A Mind, a Machine, and a Game in Between

Claude Shannon and the Origin of the Information Age

Mithra Karamchedu | Harvey Mudd College

Supervising Instructor: Jemma Lorenat, Pitzer College

"The specter of information is haunting sciences." In 1989, the quantum physicist Wojciech H. Zurek invited his colleagues at the Santa Fe Institute to respond to this simple yet profound prompt. Zurek's colleagues, each of whom felt the "specter of information" in very deep ways within their work, spread far across the scientific community. They were mathematicians, computer scientists, statisticians, physicists, cosmologists, and biologists; even more broadly, ecologists, economists, and psychologists could feel the reverberations of information theory. As Zurek suggested, information theory was like an inexplicable spark that seemed to inspire every new idea in the sciences. But what inspired Claude Shannon, the pioneer of information theory, also toed the line between the simple and the profound: games, puzzles, and toys.

It is always challenging to encapsulate someone in a single word, but many would describe Claude Shannon as *playful*. Jimmy Soni and Rob Goodman's 2017 biography of Shannon, *A Mind at Play*, does just that, showing links between Shannon's intellectual work and his playful and creative side, from his mathematical work on juggling to his development of chess-playing machines. In fact, Shannon's relationship with play might go even further, some of his most technical work being not only connected to but also a direct product of his playful experimentation. For Shannon, "play" usually meant conceptualizing toys and games, or finding an abstract property guiding a practical engineering system.

As a child growing up in the small town of Gaylord, Michigan in the 1920s and 30s, Shannon immersed himself in creating machines and contraptions that exercised his inner whimsy. According to Soni and Goodman,

> One neighbor, Shirley Hutchins Gidden, offered to the *Otsego Herald Times* that Shannon and her brother, Rodney Hutchins, were a conspiratorial pair . . . One experiment stood out: a makeshift elevator built by the two boys inside the Hutchins family barn. Shirley was the 'guinea pig,' the first to take a ride on the elevator, and it says something about the quality of the boys' handiwork (or her luck) that she lived to tell the tale to a newspaper seven years later (Soni and Goodman 11).

Shannon and his close friend Rodney Hutchins were often up to something, and their most famous invention hints at the direction Shannon's young mind was blossoming.

Gaylord was too small a town for major telephone and telegraph companies to justify expanding to, but luckily, its citizens were both industrious and enterprising. Like many other small towns, they installed their own wire across the town fences, forming a communications network that Shannon not only took notice of, but took part in: "Claude's stretch is electric. He charged it himself: he hooked up dry-cell batteries at each end, and spliced spare wire into any gaps to run the current unbroken. Insulation was anything at hand: leather straps, glass bottlenecks, corncobs, inner-tube pieces. Keypads at each end—one at his house on North Center Street, the other at his friend's house [Rodney Hutchins] a mile away—made it a private barbed-wire telegraph" (Soni and Goodman 4). With this invention in particular, we see Shannon's mind begin to embrace the problems of information not just theoretically but mechanically. The question here is, how can we *build* something that sends information? But Shannon was just as fascinated by the more mathematical aspects of communication, being especially fond of Morse code and cryptography: "On April 17, 1930, thirteen-year-old Claude attended a Boy Scout rally and won 'first place in the second class wig-wag signaling contest.' The object was to speak Morse code with the body, and no scout in the county spoke it as quickly or accurately as Claude" (Soni and Goodman 10–11). Shannon's favorite Edgar Allan Poe story was "The Gold-Bug," which ends with a brief lesson on cryptography and breaking ciphers with frequency analysis (Gleick 171). Although it was many years later when Shannon would write his seminal paper "A Mathematical Theory of Communication" (abbr. *AMToC*), his childhood pastimes, driven by his whimsy and desire to play, leave clues about the specter of information that may have been brewing inside.

Both mechanical and mathematical, Shannon struggled to decide whether he should major in engineering or mathematics; he knew whatever field he was interested in touched both, but that field didn't have a complete name quite yet. After starting in the fall of 1932, he majored in both electrical engineering and mathematics at the University of Michigan, leaving him "trained in two fields that would prove essential to his later successes" (Soni and Goodman 17).

