Claude Shannon and George Boole, Enablers of the Information Age

Robert Gallager
Prof. Emeritus, EECS, MIT

October 30, 2015

George Boole (1815-1864) developed Boolean logic

The principles of logical thinking have been understood (and occasionally used) since the Hellenic era.

Boole's contribution was to show how to systemize these principles and express them in equations (called Boolean logic or Boolean algebra).

Claude Shannon (1916-2001) showed how to use Boolean algebra as the basis for switching technology. This conribution systemized logical thinking for computer and communication systems, both for the design and programming of the systems and their applications.

Logic continues to be abused in politics, religion, and most non-scientific areas.

Claude Shannon also created information theory. This was a 'beautiful and fascinating theory' for many years, but eventually, almost when no one was looking, it became the conceptual architecture of virtually all commercial communication systems.

Would modern communication technology, computer technology, and their synthesis have developed as quickly and in the same way without Shannon and Boole?

Historians try to answer these questions, but I can not. There can be little question, however, that their contributions were extraordinary.

For us, as scientists and educators, it is more important to understand the characteristics that made these giants great.

Claude Shannon gave a talk entitled ‘Creative Thinking’ in 1952 to a small group of researchers. He started with 3 main attributes:

- Training and experience
- Intelligence
- Motivation (the inner drive to formulate questions and find answers; curiosity about fundamental characteristics; need to understand in multiple ways; satisfaction from understanding)

He then continued with a number of ‘tricks’ that he often found useful. These tricks appear to be the major principles of theoretical research.

Tricks for formulating and solving problems

1. Simplification: get rid of enough detail (including practical aspects) for intuitive understanding.
2. Similarity to a known problem (experience helps)
3. Reformulate (avoid getting in a rut)
4. Generalize (more than opposite of simplify)
5. Structural analysis (break problem into pieces)
6. Inversion (work back from desired result)

Other tricks that Shannon often used

1. Be interested in several interesting problems at all times. Work on the most interesting one.
2. Look for contradictions as well as proofs.
3. Study what is happening in multiple fields, but don't work on what many others are working on.
4. Ask conceptual questions about everyday things.
5. Don't write papers unless you really want to share something fascinating.
6. Don't assume your readers know everything you do. Spoon feeding is not a bad idea.

QUICK BIOGRAPHY OF SHANNON

Normal but bright nerd in high school (Gaylord, Mich).

Double degree (EE, Math) at U. Mich. at age 20.

Grad student at MIT with RA baby-sitting for Vannevar Bush's Differential Analyzer.

His SM thesis, "A Symbolic Analysis of Relay and Switching Circuits," gave practical importance to Boolean algebra; essential impact on digital technology.

This created a new field. It started as a simple elegant idea and became central to the new switching systems at AT&T.

His PhD Thesis was "An Algebra of Theoretical Genetics"

The results were important, but Shannon lost interest before publishing; the main results were rediscovered independently over the years.

Shannon never liked to write, and he became fascinated by telecommunication while finishing his PhD.

Claude worked on his mathematical theory of communication at Princeton's Advanced Study Institute in 1940–41.

During the war he worked on Fire Control at Bell Labs; he continued work on communication, and also on cryptography.

He established a mathematical basis for cryptography in 1945 based on his nascent communication theory.

By 1948, everything came together in his mathematical theory of communication.

Sources are characterized by the bit rate per symbol or per second needed to reproduce the source exactly or within a given distortion allowance.

Channels can be characterized by an essentially error free bit rate called capacity.

A standard binary interface between sources and channels loses essentially nothing. Think how central a binary interface is in the information age.

‘The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. 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. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.’

C. E. Shannon, 1948

We review the source representation briefly.

Claude looked at the example of English text and first modeled it as a stochastic process with independent identically distributed (IID) letters.

Why stochastic? It makes sense for the telecommunication system designer.

Why IID? It explains the basic idea behind compression; it builds the framework for 'better' models.

Let $p(i)$ be the probability of the letter i; the probability of an IID letter sequence $\mathbf{x}^{(\tau)} = x_1, \ldots, x_\tau$ is then

$$
\begin{aligned}
\mathsf{Pr}\{\mathbf{x}^{(\tau)}\} &= p(x_1)p(x_2)\cdots p(x_\tau) \\
\mathsf{Pr}\{Shannon\} &= p(S)p(h)p(a)p(n)p(n)p(o)p(n) \\
&= p^3(n)p(S)p(h)p(a)p(o)
\end{aligned}
$$

From the law of large numbers (LLN), typical sequences $\mathbf{x}^{(\tau)}$ with $\tau >> 1$ have about $\tau p(i)$ appearances of letter i for each i, and thus

$$
\begin{aligned}
\mathsf{Pr}\{\mathbf{x}^{(\tau)}\} &\approx \prod_i p(i)^{\,\tau p(i)} \\
&= 2^{\tau\left[\sum_i p(i) \log_2 p(i)\right]} = 2^{-\tau H} \qquad \text{where} \\
H &= \sum_i -p(i) \log_2 p(i)
\end{aligned}
$$

All typical sequences have about the same probability.

