyce's book *Finnegans Wake*." Basic English vocabulary, he noted, is limited to 850 words and has many repetitive patterns—the redundancy is very high. Joyce's book is full of new words that expand the language structure, reducing its redundancy.

Shannon also observed that the redundancy of a language is related to the existence of crossword puzzles. If the redundancy is zero, the language has no patterns—there are no constraints. So any sequence of letters is a word and any combination of them can form a crossword. If the redundancy is too high, the language has too many constraints and it is difficult to form arrays of words—they don't "cross" easily with each other. Shannon estimated the redundancy of English to be roughly 50 percent, about the right amount to make possible large crosswords (later he revised this value to 80 percent). "The redundancy of English," he wrote in an article for the *Encyclopaedia Britannica*, "is also exhibited by the fact that a great many letters can be deleted without making it impossible for a reader to fill the gaps and determine the original meaning. For example, in the following sentence the vowels have been deleted: MST PPL HV LTTL DFFCLTY N RDNG THS SNTNC."[102]

But how do you know *how much* a message can be compressed? How do you know you have the most efficient code? Shannon's measure of information gives the answer. His formula "captures" the statistical structure of a source and translates it into a single number given in bits. This number tells how many bits you have to use to encode the information in the most efficient way. What is the best code, say, for English text? Consider the case of letters written in pieces of paper and selected randomly—the monkeys in the typewriters. For this situation, using Shannon's formula, H equals to 4.7 bits. Since a communication system cannot transmit a fraction of a bit, you would need 5 bits on average for each letter. To send a text with 1,000 letters you would need 5,000 bits. Now consider you are encoding not single letters but groups of letters that form words. In this case, according to Shannon's calculations, H would be equal to just one 1 bit. That means that some words would be represented by long sequences of bits, other words by short ones, and on average, when you count the total number of bits and letters, it turns out you used just 1 bit for each letter. To send the same text with 1,000 letters you would need now 1,000 bits. So when you take into

account the statistical patterns of English you can obtain codes that represent the same amount of text with fewer bits.

These statistical gains are important not just for text. In fact, as Shannon recalled later, one of the motivations for his work on information theory was to determine if television signals could be compressed.[103] These signals required a lot of bandwidth—or a "huge" channel. Compressed signals would transmit the images faster, saving transmission time—and money for the broadcasting companies. Today we face a similar situation with other sources and channels. Anyone who uses a modem to access the Internet knows that transferring large images, high-quality music, or video might take a long time. And it would take much longer if it weren't for compressing codes based on Shannon's ideas. Multimedia files usually are huge, but like text, they also contain redundancy and other patterns that can be removed.

It was an efficient coding technique that made possible the recent music frenzy on the Internet, when millions of people suddenly started sharing their favorite songs with a few mouse clicks. The so-called MP3 audio format can transform a large audio file into a much smaller one. Usually, a ten to one compression rate can be achieved.[104] Take a high-quality audio file with, say, 50 megabytes, like the ones stored in music CDs. Over a modem connection that most people have at home, the transmission would take a seemingly endless two hours. Compressing the file with MP3, you get a 5-megabyte file that might be transmitted in about ten minutes. With coding, the message could match the medium.

* * *

BUT REDUNDANCY IS not always undesirable. In some cases, according to Shannon's theory, you want to *add* redundancy to a message. The reason is that these repetitive patterns can "protect" the message from errors. Before Shannon, engineers thought that, to reduce communication errors it was necessary to

increase the power of the transmission. When you make your signal stronger, it is more difficult for noise to affect the communication. But doing so has a high cost: it demands more energy, which means larger batteries or other sources of electricity. Another way they considered to deal with errors was to send the message repeatedly. If you send the letter A three times and because of an error one of them turns into, say, a B, you would still be able to get the right message: with a kind of "majority vote" you still could conclude that an A was transmitted. If instead of three you send the same letter five, ten, or one thousand times you can improve the reliability of the communication even more. But notice what is happening: now you need to send the same letter several times while you could be sending *other* letters. In other words, you are squandering your precious transmission time. The transmission rate is reduced by three, five, or one thousand times when you repeat the messages. To make the error rate go to zero, the repetition must increase indefinitely, which in turn makes the transmission rate go to zero—you are forever sending the same letter.

Shannon showed that the idea of adding redundancy was right, but the way engineers were doing it wasn't. He showed there was a much more efficient way to encode a message to have not just good, but perfect communication. He proved mathematically that using the proper codes data could be transmitted virtually free from errors. While the methods to compress information form what is called "source coding," this part of Shannon's theory dealing with error-correcting methods is known as "channel coding."

Imagine that you want to transmit four letters, A, B, C, and D. You decide to represent A as 00, B as 01, C as 10, and D as 11—the simplest binary code for four letters. But then you learn that the channel is "noisy." When you send a stream of bits over this channel, unpredictable errors affect the transmission, "flipping" the bits—a 1 arrives as a 0 or vice-versa. You might send an A (00), but in the channel the second bit is flipped, and what you get is 01. But this is a B. In other words: you got a transmission error. Is there a way to overcome this kind of problem?

Before Shannon, engineers recurred to the repetition method. You can decide, say, to send each letter three times repeatedly. To transmit B you send 010101. In this case, if one bit is flipped you don't get a wrong letter: you can correct the error. For example, if the first bit is flipped you receive 110101, which is closer to 010101 than any of the other three possibilities (000000, 101010, and 111111). What if more than one bit if flipped? Then there is no escape: you will get errors. This is, therefore, a one-error correcting code. It requires a total of six bits to transmit two "useful" bits (the encoded letters), which represents a transmission rate of two to six, or 33 percent.

Can we do better? Coding theory gives the answer: yes. The solution is to encode each letter with specially constructed sequences of bits. Suppose now that you represent each letter with *five* bits instead of six: you represent A as 00000, B as 00111, C as 11100, and D as 11011 (each sequence of zeroes and ones is called a codeword). Again, this is a one-error correcting code. If one bit is flipped, you still can get the right letter. But note that you have a more efficient code than before. Now you use five-bit codewords to transmit each letter. So the transmission rate is two to five, or 40 percent. In this case, you introduced the redundancy in a more intelligent way. Code designers study the noise in a channel to create very efficient codes that make erroneous situations very unlikely. They know they can't always win—but they know how to win most of the time.

In coding theory, the field that deals with source and channel coding, there is a measure of how similar two codewords are: the number of bits you have to flip to change a codeword into another is called Hamming distance, after Richard Hamming, a pioneer in the field and also a Bell Labs researcher. The Hamming distance for A and B in this last case is three, because you need to flip three bits of an A to make it a B—and vice versa. A Hamming distance of three is usually considered a good distance; that means the codewords have a good number of distinct bits. A basic rule that coding engineers learn is that, codes with a short Hamming distance can be messed up with each other. In systems like CD players and hard drives, special codes are used to expand the Hamming distance so that

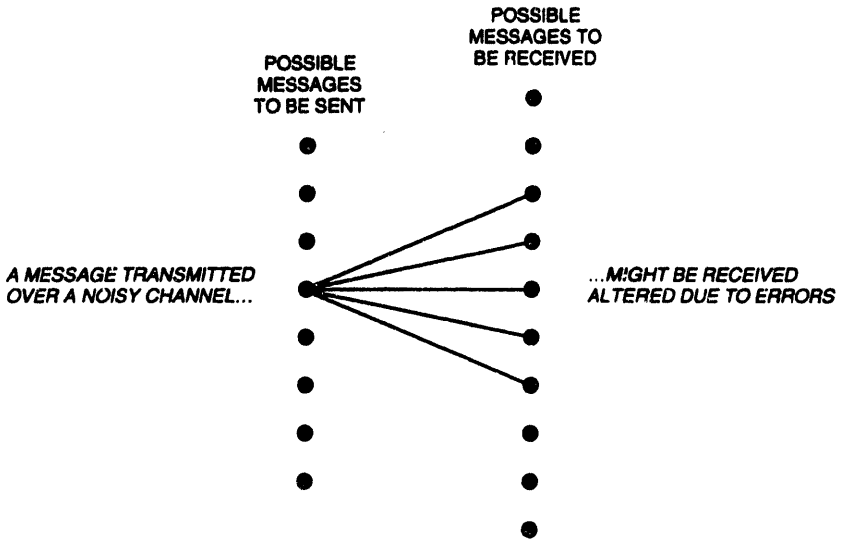
several errors can be detected and corrected. That is why you can scratch a CD and still get crystal-clear music. The coding schemes embedded in this equipment create complicated sequences of bits, usually adding "error-check bits" used to fix corrupted data.

Another important—and surprising—concept in Shannon's paper was that of channel capacity. Shannon showed that every communication medium has an upper limit for transmitting data, a quantity given in bits per second. If you try to send data above this threshold, you would necessarily get errors in the transmission. Towards the end of the paper, Shannon related bandwidth and error rate in an equation that gives the capacity of a channel. He put together the ideas developed by Hartley and Nyquist into a single mathematical expression that could measure the transmission limit of a telephone wire, a wireless connection, or an optical fiber. Shannon gave engineers an ultimate measure of how well a communication system could work. "Once you know you're close to the capacity, you know you're doing a good job," said MIT professor of electrical engineering Hermann Haus, a leading researcher in optical communication. "So you don't waste money trying to improve your system—you've reached the limits."[105]

Shannon's information theory showed that coding provided the perfect means to overcome error—a result that surprised many experienced communication engineers. "People were thinking of a better way to communicate in the presence of noise," MIT professor Robert Fano says. "But the notion that it was possible to eliminate completely the effects of noise was totally unknown, unthinkable."[106] Now engineers learned that, with coding, they could have error-free communication without reducing the transmission speed nor increasing the power of the transmitter. If you had a channel with capacity of, say, 1 million bits per second, you could have an error rate as low as desired and still use the channel at 1 million bits per second—all you needed was the right code. Shannon called this result the "fundamental theorem" for a discrete channel with noise.

To prove the theorem he used schematic representations of the relations between transmitted and received messages—the "fan diagrams" (*see figure*

below). A message sent over a noisy channel can arrive at the receiver altered. But if this set of altered messages refers to one—and only one—of the original messages, there will be no confusion in the receiving end. Roughly speaking, the code used has to guarantee that the fan diagrams will not overlap.



FAN DIAGRAM The figure shows a typical "fan diagram," as this kind of schematic drawing was known to MIT graduate students in the 1950s. The points on the left represent the messages to be sent. Due to errors in the channel, a transmitted message can arrive altered (points on the right connected by lines).

Using a few laws of probability theory, Shannon elegantly proved his fundamental theorem and demonstrated that such codes must exist. But he didn't showed how to obtain them—a task that would keep thousands of coding theorists and communication engineers busy for several decades. Moreover, the possibility of error-free transmission up to the channel capacity had an intrinsic "side effect." The tradeoff is that the more errors you want to correct and the closer you get to the capacity, the more complicated is the encoding of the information. That is, the codewords become longer and longer. And you need to analyze extremely huge chunks of data to find statistical patterns that optimize these codes. So you spend

a long time encoding and decoding very long sequences of bits, which translates into a delay that in practice might be unacceptable. You still send your data at 1 million bits per second—but you had to sit and wait for the messages transmitted to be reconstructed.