In his 2011 book *The Information*, James Gleick highlights a pivotal moment in Shannon's academic journey that brought him closer to uniting these two interests. In 1936, Shannon's senior year, "he saw a postcard on a bulletin board advertising a graduate-student job at the Massachusetts Institute of Technology. Vannevar Bush, then the dean of engineering, was looking for a research assistant to run a new machine with a peculiar name: the Differential Analyzer" (Gleick 171). Bush's Differential Analyzer was monumental in concept and in scale: the size of a whole room, the Analyzer would solve differential equations by using wheels and rotors to perform integration. It was an *analog* machine; to solve a differential equation, it would create an analogous model of that equation inside itself, a microcosm of that equation's continuity and detail—just like how film captures images and vinyl stores sound. But whenever the input differential equation changed, the machine's analog structure meant that its entire internals would have to as well. Bush hired Shannon to help with this problem, beginning Shannon's master's degree at MIT.

At the time, circuit design was seen as more of an art than a science, in that devising circuits to solve a problem was driven by a deep intuition—a trade knowledge—rather than formal, systematic logic. Working on Bush's Differential Analyzer, Shannon found this slightly perturbing. In the summer of 1937, Shannon also began an internship at AT&T's Bell Laboratories in New York, "the heart of the phone company, the owner of the most complicated, far-reaching network of circuits in existence" (Soni and Goodman 38). While working there, he couldn't help but feel that there was a real *logic* behind the circuitry of Bell Labs' telephone networks and Bush's Differential Analyzer, that there was something mathematical that could always drive the mechanical. His mind turned to Boolean algebra, the formal system that lay at the heart of mathematical logic.

To better understand Shannon's work, let us briefly explore the essence of Boolean algebra, which we can think of as the theory of three operations, known as And, Or, and Not, given by the "truth tables" below:

And ($\wedge$)		
0	0	0
0	1	0
1	0	0
1	1	1

Or ($\vee$)		
0	0	0
0	1	1
1	0	1
1	1	1

Not ($\neg$)	
0	1
1	0

For example, let us consider the operation And. And takes in two bits as input and gives one bit as output—we can think of a 0 bit as being "false" and a 1 bit as "true." Since And accepts two inputs, each of which can be either 0 or 1, there are $2^2 = 4$ different 2-bit inputs we can give the AND operator: 0 0, 0 1, 1 0, and 1 1.

We see that AND returns a 1 only when both its inputs are 1; otherwise, AND is 0. If x is the first input and y is the second input, then x AND y is true precisely when x is true *and* y is true. That is, AND simply tells us whether both its inputs are true, just like its name would suggest. OR, on the other hand, returns true when either x *or* y is true; when both inputs are false, OR is false too.

But the third operation, NOT, is a little different from the other two. NOT accepts only one bit as input and gives one bit as output. If our input x is true, then NOT x is false; if x is false, then NOT x is true. NOT flips the truth or falsehood of its input. In Boolean algebra, our inputs x, y, z, ... are propositions, statements that we would like to evaluate as either true or false. As an example, if we would like to make sure that (1) today is Monday, and (2) it is either cloudy or not raining, then we would write the Boolean expression

$$x \,\text{AND}\, (y \,\text{OR}\, (\text{NOT}\, z)) = x \wedge (y \vee \neg z),$$

where x is the proposition "Today is Monday," y is "It is cloudy," and z is "It is raining." The Boolean expression $x \wedge (y \vee \neg z)$ is true precisely when today is Monday, and it is either cloudy or not raining.

The reason why these three operations are so powerful is quite surprising! If we were looking at Boolean operations that accept two bits as input and return one bit as output, then we would have $2^4 = 16$ possible Boolean operations to consider (as we need to assign an output of 0 or 1 to each of the four 2-bit input combinations). Interestingly, every single one of these 16 operators can be expressed as some combination of ANDs, ORs, and NOTs chained together, and this result is a deep foundation of Boolean algebra. Let us consider an arbitrary operation from these 16:

