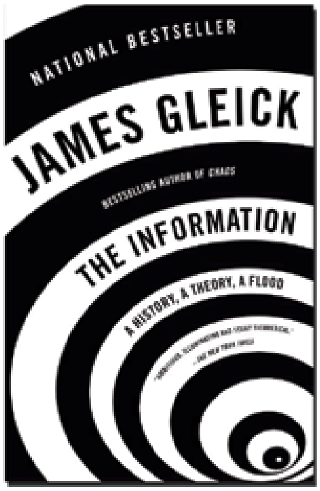
A Natural History of Information

A review of 'The Information: A History, a Theory, a Flood' by James Gleick

review

KESHAV MUKUNDA



In 1948, a paper was submitted to a Bell Labs technical journal, in which the author proposed a new theoretical framework to analyse problems in communication. Suggesting that it be rejected for publication, the reviewer of the paper said it was "poorly motivated and excessively abstract", and went on to say, "it is unclear for what practical problem it might be relevant ... the author mentions computing machines – I guess one could connect such machines, but a recent IBM memo stated that a dozen or so such machines will be sufficient for all the computing that we'll ever need in the foreseeable future, so there won't be a whole lot of connecting going on" [1]. While this clearly mistaken reviewer remains anonymous, the author of the submitted paper entitled "A Mathematical Theory of Communication", a relatively young mathematician and engineer named Claude Shannon, went on to become one of the founding pioneers of the new field called *information theory*. In fact, Shannon's 1948 paper – published around the same time that the transistor was invented – essentially created the field of information theory, which has deeply influenced the development of engineering and computer science. Among

Keywords: *information, bit, binary, digit, communication*

numerous other ideas, Shannon is credited with conceptualizing the digital computer and circuit design theory, as well as being the first to use the word *bit* as a contraction of the phrase 'binary digit', in the same 1948 paper.

The story of Shannon's work, as well as a wider look at the evolution of the idea of information, appears in a recent book called *The Information: A History, a Theory, a Flood.* The book is by the well-known science writer James Gleick, who has written several bestsellers. Published in 1987, his first book *Chaos: Making a New Science* described the development of chaos theory, the mathematical study of sensitive dynamical systems, and was important in helping popularize chaos theory and fractals. Among other things, the book spread greater awareness of the phrase "the butterfly effect", making it a cultural meme that has since appeared in movies and pop culture generally. *Chaos* won a Pulitzer Prize in 1988, and has sold millions of copies since then. Among Gleick's other widely acclaimed books are two biographies, *Genius: The Life and Science of Richard Feynman,* and *Isaac Newton,* both finalists for the Pulitzer Prize as well.

Unlike a more straightforward historical biography of a person, *The Information* describes the evolution of an idea, bringing together strands of history and culture to show how a crucial new construct emerged in our understanding of the world. This new construct was defined, measured and articulated as 'information'. It now appears to be everywhere we look, and this ubiquity is perhaps the reason why the idea of information was traditionally overlooked. After all, anything that is considered self-evident and 'obvious' usually hides a deeper and richer understanding of the world; as the mathematician E. T. Bell put it, *"obvious* is the most dangerous word in mathematics" [2].

On the face of it, you might wonder what a book describing the history of the concept of information has to do with mathematics. In fact, the theoretical foundations that help to describe and measure information are mathematical in nature. When Shannon thought about transmitting information, he did not consider the meaning or sense of the message; as he put it, "semantic aspects of communication are irrelevant" to the engineering problem, although he conceded that "frequently the messages have *meaning*" [3]. Instead, he formulated information as a mathematical measure of the number of possible states that a message could take, using symbols from a finite underlying alphabet. In this sense information is directly related to the idea of entropy, and can be measured in a similar way. Many of the other concepts described in this book also have a solid mathematical basis, and the whole history of the conceptualization of information shows us the ways in which mathematics can be applied to all sorts of questions about the world. Gleick is at ease writing and describing these mathematical ideas, and as a textured background to these ideas he provides a narrative that takes in a wide sweep of history, linking the technical to the cultural.