Cumulatively, their probability is $\approx$ 1.

There are about $2^{\tau H(P)}$ typical sequences. They can be represented by about τH bits.

Note that Shannon used the 'tricks' of simplification to IID, then similarity to LLN, then reformulation to look at typical sequences. Then structural analysis to separate finding entropy from actual source coding.

The above typical sequence argument extends naturally from IID sequences to ergodic Markov chains. Shannon explained this by looking at digrams and trigrams of letters and then of words.

This generalization starts to approximates natural language.

Shannon also devised a simple algorithm for encoding sequences into almost the minimum number of bits.

In 1952, Dave Huffman beat Shannon at his own game by reformulating Shannon's approach into a beautifully simple optimal source coding algorithm.

Jacob Ziv and Abe Lempel in 1978-9 extended source coding to 'universal source coding' where the sequence probabilities were simultaneously measured and used.

This turned source coding into something very practical, since real data sources usually have slowly changing statistics.

All of Shannon's work on source coding might have been done by a well-trained, bright, motivated graduate student by making very good guesses and by using Shannon's tricks.

His work was brilliantly simple and simply brilliant.

These same typical sequence arguments work for noisy channels.

Here Claude looked at jointly typical input/output sequences with an arbitrary simple input model.

The channel was modeled by stochastic outputs given inputs.

The trick here was a randomly chosen code of input sequences.

Shannon's genius lay in finding the "right way," the "simple way" to look at everyday technological problems.

Examples: communication systems, switching systems, crypto systems, chess playing machines, solving mazes, controlling unicycles, gambling strategies, etc.

He built mathematical (and physical) models to help understand these problems, but his focus was on the underlying problem (the architecture), not in mathematics per se nor in problem details.

Shannon was almost the opposite of an applied mathematician.

Applied mathematicians solve mathematical models formulated by others (perhaps with minor changes to suit the tools of their trade).

Shannon was a creator of models — his genius lay in determining the core of the problem, removing details that could be reinserted later.

QUICK BIOGRAPHY OF GEORGE BOOLE

Son of a cobbler who was more interested in mathematics and optics than cobbling; brought up in Lincoln, England (120 miles N. of London)

George was largely self-taught, first in religion and multiple languages, then mathematics. He supported himself (from age 16) by teaching in day schools and boarding schools.

He started a prolific career in writing mathematics papers, and won a Royal Medal from the Royal Society in 1944 for a paper on symbolic algebra. After the Royal Medal, his life was a sequence of successes.

His “Mathematical Analysis of Logic” came in 1847, followed by becoming Professor of Mathematics at University College, Cork, Irland, and finally “An Invesigation into the Laws of Thought” in 1854. He had no academic degrees.

Boole was respected in his time for many contributions, but he is remembered for the two papers on logic, now known as Boolean algebra.

The principles of logic have been known since Aristotle, but Boole succeeded in expressing logical propositions by equations.

This reduction to equations brings a clarity and simplicity to logic which is absent with the fuzziness of natural language.

SHANNON AND BOOLE SIMILARITIES

Recognized when very young (Shannon 22, Boole 29)

The magnum opus of each opened up a new field and required about 8 years

Each magnum opus was quite simple in retrospect.

Boole's research appears to indicate that he understood Shannon's 'tricks' of creative research.

Simplification is probably the most difficult to understand of Shannon's 'tricks.' There are many quotes on the internet about simplicity, but most of them seem to promote ignorance rather than what Shannon meant. The following get close to Shannon's 'trick.'

Steven Weinberg: "In the study of anything outside human affairs, including the study of complexity, it is only simplicity that can be interesting."

Einstein: "Everything should be as simple as possible, but no simpler."

Alfred N. Whitehead: "Search for simplicity, but mistrust it."

Whitehead's version is a home run, expressing the idea even better than Shannon.

Searching for simplicity is really searching for intuitive understanding of a simplified version of the problem that doesn't ignore the underlying original issues.

Mistrusting that simplicity means critical questioning of that intuitive understanding, hopefully leading to generalization or to alternative simplifications.

This search and mistrust leads to a process of successive probing, varying the simplification, generalizing, and reformulating, each step based on what has been learned before and leading to greater understanding.

Shannon was a grand master of this process. His writing often left out unsuccessful steps, but his explanation of source coding makes the process very clear.

We are not all grand masters, but Shannon's tricks can be used to advantage by all of us. In these days when we are all too busy to think, perhaps slowing down a little and giving understanding a chance would be fun.

Perhaps we even might teach our students about creative research instead of pushing them to program more and more complex problems. Computational resources let us solve incredibly complex problems, but do we learn anything from those solutions?

Graduate students often mistake simplicity for triviality. They stumble on some simple and elegant result, and immediately try to complicate it as much as possible.

The misconception is that it takes the best students to solve the most complex problems.

Actually, it takes the best students to *find* **the simplest open problems.**

Perhaps more familiarity with Shannon's 'tricks' would be helpful to them.