XOR ($\oplus$)		
0	0	0
0	1	1
1	0	1
1	1	0

In context, XOR is the "exclusive or": when we say "It's either heads or it's tails," we're really saying "It's either heads xor it's tails," since we don't have the situation where it's heads and tails at the same time. In principle, we could define XOR as this "hard-coded" truth table, but we also have another way now! XOR is true precisely when either x or y is true, but not both. That is, we need to have x OR y, but also not x AND y. In other words, $x \,\text{XOR}\, y = (x \,\text{OR}\, y) \,\text{AND}\, (\text{NOT}\,(x \,\text{AND}\, y))$. Using the logical symbols for AND, OR, NOT, and XOR, we find that

$$x \oplus y = (x \vee y) \wedge \neg(x \wedge y).$$

And we have done it! We have defined XOR using our existing building blocks. Not only can each of these 16 Boolean operators be written as a combination of ANDs, ORs, and NOTs, but *every* Boolean operator with any number of inputs and outputs can too. In the parlances of electrical engineering and mathematical logic, we say that AND, OR, and NOT are a *universal* gate set or *functionally complete*, respectively.[1] This is a powerful result. With enough effort, Boolean algebra can capture essentially every logical statement we can think of, but everything in Boolean algebra can also be written as ANDs, ORs, and NOTs. As a result, logic itself can be framed as the story of these three operations, and Shannon knew that a beautiful and fascinating world lay here.

Up until this point, Boolean algebra mostly lay in the domain of mathematics and philosophy—Shannon actually learned it in his senior-year philosophy class, being especially fond of the word "Booooooooolean" (Soni and Goodman 37). By simply *playing* with circuits and logic—fields that were considered mostly nonoverlapping—and seeing what he could find, Shannon discovered a startling fact that transformed an art into a science. He realized that circuit design is not just driven by logic, but in a sense, circuit design *is* Boolean algebra. In Shannon's 1937 master's thesis, "A Symbolic Analysis of Relay and Switching Circuits"[2] (abbr. *ASAoRaSC*), he showed that the AND, OR, and NOT operators from Boolean algebra can be implemented as three very simple circuits or "logic gates." To implement an AND gate, Shannon found a circuit that would produce output current if both its inputs had current flowing through them. His OR gate produced output current when either input had current, and his NOT gate produced output current only when there was no input current. And these three gates were enough, showing that machines could perform logic and were even *logical*

[1] In fact, we don't even need to use all three! Since $x \vee y = \neg(\neg x \wedge \neg y)$ and $x \wedge y = \neg(\neg x \vee \neg y)$ (known as *De Morgan's laws*), we can write AND as a combination of ORs and NOTs, and OR as a combination of ANDs and NOTs. Both AND and NOT as well as OR and NOT are universal.

[2] Later published in the 1938 issue of *Transactions of the American Institute of Electrical Engineers.*

in design. Just like that, he made the mechanical mathematical.

But there was a part of Shannon's thesis that was characteristically "Claude." After Shannon introduces these major ideas, he first spends time showing how to convert circuits with a nonstandard structure (a non-*series-parallel* structure, to be precise) to his algebraic form by applying certain transformations; next, he covers some general mathematical results[3] and gives helpful examples where he translates different functions to circuit form; and then, he spends the rest of his thesis just having fun! According to Gleick, Shannon "followed this tower of abstraction with practical examples—inventions, on paper, some practical and some just quirky" (Gleick 174). The language stays technical, but Shannon shows us various contraptions we can build using these ideas: an electric adder circuit, a "Selective Circuit," and even an electric combination lock. "An electric lock is to be constructed with the following characteristics," he writes. "There are to be five pushbutton switches available on the front of the lock. These will be labeled a, b, c, d, e ..." (Shannon, *ASAoRaSC* 722). Shannon's discovery of the logical foundation of circuit design started with him playing with two different fields until they fused into a more complete whole—and very fittingly, he injected the same playful spirit into his thesis. Now he had gone the other way, making the mathematical mechanical.

Shannon's master's thesis earned him widespread acclaim, being called by his contemporaries "possibly the most important, and also the most famous master's thesis of the century," "one of the greatest master's theses ever," and "the most important master's thesis of all time" (Soni and Goodman 42). Interestingly, in the same year Alan Turing published his groundbreaking paper "On Computable Numbers, with an Application to the Entscheidungsproblem," where he introduced the *Turing machine*, a theoretical model that now defines what it means for something to be a proper "computer." Although important in different ways, Shannon and Turing's work shared a key aspect in common: an emphasis on *digital* over analog machines, based on discrete symbols (i.e. bits) rather than true continuity. The impact was sudden: "Less than a decade after Shannon's paper, the great analog machine, the differential analyzer, was effectively obsolete, replaced by digital computers

[3] These results were mainly about how to handle systems of simultaneous equations, as well as understanding the behavior of "star" and "mesh" networks (Shannon, *ASAoRaSC* 716).

that could do its work literally a thousand times faster" (Soni and Goodman 43).

