CLAUDE SHANNON

(1916-2001)

Shannon invented Information Theory

He created the architecture and concepts governing digital communication.

More than anyone, he created the foundations for the Information Age.

Reading his work is an intellectual delight.

Claude looked for problems with puzzling behavior.

He found these problems in technology and everyday life, not in journals.

He had a genius for finding appropriate models and simple examples to build up understanding.

He was solving generic puzzles rather than proving theorems.

Shannon was always toying with a number of different problems in his head at the same time.

He would focus on whichever seemed the most fascinating, amusing, or insightful.

When something gelled in his head, and when he was pushed, he would write a paper, often by dictation of a first and only draft.

Is this a style of research we should emulate?

Should universities encourage researchers with this style?

I contend that, with qualifications, the answers are yes and yes!

This is not the scholar style, nor the programmatic style, nor the techno-jock style.

QUICK BIOGRAPHY

His SM thesis, ”A Symbolic Analysis of Relay and Switching Circuits,” was his first major accomplishment.

This provided a mathematical basis for switching. It was simple, elegant, and important.

His PhD Thesis was a mathematical theory of genetics.

The results were important, but Shannon lost interest before publishing; the main results were discovered 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 40-41.

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 45 based on his nascent communication theory.

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

Sources are characterized by bit rate.

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

A standard binary interface between sources and channels loses essentially nothing.

We review the source representation briefly.

He looked at the example of English text and first modeled it as a stochastic process with 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 letter i and consider n successive source outputs.

For n large, typical sequences x have about nP_i appearances of letter i, and thus

$$\begin{aligned}\Pr\{x\} &\approx \prod_i P_i^{nP_i} \\ &= 2^{n\sum_i P_i \log_2 P_i} \\ &= 2^{-nH(P)} \\ H(P) &= \sum_i -P_i \log_2 P_i\end{aligned}$$

All typical sequences have about the same probability.

Cumulatively, their probability is $\approx$ 1.

There are about $2^{nH(P)}$ typical sequences.

They can be represented by $nH(P)$ bits.

Hidden assumption: typicality is based on LLN. Long delays necessary for LLN behavior.

Shannon's entire theory was based on the LLN regime.

Was this an oversight?

No, it was a stroke of genius.

The theory fit together this way, and all major results depend on it.

Later research has made extensions for finite delay, feedback, and non-ergodicity.

The above typical sequence argument extends naturally to Markov sources.

Again, there are about 2^{nH} typical sequences, each with probability 2^{-nH}.

Universal coding techniques were developed later.

They depend on typical sequence arguments, but work for ‘any’ statistics.

These same typical sequence arguments work for noisy channels.

The channel is modeled by conditional probabilities $\Pr(y|x)$**.**

$$C = \sup_{P(X)} H(X) - H(X|Y)$$

For memoryless channels, optimum is easy to find.

We look at typical input/output sequences with optimum input distribution.

The trick here is to select a subset of typical inputs randomly as a code.

If the input bit rate is less than capacity, error probability can be made arbitrarily small.

If the entropy of the input is larger than capacity, error probability is large.

These results say nothing about information. They relate to possibilities, probabilities, and communicatation.

It took 30 years for these results to become central to communication systems.

Data compression, coding, better modems, market, and digital technology were all needed.

Engineers also had to understand the architecture of Shannon's view.

Communication theory, as a mathematical theory, is a little like electromagnetism.

We know how to formulate the central problems mathematically and understand what is involved.

There are simple examples that provide insight.

There are complex examples that are understood in principle.

In the network field, for example, there is no central mathematical theory.

We apply mathematics to queueing, to routing, to congestion, but the overall structure is missing.

Change is very rapid, but progress is slow.

Is a more cohesive structure possible? No one knows, but we ought to try.

The pace of truly basic research is very slow.

Shannon thought about communication for 8 years before writing anything.

It took another 30 years before it became central.

Many think the pace of research is accelerating, but it is probably getting slower.

When technology and architecture are all in place except for one missing link, rapid progress is usually made.

People know what to focus on.

Product cycle ‘research’ works fine.

Often, however, many links are missing.

Technology then stumbles along, year after year, with ad hoc solutions.

Multi-link problems take a very long time, even if they are looked at seriously.

Experience plus intuition helps locally, but not long term.

‘Engineering research’ is not well tuned to these problems, and is becoming less tuned.

We teach our students to be ‘problem solvers’.

That is, we teach them a lot of mathematics to analyze well-posed problems.

We reward them for their speed and ability to handle detail.

Modeling is left to industry practitioners who teach the models they have been taught.

Modeling to Shannon was not a compromise between reality and tractibility.

There was also the need to provide insight into similar but more complex models - to suggest a general structure.

We must help students to learn this art.

We must also learn how to reward excellence in this art.

We also teach students to read and understand the literature.

They view their role as creating additional chunks of this literature.

They (and we) view their success in number of papers.

They view their fields as very complex. Simulation tools and machine analysis help cope with the complexity.

Information technology makes it easier to do this complex research.

It seems to be drawing the focus away from Shannon's approach to research:

Trivialize the problem to the point where it can be grasped.

Then explore that and many other ways to trivialize the problem.

Then build understanding from these simple models.

Can we encourage students to choose and explore simple models more?

Several approaches:

Cover less material in class - expect greater understanding.

Give more exercises requiring explanation and discussion.

More exercises finding counter-examples.

Research on information systems is a fairly new field.

It does not have the traditions that carry physical sciences through fads.

In a sense, it doesn't have physical reality to refine the concepts.

That is, we can build more and more complex systems, and pretend that they simply need a little added debugging.

As scientists, we don’t like to spend time on flaky topics like studying styles of research.

We also don’t like to offend our colleagues by saying one style (ours) is better than another (theirs).

When a very successful style is getting lost, however, we at least must argue for the need for diversity in research styles.