University of Pennsylvania
ScholarlyCommons

Departmental Papers (ASC) | Annenberg School for Communication

2009

Mathematical Theory of Communication

Klaus Krippendorff
University of Pennsylvania, kkrippendorff@asc.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/asc_papers

Part of the Communication Technology and New Media Commons

Recommended Citation
Krippendorff, K. (2009). Mathematical theory of communication. In S. W. Littlejohn & K. A. Foss (Eds.), *Encyclopedia of communication theory* (pp. 614-618). Los Angeles, CA: Sage. Retrieved from http://repository.upenn.edu/asc_papers/169

This is the corrected version of the original entry.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/asc_papers/169
For more information, please contact repository@pobox.upenn.edu.

Mathematical Theory of Communication

Disciplines
Communication Technology and New Media

Comments
This is the corrected version of the original entry.

This book chapter is available at ScholarlyCommons: http://repository.upenn.edu/asc_papers/169

Corrected pages 614-618 in S. W. Littlejohn & K. A. Foss (Eds.).
Encyclopedia of Communication Theory. Los Angeles: Sage, 2009.

MATHEMATICAL THEORY OF COMMUNICATION

Claude Shannon's mathematical theory of communication concerns quantitative limits of mediated communication. The theory has a history in cryptography and of measuring telephone traffic. Paralleling work by U.S. cybernetician Norbert Wiener and Soviet logician Andrei N. Kolmogorov, the theory was first published after declassification in 1948. Due to Wilbur Schramm's initiative, it appeared in 1949 as a book with a brief commentary by Warren Weaver. The theory provided a scientific foundation to the emerging discipline of communication but is now recognized as addressing only parts of the field.

For Shannon, "the fundamental problem of communication is reproducing at one point either exactly or approximately a message selected at another point" (Shannon & Weaver, 1949, p. 3). Shannon did not want to confound his theory by psychological issues and considered meanings irrelevant to the problem of using, analyzing, and designing mediated communication. The key to Shannon's theory is that messages are distinguished by selecting them from a set of possible messages – whatever criteria determine that choice. His theory has 22 theorems and seven appendices. Its basic idea is outlined as follows.

The Basic Measure

Arguably, informed choices are made to a degree better than chance and selecting a correct answer from among many possible answers to a question is more difficult and requires more information than selecting one from among few. For example, guessing the name of a person is more difficult than guessing its gender. So, its name would provide more information than its gender, the former often implying information about the latter. Intuitively, communication that eliminates all alternatives conveys more information than one that leaves some of them uncertain. Furthermore, two identical messages should provide the information of any one and two different messages should provide more information than either by itself..

To define quantities associated with selecting messages, in his 2nd theorem, Shannon proved that the logarithm function was the only one that conforms to the above intuitions. Logarithms increase monotonically with the number of alternatives available for selection, and are additive when alternatives are multiplicative. While the base of this logarithm is arbitrary, Shannon set it to two, thereby acknowledging that the choice among two equally likely alternatives—answering a yes or no question or turning a switch on or off—is the most elementary choice conceivable. His basic measure, called entropy H is

$$H(X) = -\sum_{x \in X} p_x \log_2 p_x$$

where p_x is the probability of message x occurring in the set of possible messages X. The minus sign assures that entropies are positive quantities. With N_X as the size of the set X of possible messages, H's range is

$$0 \leq H(X) \leq \log_2 N_X.$$

H averages the number of binary choices needed to select one message from a larger set, or the number of binary digits, *bits* for short, needed to enumerate that set. H is interpretable as a measure of uncertainty, variation, disorder, ignorance, or lack of information. When alternatives are equally likely,

No alternative
= no choice, p_x=1 and H = 0 bits

Two alternatives
= one binary choice, p_x=0.5 and H = 1 bit

Four alternatives
= two binary choices, p_x=0.25 and H = 2 bits

Eight alternatives
= three binary choices, p_x=0.125 and H = 3 bits

…

N alternatives
= $\log_2 N$ binary choices, $p_x = 1/N$ and $H = \log_2 N$ bits

2^N alternatives
= N binary choices, $p_x = 2^{-N}$ and $H = N$ bits.

Entropies and Communication

The additivity of H gives rise to a calculus of communication. For a sender S and a receiver R one can measure three basic entropies.

1. The uncertainty of messages s at sender S, occurring with probability p_s

$$H(S) = -\sum_{s \in S} p_s \log_2 p_s .$$

2. The uncertainty of messages r at receiver R, occurring with probability p_r

$$H(R) = -\sum_{r \in R} p_r \log_2 p_r .$$

3. The total uncertainty of s-r pairs of messages in a $S \times R$ table, occurring with probability p_{sr}

$$H(SR) = -\sum_{s \in S} \sum_{r \in R} p_{sr} \log_2 p_{sr} .$$

These lead to the sender-unrelated uncertainty entering a communication channel, colloquially called *noise*, and expressed as:

$$H_S(R) = H(SR) - H(S),$$

the uncertainty in the sender lost due to simplifications or omissions during communication, called *equivocation*, and expressed as

$$H_R(S) = H(SR) - H(R).$$

and the amount of *information transmitted* between sender and receiver, which can be obtained in at least four ways:

$T(S:R) = H(R) - H_S(R)$, the uncertainty at receiver R minus noise

$= H(S) - H_R(S)$, the uncertainty at sender S minus equivocation

$= H(S) + H(R) - H(SR)$, the sum of the uncertainties at S and R minus the total

$= H(SR) - H_S(R) - H_R(S)$, the total uncertainty minus noise and equivocation

The algebraic relationships between these quantities are visualized in the center of Figure 1.