Engineers had always sought a balance between transmission speed and error rate. Shannon's theory gave them full control of these parameters. An optimum transmission could be achieved in two clear steps. First, removing the unnecessary redundancy of a message, the unnecessary bits. And then adding the right kind of redundancy—"error-correction bits"—to make the transmission error-proof. Shannon showed that, with the proper source and channel coding, it was possible to construct the perfect message for each communication system. These ideas were so surprising and counterintuitive that ETH professor James Massey compared Shannon's information theory to the Copernican revolution. Copernicus's astronomical theory showed that the earth wasn't the center of the universe, but was one of the many planets orbiting the sun. Shannon's also introduced radically new concepts that turned the communication world upside down. And despite his theory being clearly superior in scientific terms, it was long and bitterly resisted by the advocates of the old school of thought.[107]

* * *

WHEN SHANNON DEDUCED his formula for the measure of information, he noticed something interesting. The combination of the logarithmic function and probabilities had appeared in other fields before. His formula was very similar to those used in statistical mechanics and thermodynamics to measure what is called entropy. In a wider sense, entropy is a measure of disorder. The higher the entropy, the greater the randomness—or "shuffleness"—of a system. A well-shuffled deck of cards has more entropy than a deck in which, say, the cards are grouped by suit. The concept of entropy was proposed by the German physicist Rudolf Clausius in the late 19th century and further developed by Ludwig

Boltzmann and Willard Gibbs. Shannon realized his formula was similar to those obtained by Boltzmann and Gibbs.[108] That was no coincidence; in a sense, his measure of information was also a measure of disorder. When we take a card from a well-shuffled deck we are more uncertain about the outcome than when we take a card from an ordered deck. So the shuffled deck has more entropy and produces more information when cards are taken. For Shannon, the connection seemed clear: he called his measure of information "entropy."[109]

But a connection between entropy and information had been established much before Shannon's work appeared. In 1929, in a paper titled "On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings," the physicist Leo Szilard analyzed a problem known as the "Maxwell's demon." The demon was an imaginary being created by physicist James Clerk Maxwell in 1867 to contradict the second law of thermodynamics. This law states that entropy never decreases. In fact, in most everyday processes—as when an ice cube melts into liquid water—entropy always *increases*. The liquid water is more disorganized, has more entropy than the ice cube. British astronomer Arthur Eddington considered that the second law defines the "arrow of time." It is such a fundamental law that, for him, it held "the supreme position among the laws of Nature."[110] In fact, the philosophical implications of the second law were vast; every physicist had his own interpretation. For William Thomson, better known as Lord Kelvin, the second law was a confirmation of the biblical view of the universe's impermanence, for the inexorable increase of the entropy meant a degradation of the usable energy.[111]

Maxwell imagined a box filled with gas and divided in the middle by a partition with a small gate. The demon was a kind of molecular-sized intelligent gatekeeper who monitored the molecules in the two sides of the box. In the beginning, the gas was at the same temperature in both sides. Although the average speed of the gas molecules determines this temperature, the molecules don't have all the *same* speed; some move faster than the average speed, others slower. The demon's goal was to have the faster molecules in one side and the

slower ones in the other. To accomplish that, according to Maxwell, he had just to observe the molecules approaching the gate and select which had to go to either side. By simply opening and closing the small gate—and without adding energy to the molecules—the demon could make one side of the box hot and the other cold. That was like getting boiling water and cubes of ice just by separating the faster and slower molecules from a bowl with warm water. Maxwell's demon could put the gas into a more organized state—thus making the entropy *decrease*. The second law of thermodynamics was threatened by a mere imaginary being.

Szilard was the first to understand how to get around the problem. He realized that the process of *observing* the molecules necessarily required an energy exchange with the system. To measure the speed of the molecules, the being had in some way to interact with them; one way to accomplish this measurement, physicists suggested later, was to send rays of photons that hit the molecules and bounced back, giving the being certain information about the molecules' speed. These measurements accounted for the apparent decrease of entropy, whereas the overall entropy was actually increasing—the second law reigned again. Szilard's explanation of the interaction between the demon and the molecules in terms of an acquisition of information established a direct link between thermodynamics' entropy and what we understand today as information theory's entropy.[112]

But when Shannon wrote his paper in 1948, he wasn't aware of Szilard's work.[113] So why did he decide to call his measure of information "entropy"? According to physicist Myron Tribus, Shannon said to him in a private conversation in 1961 that the suggestion came from John von Neumann. The Princeton mathematician suggested to Shannon to adopt the term because most people didn't know what entropy really was, and if he used the term in an argument he would will every time.[114] Von Neumann himself had explored the concept of information in quantum mechanics in the early 1930s[115], but Shannon told Bob Price he didn't recall the conversation with Tribus. "Now, I'm not sure where I got that idea," Shannon said to Price. "I think somebody had told me that.

But anyway, I'm quite sure that it didn't happen between von Neumann and me." It seems that the connection occurred to Shannon later in his development of information theory. In his 1945 cryptography report, Shannon regarded the logarithmic formula as a measure of *uncertainty*—not entropy. But later, the section titled "Choice, Information and Uncertainty" in that report became "Choice, Uncertainty and Entropy" in the 1948 paper.

In that year, a few months after Shannon's paper appeared, Norbert Wiener published his influential book *Cybernetics: or Control and Communication in the Animal and the Machine*. In that work Wiener presented a formula for the measure of information almost identical to Shannon's and also observed its connection to the thermodynamics concept of entropy.[116] At MIT, not long after the war, Wiener used to storm into a colleague's office—a huge cigar in one hand—burble his latest theory and leave without further explanations. In one of these occasions, Wiener entered Robert Fano's office and declared, "Information is entropy."[117] Shannon, while a student at MIT, took a course with Wiener on Fourier theory, but was not very close to him in research or in personal terms.[118] But they did exchange some correspondence and in October 1948 Shannon wrote Wiener saying he had read *Cybernetics* and found it "full of new and interesting ideas," adding that, "It was interesting to note how closely your work has been paralleling mine in a number of directions." In the end of the letter, Shannon wrote: "I would be interested in any comments you might have on my paper 'A Mathematical Theory of Communication' ... The second installment, dealing with the continuous case, is closely related to some of your work, and contains several references to your papers."[119] Wiener answered the letter. He thanked Shannon for his interest, said he also valued his colleague's work in the field, and commented on some other issues. About Shannon's paper specifically? Not a word.[120]

"The question of separating what is Wiener's contribution and what is Shannon's raises a lot of debate," says Fano. For him, Wiener—especially with his work on the aircraft trajectory prediction—was the first to address the question

of noise and also to point out that information was in a sense statistical. But it was Shannon, Fano says, who really realized the discrete nature of information—the necessary step to arrive at the crucial idea of coding. So while Wiener's work dealt with filtering noise out of a received signal, Shannon's dealt with overcoming noise in a signal transmission.[121] Despite their similar ideas, no collaboration ever took place, especially in the case of information and entropy.[122] "I think actually Szilard was thinking of this, and he talked to von Neumann about it, and von Neumann may have talked to Wiener about it," Shannon recalled years later. "But none of these people actually talked to me about it before I was thinking of this by myself."[123]

What explains the connection between information and entropy appearing in the works of Szilard, Wiener, Shannon, and others?[124] What is this mysterious measure that can be applied to boiling water and ice cubes, the toss of a coin, and language? Some think Shannon's quantity was badly named; thermodynamics' entropy and Shannon's entropy are two different things.[125] Others think that entropy is just a kind of information.[126] And Shannon himself thought there is a deep, underlying connection between the two.[127] "Some scientists believe," he wrote in the late 1960s, "that a proper statement of the second law of thermodynamics requires a term relating to information."[128] That would mean that Shannon's theory is more than just a theory of communication that tells how engineers should transmit data. Indeed, the results were so profound that, in 1983, the great Russian mathematician Andrey N. Kolmogorov wrote that "information theory must precede probability theory, and not be based on it."[129] Shannon built his work over probability theory and what he found was even more fundamental than probability theory itself. Should it also precede one of the most supreme laws of Nature, the second law of thermodynamics?

"The discussion about entropy and information can get very philosophical," says Gerhard Kramer, a researcher in the mathematics department at Bell Labs.[130] Like Maxwell's demon, this question has puzzled engineers, mathematicians, physicists, and philosophers. "What is information?" asked MIT

professor Seth Lloyd, who teaches the course "Information and Entropy" together with professor Paul Penfield. "It is knowledge that you can pass back and forth," said one of the freshmen (yes, the course is taught for freshmen). "It is the description of stuff," said another. "Encoding for meaning," guessed a third.[131] One of the goals of the course, conceived by Penfield three years ago, is to study the second law of thermodynamics as a kind of information processing. "Information and entropy are related in an intimate way," said Penfield to the students. "They are the same thing but in different contexts."[132] The course begins with the concepts of bits and codes and advances to the study of various situations in which information can be manipulated, stored, and transmitted, such as in communication systems and in quantum computers—the use of quantum phenomena to process information.

So information can exist in a variety of situations. But is it just an abstract entity or has it a physical reality of its own? Information is entropy or entropy is information? Is there a minimum amount of energy to store a bit? And to erase one? Scientists are looking for the answers and their exploratory ideas take several directions.[133] In a paper published in 2002, Lloyd calculated the amount of information that the Universe can store: 10^{90}—or ten followed by ninety zeros—bits.[134] "Is the Universe a computer? It is certainly not a digital computer running Linux or Windows," he concluded in the end. "But the Universe certainly does represent and process quantifiable amounts of information in a systematic fashion."[135] Ed Fredkin, a former MIT professor and a friend of Shannon, believes that the laws of nature at its most fundamental level result from a sort of digital information processing—what he calls "digital mechanics," a substrate for quantum mechanics.[136] And Princeton physicist Freeman Dyson wonders whether life is analog or digital. "We don't yet know the answer to this question," he wrote in an essay on the subject. "If we are partly analog, the downloading of a human consciousness into a digital computer may involve a certain loss of our finer feelings and qualities. I certainly have no desire to try the experiment myself."[137]

* * *

LETTERS CAME FROM all parts of the country. And also from Canada, England, France, and Japan. They came from universities—MIT, Harvard, Princeton, Pennsylvania, Johns Hopkins—companies' laboratories—Westinghouse, RCA, General Electric—and governmental and military institutions—Navy's Bureau of Ordnance, Los Alamos Laboratory, Brookhaven National Laboratory. Shannon, then at age 32, felt the reaction to his 1948 paper in his mailbox. Not only engineers and mathematicians were writing, but also economists, psychologists, sociologists, linguists. Some posed questions; others shared ideas. And most asked for copies of his stunning paper. Just a month after the first part was published, a researcher from a company's engineering department wrote asking for "half dozen" copies. Yet another wondered if Shannon could send him the unpublished portion.[138]

Many papers in the *Bell System Technical Journal* discussed technologies that are now outdated. Shannon's paper, on the contrary, brought ideas that are not just useful today—they are the *basis* of modern communication theory. He formulated information theory in general terms, detaching his concepts from the specific systems of the time and thus transcending the inevitable obsolescence of technology. "I have on occasion skimmed through other journals of 1947 to form a basis for the calibration of Shannon's work," wrote Robert Lucky, a former Bell Labs director of research. "What kind of world was it then? What I found was a lot of forgettable hopping, chirping, and flying engineering things that suffered extinction in the survival of the fittest in the decades that followed. . . . No museum would ever be interested in them."[139]

But as with most revolutionary theories, those that force scientists and researchers to rethink what they considered true and fundamental, Shannon's ideas took some time to become fully accepted. "When the paper first appeared, few people really understood it," says David Slepian. "The old school of

engineers thought it was nonsense."[140] In 1949, the mathematician Joseph Doob wrote a review of Shannon's paper that caused—and still causes—a lot of debate. "The discussion is suggestive throughout, rather than mathematical," he wrote at one point in his review, "and it is not always clear that the author's mathematical intentions are honorable." Doob, a great mathematician soon to become an authority in stochastic processes, argued that the paper wasn't mathematics, that Shannon's proofs weren't rigorous enough.