After Shannon completed his master's thesis, Bush saw it as an obligation to direct Shannon's originality toward the truly meaningful scientific problems of the time. He encouraged Shannon to pursue a Ph.D. thesis that would extend these logical ideas to genetics, developing a mathematical system—just like Boolean algebra—that could formally model the basis of genetic variation. In 1940, Shannon received his doctorate in mathematics from MIT with a dissertation titled "An Algebra for Theoretical Genetics"; although Shannon's work was deeply insightful and recognized by some contemporaries, it was sadly undervalued, even by Shannon himself. His attention was shifting, and the real idea he wanted to begin playing with was *communication*. In a 1939 letter to Bush, he wrote, "Off and on 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. . . ." (Soni and Goodman 60).

In the ten years that followed after starting his Ph.D., Shannon spent another summer at Bell Labs, worked as a postdoctoral fellow at the Institute for Advanced Study, and reunited with Bell Labs as part of the National Defense Research Committee during World War II, before finally joining Bell Labs as a full-time researcher within their mathematics department. Even though it was funded by AT&T, an industrial telephone and telegraph company, Bell Labs offered an intellectual haven whose scientific legacy seems almost mythical: Soni and Goodman cite Jon Gertner, who writes that Bell Labs "was where the future, which is what we now happen to call the present, was conceived and designed" (Soni and Goodman 65). In Bell Labs, engineers and especially mathematicians were given the freedom to pursue any projects generally related to the Labs' focus on communication, allowing Shannon to play with the ideas that truly interested him. In his summer at Bell Labs as a Ph.D. student, Shannon worked on two problems that let him connect his mathematical interests in communication with "the part of his nature that was hardheaded and practical" (Soni and Goodman 72). In one of them, for example, Shannon needed to find the smallest number of wire colors that would be needed so that different wires that connect to the same vertex in a network (in graph-theoretic terms, a *Shannon multigraph*) would have different colors.[4] At Bell Labs, he was in an environment where he could get his hands dirty, a type of place that let his mind flourish.

During World War II, Shannon worked with Bell Labs as part of the National Defense Research Committee, tasked with handling noise and interference within an "antiaircraft gun," a system designed to fire projectiles at fast-moving enemy planes. Even though Shannon was not especially passionate about his wartime research, it would provide a very robust foundation for his future academic plans in the theory of communication. After the war, Shannon was offered a full-time position at the Labs, and he was finally able to devote himself to the problem of the "transmission of intelligence" that he mentioned to Bush in 1939. Once he finished his Bell Labs workday, Shannon regularly spent many hours late into the night grappling with this problem; he found key questions at its center: what does it mean to send or *communicate* information? Moreover, what really is *information*?

Shannon's solution was staggeringly insightful—but also rooted in playfulness at each step along the way. According to Shannon, information is a measure of *surprise* or *unpredictability.* For example, if we are writing a message in English and we use the word "I," then there's a very reasonable chance that the next word will be "am", "do," or "have." In a sense, the words "am," "do," and "have" are slightly *redundant* because we expected to see them with high probability. English, like several other languages, has many of these redundancies built-in: "If the letter *u* ends a word, the word is probably *you.* If two consecutive letters are the same, they are probably *ll*, *ee*, *ss*, or *oo.* And structure can extend over long distances: in a message containing the word *cow*, even after many other characters intervene, the word *cow* is relatively likely to occur again. As is the word *horse*" (Gleick 226).

Shannon experimented with this idea by creating a wonderful toy example: according to Gleick, he pulled out a mystery novel from his bookshelf and took note of the frequencies of different letter combinations. First, he made a "zero-order approximation," made with random letters and no pattern/redundancy:

```
XFOML RXKHRJFFJUJ ZLPWCFWKCYKJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.
```

[4] Shannon's result appears in his 1949 paper "A Theorem on Coloring the Lines of a Network" (abbr. *AToCtLoaN*) in the *Journal of Mathematics and Physics.* His solution is quite simple: we take the maximum number of wires connected to a vertex in the graph (the maximum *degree*), multiply it by 1.5, and find the greatest integer at or below this value. Formally, $\lfloor 1.5m \rfloor$ is an upper bound on the number of wire colors needed, where m is the maximum number of wires connected to a single vertex (Shannon, *AToCtLoaN* 1).

