Font Size: A A A

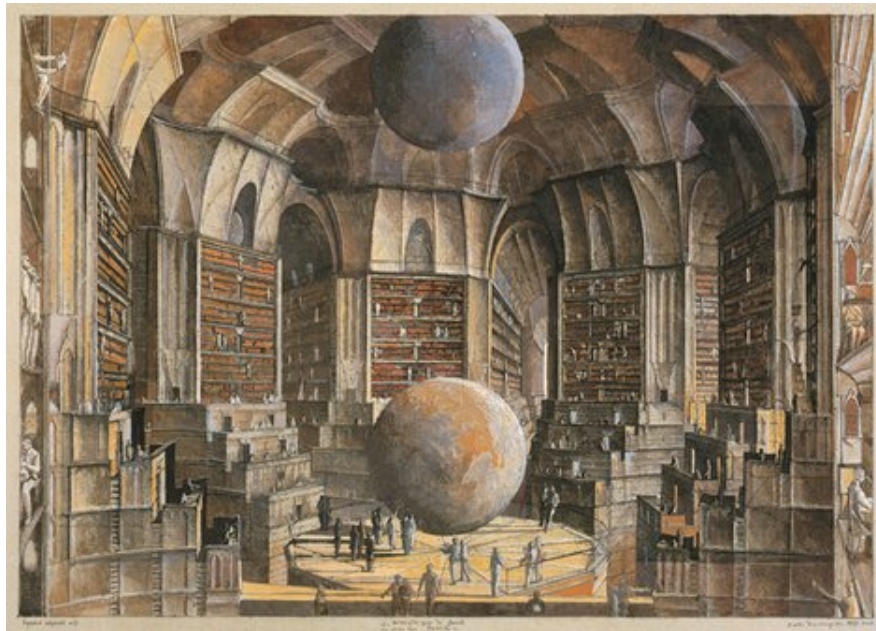# The New York Review of Books

How We Know

MARCH 10, 2011

Freeman Dyson

The Information: A History, a Theory, a Flood
by James Gleick
Pantheon, 526 pp., $29.95

Fitch-Febvrel Gallery

Érik Desmazières: La Salle des planètes, *from his series of illustrations for Jorge Luis Borges's story 'The Library of Babel,' 1997–2001. A new volume of Desmazières's catalogue raisonné will be published by the Fitch-Febvrel Gallery later this year. Illustration © 2011 Artists Rights Society (ARS), New York/ADAGP, Paris.*

James Gleick's first chapter has the title "Drums That Talk." It explains the concept of information by looking at a simple example. The example is a drum language used in a part of the Democratic Republic of Congo where the human language is Kele. European explorers had been aware for a long time that the irregular rhythms of African drums were carrying mysterious messages through the jungle. Explorers would arrive at villages where no European had been before and find that the village elders were already prepared to meet them.

Sadly, the drum language was only understood and recorded by a single European before

it started to disappear. The European was John Carrington, an English missionary who spent his life in Africa and became fluent in both Kele and drum language. He arrived in Africa in 1938 and published his findings in 1949 in a book, *The Talking Drums of Africa*.[1] Before the arrival of the Europeans with their roads and radios, the Kele-speaking Africans had used the drum language for rapid communication from village to village in the rain forest. Every village had an expert drummer and every villager could understand what the drums were saying. By the time Carrington wrote his book, the use of drum language was already fading and schoolchildren were no longer learning it. In the sixty years since then, telephones made drum language obsolete and completed the process of extinction.

Carrington understood how the structure of the Kele language made drum language possible. Kele is a tonal language with two sharply distinct tones. Each syllable is either low or high. The drum language is spoken by a pair of drums with the same two tones. Each Kele word is spoken by the drums as a sequence of low and high beats. In passing from human Kele to drum language, all the information contained in vowels and consonants is lost. In a European language, the consonants and vowels contain all the information, and if this information were dropped there would be nothing left. But in a tonal language like Kele, some information is carried in the tones and survives the transition from human speaker to drums. The fraction of information that survives in a drum word is small, and the words spoken by the drums are correspondingly ambiguous. A single sequence of tones may have hundreds of meanings depending on the missing vowels and consonants. The drum language must resolve the ambiguity of the individual words by adding more words. When enough redundant words are added, the meaning of the message becomes unique.