In fact, the paper did have holes. Certain steps and assumptions were not clear. Some theorems and proofs weren't perfect. But this very approach to the problem of communication was perhaps the main reason why Shannon's work was so successful. Shannon was a mathematician *and* an electrical engineer. He always worked in the intersection of both fields. He was employed in the mathematics department of a telephone company. He was writing a theoretical paper for an engineering journal. For these reasons, he had practical considerations in mind. He had to avoid certain mathematical formalisms and move ahead in his theory. "The occasional liberties taken with limiting processes in the present analysis," he wrote halfway in the paper, "can be justified in all cases of practical interest."[141] Those who understood the essence of Shannon's theory—and its implications—quickly recognized the need for a balance between formalism and pragmatism. Shannon had showed the way; from there, they had to work the rest of the results. The next step was to refine the proofs to make sure the theorems were valid, that the engineers weren't stepping in muddy territory, that they could trust the theory to build what really matters: the applications.

Not long after the paper came out, some began to work on the math. "When I read the paper I finally understood the problem Shannon was trying to explain," says Brockway McMillan, to whom Shannon had shown a sketch of the fan diagrams, an important element in the formulation of information theory. A few years later, McMillan wrote a paper explaining the theory to statisticians and putting its results into more rigorous terms.[142] Finally mathematicians began to

understand and accept Shannon's ideas. "I gave a lot of talks about that paper," says McMillan.[143]

The science of communication was finally becoming a unified field, with its diverse disciplines brought together by integrating forces such as Wiener's theory on prediction and filtering of noisy signals, Shannon's information theory, and other works appearing at that time in the United States, England, Germany, and the Soviet Union.[144] A symposium organized in London in September 1950 brought together, according to one attendee, "an intriguing mixture of mathematicians, statisticians, physicists, biologists, physiologists, psychologists and communication engineers."[145]

Groups of information theorists formed in several parts of the world, but the leading place was MIT. In the spring of 1951, Robert Fano started teaching course 6.574, "Transmission of Information," a seminar on information theory for graduate students. As part of the course, he challenged his graduate students to come up with better proofs for Shannon's theorems. To his delight, the proofs did come. A student figured out the most efficient method to represent information with zeroes and ones. Another extended McMillan's formulation and proved definitely that Shannon's idea of virtually error-free communication up to the channel capacity was right. And in fact, there wasn't a single theorem flawed. "We should say that after fifty years," Stanford University professor Thomas Cover wrote in 1998, "it is clear that Shannon was correct in each of his assertions and that his proofs, some of which might be considered outlines, could eventually be filled out along the lines of his arguments."[146]

In the mid-1960s, after nearly two decades since Shannon's foundational paper was published, most theoretical results were exhaustively explored, and all proofs were, well, proved. People had been working for a while on the development of codes and satisfactorily ones were available. And it seemed they were good enough for any application. At that point some began to say that information theory as a field was dead. There were no more problems to solve.

Not everybody agreed with that of course. Microchips were on their way, some noted, and they would certainly open new possibilities for information theory. At MIT the order was to attack practical problems. "Their advice was, 'Don't work on theory, go to the applications'," says MIT electrical engineering professor David Forney, then a graduate student.[147] But it wasn't easy. Computers were still the size of refrigerators and communication was more analog than ever. Information theorists had the codes to compress information and correct errors—the "software"—but they didn't have the hardware to *execute* those instructions. Building any system with advanced coding schemes would require new electronics—which would certainly cost lots of money. Not surprisingly, the first complete application of Shannon's ideas was part of a government project.

In 1958, following the Soviet Union's launch of Sputnik, the first artificial satellite put in orbit, the U.S. government created NASA. The space race accelerated throughout the years and several spacecraft and probes were developed. In 1968 NASA sent to space the solar-orbiting spacecraft Pioneer 9, designed to study the flow of the solar wind and the high-energy particle streams coming from the sun. The144-pound spacecraft was the first to carry a digital system for encoding and transmitting information. Shannon's theory said that the better the codes you use, the more noise you can overcome. In space missions, that translated into an invaluable result: spacecrafts could go farther. In fact, with Pioneer 9's digital coding scheme, the spacecraft could travel 40 percent farther and still be "heard" on earth. To do the same thing using more transmission power, NASA would have to spend millions of dollars in better transmitters and receivers.[148]

Deep space communication and coding, according to James Massey, was "a marriage made in heaven."[149] "It is no exaggeration," Massey wrote in a paper, "to say that the Pioneer 9 mission provided communications engineers with the first incontrovertible demonstration of the practical utility of channel coding techniques and thereby paved the way for the successful application of coding to many other channels." Pioneer 9 circled the sun twenty two times, covering

eleven billion miles, and sent 4.25 billion bits of scientific data back to earth during its operational lifetime.[150] Succeeding spacecrafts and probes carried even more sophisticated codes and could transmit not only scientific data but also clear pictures of Mars, Saturn, and Jupiter. One of those error-correcting codes is the same used now in CD players.

In the 1970s, microchips and computers began to be mass-produced. The hardware to run information theory's coding schemes was finally available. "With Shannon's remarkable theorems telling communications engineers what ultimate goals to strive for, and integrated circuits providing ever-improving hardware to realize these goals, the incredible digital communications revolution has occurred," Solomon Golomb, a professor at University of Southern California, wrote recently.[151] Incidentally, the microelectronic revolution started very close to Shannon, also at Bell Labs, when researchers Walter Brattain, John Bardeen, and William Shockley invented the transistor in 1947. One day Shannon was chatting with Shockley and saw on his desk a small plastic object with three wires extending from it. "This was my first glimpse of a transistor, quite possibly the greatest invention of the 20th century," Shannon recalled later.[152] A tiny electronic device to amplify and switch an electric current, the transistor was much better than the clumsy vacuum tubes. Several of them can command a complex flow of bits in a circuit. And the more transistors, the more bits can be manipulated. In 1958 the first integrated circuit—or simply microchip—was invented. Since then, every few years long more powerful microchips are produced. Today's chips contain thousands or millions of transistors. According to an empirical trend observed in the semiconductor industry—known as Moore's Law, after Gordon Moore, who worked with Shockley and later became one of the founders of Intel[153]—the number of transistors that can be packed in a chip doubles every eighteen months. Microchips can now execute the complex codes required by information theory.

The search for better codes that can approach the channel capacity has kept engineers busy since Shannon's paper appeared in 1948. But in a conference

in Geneva in 1993, French coding theorists surprised their colleagues when they presented a code—dubbed the turbo code—that could get extremely close to the Shannon limit—closer than anyone had gotten before. At first, veterans in the coding field were skeptical, but the results turned out to be correct, and the turbo codes revolutionized coding for reliable communications. "Other researchers have refined these results and have been able to construct simpler codes that approach capacity basically as closely as desired," says Kramer, from Bell Labs. "In the future, you will probably be using such codes on a daily basis if you have a cell phone."[154]

From multi-million dollar space probes to cell phones and CD players that many people can afford, Shannon's legacy is nearly everywhere today. His information theory is invisibly embedded in the hardware and software that make our lives easier and more comfortable. We don't see the error-correction schemes embedded in our hard drives, modems, and computer networks. Or the coding technology that makes mobile phones work. Or the compression methods that shrink sound, images, and video sent over the Internet. Or the encryption schemes that make online shopping secure. These and other technologies are based on the laws of communication Shannon established 55 years ago. Today's digital machines are fueled by bits, which made possible a simpler and cheaper representation of information. In digital systems, all sorts of data are transformed into two-value elements—the mathematical bit becomes a physical bit. In a CD bits are stored with tiny pits and bumps in the plastic surface. In an optical fiber, bits are transmitted with pulses of light and "blank spaces" (absence of light). In certain chips, bits flow as zero- and five-volt signals. With bits we can manipulate, copy, and communicate information in an unprecedented way. "I think it can truly be said that Claude Shannon laid the cornerstone for the field of digital communication," wrote communication engineer Robert Kahn in an article in 2001. "He provided the goals which generations aspired to attain. He provided the formalism that shaped the way an entire field thought about their discipline,

and his insights forever altered the landscape and the language of communications."[155]

* * *

JANUARY 1949. WARREN Weaver, the man who had directed the National Defense Research Committee's gun fire control division, wrote Shannon saying he had recently met Chester Barnard, the new president of the Rockefeller Foundation, and described to him what was information theory. Barnard became very interested and asked Weaver to write something less mathematical for him. "This turned out to be a real job," Weaver wrote, referring to a text in which he tried to explain information theory in more accessible terms and discuss its possible applications. Weaver asked for Shannon's feedback on that piece and wondered about getting it published. "Having written this out, and assuming that it does not turn out to be too horribly inaccurate, I have tentatively considered attempting a rewriting, briefer and more popular in form, which I might submit somewhere for publication—probably the Scientific American," he wrote. "But I couldn't possibly do that unless you are entirely willing to have me do so. Comments?" Weaver's article appeared in *Scientific American* in July. And in that same year, his introductory piece and Shannon's original paper were published in book form. Information theory wasn't just technical reading anymore. It was available in bookstores.