Then, he made a first-order approximation, where "each character is independent of the rest, but the frequencies are those expected in English: more *e*'s and *t*'s, fewer *z*'s and *j*'s, and the word lengths look realistic" (Gleick 226):

`OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.`

Although it's hard to articulate why, this feels mysteriously closer to English—the vowels and consonants somehow feel more well-balanced. Shannon repeated the process with a second-order approximation, taking into account the frequencies of *bigrams*, or letter pairs:

`ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.`

And then a third-order approximation, incorporating *trigram* frequencies:

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

Very quickly, the text becomes startlingly English-like! When we instead try to approximate English with word instead of letter frequencies, we get the following second-order approximation:

`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.`[5]

From these approximations, Shannon realized that languages are filled with pattern, and patterns are redundancies—patterns are *expected* pieces of language—so they don't give us much new information. If we strongly expect to see a certain character or symbol, we don't learn too much when we actually see it. As an example, if we saw the words "the sky is," then we wouldn't be so surprised if the next word was "blue": this is a quartet of words we encounter quite often. On the other hand, we *would* be very surprised if the next word was "mischievous," and this surprise teaches us something new that we wouldn't have expected before. In another example, Shannon asked his wife and fellow researcher Betty to be his "test subject":

[5]See Gleick 226–227, and Soni and Goodman 146–148, for text approximation examples.

> He pulled a book from the shelf (it was a Raymond Chandler detective novel, *Pickup on Noon Street*), put his finger on a short passage at random, and asked Betty to start guessing the letter, then the next letter, then the next. The more text she saw, of course, the better her chances of guessing right. After "A SMALL OBLONG READING LAMP ON THE" she got the next letter wrong. But once she knew it was *D*, she had no trouble guessing the next three letters (Gleick 230).

The letters "ESK" came for free—with little surprise—so they didn't offer much information. But the first letter "D" *did* surprise us, so it does in fact give us new information. In both these examples, we see that information is a measure of the surprise or unlikelihood of a message. Curiously enough, this definition of Shannon doesn't depend on the meaning of the message, just its statistical structure: "Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem" (Shannon, *AMToC* 1). Once again, we see Shannon balancing his desire for mathematical abstraction with the practicality of building something. But Shannon's removal of meaning from information raises an important question: when we have a message, how can we measure its information content?

In his 1948 paper "A Mathematical Theory of Communication," Shannon lays forth all these ideas, defining the problem of communication, what it means for a message to have information, and how we can measure information. After showing us his approximations of the English language, Shannon mentions that "The redundancy of ordinary English, not considering statistical structure over greater distances than about eight letters, is roughly 50%" (Shannon, *AMToC* 14). In a paper published about two years later, "Prediction and Entropy of Printed English,"[6] Shannon "raised that estimate to 75 percent—warning, however, that such estimates become 'more erratic and uncertain, and they depend more critically on the type of text involved"' (Gleick 229). Assuming that the redundancy of English is 75%, we should be able to shrink or *compress* our messages to be give-or-take 75% shorter while still retaining the same information. Specifically, there will exist a scheme that compresses or *encodes* our messages to a fraction of their original size (here, about 1/4), and a scheme that reconstructs or *decodes* them.[7] The length of a message, when it has been

[6] Published in the *Bell System Technical Journal* in 1951.

[7] Numerous algorithms exist for this purpose, such as Shannon-Fano coding, Huffman coding, Lempel-Ziv

compressed as much as possible, is its information content—known as its *entropy*. And the units of entropy: bits.[8]

Shannon's predecessors in this style of communications engineering, Ralph Hartley and Harry Nyquist, had answered this question about entropy for a specific case, when each symbol or character in the message has the same probability of occurring. For this case, Hartley and Nyquist's solution was that the entropy H of the message would be

$$H = n \log_2 s,$$

where n is the original length of the message and s is the number of possible symbols in the alphabet we are drawing from. Shannon expanded on Hartley and Nyquist's formula; to fully understand Shannon's formula, let us assume that we have a family of symbols that can occur, consisting of Symbol 1, Symbol 2, Symbol 3, ... that we chain together to form messages. Formally speaking, we say that we have a *random variable* X that represents the symbol that ultimately occurs; X will be equal to different symbols with different probabilities. These symbols might be characters, words, or even sequences of words and characters (like the bigrams and trigrams from before). Let p_i denote the probability that X equals Symbol i—that Symbol i is the next symbol in our message. Then, the entropy of the random variable X, our source of symbols, is given by

$$H[X] = -\sum_i p_i \log_2 p_i,$$

the negative sum of each symbol's probability times the binary logarithm of its probability. That is, H is the *average* binary logarithm of the symbols' probabilities, just negated. But this is a statement about the entropy of a random variable, not the entropy of a specific message. Amazingly though, this is essentially all we need! In his paper, Shannon proves a result known as his "source coding theorem," which states that for any message M, the entropy formula above places a limit on how much we can compress M.[9] In particular,

compression, and many others.