In 1954 a visitor from the United States came to Carrington's mission school. Carrington was taking a walk in the forest and his wife wished to call him home for lunch. She sent him a message in drum language and explained it to the visitor. To be intelligible to Carrington, the message needed to be expressed with redundant and repeated phrases: "White man spirit in forest come come to house of shingles high up above of white man spirit in forest. Woman with yam awaits. Come come." Carrington heard the message and came home. On the average, about eight words of drum language were needed to transmit one word of human language unambiguously. Western mathematicians would say that about one eighth of the information in the human Kele language belongs to the tones that are transmitted by the drum language. The redundancy of the drum language phrases compensates for the loss of the information in vowels and consonants. The African drummers knew nothing of Western mathematics, but they found the right level of redundancy for their drum language by trial and error. Carrington's wife had learned

the language from the drummers and knew how to use it.

The story of the drum language illustrates the central dogma of information theory. The central dogma says, "Meaning is irrelevant." Information is independent of the meaning that it expresses, and of the language used to express it. Information is an abstract concept, which can be embodied equally well in human speech or in writing or in drumbeats. All that is needed to transfer information from one language to another is a coding system. A coding system may be simple or complicated. If the code is simple, as it is for the drum language with its two tones, a given amount of information requires a longer message. If the code is complicated, as it is for spoken language, the same amount of information can be conveyed in a shorter message.

Another example illustrating the central dogma is the French optical telegraph. Until the year 1793, the fifth year of the French Revolution, the African drummers were ahead of Europeans in their ability to transmit information rapidly over long distances. In 1793, Claude Chappe, a patriotic citizen of France, wishing to strengthen the defense of the revolutionary government against domestic and foreign enemies, invented a device that he called the telegraph. The telegraph was an optical communication system with stations consisting of large movable pointers mounted on the tops of sixty-foot towers. Each station was manned by an operator who could read a message transmitted by a neighboring station and transmit the same message to the next station in the transmission line.

The distance between neighbors was about seven miles. Along the transmission lines, optical messages in France could travel faster than drum messages in Africa. When Napoleon took charge of the French Republic in 1799, he ordered the completion of the optical telegraph system to link all the major cities of France from Calais and Paris to Toulon and onward to Milan. The telegraph became, as Claude Chappe had intended, an important instrument of national power. Napoleon made sure that it was not available to private users.

Unlike the drum language, which was based on spoken language, the optical telegraph was based on written French. Chappe invented an elaborate coding system to translate written messages into optical signals. Chappe had the opposite problem from the drummers. The drummers had a fast transmission system with ambiguous messages. They needed to slow down the transmission to make the messages unambiguous. Chappe had a painfully slow transmission system with redundant messages. The French language, like most alphabetic languages, is highly redundant, using many more letters than are needed to convey the meaning of a message. Chappe's coding system allowed messages to be transmitted faster. Many common phrases and proper names were

encoded by only two optical symbols, with a substantial gain in speed of transmission. The composer and the reader of the message had code books listing the message codes for eight thousand phrases and names. For Napoleon it was an advantage to have a code that was effectively cryptographic, keeping the content of the messages secret from citizens along the route.

After these two historical examples of rapid communication in Africa and France, the rest of Gleick's book is about the modern development of information technology. The modern history is dominated by two Americans, Samuel Morse and Claude Shannon. Samuel Morse was the inventor of Morse Code. He was also one of the pioneers who built a telegraph system using electricity conducted through wires instead of optical pointers deployed on towers. Morse launched his electric telegraph in 1838 and perfected the code in 1844. His code used short and long pulses of electric current to represent letters of the alphabet.

Morse was ideologically at the opposite pole from Chappe. He was not interested in secrecy or in creating an instrument of government power. The Morse system was designed to be a profit-making enterprise, fast and cheap and available to everybody. At the beginning the price of a message was a quarter of a cent per letter. The most important users of the system were newspaper correspondents spreading news of local events to readers all over the world. Morse Code was simple enough that anyone could learn it. The system provided no secrecy to the users. If users wanted secrecy, they could invent their own secret codes and encipher their messages themselves. The price of a message in cipher was higher than the price of a message in plain text, because the telegraph operators could transcribe plain text faster. It was much easier to correct errors in plain text than in cipher.

Claude Shannon was the founding father of information theory. For a hundred years after the electric telegraph, other communication systems such as the telephone, radio, and television were invented and developed by engineers without any need for higher mathematics. Then Shannon supplied the theory to understand all of these systems together, defining information as an abstract quantity inherent in a telephone message or a television picture. Shannon brought higher mathematics into the game.