Throughout the years, Shannon's paper influenced—and continues to influence—scientists, engineers, and many other individuals. A graduate student named his golden retriever "Shannon." An undergraduate at Cornell captivated by Shannon's paper changed his major to mathematics. Another began referring to himself as "third generation Shannon" as he took the legacy on and started passing out copies of Shannon's paper.[156] Indeed, some say there is no better introduction to the field than Shannon's paper itself. "Since I am an information theorist, Shannon gave me my life's work," wrote Robert McEliece, an eminent

figure in the field. "I first read parts of his 1948 paper in 1966, and I reread it every year, with undiminished wonder. I'm sure I get an IQ boost every time."[157]

Published in book form with a subtle—but significant—change in the title, Shannon's work now wasn't just "A Mathematical Theory...," but "*The* Mathematical Theory of Communication." Soon, concepts, terms, and ideas of information theory began to be employed in fields other than engineering and mathematics. In the early 1950s, scientists studied the human ear and eye as information channels and calculated their supposed "capacities." The ear? Ten thousand bits per second. The eye? Four million bits per second.[158] A person speaking? About thirty bits per second.[159] A group of psychologists measured a person's reaction time to various amounts of information. The subject sat before a number of lights with associated buttons. The lights went on according to some patterns—conveying different amounts of information—and the subject had to push the corresponding buttons as quickly as possible. For one bit, the reaction time was 0.35 second. For two bits, 0.5 second. For three bits, 0. 65 second. Every bit added increased the reaction time in 0.15 second. The experiment showed that the reaction time increases linearly with an increase in the amount of information conveyed by the lights.[160] "These results," Shannon wrote in the *Britannica* article, "suggest that under certain conditions the human being, in manipulating information, may adopt codes and methods akin those used in information theory."[161] Could the understanding of how machines process information help us to finally understand how we human beings process information?

Another field that adopted information theory ideas was linguistics. The concept of a stochastic process as an information source that could produce language was a matter of a lot of debate. If a mathematical model could generate language, and if this model was implemented into a computer, could a machine speak? In a paper published in 1956 in the *Transactions of Information Theory*, a young linguistics professor recently admitted to MIT argued with vehemence—and math as well—that Markov processes couldn't generate language. For Noam Chomsky, certain concepts from information theory could be useful for linguists,

but language had a grammatical pattern that couldn't be represented by such processes.[162]

As the years passed, Shannon's ideas continued to disseminate to other fields—economics, biology, even art and music. "Many were honest attempts to apply the new exciting ideas to fields in need of help," wrote David Slepian in 1998. "Time has shown that for the most part only mirages were seen."[163] In fact, information theory didn't have a major impact on any of these fields. But its concepts proved valuable anyway. Today, linguists use the notion of entropy and Shannon's fundamental theory to determine the number of words of a language and their lengths.[164] Geneticists and molecular biologists use information theory to study the genetic code and to investigate things such as why sex is a winning "evolutionary strategy" for many species.[165] And investors and market analysts use information theory to study the behavior of the stock market and optimize a portfolio of stocks.[166]

But the incursions of information theory into fields other than communication as it happened in the 1950s and 1960s didn't please its inventor. "Information theory has, in the last few years, become something of a scientific bandwagon," Shannon wrote in an editorial for the Institute of Radio Engineers information theory journal in 1956.[167] "Our fellow scientists in many different fields, attracted by the fanfare and by the new avenues opened to scientific analysis, are using these ideas in their own problems. Applications are being made to biology, psychology, linguistics, fundamental physics, economics, the theory of organization, and many others." Shannon called for moderation, especially outside the engineering domains. More than that, within the field itself, Shannon called for more rigor and diligence: "Research rather than exposition is the keynote, and our critical thresholds should be raised. Authors should submit only their best efforts, and these only after careful criticism by themselves and their colleagues. A few first rate research papers are preferable to a large number that are poorly conceived or half-finished. The latter are not credit to their writers and

a waster of time to their readers." It seems Shannon followed that philosophy—just "first rate research papers"—all his life.

* * *

CLAUDE SHANNON NEVER liked to write. He even avoided writing.[168] At home, he used to keep unanswered correspondence in a box in which he wrote: "Letters I've procrastinated too long on."[169] Shannon left few records at Bell Labs as well. He simply seems not to have kept paper records. On joining Bell Labs every researcher receives a lab notebook. Apparently, he didn't turn his in to the files."[170] At the Institute for Advanced Study in Princeton, where he started working on information theory, he left no records either.[171]

Shannon was always looking for a new problem. But once he found the solution, he didn't care about writing it down or getting it published. Many times he started and soon abandoned a draft, putting it in a file on his attic. "After I had found answers, it was always painful to publish, which is where you get the acclaim. Many things I have done and never written up at all. Too lazy, I guess. I have got a file upstairs of unfinished papers!" he said in an interview with *Omni* magazine in 1987. "But that's true of most of the good scientists I know. Just knowing for ourselves is probably our main motivation."[172]

And that was how Shannon worked. He always kept most things to himself.[173] He was a quiet, introspective person. Not shy, not someone who tries to hide in a corner, avoiding contact. Shannon was very warm and cheerful and enjoyed talking to other people. Among his family and friends he used to laugh heartily in a good conversation. He loved puzzles, pranks, and gadgets. He could ride a unicycle and juggle at the same time. He was tall and thin—"a frail man," he once described himself. He had a long face and gray eyes with a very narrow, light gray rim around the iris.

Shannon was just self sufficient, content to be alone immersed in his solitary research. For those who didn't know him, he seemed serious, sometimes

cold. But despite his fame, he treated his colleagues and students as his equal.[174] "We all revered Shannon like a god," said Hermann Haus, who once gave a talk at MIT in which Shannon was in the audience. "I was just so impressed, he was very kind and asked leading questions. In fact, one of those questions led to an entire new chapter in a book I was writing."[175]

In 1948, Shannon met Mary Elizabeth Moore, a mathematician at Bell Labs working in John Pierce's group. On March 27, 1949 they got married and went to live first in New York and later in New Jersey. Like Shannon, Betty was bright, sharp, and had a great sense of humor. "A perfect couple," some say.[176] Among the benefits of marriage was one solution to Shannon's problems. He didn't have trouble memorizing numbers, formulas, theorems, and mathematical constants. That was his alphabet.[177] But to overcome the pain of writing, Shannon found he could dictate his papers to Betty, who would write down everything on paper.[178] The manuscript was then sent to a pool of secretaries at Bell Labs who typed everything (a "noisy channel" of its own, as errors appeared here and there, as when in a paper on coding, the phrase "signaling alphabet" became "signaling elephant"[179]). And that was not just a dictating-writing process. Betty, as a mathematician and a Bell Labs scientist in her own right, would engage in the elaboration of Shannon's ideas. Many of Shannon's papers from 1949 until the end of his life were created this way.[180]

After the publication of the 1948 paper, Shannon continued contributing to the field he had created—but not for too long. As information theory grew in importance, Shannon became more and more distant and uninterested, culminating in his 1956 "bandwagon" editorial. In that same year, after fifteen years at Bell Labs, Shannon decided to leave the laboratories. He received offers from a number of universities to teach.[181] But he chose the place he knew could offer the freedom he had always sought: MIT.

Shannon went back to MIT first as a visiting professor, and as a faculty member in 1958. At about that time, professor Robert Fano was teaching his information theory seminar. Shannon then began teaching an advanced topics

course on information theory for graduate students. He would come to class and talk about research problems he was working on—how to build reliable machines with unreliable components; how to maximize a portfolio of stocks;[182] how to build an unbreakable cryptography system. He was a good lecturer, but not the kind who tries to entertain his students. He would put a problem in the blackboard and look at it for a while. Then he would put down something else or grab a piece of paper and check a passage on it. "I have the indelible recollection of a person who could reason in theorem-sized steps, the logic of which were never easy for other to grasp," wrote Robert Kahn, who took Shannon's course in the 1960s.[183] Sometimes Shannon could be very halting, and he didn't solicit much interaction; but he was always open to questions. "For some problems, he had good results. For others, he made no progress beyond formulating the problem," says David Forney, who attended the seminar in the spring of 1964. "It was great for graduate students looking for a thesis topic."[184]

Shannon's lecture style revealed much of his own research habits: he chose topics according to his curiosity and was always trying to simplify the problems, often using simple games or toy examples. "He was a different model of scientist," says Robert Gallager, from MIT. "He wasn't a scholar who wants to know everything in a field. He liked puzzles, he liked to solve complicated things making them simple."[185] Also, Shannon always sought broad, fundamental results instead of technical details—a single question from him could lead to an entire chapter in a book. "His way of attacking a problem and seeing how he thought and addressed research were strong influences on my own research," says Leonard Kleinrock, a professor of computer science at the University of California at Los Angeles. Because of that influence, Kleinrock says he always looked for the global behavior, for the underlying principles, for the extreme and simple cases, for intuitive understanding. "Shannon was the best," he says.[186]

Information theory had had its roots at Bell Labs, but now MIT became, according to an observer, the "mother church" of the field.[187] From the groups of information theorists at Bell Labs and MIT came significant contributions to

communication, computer science, and artificial intelligence. Students graduated from MIT and started teaching information theory at places such as Cornell, Berkeley, and Notre Dame. Others went to companies. Gallager and Massey became consultants to Codex, which produced many coding systems for the military and where Forney created the first commercially successful modems (the company was later acquired by Motorola). Former students Andrew Viterbi and Irwin Jacobs started Qualcomm and developed wireless technologies widely used in the cell phone system nowadays. Jacob Ziv, an Israeli electrical engineer who graduated from MIT in 1962, invented a method of "squishing" bits that transformed the field of data compression and which is widely used now in computers and other digital systems. And in 1963, Robert Fano became the head of MIT's Project MAC, an initiative to share computing resources that later became the famed Laboratory for Computer Science.[188]

During those bright years of information theory at MIT, Fano says that at a certain point Shannon was asked to be an advisor to students. But he reacted saying, "I can't be an advisor. I can't give advice to anybody. I don't feel the right to advise." In fact, Shannon didn't have many advisees, but the few he advised made significant contributions in several fields. In the early 1950s, while still at Bell Labs, Shannon supervised John McCarthy and Marvin Minsky during summer jobs at the laboratories.[189] Both went on to be pioneers in the field of artificial intelligence, an area in which Shannon also made fundamental contributions. While at Bell Labs, he published an influential paper about programming a machine to play chess—the first to speculate about such possibility.[190] At about the same time, using telephone relays, he built a mouse that could find its way through a maze (actually, a mechanism under the maze guided the mouse with magnets). "Since the drive mechanisms and relay computing circuit were all under the maze floor," Shannon recalled later, "some of my persnickety friends complained that the mouse was not solving the maze, but the maze was solving the mouse."[191] At MIT, in the 1960s, he advised Leonard Kleinrock, who would become an Internet pioneer, and also Ivan Sutherland,

whose master's thesis opened the field of computer graphics.[192] In the late 1960s, another student, Edward Thorp teamed with Shannon to develop a cigarette pack sized computer to predict roulette. Thorp, Shannon, and their wives even went to Las Vegas to test the device, which worked as expected, except for the fact that a minor hardware problem prevented any serious betting.[193]

At MIT and elsewhere, Shannon gained the status of celebrity, and for his work on information theory he received innumerable honors. In 1966 he was awarded the U.S. National Medal of Science. (Russian information theorists tried to elect him to the Soviet Academy of Sciences, but the proposal didn't find enough support.)[194] In 1985 he was awarded the Gold Medal of the Audio Engineering Society; the plaque he received read simply, "For contributions that made digital audio possible."[195] That same year, he won Japan's Kyoto Prize, known as the Nobel for mathematical achievements. In his discourse, Shannon showed his classical block diagram of communication and said: "Incidentally, a communication system is not unlike what is happening right here. I am the source and you are the receiver. The translator is the transmitter who is applying a complicated operation to my American message to make it suitable for Japanese ears. This transformation is difficult enough with straight factual material, but becomes vastly more difficult with jokes and double entendres. I could not resist the temptation to include a number of these to put the translator on his mettle."[196]

Not long after receiving the Kyoto Prize, Shannon received a letter from Sweden. Some good news from the Nobel Committee? Not, actually. "Dear Professor Claude Shannon," the letter began. "I'm a collector of autographs of persons who really have done something positive for their people and the future of mankind." The person wondered if Shannon could send him an autographed photo.