[8] Shannon's 1948 paper is actually the first written use of the term *bit*, "a word suggested by J. W. Tukey" (Shannon, *AMToC* 1). In his 1938 master's thesis, Shannon doesn't use the word "bit" to describe logic gate inputs and outputs but instead phrases everything in the language of electrical engineering; implicitly, though, he is referring to the same idea.

[9] Strictly speaking, the source coding theorem applies in the limit; certain small messages might not be able to be compressed fully according to the entropy of X, but this error term diminishes for larger messages.

the entropy of the random variable X is the average entropy of any symbol that we pick in our family—if we have a message M formed from these symbols, some symbols in M might be compressed slightly more, some slightly less, but the average compression across all the symbols in the family is exactly $H[X]$. And so, for any message we are interested in, Shannon's formula will effectively tell us its information content based on the family of messages it belongs to, making it a shockingly profound result in such a concise package.

There was one other question too: what is "the problem of communication"? Shannon tackles this question immediately, writing, "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point" (Shannon, *AMToC* 1). Specifically, Shannon says that a communication system consists of five parts: (1) an *information source*, (2) a *transmitter* of that information, (3) a *channel*, or the medium we send our messages through, (4) a *receiver* on the other end of the channel that reverses the work of the transmitter, and (5) a *destination*. Optionally, we can add a sixth element, a *noise source*, which pollutes and possibly removes information from the transmitted message along the channel. In Shannon's formulation, communication happens precisely when the destination can reproduce messages created at the information source.

With that, Shannon could give a name to a new field: information theory.

"Most mathematical theories take shape slowly," Gleick explains. "Shannon's information theory sprang forth like Athena, fully formed" (Gleick 233). Since Shannon published "A Mathematical Theory of Communication" in the somewhat esoteric *Bell System Technical Journal* and his paper was accessible to "a small clutch of communications engineers and mathematicians," the academy didn't immediately take notice (Soni and Goodman 165). But within a few months, "word of a breakthrough propagated through the community of communications engineers" (Soni and Goodman 166). Soon after, numerous communications engineers and newly-called "information theorists" wrote papers that built on Shannon's framework: he had once again sent ripples cascading through the engineering circles and even beyond. In 1948, Shannon met with Warren Weaver, the director of the Division of Natural Sciences at the Rockefeller Foundation, to discuss the implications of these ideas. The next year, Shannon and Weaver republished the paper in book form, but they changed the original title by a single word, giving it a whole new gravity: <u>The</u> *Mathematical Theory of Communication*.

The implication of Shannon's work really speaks for itself in the electronics and communication technologies we use today. According to Soni and Goodman,

> [Shannon] would live to see "information" turn from the name of a theory to the name of an era. "The Magna Carta of the Information Age," *Scientific American* would call Shannon's 1948 paper decades later. "Without Claude's work, the internet as we know it could not have been created," ran a typical piece of praise. And on and on: "A major contribution to civilization." "A universal clue to solving problems in different fields of science." "I reread it every year, with undiminished wonder. I'm sure I get an IQ boost every time." "I know of no greater work of genius in the annals of technological thought" (Soni and Goodman 165).

But ultimately, what really excited Shannon was that information theory gave him new ways to play. For the most part, the types of games that we've seen Shannon playing with are ones you can do with pencil and paper; but by applying information theory and circuit design, Shannon was able to make games and toys with profound ideas lurking under their surface.

