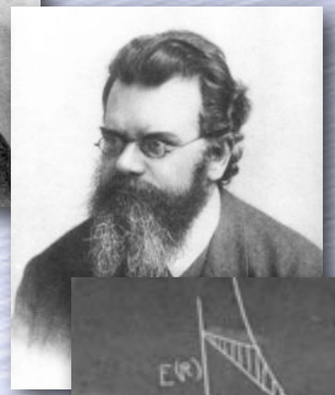
Stories About

Information, Secrets & Knowledge

What is information?

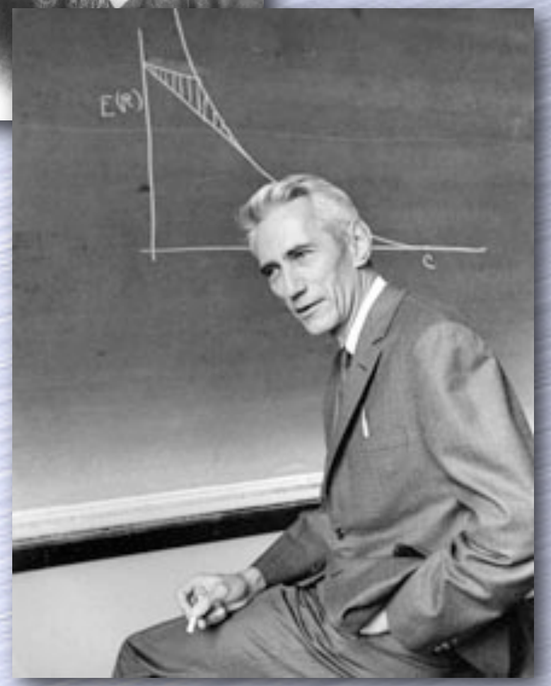
- Or rather the lack of it?
  - Uncertainty
  - Measured using Entropy
    - Borrowed from thermodynamics
    - An inherently “probabilistic” notion

Rudolf Clausius
(1822-1888)

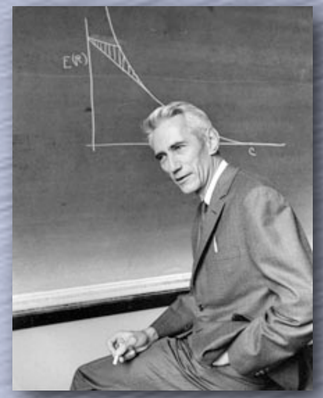


Ludwig Boltzmann
(1844-1906)

Claude Shannon
(1916-2001)



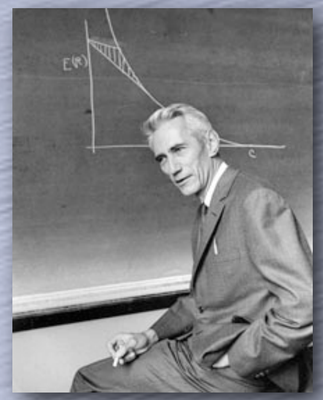
Entropy



Claude Shannon
(1916-2001)

- Suppose Alice flips a fair coin
  - Heads with probability 0.5, and Tails with probability 0.5
  - To communicate it to Bob, she needs to send 1 bit
- Now suppose the coin is biased: 0 with probability 0.9 and 1 with probability 0.1
  - Alice still needs to send 1 bit
  - But shouldn't it be easier to communicate the biased bit, as Bob already has a good idea what it would be?
    - Yes, if you consider communicating several bits

Entropy

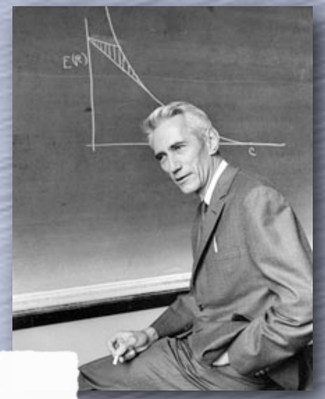


Claude Shannon
(1916–2001)

- Consider communicating 10 coin-flips
  - There are $2^{10} = 1024$ possibilities
  - Fair coin: each possibility has probability $1/1024 \approx 0.001$
  - Not so with the biased coin: E.g., $111\cdots1$ has probability only $(0.1)^{10} = 0.0000000001$, but $000\cdots0$ has $(0.9)^{10} \approx 0.35$
    - Idea: Use short strings to encode more likely possibilities and longer strings to encode less likely possibilities.
      - The "average" number of bits sent is much less than 10 bits. Entropy measures this (≈ 4.7).

Information Theory

- Ways to quantify information
  - Application 1: to study efficiency of communication (compression, error-correction)
  - Application 2: to study the possibility of secret communication
    - But what is secrecy?



Claude Shannon
(1916-2001)

Reprinted with corrections from *The Bell System Technical Journal*,
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

Communication Theory of Secrecy Systems*

By C. E. SHANNON

1 INTRODUCTION AND SUMMARY

The problems of cryptography and secrecy systems furnish an interesting application of communication theory[1]. In this paper a theory of secrecy systems is developed. The approach is on a theoretical level and is intended to complement the treatment found in standard works on cryptography[2]. There, a

A Game

- A “dealer” and two “players” Alice and Bob
- Dealer has a message m
- She wants to “share” it among the two players so that neither player by herself/himself learns anything about the message, but together they can find it
- Send half the bits of m to Alice and the rest to Bob?
  Send alternate bits to Alice and the rest to Bob?
  - In each case, Alice/Bob learns something about the message
- Other ideas?