The prizes and honors, however, didn't interest Shannon very much. Nor did his celebrity status. His colleagues say he hated giving speeches. But in 1985, for some reason, Shannon decided to go to the international symposium on information theory—something he didn't do for years—held in Brighton,

England. His presence caused turmoil and the audience lined up for autographs. "It was as if Newton had showed up at a physics conference," recalled Robert McEliece, the chairman of the symposium.[197]

By the early 1980s, direct and indirect applications of information theory could be found in several fields—from deep-space probes in the far reaches of the solar system to modems that transmitted data over a telephone line. Shannon's theory was known among electrical engineers, mathematicians, computer scientists, physicists, linguists, psychologists, and Wall Street investors. But by that point, Shannon had lost most of his interest in information theory. He rarely could be found in his office at MIT.

Shannon became an emeritus professor in 1978 when he was already practically retired. He spent his time at home on his own things. He liked to play with his kids—Andy, Bob, and Peggy—and work on gadgets. He usually concentrated on one thing at a time. He could stay up all night building a machine and get up the next day and immediately get back to it.[198] But then he would start working on another thing, and then move to another, and another.[199] In Christmas of 1951, Betty gave him a unicycle. He loved it. Soon he was riding it in the corridors of Bell Labs—now a legendary moment in the culture of the labs—and also building at home unicycles of different shapes and sizes. "Well, let me put it this way," Shannon told Bob Price, "that my mind wanders around and I will conceive of different things day and night, as Betty will attest—like the science fiction writer or something like that. I'm thinking what if it were like this, or what is this problem, or is there an interesting problem of this type? And I'm not caring whether somebody's working on that or whether Washington would care one way or the other, or anything of that sort. I just like to solve a problem. And I work on these all the time." The kind of problems Shannon liked involved a physical, an engineering situation for which he could sort of tailor make a mathematics. He did that for switching circuits, genetics, cryptography, communication, chess-playing machines, and juggling.[200]

At home he kept a large room to store his stuff—the "toy room," he and Betty called it. The largest wall was devoted to his diplomas, plaques, and prizes. Another wall was filed by posters of Darwin, Einstein, Newton, Hilbert—and a Smurf. Shelves were packed with collections of Swiss knives—one with one hundred blades—hats, chess sets, musical instruments, juggling balls, rings, clubs, and Rubik's cubes. He tried to build a machine to solve a cube—but as with many of his theories and inventions, he never finished it. But he wrote a song:

Ta! Ra! Ra! Boom De ay!
Cu-bies in disarray?
First twist them that-a-way,
Then turn them this-a-way.

Respect your cube and keep it clean.
Lube your cube with Vaseline.
Beware the dreaded cubist's thumb,
The callused hand and fingers numb.
No borrower nor lender be.
Rude fold might switch two tabs on thee,
The most unkindest switch of all,
Into insolubility.

In-sol-u-bility.
The cruelest place to be.
However you persist
Solutions don't exist.[201]

The mouse-maze system he built at Bell Labs, the roulette prediction portable computer, and other incredible machines Shannon built were also there, including a robotic puppet in the form of comedian W. C. Fields—his favorite gadget.[202] Made of Erector set pieces, it could bounce juggle three steel balls over the head of a drum, making a sonorous "thunk" with each hit.

Juggling fascinated Shannon, and he spent a lot of time practicing—and also figuring out mathematical laws relating the time the objects stay in the air and the height they had to be thrown. One Sunday in the mid-1980s, members of the MIT Juggling Club were practicing in the field just across the street from the Institute's main entrance when a gray haired man stopped by and asked, "Can I measure your juggling?" "He wanted to know how long a ball stays in the air and in the juggler's hand," says Arthur Lewbel, the founder of the club and now a professor of economics at Boston College. "He made some measurements that day using a stopwatch and later came for more measurements." Lewbel says that they had no idea that the cheerful gray haired man was Claude Shannon, a professor emeritus at MIT, a tinkerer who loved to build all sorts of gadgets, and a man who made a lot of money in the stock market investing in high-tech companies started by friends, such as Teledyne, Hewlett-Packard, and Codex.[203] Not long after, members of the club found themselves in his living room, in a large house in Winchester, in the suburbs of Boston, to see his machines, watch juggling videos, and eat pizza. "Shannon was the first to apply mathematics to juggling," says Lewbel. After his juggling measurements, Shannon came up with a juggling theorem and wrote an article for *Scientific American*. The magazine asked for revisions, which Shannon never did, and the article was never published.

In 1993 Robert Fano met Shannon at a friend's memorial. "I asked him something about the past, nothing technical or mathematical," says Fano, "and Claude answered just 'I don't remember.'"[204] In the early 1990s, Shannon noticed some memory lapses. Sometimes he couldn't drive back home. Later he wouldn't recognize his own writings. And then his friends. It was a long battle against Alzheimer's disease. "The last time I saw Claude, Alzheimer's disease had gotten the upper hand. As sad as it is to see anyone's light slowly fade, it is an especially cruel fate to be suffered by a genius," wrote Lewbel in an article on Shannon in late 2001. "He vaguely remembered I juggled, and cheerfully showed me the juggling displays in his toy room, as if for the first time. And despite the loss of

memory and reason, he was every bit as warm, friendly, and cheerful as the first time I met him." Shannon died in February 24, 2001, at age 84.

* * *

TODAY, THE MOVEMENT in the corridors of Bell Labs at Murray Hill is not so intense as in the golden years of the post-war. But the cafeteria is still the place where scientists and researchers get together to chat, solve problems, and exchange ideas with their colleagues. Bell Labs is now the research and development arm of telecom company Lucent Technologies. The old buildings in the Murray Hill campus are now filled with high-tech equipment being tested and developed as the labs seek to remain a leading place for innovative technology. The transistor was invented here. And so were the orbiting communication satellite, the solar battery cell, the Unix operating system, and many other significant inventions.[205] Over the years Bell Labs has collected six Nobel prizes and over 30,000 patents.[206]

The members of the mathematics department don't fly kites or play word games during lunch as in the 1940s. Now they have their own games, puzzles, and jokes. "The other day I dialed an 1-888 number to get some information on passports," says a Russian researcher. "But I think I missed one digit and I reached an erotic line!" Says one of his colleagues: "The number has a short Hamming distance." Among this new generation of information theorists, Shannon's legacy is enormous. Posters and pictures of him hang in the offices. Papers on information theory and editorials such as Shannon's "The Bandwagon" fill the boards on the corridors. One researcher reads Shannon's papers when she is depressed. "Those who knew Shannon say he wasn't showy. He wasn't, say, a Dick Feynman," says Kramer, a young researcher in the department who never met Shannon but keeps a venerable poster of him in his office. "He must have been captivating in his own way."[207]

At the main building in Murray Hill, on the right side of the entrance lobby, a tall hall illuminated by a skylight, there stands a ten-foot tall bust of Alexander Graham Bell, the inventor of the telephone and after whom Bell Labs was named. Bell's figure, with opulent moustache and sideburns, gazes into the horizon, while its pedestal bears the following quote: "Leave the beaten track occasionally and dive into the woods. You will be certain to find something you have never seen before." On the opposite side of the hall from Bell's bust, another bust stands by a column. Claude Shannon's bust is not as tall as Bell's, but it shares a privileged location on the hall. His figure has a thoughtful look, his head turned a little downwards, his hand holding his chin. On Shannon's there is no quotation, just the message:

$$H = -p \log p - q \log q$$

ACKNOWLEDGEMENTS

Several times throughout the course of this work, I met Robert Kanigel in his office on the fourth floor of the Humanities building, where he receives his students sitting in a comfy easy upholstered chair and offering his guests a hard wooden slatted chair. I wonder what his true intentions were; perhaps he wanted to create an uneasy atmosphere, and provoke, inspire, push his students. Whether that was true or not, it worked. I am indebted to Rob for his help, advice, support—for provoking, inspiring, and pushing me during my year at MIT. He helped me to choose the topic of this article, shape its structure, refine and polish the text. I thank him for his help, patience, and constant cheerfulness in all—easy and hard—moments.

The faculty of the science writing program surpassed all expectations and provided memorable moments of unique writing and editing wisdom during our "mega seminars." I want to thank B. D. Colen, Alan Lightman, and Boyce Rensberger for an incredible year at MIT. Thanks go also to Jim Paradis, head of the writing program, for all his support—and a great Thanksgiving dinner at his home. Sarah Merrow went beyond her duties as the program administrator to help me in all sorts of ways, becoming a good friend even before I arrived in Cambridge. Sarah helped to make the "MIT experience" a very enjoyable and agreeable one.

At MIT I also had the opportunity to take great courses on the history of science and technology. I wish to thank David Kaiser, David Mindell, and Rosalind Williams for their effort and dedication to the art of teaching. Additional thanks go to Paul Penfield, who let me audit the interesting freshmen course on information and entropy he teaches with Seth Lloyd. (What is information anyway?)

I am thankful to all who contributed to this work explaining to me the toughest parts of Shannon's theory, submitting to interviews, providing invaluable material, and referring me to various important sources. I am specially grateful to Betty Shannon, who kindly received me at her home in Winchester for an evening of wonderful stories and a tour in the famed "toy room." For their guidance and encouragement, I am indebted to Robert Fano, Robert Gallager, and Robert Price. For sharing their recollections and providing invaluable help, special thanks are due to David Forney, the late Hermann Haus, Amos Joel, Leonard Kleinrock, Arthur Lewbel, James Massey, Brockway McMillan, Joel Moses, Ronald Rivest, and David Slepian.

At Bell Labs in Murray Hill, I am thankful to Gerhard Kramer, James Mazo, and Debasis Mitra. Gerhard proved to be not only an excellent host and "tour guide" at the labs, but also an excellent teacher, patiently explaining to me the "fundamental theorem," turbo codes, one-time pad cryptography, and many other tough things.