Stories of Shannon's elaborate mechanical contraptions are legion: "He was the rare scientific genius who was just as content rigging up a juggling robot or a flamethrowing trumpet as he was pioneering digital circuits" (Soni and Goodman xv). In the 1950s, Shannon grew increasingly involved with the *cybernetics* movement, a newly emerging field based on the power of feedback systems in shaping intelligence, complex natural behavior, etc. During one cybernetics meeting, Shannon brought a small contraption he had built, an electronic rat he had named "Theseus." Theseus could navigate a 5×5 maze, searching for a goal: inside its "mind," it stored 50 bits of information, two bits per square of the board, indicating the direction it left that square in. The behavior of Theseus was surprisingly human-like: it could "learn" the structure of the maze so that it would be able to reach the goal starting from any point of the board (but would have to reset if the maze configuration was changed). Once, Theseus became stuck in a loop, and the neurophysiologist Ralph Gerard exclaimed, "A neurosis!" In response, Shannon added an "antineurotic circuit" to prevent any loops from occurring (Gleick 250).

Theseus showed just how important it was to see the "inside" of the mind, challenging some of the contemporary beliefs of behavioral psychology that focused on measuring external behavior. Shannon's contraption played a unique role in shifting psychology from behaviorism to what we see today, a transition known as the "cognitive revolution."

Shannon had first applied pen-and-paper games to create circuits and define information, and now, he applied circuit design and information theory to create physical toys. The mechanical became the mathematical, and vice versa. As another example of his fondness for games and toys, Shannon was an avid chess player, and he would apply information theory to estimate the number of chess games and positions.[10] Not only that, he also built one of the first chess-playing machines.

More importantly, though, others were finding Shannon's ideas and adapting them for their own games. It is impossible to overstate how many fields Shannon's work has touched. Information entropy—now sometimes called "Shannon entropy"—underpins the modern understanding of genetic variation, algorithmic randomness, and the nature of thermodynamic entropy. Nowadays, information theory grows almost as fast as the Internet it helped create.

Even as Shannon grew older and suffered from the effects of Alzheimer's disease, he still retained his playful spirit. He loved to unicycle and to juggle—having actually written an unpublished paper on the mathematics of juggling—and preserved these passions in his later years.

Zurek gathered all the responses to his prompt—"the specter of information is haunting sciences"—into a collection titled *Complexity, Entropy, and the Physics of Information.* The book begins with a "manifesto":

> The specter of information is haunting sciences. Thermodynamics, much of the foundation of statistical mechanics, the quantum theory of measurement, the physics of computation, and many of the issues of the theory of dynamical systems, molecular biology, and genetics and computer science share information as a common theme (Zurek vii).

More informally, information theory is now a very serious thing—and it actually always was. According to Soni and Goodman, Shannon's "style of work was characterized by such

[10] He estimated that there are about 10^{120} chess games, now known as *Shannon's number*, and on the order of 10^{43} possible chess positions.

lightness and levity, in fact, that we can sometimes forget the depth and difficulty of the problems he took on" (Soni and Goodman 277). But for Shannon, the problems of circuit design and information theory were inspired by a unique kind of game and led to a unique kind of game. Soni and Goodman say it best:

> Maybe it is too much to presume that the character of an age bears some stamp of the character of its founders; but it would be pleasant to think that so much of what is essential to ours was conceived in the spirit of play (Soni and Goodman xv).

Gleick, James. *The Information: A History, a Theory, a Flood.* First edition, Pantheon Books, 2011.

Shannon, Claude E. "A Mathematical Theory of Communication" (reprint with corrections). *Bell System Technical Journal*, vol. 27, no. 4, Oct. 1948, pp. 1–55 (original pp. 379–423, 623–656). *Harvard Mathematics Department*, people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf.

Shannon, Claude E. "A Symbolic Analysis of Relay and Switching Circuits." *Transactions of the American Institute of Electrical Engineers*, vol. 57, no. 12, Dec. 1938, pp. 713–723, doi.org/10.1109/T-AIEE.1938.5057767.

Shannon, Claude E. "A Theorem on Coloring the Lines of a Network." *Journal of Mathematics and Physics*, vol. 28, no. 1–4, Apr. 1949, pp. 148–152, doi.org/10.1002/sapm1949281148.

Soni, Jimmy, and Rob Goodman. *A Mind at Play: How Claude Shannon Invented the Information Age.* First trade paperback edition, Simon & Schuster, 2018.

Zurek, Wojciech H., editor. *Complexity, Entropy, and the Physics of Information.* First edition, *Santa Fe Institute Studies in the Sciences of Complexity*, vol. VIII, Addison-Wesley Publishing Company, 1990.