Accordingly,

- *Communication* is the extent a sender is able to limit the receiver's choices, and
- *Information* is the extent a receiver knows the sender's choices.

Both are interpretations of the same quantity, $T(S:R)$, the amount of information transmitted. Both express differences between two uncertainties, with and without knowing the choices made at the other point of a communication channel. They increase with the number of choices or the improbability of the alternatives reduced in the cause of transmission. $T(S:R)$ quantifies a symmetrical relationship, not the property of a message.

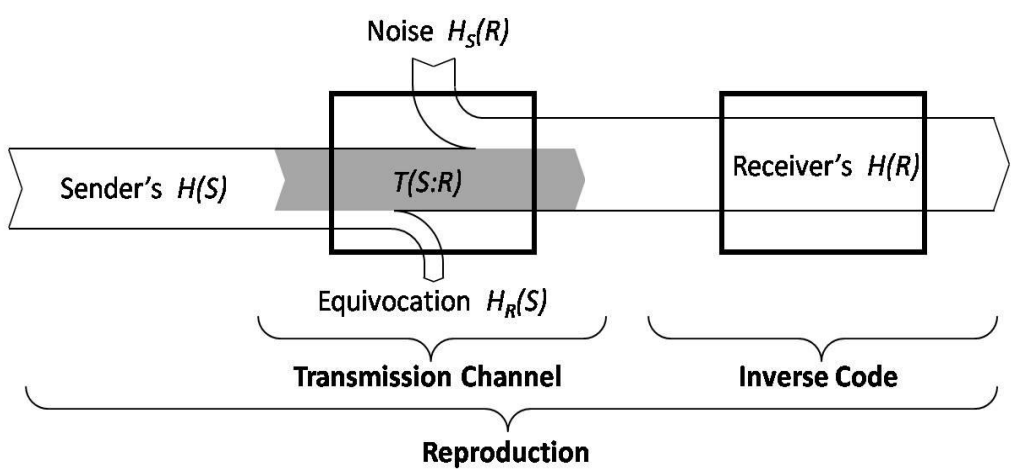

Figure 1 Transmission, Noise, and Equivocation in a Communication Channel

Relation to Thermodynamics

The well-known second law of thermodynamics states that for any closed system, utilizable energy differences such as of temperature, pressure, and chemical potential decrease over time. Only outside resources may counteract this natural tendency. Thermodynamic processes converge to a state of maximum entropy at which all utilizable energy is exhausted and everything stops. Shannon's theory of communication has been considered a more general formulation of this law. It states that noise or disorder can only increase, in communication terms eroding information, i.e., the ability of the receiver to relate what is received to what was sent. Without outside intervention, the process converges to where only noise prevails, equivocation has irrecoverably omitted all details of the original, and communication has ceased. One can experience the beginning of this process by repeatedly photographing a photograph or making Xerox copies of Xerox copies *ad infinitum*. After each iteration, the grain of an image becomes rougher, distinctions become blurred, ultimately disappear, and the chance to recover the original becomes increasingly unlikely.

Coding

Figure 1 distinguishes between the transmission of information and the reproduction of original messages. Transmission involves translating messages from one medium to another, not necessarily readable along its way. For example, the signals that transmit a document to a fax machine may be overheard without making sense, yet they convey all the information needed to reproduce the original or a close approximation of it. Fax machines embody a code. A *code* is a formal rule by which patterns are translated from one medium to another. To reproduce an original message, however, the last code must invert the aggregate of the preceding ones.

According to Shannon's 11th theorem, when the channel capacity

$$Max[T(S:R)] \geq H(S),$$

one can devise a code that reconstructs – up to a small error – the original messages from what was transmitted.

The distinction between transmission and reproduction is central for cryptography. Cryptographers pursue two tasks, (1) finding a code to decipher intercepted messages whose apparent gibberish is presumed to transmit valuable information, and (2) developing codes by which messages with sensitive information may be intercepted by unauthorized persons but cannot be read by them. During WWII, Shannon proposed unbreakable codes, now outdated and replaced by pairs of encoding and decoding algorithms whose reconstruction exceeds computational limits.

Hence, the mathematical theory of communication also addresses limits on the ability to find codes to reproduce a sender's originals. As such, the transmission of choices and the reproduction of original messages are prerequisites of all mediated communication. The readability of reproduced messages is a cultural issue, however, and goes beyond the theory. To understand each others' messages, communicators must be literate in each others' language communities.