Shannon didn't like to write. And he didn't like to give interviews neither. But in the few occasions he agreed, the material revealed great insight on his personality, interests, and genius. Friedrich-Wilhelm Hagemeyer interviewed Shannon and other leading figures in the field of communication in the late 1970s. Bob Price interviewed Shannon in the early 1980s. I would like to express my appreciation for their efforts in trying to obtain the message straight from the source. Price kindly shared his transcript and the original audio recording and also helped me to find and follow other leads. Hagemeyer referred me to the most relevant parts in his comprehensive PhD dissertation on information theory. I owe a great debt of gratitude to Hermann Rotermund, who transformed Hagemeyer's tape-recorded analog interviews into digital MP3 files and mailed them to me on a CD—Shannon's voice signals encoded in bits.

The three "Bobs"—Fano, Gallager, and Price—and also David Slepian and Gerhard Kramer read the manuscript and provided invaluable comments and suggestions, helping me to "debug" the original text and fix errors here and there.

Librarians and archivists provided invaluable help in several ways; my thanks to Leonard Bruno and Joseph K. Brooks of the Manuscript Division at the Library of Congress, Nora Murphy of the MIT Archives, Jenny O'Neill of the MIT Museum, the staff of the MIT libraries, and Mary Ann Hoffman of the IEEE History Center.

Coming to MIT was certainly a unique experience. It wouldn't be possible without the training and education I received at the University of Sao Paulo, in Brazil. Special thanks go to Jose Jaime da Cruz, Roseli Lopes, Ricardo Paulino Marques, Henrique Del Nero, Felipe Pait, Paulo Sergio Pereira da Silva, and Marcelo Zuffo. Also in Brazil, I want to express my gratitude to Helio Gurovitz, who taught me everything I knew about journalism before coming to MIT.

My classmates in MIT's first science writing class also helped to make this a memorable year; for their friendship, I wish to thank Timothy Haynes, Matthew Hutson, Karen MacArthur, Maywa Montenegro, Sorcha McDonagh, and Anna Lee Strachan. Thanks go also to other friends at MIT and elsewhere: Paulo Blikstein, Pedro Bolle, Patrick Kann, Charles Low, Sidney Nakahodo, Fabio Rabbani, Lidia Reboucas, Matthew Traum, and the Big Head Adventure Team—Tadeu Azevedo, Daniel Garcia, Andre Kupfer, Humberto Marin, Marcus Paula, and Daniel Pezeta.

My family, despite being thousands of miles far to the south, was always present, providing comfort, confidence, and inspiration. I am infinitely thankful to my father, Joao, my first editor; my mother, Clarice, my most dedicated reader; and my little sister Maya, always a source of wise thoughts and words.

NOTES

[1] Robert Price, communication engineer, interview by author, Lexington, Mass., April 10, 2003. On July 28, 1982, Price conducted a lengthy interview with Shannon, recorded on tape. Price was interested in the origins of information theory and more specifically in the history of a communication technology known as spread spectrum. An edited version of Price's interview was published in 1984 (F. W. Ellersick, ed., "A Conversation with Claude Shannon: One Man's Approach To Problem Solving," *IEEE Communications Magazine*, May 1984, p. 123.) The transcript of the interview is available, under copyright, at the IEEE Archives in New Jersey.

[2] Mary Elizabeth (Betty) Shannon, interview by author, Winchester, Mass., February 25, 2003; Robert Fano, professor of electrical engineering and computer science, MIT, interview by author, Cambridge, Mass., December 19, 2002; Hermann Haus (deceased), professor of electrical engineering and computer science, MIT, interview by author, Cambridge, Mass., December 20, 2002; David Slepian, information theorist, former Bell Labs researcher, interview by author, Murray Hill, New Jersey, March 7, 2003; Arthur Lewbel, professor of economics, Boston College, interview by author, Chestnut Hill, Mass., January 30, 2003; Robert Gallager, professor of electrical engineering and computer science, MIT, interview by author, Cambridge, Mass., January 15, 2003.

[3] For other epiphanies in science see, for example, Paul Strathern, *Mendeleyev's Dream: The Quest For the Elements*, St. Martin's Press, 2001; David A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*, Chapter 4, "Opening Black's Box," and Chapter 9, "Analog's Finest Hour," The Johns Hopkins University Press, 2002.

[4] John Horgan, "Profile: Claude E. Shannon," *IEEE Spectrum*, April 1992, p. 74. In fact, Shannon seemed to always evade this kind of question—see, for instance, Ellersick, op. cit. Also, in 1977, Friedrich-Wilhelm Hagemeyer, a German graduate student in history of science, came to the United States and interviewed several leading scientists and information theorists, including Shannon. His thesis was published in 1979 and never translated into English (F. W. Hagemeyer, "Die Entstehung von Informationskonzepten in der Nachrichtentecknik," PhD dissertation, Free University of Berlin, Germany 1979). The author obtained several of the tape recordings of the Hagemeyer's interviews. In his interview with Hagemeyer, Shannon also avoided tracing the origins of information theory back to a single moment or idea.

[5] In 1953, just five years after Shannon's paper came out, *Fortune* magazine published an elaborate article on information theory. The theory, wrote the author, "seems to bear some of the same hallmarks of greatness" of quantum theory, but is "still almost unknown to the general public" (Francis Bello, "The Information Theory," *Fortune*, December 1953, pp. 136-141 and 149-157).

[6] "A Mathematical Theory of Communication" was published in two parts in the *Bell System Technical Journal*, Vol. 27, July 1948, pp. 379–423 and October 1948, pp. 623–656. For an introduction to information theory, see Claude E. Shannon and Warren Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949; John R. Pierce, *An Introduction to Information Theory: Symbols, Signals, and Noise*, New York: Dover, 1961; and Robert W. Lucky, *Silicon Dreams: Information, Man, and Machine*, New York: St. Martin's Press, 1989.

[7] See, for example, Sergio Verdú, "Fifty Years of Shannon Theory," *IEEE Transactions on Information Theory*, Vol. 44, No. 6, October 1998, p. 2057; Amos Joel, "Telecommunications and the IEEE Communications Society," *IEEE Communications Magazine*, 50th Anniversary Commemorative Issue, May 2002.

[8] M. Mitchell Waldrop, *The Dream Machine: J. C. R. Licklider and the Revolution That Made Computing Personal*, Penguin Books, 2001, p. 81. (An article in the September 1952 issue of *Scientific American* said about the term: "It is almost certain that 'bit' will become common parlance in the field of information, as 'horsepower' is in the motor field.")

[9] Gallager, interview by author. For more on the reaction to Shannon's paper, see: David Slepian, "Information Theory in the Fifties," *IEEE Transactions on Information Theory*, Vol. IT-19, No. 2, March 1973; James L. Massey, "Information Theory: The Copernican System of Communications," *IEEE Communications Magazine*, Vol. 22, No. 12, December 1984, p. 28; John R. Pierce, "The Early Days of Information Theory," *IEEE Transactions on Information Theory*, Vol. IT-19, No. 1, January 1973.

[10] Colin Cherry, interview by F. W. Hagemeyer, London, England, September 6,1976 (audio recording obtained by author).
[11] Solomon W. Golomb et al., "Claude Elwood Shannon (1916-2001)," *Notices of the American Mathematical Society*, Vol. 49, No. 1, p. 11.
[12] Brockway McMillan, interview by Hagemeyer, Whippany, New Jersey, February 8, 1977 (audio recording obtained by author).
[13] Fano, interview by author.
[14] No wonder *Scientific American* called the paper "the Magna Carta of the information age" (John Horgan, "Claude E. Shannon: Unicyclist, Juggler and Father of Information Theory," *Scientific American*, January 1990).
[15] Horgan, "Profile: Claude E. Shannon," p. 72. A similar story was told by Shannon to Arthur Lewbel (Arthur Lewbel, "A Juggler's Tribute to Claude Shannon." *IEEE Information Theory Society Newsletter*, December 2001, pp. 9–12).
[16] Claude Shannon, interview by F. W. Hagemeyer, Winchester, Mass., February 28, 1977 (audio recording obtained by author).
[17] Shannon, interview by Hagemeyer.
[18] N. J. A. Sloane and Aaron D. Wyner, ed., *Claude Elwood Shannon: Collected Papers*, John Wiley & Sons, 1993, p. xi ("Biography of Claude Elwood Shannon"). This book is a comprehensive collection of most of the papers written by Shannon.
[19] Shannon, interview by Hagemeyer.
[20] Claude Shannon, Kyoto Prize speech draft, 1985, Claude Elwood Shannon Papers, Manuscript Division, Library of Congress, Washington, D.C.
[21] Sloane and Wyner, op. cit., p. xi.
[22] Claude Elwood Shannon, *A Symbolic Analysis of Relay and Switching Circuits*, S.M. Thesis, MIT, 1940, p. 14.
[23] Letter from Charles S. Rich, secretary of the American Institute of Electrical Engineers, to Vannevar Bush (May 4, 1938), Claude Elwood Shannon Papers, Manuscript Division, Library of Congress, Washington, D.C.
[24] Amos Joel, telephone engineer, former Bell Labs researcher, telephone interview by author, May 30, 2003.
[25] Sloane and Wyner, op. cit., p. xii.
[26] Letter from R. H. Smith to Karl Compton (April 11, 1939), President Letters, MIT Archives.
[27] Letter from Karl Compton to R. H. Smith (April 13, 1939), President Letters, MIT Archives.
[28] Shannon, interview by Hagemeyer.
[29] Sloane and Wyner, op. cit., pp. 455–456.
[30] Ibid.
[31] Letter from Frank Aydelotte, director of the Institute for Advanced Study at Princeton, to Shannon (April 11, 1940), Claude Elwood Shannon Papers, Manuscript Division, Library of Congress, Washington, D.C.
[32] Shannon, interview by Hagemeyer; Shannon would say the same thing in his interview with Price.
[33] Shannon, interview by Price.
[34] Shannon, interview by Hagemeyer.
[35] Ibid.
[36] Ibid.
[37] Brockway McMillan, mathematician, former Bell Labs researcher, telephone interview, June 3, 2003.
[38] Shannon, interview by Price.
[39] Mindell, op. cit., p. 187
[40] Ibid., pp. 186–187
[41] Ibid., p 258
[42] Pesi R. Masani, *Norbert Wiener: 1894-1964*, Birkhauser Verlag, 1990, p. 39.
[43] Wiener has two autobiographies: *Ex-Prodigy* (MIT Press, 1953) and *I Am a Mathematician: The Later Life of a Prodigy* (Doubleday & Company, 1956).
[44] Pierce, *An Introduction to Information Theory*, pp. 41–42.
[45] Norbert Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Technology Press and John Wiley & Sons, 1949, pp. 1–10.