When Shannon was a boy growing up on a farm in Michigan, he built a homemade telegraph system using Morse Code. Messages were transmitted to friends on neighboring farms, using the barbed wire of their fences to conduct electric signals. When World War II began, Shannon became one of the pioneers of scientific cryptography, working on the high-level cryptographic telephone system that allowed Roosevelt and Churchill to talk to each other over a secure channel. Shannon's friend

Alan Turing was also working as a cryptographer at the same time, in the famous British Enigma project that successfully deciphered German military codes. The two pioneers met frequently when Turing visited New York in 1943, but they belonged to separate secret worlds and could not exchange ideas about cryptography.

In 1945 Shannon wrote a paper, "A Mathematical Theory of Cryptography," which was stamped SECRET and never saw the light of day. He published in 1948 an expurgated version of the 1945 paper with the title "A Mathematical Theory of Communication." The 1948 version appeared in the *Bell System Technical Journal*, the house journal of the Bell Telephone Laboratories, and became an instant classic. It is the founding document for the modern science of information. After Shannon, the technology of information raced ahead, with electronic computers, digital cameras, the Internet, and the World Wide Web.

According to Gleick, the impact of information on human affairs came in three installments: first the history, the thousands of years during which people created and exchanged information without the concept of measuring it; second the theory, first formulated by Shannon; third the flood, in which we now live. The flood began quietly. The event that made the flood plainly visible occurred in 1965, when Gordon Moore stated Moore's Law. Moore was an electrical engineer, founder of the Intel Corporation, a company that manufactured components for computers and other electronic gadgets. His law said that the price of electronic components would decrease and their numbers would increase by a factor of two every eighteen months. This implied that the price would decrease and the numbers would increase by a factor of a hundred every decade. Moore's prediction of continued growth has turned out to be astonishingly accurate during the forty-five years since he announced it. In these four and a half decades, the price has decreased and the numbers have increased by a factor of a billion, nine powers of ten. Nine powers of ten are enough to turn a trickle into a flood.

Gordon Moore was in the hardware business, making hardware components for electronic machines, and he stated his law as a law of growth for hardware. But the law applies also to the information that the hardware is designed to embody. The purpose of the hardware is to store and process information. The storage of information is called memory, and the processing of information is called computing. The consequence of Moore's Law for information is that the price of memory and computing decreases and the available amount of memory and computing increases by a factor of a hundred every
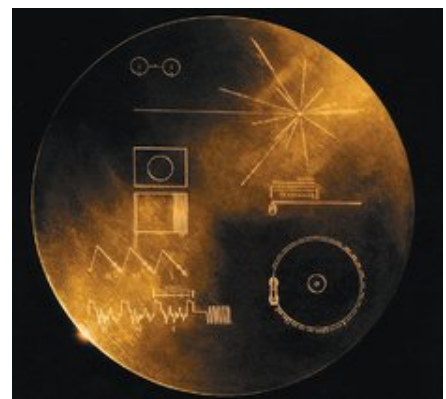
NASA/JPL-Caltech
The cover of the 'Golden Record,' stowed aboard the Voyager *spacecraft and sent into space in 1977.*

decade. The flood of hardware becomes a flood of information.

In 1949, one year after Shannon published the rules of information theory, he drew up a table of the various stores of memory that then existed. The biggest memory in his table was the US Library of Congress, which he estimated to contain one hundred trillion bits of information. That was at the time a fair guess at the sum total of recorded human knowledge. Today a memory disc drive storing that amount of information weighs a few pounds and can be bought for about a thousand dollars. Information, otherwise known as data, pours into memories of that size or larger, in government and business offices and scientific laboratories all over the world. Gleick quotes the computer scientist Jaron Lanier describing the effect of the flood: "It's as if you kneel to plant the seed of a tree and it grows so fast that it swallows your whole town before you can even rise to your feet."

On December 8, 2010, Gleick published on the The *New York Review's* blog an illuminating essay, "The Information Palace." It was written too late to be included in his book. It describes the historical changes of meaning of the word "information," as recorded in the latest quarterly online revision of the *Oxford English Dictionary*. The word first appears in 1386 a parliamentary report with the meaning "denunciation." The history ends with the modern usage, "information fatigue," defined as "apathy, indifference or mental exhaustion arising from exposure to too much information."

The consequences of the information flood are not all bad. One of the creative enterprises made possible by the flood is Wikipedia, started ten years ago by Jimmy Wales. Among my friends and acquaintances, everybody distrusts Wikipedia and everybody uses it. Distrust and productive use are not incompatible. Wikipedia is the ultimate open source repository of information. Everyone is free to read it and everyone is free to write it. It contains articles in 262 languages written by several million authors. The information that it contains is totally unreliable and surprisingly accurate. It is often unreliable because many of the authors are ignorant or careless. It is often accurate because the articles are edited and corrected by readers who are better informed than the authors.