The book begins with the description of the talking drums in sub-Saharan Africa that are used to transmit complex messages across long distances, being sent from village to village by relay. The language of the drum involves sound combinations that have a few different dimensions – tones as well as vowels and consonants - which are used to encode detailed messages. This sets the stage for the investigation of the 'amount' of a message that can be reliably transmitted using relatively simple alphabets; in fact, the example shows that the 'size' of a message is not always directly correlated with the amount of information it carries. Gleick then moves on to investigate the early attempts at creating long distance telegraph systems, and then telephones, bringing out curious stories of the many people and inventions that flourished in the early years of each invention. There are the Chappe brothers in 18th century France, for example, who devised an early form of telegraph using a network of tall towers with men communicating between them using flags. Even the famous mathematician C.F. Gauss, together with the physicist Weber, came up with a scheme involving electric currents that travelled through wires to deflect small needles left or right. We clearly see two separate and interconnected problems emerging: creating a useful language or alphabet of the message, and inventing effective communication technology itself.

Gleick then discusses how language, when moving from the oral to the written, brought with it questions of representation and standardization. How exactly were the letters in the alphabet to be arranged to spell a given word: for example, would it be *wordes* or *words, colume* or *column*? This leads to the story of how the Oxford English Dictionary was originally compiled, and to its subsequent development. Moving from the written to the published word to transmitting them through wires, the book then considers how language was encoded in electronic switches, bringing logic and language together in the early computers. Again, through this journey we meet several people involved in these ideas through history: Charles Babbage, Ada Byron, and their early mechanical computer called the 'analytical engine'; Augustus De Morgan, George Boole and their symbolic 'algebra of logic'; and of course Alan Turing, with his formalization of the meanings of ideas such as 'algorithm' or 'computation'.

From here the journey to Shannon's 1948 paper was not self-evident. Gleick describes Shannon's early ideas, and the people he worked with. Some were more responsive to his ideas than others. For example, Turing and Shannon often had lunch together and discussed their work, and in an interview given in 1982, Shannon said that Turing "didn't always believe these . . . my ideas . . . he didn't believe they were in the right direction" [4]. Once the idea of information emerged, it spread quickly to various disciplines to different levels of success. In the 1950s, as the structure of DNA was discovered, the biologist Francis Crick described the copying of a sequence of nucleic acids as a transfer of information. At the time, this was meant as a metaphoric description, but soon biologists and geneticists would talk of information, alphabets, and the transcription of codes in a literal sense. Information theory permeated economics, philosophy and physics, while it also remained significant and useful in the growing computer industry.

Gleick eventually argues that the idea of information is more universally fundamental than we might think. In fact, some theoretical physicists now suggest that space and time are themselves simply constructed by the exchange of discrete bits of information. In this view, information is the essence out of which everything else in the physical universe arises; or, as the physicist John Archibald Wheeler put it, "all things physical are information-theoretic in origin" [5].

The book doesn't move linearly through history, but instead weaves between different times and different discoveries to tease out the threads of the various insights that led to the concept of information. In hindsight, it might seem obvious to us now that the idea of information would emerge in certain historical contexts, and we can now easily see and name these ideas in those contexts; but it would have taken a great leap of understanding at the time to see how all the pieces fit together. By giving us this non-linear narrative, Gleick delightfully shows not just that our human scientific understanding of the world meanders in several different directions with no evident direction of 'progress', but also that each human idea does not constructively build on the ones that came before. We almost get the sense that there are several plot lines evolving in what is a large detective story, and Gleick brings them together in a satisfying way.

In all of the discussion on information, however, Gleick sidesteps issues of the control of access to information, steering clear of any political analysis or discussion of how information and state power are closely related. Any history of information would surely have to acknowledge these relations, and it would have been interesting if the book considered this. Still, he does mention Wikipedia and the ways in which entries can be silenced or censored by vandals and deletionists, as the community of editors struggle to reach an elusive compromise. In fact, this struggle is not just between opposing points of view, but also with our management of the sheer quantity of information available to us now. To Gleick, the bold Wikipedia project is one attempt to deal constructively with the new flood of information we are continually exposed to. One factor in this flood is the curious fact that it apparently takes more energy to actually delete electronic information than to simply store it, in an entropic sense. Why delete or forget anything, in that case? However, as we electronically preserve more information now than ever before, this 'information overload' makes it difficult for us to decide on the value of any given piece of knowledge. The more information we have access to, the harder it is to filter out irrelevant noise to find what we want, and then understand what it means. The challenge of making sense of it all is more relevant than ever, and Gleick is optimistic about our collective ability to manage the challenges and even to create meaning in what could become a bewildering jumble.