[46] David Jerison and Daniel Stroock. "Norbert Wiener," *Notices of the American Mathematical Society*, Vol. 42, No. 4, April 1995, p. 438.
[47] Mindell, op. cit., p. 280.
[48] Shannon, interview by Hagemeyer.
[49] Mindell, op. cit., p. 280.
[50] R. B. Blackman, H. W. Bode, and C. E. Shannon, "Data Smoothing and Prediction in Fire-Control Systems," Summary Technical Report of the National Defense Research Committee, Division 7, Gunfire Control, Washington, D. C., 1946, p. 80 (Chapter 7: General Formulation of the Data-Smoothing Problem).
[51] Mindell, op. cit., pp. 254–258.
[52] For a comprehensive account of the emergence of electrical signals as a way to represent the world in control and communication systems throughout the two world wars, see Mindell, op. cit.
[53] M. D. Fagen, ed., *A History of Engineering and Science in the Bell System: National Service in War and Peace (1925–1975)*, Bell Telephone Laboratories, 1978, p. 301.
[54] The cryptography scheme used in the X System was known as the "Vernam cipher" or "one-time pad" (in the case when used keys were discarded). The one-time pad is not only practically unbreakable, but theoretically unbreakable as well. There are no patterns in the messages to make possible a "statistical attack." A "brute force" attack doesn't work either because for each key the enemy tries, he gets a different deciphered message. In the case, say, of a fifteen-character message, the enemy will generate all possible fifteen-character text strings when trying the keys; he might obtain "ATTACK TOMORROW," or "NO ATTACK TODAY," or any other fifteen-character text, from gibberish to pieces of Shakespeare's plays. The enemy has no way to decide which of the messages is the "right" one. For a complete description of this method, see: David Kahn, *The Codebrakers: The Story of Secret Writing*, New York: Macmillan Company, 1967, pp. 394–403.
[55] Fagen, op. cit., p. 297; William R. Bennett, "Secret Telephony as a Historical Example of Spread-Spectrum Communication," *IEEE Transactions on Communication*, Vol. COM-31, No. 1, January 1983, p. 98.
[56] Shannon, interview by Hagemeyer.
[57] Fagen, op. cit., p. 317.
[58] Andrew Hodges, *Alan Turing: The Enigma*, New York: Simon and Schuster, 1983, pp. 242–255.
[59] Shannon, interview by Hagemeyer.
[60] Ibid.
[61] Hodges, op. cit., p. 250.
[62] Robert G. Gallager, "Claude E. Shannon: A Retrospective on His Life, Work, and Impact," *IEEE Transactions on Information Theory*, Vol. 47, No. 7, November 2001, p. 2682.
[63] Claude Shannon, "A Mathematical Theory of Cryptography," Technical Memorandum, September 1, 1945. MIT Archives, p. 7.
[64] Ibid., p. 74. It seems this was one of the first occurrences of the expression—an expression Shannon would repeat exactly once in his 1948 paper (p. 393 in the original). It is also worth to mention that the term was used before Shannon's paper came out. On March 24, 1948, during the Institute of Radio Engineers National Convention at the Hotel Commodore in New York, Shannon gave a talk titled "Information Theory." The short abstract for the talk read, "Limitations on the transmission of information imposed by bandwidth, time, and signal-to-noise ratio" (Proceedings of the IRE, 1948, p. 366). In 1949, Shannon's talk appeared in print as "Communication in the Presence of Noise" (Reprinted in Sloane and Wyner, op. cit., pp. 160–172).
[65] McMillan, interview by author.
[66] Gallager, op. cit., p. 2682.
[67] Betty Shannon, interview by author.
[68] Shannon, interview by Hagemeyer; Shannon, interview by Price; Betty Shannon, interview by author.
[69] Slepian, interview by author; McMillan, interview by author.
[70] McMillan, interview by author.
[71] S. Millman, ed., *A History of Engineering and Science in the Bell System: Communication Sciences (1925–1980)*, AT&T Bell Laboratories, 1984, p. 3.
[72] McMillan, interview by author.
[73] Shannon, interview by Price.

[74] Ibid.; Shannon also mentioned Hartley's paper in his interview with Hagemeyer and in his Kyoto Prize speech draft.
[75] R. V. L. Hartley, "Transmission of Information," *Bell System Technical Journal*, July 1928, pp. 535–563.
[76] Ibid., p. 536.
[77] Ibid., p. 536.
[78] Ibid., p. 540.
[79] Millman, op. cit., pp. 152–160.
[80] John R. Pierce, *Signals: The Telephone and Beyond*, San Francisco: W. H. Freeman and Company, 1981, p. 41.
[81] Hendrik W. Bode, "Obituary Statement: Harry Nyquist," *IEEE Transactions on Automatic Control*, Vol. AC-22, No. 6, December 1977.
[82] In the 1920s, Harry Nyquist published two important papers on the transmission of signals in telegraph systems, both of them cited in Shannon's 1948 paper: "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, April 1924, pp. 324–346, and "Certain Topics in Telegraph Transmission Theory," *Transactions of the A.I.E.E.*, Vol. 47, April 1928, pp. 617–644 (reprinted in: *Proceedings of the IEEE*, Vol. 90, No. 2, February 2002, pp. 280–305).
[83] Nyquist developed a formula to measure the transmission speed of a telegraph system that also used the logarithmic function.
[84] Massey, op. cit., p. 27.
[85] Mindell, op. cit., pp. 134–135.
[86] Shannon, Kyoto Prize speech draft, op. cit.
[87] Sloane and Wyner, op. cit., pp. 455 and 456.
[88] Shannon, interview by Price.
[89] Shannon, interview by Hagemeyer.
[90] Pierce, *An Introduction to Information Theory*, p. 40.
[91] Fano, interview by author.
[92] David Slepian, personal communication with author, June 15, 2003.
[93] Shannon, interview by Price.
[94] Waldrop, op. cit., p. 81.
[95] Shannon and Weaver, op. cit., pp. 43 and 44.
[96] Ibid., p. 44.
[97] For technical and historical details on the sampling theorem, see Hans Dieter Lüke, "The Origins of the Sampling Theorem," *IEEE Communications Magazine*, April 1999.
[98] Millman, op. cit., pp. 399–418.
[99] Mindell, op. cit., p. 318; for examples of digital computers, see, for example, H. H. Goldstine and Adele Goldstine, "The Electronic Numerical Integrator and Computer (ENIAC)," *IEEE Annals of the History of Computing*, Vol. 18, No. 1, 1996, pp. 10–16; M. M. Irvine, "Early Digital Computers at Bell Telephone Laboratories," *IEEE Annals of the History of Computing*, July–September 2001, pp. 22–42.
[100] James Massey, information theorist, former professor of electrical engineering at ETH, Zurich, personal communication, June 3, 2003.
[101] This is the code according to the International Morse Code. In the American Morse Code, an "X" is a dot-dash-dot-dot sequence.
[102] The Encyclopaedia Britannica article is reprinted in Sloane and Wyner, op. cit., pp. 212–220.
[103] Shannon, interview by Hagemeyer; also Bello, op. cit., p. 140.
[104] Fraunhofer Institute for Integrated Circuits, "Audio & Multimedia MPEG Audio Layer-3," URL: `http://www.iis.fraunhofer.de/amm/techinf/layer3/`
[105] Haus, interview by author.
[106] Fano, interview by author.
[107] Massey, op. cit., p. 27.
[108] Shannon, interview by Hagemeyer.
[109] Shannon and Weaver, op. cit., p. 51.
[110] Arthur S. Eddington, *The Nature of the Physical World*, Macmillan, 1948, p. 74.

[111] Anson Rabinbach, *The Human Motor: Energy, Fatigue, and the Origins of Modernism*, University of California Press, 1992, p. 47.
[112] For a comprehensive discussion about Maxwell's demon, entropy, and information see Harvey S. Leff and Andrew F. Rex, ed., *Maxwell's Demon: Entropy, Information, Computing*, Princeton University Press, 1990.
[113] Shannon, interview by Hagemeyer.
[114] Raphael D. Levine and Myron Tribus, ed., *The Maximum Entropy Formalism (A Conference Held at the Massachusetts Institute of Technology on May 2-4, 1978)*, MIT Press, pp. 2 and 3.
[115] Shannon and Weaver, op. cit., p. 3.
[116] Norbert Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press and John Wiley & Sons, 1961, pp. 62–64.
[117] Fano, interview by author. Fano also mentions Wiener's statement on information and entropy in the preface of his book, *Transmission of Information: A Statistical Theory of Communications*, MIT Press, 1961, p. vii.
[118] Shannon, interview by Hagemeyer. Wiener says the same about Shannon in *I Am a Mathematician*, op. cit., p. 179.
[119] Letter from Shannon to Wiener (October 13, 1948), Wiener Papers, MIT Archives.
[120] Letter from Wiener to Shannon (no date), Wiener Papers, MIT Archives.
[121] Shannon cited and acknowledged Wiener's work several times in his 1948 paper. In a footnote, he observed: "Communication theory is heavily indebted to Wiener for much of its basic philosophy and theory. His classic NDRC report, *The Interpolation, Extrapolation and Smoothing of Stationary Time Series* (Wiley, 1949), contains the first clear-cut formulation of communication theory as a statistical problem, the study of operations on time series. This work, although chiefly concerned with the linear prediction and filtering problem, is an important collateral reference in connection with the present paper. We may also refer here to Wiener's Cybernetics (Wiley, 1948), dealing with the general problems of communication and control." In the acknowledgements, Shannon wrote: "Credit should also be given to Professor N. Wiener, whose elegant solution of the problems of filtering and prediction of stationary ensembles has considerably influenced the writer's thinking in this field." And Wiener himself claimed part of the credit for the creation of information theory. "Information theory has been identified in the public mind to denote the theory of information by bits, as developed by Claude E. Shannon and myself," Wiener wrote in 1956 (Norbert Wiener, "Editorial: What Is Information Theory?" *IEEE Transactions on Information Theory*, Vol. 2, Issue 2, Jun 1956, pp. 48).
[122] Fano, interview by author.
[123] Shannon, interview by Hagemeyer.
[124] For a short story on how the concepts of information and entropy appeared in the works of Wiener, Shannon, and others, see Jeremy Campbell, *Grammatical Man: Information, Entropy, Language, and Life*, Simon and Schuster, 1982. And Wiener and Shannon weren't alone; others were publishing works with similar ideas. See, for example, Dennis Gabor, "Theory of Communication," J. IEE (London), Vol. 93, 1946, pp. 429–457; William G. Tuller, "Theoretical Limitations on the Rate of Transmission of Information," PhD dissertation, MIT, June 1948.
[125] See, for example, Thomas Schneider's web site on information theory and molecular biology, URL: `http://www.lecb.ncifcrf.gov/~toms/pitfalls.html#shannon_entropy_is_a_misnomer`
[126] Paul Penfield, Jr., *Information and Entropy, Course Notes*, MIT, version 1.0.2, January 30, 2003, URL: `http://www-mtl.mit.edu/Courses/6.050/2003/`
[127] Shannon's *Encyclopaedia Britannica* article (Sloane and Wyner, op. cit., p. 213).
[128] Ibid., p. 215.
[129] Thomas M. Cover et al., "Kolmogorov's Contributions to Information Theory and Algorithmic Complexity," *The Annals of Probability*, Vol. 17, No. 3, July 1989, p. 840.
[130] Gerhard Kramer, Bell Labs researcher, interview by author, Murray Hill, New Jersey, February 21, 2003.
[131] "Information and Entropy," MIT course, lecture on February 6, 2003.
[132] Ibid., February 4, 2003.
[133] For more on information in physical systems see Neil Gershenfeld, *The Physics of Information Technology*, Cambridge University Press, 2000.