Jimmy Wales hoped when he started Wikipedia that the combination of enthusiastic volunteer writers with open source information technology would cause a revolution in human access to knowledge. The rate of growth of Wikipedia exceeded his wildest dreams. Within ten years it has become the biggest storehouse of information on the planet and the noisiest battleground of conflicting opinions. It illustrates Shannon's law of reliable communication. Shannon's law says that accurate transmission of information is possible in a communication system with a high level of noise. Even in the noisiest

system, errors can be reliably corrected and accurate information transmitted, provided that the transmission is sufficiently redundant. That is, in a nutshell, how Wikipedia works.

The information flood has also brought enormous benefits to science. The public has a distorted view of science, because children are taught in school that science is a collection of firmly established truths. In fact, science is not a collection of truths. It is a continuing exploration of mysteries. Wherever we go exploring in the world around us, we find mysteries. Our planet is covered by continents and oceans whose origin we cannot explain. Our atmosphere is constantly stirred by poorly understood disturbances that we call weather and climate. The visible matter in the universe is outweighed by a much larger quantity of dark invisible matter that we do not understand at all. The origin of life is a total mystery, and so is the existence of human consciousness. We have no clear idea how the electrical discharges occurring in nerve cells in our brains are connected with our feelings and desires and actions.

Even physics, the most exact and most firmly established branch of science, is still full of mysteries. We do not know how much of Shannon's theory of information will remain valid when quantum devices replace classical electric circuits as the carriers of information. Quantum devices may be made of single atoms or microscopic magnetic circuits. All that we know for sure is that they can theoretically do certain jobs that are beyond the reach of classical devices. Quantum computing is still an unexplored mystery on the frontier of information theory. Science is the sum total of a great multitude of mysteries. It is an unending argument between a great multitude of voices. It resembles Wikipedia much more than it resembles the *Encyclopaedia Britannica*.

The rapid growth of the flood of information in the last ten years made Wikipedia possible, and the same flood made twenty-first-century science possible. Twenty-first-century science is dominated by huge stores of information that we call databases. The information flood has made it easy and cheap to build databases. One example of a twenty-first-century database is the collection of genome sequences of living creatures belonging to various species from microbes to humans. Each genome contains the complete genetic information that shaped the creature to which it belongs. The genome data-base is rapidly growing and is available for scientists all over the world to explore. Its origin can be traced to the year 1939, when Shannon wrote his Ph.D. thesis with the title "An Algebra for Theoretical Genetics."

Shannon was then a graduate student in the mathematics department at MIT. He was only dimly aware of the possible physical embodiment of genetic information. The true physical embodiment of the genome is the double helix structure of DNA molecules,

discovered by Francis Crick and James Watson fourteen years later. In 1939 Shannon understood that the basis of genetics must be information, and that the information must be coded in some abstract algebra independent of its physical embodiment. Without any knowledge of the double helix, he could not hope to guess the detailed structure of the genetic code. He could only imagine that in some distant future the genetic information would be decoded and collected in a giant database that would define the total diversity of living creatures. It took only sixty years for his dream to come true.

In the twentieth century, genomes of humans and other species were laboriously decoded and translated into sequences of letters in computer memories. The decoding and translation became cheaper and faster as time went on, the price decreasing and the speed increasing according to Moore's Law. The first human genome took fifteen years to decode and cost about a billion dollars. Now a human genome can be decoded in a few weeks and costs a few thousand dollars. Around the year 2000, a turning point was reached, when it became cheaper to produce genetic information than to understand it. Now we can pass a piece of human DNA through a machine and rapidly read out the genetic information, but we cannot read out the meaning of the information. We shall not fully understand the information until we understand in detail the processes of embryonic development that the DNA orchestrated to make us what we are.

A similar turning point was reached about the same time in the science of astronomy. Telescopes and spacecraft have evolved slowly, but cameras and optical data processors have evolved fast. Modern sky-survey projects collect data from huge areas of sky and produce databases with accurate information about billions of objects. Astronomers without access to large instruments can make discoveries by mining the databases instead of observing the sky. Big databases have caused similar revolutions in other sciences such as biochemistry and ecology.

The explosive growth of information in our human society is a part of the slower growth of ordered structures in the evolution of life as a whole. Life has for billions of years been evolving with organisms and ecosystems embodying increasing amounts of information. The evolution of life is a part of the evolution of the universe, which also evolves with increasing amounts of information embodied in ordered structures, galaxies and stars and planetary systems. In the living and in the nonliving world, we see a growth of order, starting from the featureless and uniform gas of the early universe and producing the magnificent diversity of weird objects that we see in the sky and in the rain forest. Everywhere around us, wherever we look, we see evidence of increasing order and increasing information. The technology arising from Shannon's discoveries is only a local acceleration of the natural growth of information.