Redundancy

Redundancy or inefficient transmission is the difference between the capacity of a communication channel and how much of it is utilized

$$R(S:R) = Max[T(S:R)] - T(S:R).$$

Redundancy may be due to (1) unused channel capacity, (2) duplicate transmission of messages, and (3) restrictions on the set of possible messages, for example, by a grammar or specialized vocabulary. Redundancy seems wasteful but is of considerable importance in human communication.

Much of Shannon's theory concerns the ability to devise codes that identify or correct corrupted communications. Such codes depend on the existence of redundancy. This can be experienced, for example, when proofreading text. Identifying typos is possible only when a language does not employ all combinatorially possible words. "Informition" is not an English word and assuming the writer is using Standard English, it can be identified as an error and corrected without uncertainty. English has been estimated to be about 70% redundant, which makes speech quite resistant to corruptions in the form of unclear pronunciation or acoustical interferences, and writing proof readable text, amenable to spellcheckers. Many technical communication processes avoid costly redundancy, often to the detriment of their users. Telephone numbers, Passwords, and Zip Codes, for

example, are designed without it and tolerate no human error.

In his 10th theorem, Shannon proved that the correctability of corrupted communication channels is limited by the amount of redundancy available in either the same or an additional channel. If redundancy is unavailable, communication erodes – analogue to the second law of thermodynamics. When noise occurs, accurate communication must be paid for by additional redundancy.

Digitalization

A good deal of Shannon's work addresses the problem of measuring communication of continuous variables—sound, images, and motion—with entropies that are defined for discrete phenomena. Meanwhile, technology has caught up with Shannon's methods of digitalizing and quantifying continuous phenomena. Today we are constantly confronted with Shannon's quantities. When buying a computer, we need to know its memory capacity and speed, when attaching a file to an email, we need to be concerned for its size, and when signing up for internet service, we need to be sure of a high transmission rate. The bits of Shannon's measures or bytes in computer terms— 1 byte = 8 bits —have become indispensable in contemporary life.

Human Communicators

The publication of Shannon's theory encouraged many researchers to treat humans as channels of communication and measure their capacity of processing information. George A. Miller suggested that people reliably handle no more than seven plus/minus two bits simultaneously. There are estimates that reading comprehension cannot exceed 16 bits/sec. Such limits are soft, however. Generally, experiences of information processing overload cause stress, which results in errors that reduce the capacity of humans to process information.

Human information processing capacities typically are a fraction of the amounts of information various media transmit to us, say on television screens. This does not render information measures irrelevant to understanding human communication. Miller observed that humans process information in chunks of familiar categories. So, on a high resolution photograph one may recognize people and the ruins of a house destroyed by a storm, without further distinguishing who these people are and identifying the pieces that led one to conclude they belonged to a house the storm destroyed. In communication theoretical terms, mapping a sender's information into fewer meaningful categories amounts to equivocation. However, the information lost to equivocation may not be meaningless entirely. It may contribute to the aesthetic appreciation of high fidelity images. Conversely, taking a photograph and emailing it as an attachment requires only a few decisions. The camera captures far more, which its receiver may well appreciate. Generally, humans select among and compose messages from chunks of commonly understood packages of information, not individual bits.

Weaver sought to extend Shannon's theory, by identifying three levels of communication: (A) the technical problem of accurately reproducing symbols, (B) the semantic problem of accurately reproducing the desired meanings of symbols, and (C) the effectiveness problem of causing desired conduct as the result of conveying desired meanings. Weaver conceived different codes as operating on each level. However, the idea of choices among alternatives that make a difference in people's lives underlies all three levels of communication.

Four Misconceptions

Some claim Shannon's theory is one of signal transmission. However, his is a content-free mathematical calculus. The physicality of messages has little to do with the quantities it defines. Easy calculation may favor its application to discrete phenomena, digital media for example, but his quantifications are applicable to wherever its axioms are met and users are facing the question of what they can or cannot transmit.

Communication literature typically bypasses Shannon's mathematical conceptions, and interprets a simple schematic drawing that Shannon and Weaver (1949, p. 5 and 98) used to contextualize the theory as "Shannon's linear communication model." True, Shannon was concerned with the transmission of information and reproduction of messages from one point to another. This did not preclude extending the theory to circular communication structures (Krippendorff, 1986). Shannon's theory provides

a versatile calculus, not a particular communication model.

The pervasive use of the content metaphors in the discourse of communication research easily misleads communication theorists to interpret Shannon's entropies as measuring the information content of messages. In fact, anticipating such confusions, Shannon refused to call his calculus information theory, and named his *H*-measures "entropies." For him information is not an entity contained in a message, but manifest in patterns that are maintained during highly variable processes of communication.

Finally, the theory does not presume that communicators share the same repertoire of messages, for example, having the same preconceptions or speaking the same language. By quantifying how choices made at one point affect those made at another, the theory asserts fundamental limits on all levels of human communication.

Klaus Krippendorff

See also Cybernetics, Information Theory, Uncertainty Reduction Theory

Further Readings

Krippendorff, K. (1986). *Information Theory; Structural Models for Qualitative Data.* Beverly Hills, CA: Sage.

Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication.* Urbana, IL: University of Illinois Press.