[134] Seth Lloyd, "Computational Capacity of the Universe," *Physical Review Letters*, Vol. 88, No. 23, June 10, 2002, p. 237901-1.
[135] Ibid., pp. 237901-3 and -4.
[136] Ed Fredkin, Digital Philosophy website, URL: `http://www.digitalphilosophy.org/`
[137] Freeman Dyson, essay on EDGE website, URL:
`http://www.edge.org/3rd_culture/dyson_ad/dyson_ad_index.html`. Also, lecture at MIT, October 17, 2002.
[138] Letter from William Altar, electronics and nuclear physics department, Westinghouse Electric Corp., to Shannon (August 20, 1948), Claude Elwood Shannon Papers, Manuscript Division, Library of Congress, Washington, D.C.
[139] Robert W. Lucky, "When Giants Walked the Earth," *IEEE Spectrum*, April 2001, URL:
`http://www.spectrum.ieee.org/WEBONLY/resource/apr01/refl.html`
[140] Slepian, interview by author.
[141] Shannon and Weaver, op. cit., p. 81.
[142] Brockway McMillan, "The Basic Theorems of Information Theory," *The Annals of Mathematical Statistics*, Vol. 24, No. 2. June, 1953, pp. 196–219.
[143] McMillan, interview by author.
[144] For historical accounts of the development of the concept of information in communication and computing, see William Aspray, "The Scientific Conceptualization of Information: A Survey," *IEEE Annals of the History of Computing*, Vol. 7, No. 2, April 1985, pp. 117–140; E. Colin Cherry, "A History of the Theory of Information," *IEEE Transactions on Information Theory*, Vol. 1, Issue 1, February 1953, pp. 22–43; Hagemeyer, PhD dissertation, op. cit.; R. C. Olby et al., ed., *Companion to the History of Modern Science*, Routledge, 1990, pp. 537–553 (Chapter 34: Cybernetics and Information Theory).
[145] Willis Jackson, ed., *Communication Theory, Papers read at a Symposium on "Applications of Communication Theory" held at the Institution of Electrical Engineers, London, September 22nd-26th 1952*, New York: Academic Press, 1953, p. x.
[146] Golomb et al., op. cit., p. 11.
[147] David Forney, professor of electrical engineering, MIT, interview by author, Cambridge, Mass., Jan 14, 2003.
[148] J. Hagenauer, ed., *Advanced Methods for Satellite and Deep Space Communications, Lecture Notes in Control and Information Sciences 182*, Heidelberg and New York: Springer, 1992, pp. 1–17.
[149] Ibid.; also, Daniel J. Costello, Jr. et al., "Applications of Error-Control Coding," *IEEE Transactions on Information Theory*, Vol. 44, No. 6, October 1998, p. 2533.
[150] Pioneer 9, Space Calendar, NASA Jet Propulsion Laboratory, URL:
`http://www.jpl.nasa.gov/calendar/pioneer9.html`
[151] Golomb, op. cit., p. 10.
[152] Shannon, Kyoto Prize speech draft, op. cit.
[153] Michael Riordan and Lillian Hoddeson, *Crystal Fire: The Invention of the Transistor and the Birth of the Information Age*, W. W. Norton, 1997, p. 239.
1997.
[154] Kramer, interview by author.
[155] Kahn, op. cit., p. 20.
[156] According to comments posted on the web site "Remembering Claude Shannon," Project Echo, Center for History and New Media, George Mason University, URL:
`http://chnm.gmu.edu/tools/surveys/responses/80/`
[157] Robert J. McEliece, Project Echo's website, op. cit., March 12, 2001.
[158] Homer Jacobson, "The Informational Capacity of the Human Eye," *Science*, New Series, Vol. 113, No. 2933, March 16, 1951, pp. 292 and 293.
[159] John R. Pierce, *An Introduction to Information Theory*, p. 234.
[160] Ibid., pp. 230 and 231.
[161] *Encyclopaedia Britannica* article (Sloane and Wyner, op. cit., p. 219).
[162] Noam Chomsky, "Three models for the description of language," *IRE Transactions on Information Theory*, Vol. 2, No. 3, September 1956, p. 115.

[163] Slepian, op. cit., p. 145.
[164] See, for example, Martin A. Nowak and Natalia L. Komarova, "Towards an evolutionary theory of language," *Trends in Cognitive Sciences*, Vol. 5, No. 7, July 2001, pp. 288-295; Martin A. Nowak et al., "An error limit for the evolution of language," *Proceedings of the Royal Society of London, Series B*, Vol. 266, 1999, pp. 2131-2136; Joshua B. Plotkin and Martin A. Nowak, "Language Evolution and Information Theory," *Journal of Theoretical Biology*, Vol. 205, 2000, pp. 147–159.
[165] See, for example, Thomas Schneider's web site on information theory and molecular biology, URL: `http://www.lecb.ncifcrf.gov/~toms/`; David MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003, URL: `http://www.inference.phy.cam.ac.uk/mackay/itprnn/book.html` (Chapter 19: Why have Sex? Information Acquisition and Evolution).
[166] See, for example, Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991 (Chapter 15: Information Theory and the Stock Market).
[167] Claude E. Shannon, "The Bandwagon," *IRE Transactions on Information Theory*, March 1956, p. 3.
[168] Shannon, interview by Hagemeyer; Betty Shannon, interview by author; Fano, interview by author.
[169] M. Mitchell Waldrop, "Claude Shannon: Reluctant Father of the Digital Age," *Technology Review*, July/August, 2001, p. 71; Slepian, interview by author.
[170] Sheldon Hochheiser, historian, AT&T Archives, personal communication, February 19, 2003.
[171] Shannon, interview by Price.
[172] Anthony Liversidge, "Interview: Claude Shannon, Father of the Electronic Information Age," *Omni*, Vol. 9, No. 11, August 1987, p. 66.
[173] Betty Shannon, interview by author.
[174] Hermann Haus, Project Echo's website, op. cit., July 30, 2001.
[175] Haus, interview by author.
[176] Lewbel, interview by author.
[177] Betty Shannon, interview by author.
[178] Ibid.
[179] Slepian, interview by author.
[180] Betty Shannon, interview by author.
[181] Shannon, interview by Hagemeyer.
[182] "Information Theory," Seminar Notes, March 1956, Shannon Papers, MIT Archives.
[183] Robert E. Kahn, "A Tribute to Claude E. Shannon (1916–2001)," *IEEE Communications Magazine*, July 2001, p. 18.
[184] Forney, interview by author.
[185] Gallager, interview by author.
[186] Leonard Kleinrock, professor of computer science, University of California at Los Angeles, personal communication, June 5, 2003.
[187] Andrew Viterbi, interview by David Morton, Oral History, IEEE History Center, October 29, 1999, URL: `http://www.ieee.org/organizations/history_center/oral_histories/transcripts/viterbi.html`
[188] J. A. N Lee, J. McCarthy, and J. C. R. Licklider, "The beginnings at MIT," *IEEE Annals of the History of Computing*, Vol. 14, Issue 1, 1992, pp. 18–30; MIT Laboratory for Computer Science, Timeline, URL: `http://timeline.lcs.mit.edu/`; Institute Archives & Special Collections, MIT History, "History of the Laboratory for Computer Science (LCS)," URL: `http://libraries.mit.edu/archives/mithistory/histories-offices/lcs.html`. For other accounts on the origins and early days of computing and networking, see Howard Rheingold, *Tools for Though: The History and Future of Mind-Expanding Technology*, MIT Press, 2000 (a revised version of the original 1985 issue), or on the web, URL: `http://www.rheingold.com/texts/tft/`. Waldrop, op. cit.
[189] Shannon, interview by Hagemeyer.
[190] Claude Shannon, "Programming a Computer for Playing Chess," *Philosophical Magazine*, Series 7, Vol. 41, No. 314, March 1950, pp. 256–275 (reprinted in Sloane and Wyner, op. cit, pp. 637–656).
[191] Shannon, Kyoto Prize speech draft, op. cit.

[192] Both cite Shannon in the acknowledgements of their theses: Leonard Kleinrock, "Message Delay in Communication Nets with Storage," PhD dissertation, MIT, 1962, p. iv; Ivan Sutherland, "Sketchpad, A Man-Machine Graphical Communication System," PhD dissertation, MIT, 1963, p. 4.
[193] Edward O. Thorp, "The Invention of the First Wearable Computer," *IEEE 2nd International Symposium on Wearable Computers*, 1998.
[194] Boris Tsybakov, Project Echo's website, op. cit., June 20, 2001.
[195] Kees Immink, Project Echo's website, op. cit., July 24 2001.
[196] Shannon, Kyoto Prize speech draft, op. cit.
[197] Horgan, "Profile: Claude E. Shannon," p. 75. In the symposium, McEliece introduced Shannon as "one of the greatest scientific minds of our time." Shannon reacted saying, after the applause, "This is...ridiculous!" (Anthony Ephremides, "Historian's Column," *IEEE Information Theory Society Newsletter*, Vol. 51, No. 1, March 2001, p. 1).
[198] Betty Shannon, interview by author.
[199] Ibid.
[200] Shannon, interview by Hagemeyer.
[201] The song included several footnotes. In the first one, Shannon wrote: "We are with Elliot and will freely use footnotes to clarify and amplify our meaning." He was referring to T. S. Elliot's "The Waste Land." Published in 1922 with a "wealth of footnotes," Shannon wrote, it caused "considerable commotion among the critics."
[202] Horgan, "Profile: Claude E. Shannon," p. 73.
[203] Lewbel, interview by author; See also Lewbel, op. cit., pp. 9–12.
[204] Fano, interview by author.
[205] For different accounts on Bell Labs' history, inventions, and culture, see Fagen, op. cit.; Millman, op. cit.; Prescott C. Mabon, *Mission Communications: The Story of Bell Laboratories*, Bell Telephone Laboratories, 1975; Jeremy Bernstein, *Three Degrees Above Zero: Bell Labs in the Information Age*, Charles Scribner's Sons, 1984; Narain Gehani, *Bell Labs: Life in the Crown Jewel*, Silicon Press, 2003.
[206] Bell Labs Awards, URL: `http://www.bell-labs.com/about/awards.html`
Bell Labs Press Release, "Bell Labs receives its 30,000th patent," March 10, 2003, URL: `http://www.lucent.com/press/0303/030310.bla.html`
[207] Kramer, interview by author. Feynman, by the way, would have run into Shannon had he not turned down a job offer from Bell Labs in the spring of 1941 (James Gleick, *Genius: The Life and Science of Richard Feynman*, New York: Vintage Books, 1993. p. 137). One wonders what the two of them might have done together.