A Solution

- To share a bit m, Dealer picks a uniformly random bit b and gives a := m⊕b to Alice and b to Bob
  - Together they can recover m as a⊕b
  - Bob learns nothing (b is a random bit)
  - Neither does Alice: for each possible value of m (0 or 1), a is a random bit (0 w.p. ½, 1 w.p. ½)
    - Her view is independent of the message
- Multiple bits can be shared independently: e.g., $\underline{m_1 m_2} = \underline{a_1 a_2} \oplus \underline{b_1 b_2}$

XOR

	0	1
0	0	1
1	1	0

m = 0 → (a,b) = (0,0) or (1,1)
m = 1 → (a,b) = (1,0) or (0,1)

Secrecy

- Is the message m really secret?
- If m is a single bit, Alice or Bob can correctly find m with probability ½, by randomly guessing
  - Worse, if they already know something about m, they can do better (Note: we didn't say m is uniformly random!)
- But they could have done this without obtaining the shares
  - The shares didn't leak any additional information to either party
- Typical cryptographic goal: **preserving** secrecy

Encryption

- Typical cryptographic goal: **preserving** secrecy
- Another important example: Secrecy against an eavesdropper who sees an encrypted message
- Perfect secrecy: the eavesdropper's view is independent of the message
- Shannon: Can achieve this if and only if the sender and the receiver pre-share a key that is as long as the message
  - Very unsatisfactory
- Fix: A new notion of secrecy!

A New Notion of Secrecy!

- Information leaked may not be a problem if it is in a form that is hard to work with
  - Knowledge! Information that is computationally useful.
- Information Theory, a la Shannon, considers information that is available to (unrealistic) computationally all-powerful parties.
  - Modern theory of cryptography is based primarily on Computational Complexity Theory
    - Several consequences: Pseudorandomness, Public-Key Cryptography, …

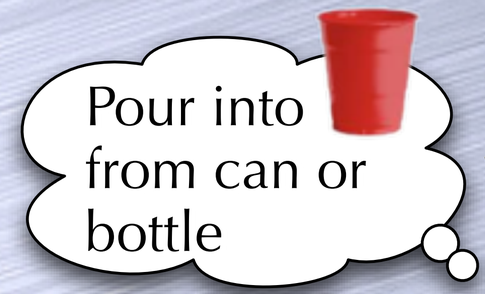
New Settings for Secrecy

- Suppose Alice claims: a soft drink in bottle and in can are chemically different
- How can she convince Bob of this?
- Option: Alice could send her laboratory equipment to Bob
- But can she avoid sharing her equipment?

An Interactive Proof

- Bob fills a cup randomly from can or bottle and sends it to Alice
- Alice tells whether cup was filled from can or bottle
- Repeat till Bob is convinced
  - If drinks are identical, Alice can win n repetitions only with probability $1/2^n$
- Bob learns nothing more: Every time he knew what Alice was going to say!

can/bottle



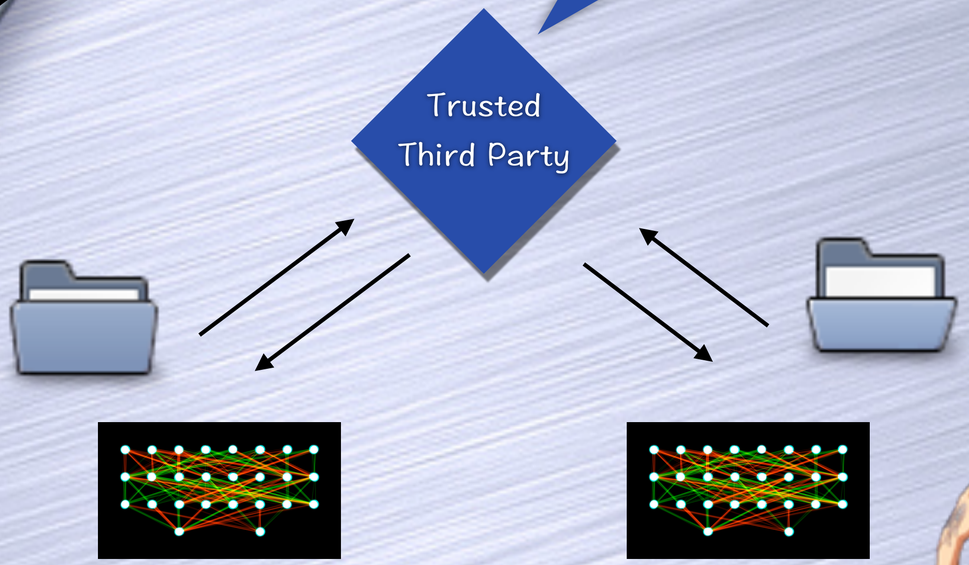
Zero-Knowledge Proof

Even More Complex Secrecy

- Alice and Bob want to collaborate: e.g., pool their data to build a machine learning model
  - But they want to keep their data private from each other — Except for what the ML model reveals
- Alice and Bob can cryptographically emulate the trusted third party!
- Secure Multi-Party Computation

Zero-Knowledge beyond output

Ideal level of secrecy: If we had a trusted third party



Information, Secrecy, Knowledge

- Information Theory gives a measure of information: **entropy**
  (And several other quantitative tools: mutual information, channel capacity, …)
  - **Secrecy** in terms of (almost) independence
- **Knowledge** as information that can be computed with
  - **Zero-Knowledge**: What will be seen can be predicted a priori (needs to look good only for computationally bounded entities)
  - **Secure Multi-Party Computation**: Zero-knowledge beyond desired outcomes