The visible growth of ordered structures in the universe seemed paradoxical to nineteenth-century scientists and philosophers, who believed in a dismal doctrine called the heat death. Lord Kelvin, one of the leading physicists of that time, promoted the heat death dogma, predicting that the flow of heat from warmer to cooler objects will result in a decrease of temperature differences everywhere, until all temperatures ultimately become equal. Life needs temperature differences, to avoid being stifled by its waste heat. So life will disappear.

This dismal view of the future was in startling contrast to the ebullient growth of life that we see around us. Thanks to the discoveries of astronomers in the twentieth century, we now know that the heat death is a myth. The heat death can never happen, and there is no paradox. The best popular account of the disappearance of the paradox is a chapter, "How Order Was Born of Chaos," in the book *Creation of the Universe*, by Fang Lizhi and his wife Li Shuxian. [2] Fang Lizhi is doubly famous as a leading Chinese astronomer and a leading political dissident. He is now pursuing his double career at the University of Arizona.

The belief in a heat death was based on an idea that I call the cooking rule. The cooking rule says that a piece of steak gets warmer when we put it on a hot grill. More generally, the rule says that any object gets warmer when it gains energy, and gets cooler when it loses energy. Humans have been cooking steaks for thousands of years, and nobody ever saw a steak get colder while cooking on a fire. The cooking rule is true for objects small enough for us to handle. If the cooking rule is always true, then Lord Kelvin's argument for the heat death is correct.

We now know that the cooking rule is not true for objects of astronomical size, for which gravitation is the dominant form of energy. The sun is a familiar example. As the sun loses energy by radiation, it becomes hotter and not cooler. Since the sun is made of compressible gas squeezed by its own gravitation, loss of energy causes it to become smaller and denser, and the compression causes it to become hotter. For almost all astronomical objects, gravitation dominates, and they have the same unexpected behavior. Gravitation reverses the usual relation between energy and temperature. In the domain of astronomy, when heat flows from hotter to cooler objects, the hot objects get hotter and the cool objects get cooler. As a result, temperature differences in the astronomical universe tend to increase rather than decrease as time goes on. There is no final state of uniform temperature, and there is no heat death. Gravitation gives us a universe hospitable to life. Information and order can continue to grow for billions of years in the future, as they have evidently grown in the past.

The vision of the future as an infinite playground, with an unending sequence of

mysteries to be understood by an unending sequence of players exploring an unending   c
supply of information, is a glorious vision for scientists. Scientists find the vision
attractive, since it gives them a purpose for their existence and an unending supply of
jobs. The vision is less attractive to artists and writers and ordinary people. Ordinary
people are more interested in friends and family than in science. Ordinary people may not
welcome a future spent swimming in an unending flood of information. A darker view of
the information-dominated universe was described in a famous story, "The Library of
Babel," by Jorge Luis Borges in 1941.[3] Borges imagined his library, with an infinite
array of books and shelves and mirrors, as a metaphor for the universe.

Gleick's book has an epilogue entitled "The Return of Meaning," expressing the concerns
of people who feel alienated from the prevailing scientific culture. The enormous success
of information theory came from Shannon's decision to separate information from
meaning. His central dogma, "Meaning is irrelevant," declared that information could be
handled with greater freedom if it was treated as a mathematical abstraction independent
of meaning. The consequence of this freedom is the flood of information in which we are
drowning. The immense size of modern databases gives us a feeling of meaninglessness.
Information in such quantities reminds us of Borges's library extending infinitely in all
directions. It is our task as humans to bring meaning back into this wasteland. As finite
creatures who think and feel, we can create islands of meaning in the sea of information.
Gleick ends his book with Borges's image of the human condition:

> We walk the corridors, searching the shelves and rearranging them, looking for lines
> of meaning amid leagues of cacophony and incoherence, reading the history of the
> past and of the future, collecting our thoughts and collecting the thoughts of others,
> and every so often glimpsing mirrors, in which we may recognize creatures of the
> information.

---

1. 1
   London: Carey Ringsgate, 1949. ↵

2. 2
   Singapore: World Scientific Publishing Co., 1989. ↵

3. 3
   Jorge Luis Borges, *Labyrinths: Selected Stories and Other Writings* (New Directions,
   1962), p. 54. ↵