Although many of the concepts in the book can be quite complicated, Gleick gives us a very readable account, going into details only as much as is necessary for us to get a sense of the mathematics and engineering involved. Even these are not presented in an abstract way, but are woven into the historical account and help to move the narrative forward. High-school students might find some of the book challenging, but it will probably help them to see the world in a new way, making connections that they had not known before. It can show them mathematics in a new light, being applied to very practical problems at the centre of modern communications and technology. Teachers of mathematics would hopefully find the book fascinating and would be able to appreciate it at a deeper level. There are a great many mathematical ideas that you might be surprised to find in a book about information: randomness and normality of numbers, Gödel and incompleteness, quantum mechanics and uncertainty – but the links that Gleick fashions between them is intellectually satisfying. In addition, a reader who has already heard a little bit about some of the people in this book – Babbage, Morse, Turing, and others – would find that this adds to the value of the unusual perspective that Gleick brings. *The Information*, perhaps the first natural history of information ever attempted, lays out for us the long course we've followed to get to where we are today.

References

[1] Santini, S., "We Are Sorry to Inform You...," Computer, v.38, no.12, Dec 2005.

[2] Bell, E. T. (1951). Mathematics, queen and servant of science. New York: McGraw-Hill.

[3] Shannon, Claude E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27(3), 379-423.

[4] Shannon, Claude E., an oral history conducted in 1982 by Robert Price. IEEE History Center, New Brunswick, NJ, USA.

[5] Wheeler, John Archibald. (1990). "Information, physics, quantum: The search for links" in W. Zurek (ed.) Complexity, Entropy, and the Physics of Information. Addison-Wesley.

Keshav Mukunda, Simon Fraser University

Appendix: Bits and Bytes and Shannon's concept of entropy

Consider an alphabet of symbols, each of which may be used to send a message, or transmission. Each symbol *i* used in the transmission could be selected with probability p_i and this probability could depend both on the symbol selected and its location within the transmission. Shannon proposed that the amount of information carried by a transmission is given by

$$H = -K \sum_i p_i \log p_i.$$

Here *K* is a positive constant and the summation is across the symbols of the alphabet. Why did he decide to use this measure? To create a way to measure the amount of information a transmission contains, Shannon set out three reasonable conditions that any such measure would have to satisfy. (See reference [1].) He then proved mathematically that there is only one possible way to define the measure, namely the one shown above.

Choosing different bases for the logarithm naturally gives us different choices of units to measure information. In his 1948 paper, Shannon suggested that using a base of 2 would be convenient for electronic devices, and that the units in this case could be called *binary digits*, or simply *bits*. He noted that this name was suggested by the mathematician J. W. Tukey, a colleague at Bell Labs. The name certainly has stuck! Note that as an electronic switch has two stable positions, ON or OFF, it carries 1 bit of information.

How are the probabilities p_i for the symbols in the alphabet found? For human languages, their values can be empirically estimated. Shannon gave the following example: the English language can be thought to contain an alphabet of 27 symbols: the usual 26 letters, plus a space. In everyday written English communications, not every symbol is equally probable, and their successive choices are not independent either. If each symbol is selected randomly with probability $\frac{1}{27}$ and each choice is made independently, then a transmission might look like this:

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMK ZAACIBZLHJQD.

Instead, if we were to use the naturally occurring frequencies of the letters in the English language, and also select each letter with a probability that depends on the previous two letters (using the naturally occurring frequencies of the various three-letter combinations) then a transmission would look like this:

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

It's clear that the resemblance to a 'meaningful' English sentence has increased, though it is still gibberish!

(Some readers may be reminded of the following lines which occur in Lewis Caroll's poem *Jabberwocky* which is part of his 'nonsense verse' work:

'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe.

The lines seem to be telling us something, though the words do not belong to any English dictionary!)

At the time of Shannon's 1948 paper, the formula for the measure of information H was already well-known in the field of statistical mechanics. In this context, the formula describes the "entropy of the system". Roughly speaking, entropy is a measure of the 'level of disorder' in a thermodynamic system, a way of measuring how far away the system is from equilibrium. If a thermodynamic system can have several microstates, each occurring with a possibly different probability p_i then the entropy of the system is defined to be

$$s = -k_B \sum_i p_i \log p_i$$

Here k_B is called the Boltzmann constant, and the summation is across all the microstates. So Shannon had created a measure of information which is an extension of the thermodynamic concept of entropy. In this sense, information can be thought of as a form of entropy.

For interested readers, [3] is the original landmark paper where Shannon introduces these notions.

References

[1] http://en.wikipedia.org/wiki/Entropy_(information_theory)#Characterization

[2] http://en.wikipedia.org/wiki/A_Mathematical_Theory_of_Communication

[3] http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf