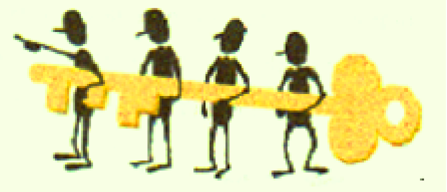


Physical roots of Information Science (IS)

Why logarithm is used in Theory of Information by Claude E. SHANNON?

DI Franz PLOCHBERGER

PreDoc at University of Vienna and TU Vienna

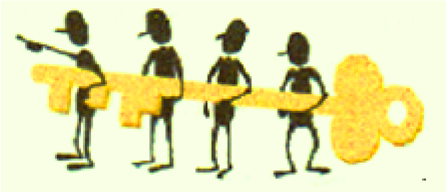
Information Scientist

Stumpergasse 6/19

1060 Wien

http://www.plbg.at

Last edition: 28. May 2013



1 Abstract

The Theory of Information or Concept of Information was an exciting new topic in science, beginning at about 1920 (by NYQUIST, HARTLEY) up to 1950 (SHANNON). It was the origin of Informatics and actual new Information Sciences.

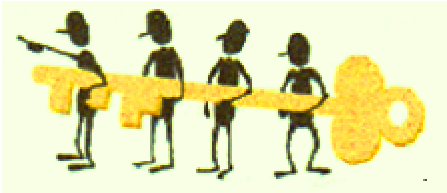
Here some physical and mathematical roots are worked out. Especially an answer to the question: Why the logarithm as mathematical function was used by these early information scientists?

Today we know what was too early for them. We can say that the word and term *data* is an important basis of word and term *information.* They saw not yet any importance in that *data.*

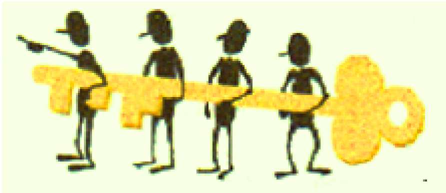
They meant *data* when they used *information!* But that we can say today, nearly 100 years later.

2 Keywords, Search Items

Energy, Entropy, Information, Data, Logarithm, Probability



3 Content



4 Why this report?

The terms thermodynamics, energy or entropy have been topics in physics as the science of matter till the end of 19th century. Then quantum mechanics got importance: Max PLANCK, Albert EINSTEIN, and Henry POINCARÈ.

Also a connection between the Viennese Physicist Ludwig Eduard BOLTZMANN (1844-1906) and the Theory of Information (HARTLEY, SHANNON) can be given.

The main question of this work is: Why the mathematical function *logarithm* is used in the Theory of Information of SHANNON?

5 Historical review

Harry NYQUIST (1889-1976) wrote 1924: **m exp n = const**, m possible different levels, n possible signal elements, as amount of all possibilities of data-transmission - he used the classical combinatorial *"variation with repeatment".* NYQUIST named his const the „amount of intelligence". His formulas were used also by the scientists HARTLEY and SHANNON.
As example for m=2 (= 2 levels) and n=8 (= 8 signal elements): const = 2 exp 8 = 256 different characters (signs) of 8 signal elements are possible. (NYQUIST H. (1924), Appendix B, p 343).
In this context no necessarity to use the logarithm was given, but he transformed his formulas to it. Why? Perhaps we can find an explanation later.

R.A.FISHER (1890-1962) used 1925 in FISHER R.(1925), p 24 of 40, for first time the term „amount of information" in connection to mathematician statistics (LOGAN R. (2012), p 70 of 91). This *term* is used today under the mathematical definition „Fisher-Information" and is really something new in statistics. In actual Information Science it's one example for using the *word* information and making by this usage a new *term* in mathematical statistics. According to PLOCHBERGER F. (2012), p. 24 it's an example for the „3rd Postulate (IP3) of Information Science". So for our aim - (Why logarithm was used in Theory of Information?) - this term isn't important.

Ralph V. L. HARTLEY (1888-1970) used 1927 in HARTLEY R. (1928) the *term information* too. In his formulas he used the mathematical function algorithm too. He for first time wanted to find a measurement based on physical facts and contrasted to psychological considerations. He made a definitive distance and separation to the *word* information, which was not measurable for him. In actual Information Science this separation can be shown correct and better detailed by using the term and word **data** instead of his term information (PLOCHBERGER F. (2012), p23 ff). Today we can say: HARTLEY for first time tried to find a common measurement of the „(storage) amount of data" which could be sent by different transmission systems – as significant property of a transmission system. He knew telegraphy, telephony, radio and television.
But that doesn't answer the question: Why HARTLEY used the algorithm in his formulas too? In HARTLEY R. (1928) p 538 he took over the formula of NYQUIST (see above) : s exp n, s he called symbols and n he called selections, const he named „the number of distinguishable sequences". He defined, that s is not so important as n for his searched term. He set H (his amount of information) proportional to n in his formula H = Kn. K is a const – including s (in reality possible levels of a technical current- or voltage-signal). For different systems he set s1 exp n1= s2 exp n2 and supposed the same amount H= K1*n1 = K2*n2. His individual definition was the following formula K = K0 * log s. This was his invention!

By setting K0 to an arbitrary constant 1 he came to the result
H = K*n = n* log s   or H = log s exp n.
On page 540 of his text (HARTLEY R. (1928)) he gives the comment: „*What we have done then is to take as our practical measure of information the logarithm of the number of possible symbols sequences*“. So in actual Information Science we can explain his „number of possible symbol sequences“ as maximal possible different signs in the storage area or data length of n bits. That was similar to NYQUIST's formula. He defined the „speed of transformation of intelligence“ and used also the same mathematical connection between amount of possible values (m or s) and amount of possible signal elements or selections n.
In my understanding I suppose that the using of algorithm by HARTLEY could come from NYQUIST. But now it lead to a realistic mathematical term, the Amount of Information by HARTLEY or in actual Information Science the amount of possible transferable signs in the data-length of n bits.

6 Information Theory by C.E. SHANNON

Claude Elwood SHANNON (1916-2001) made already 1948 a significant reference to Harry NYQUIST and R.V. HARTLEY in his work SHANNON C. (1948), p 1 or 379. SHANNON got famous as the father of the Concept of Information or Theory of Information. He used also the *word* information and investigated a new *term* around information in physics in order to get a measurement of it.

He defined real formulas and new terms.

For long time his theories were called Information Theory or Concept of Information, but actual Information Science can give more detailed integration of his terms.

For our aim – Why he used the logarithm in his definitions? – we can get now a nearly exact answer.

6.1 SHANNON entropy = Information Entropy

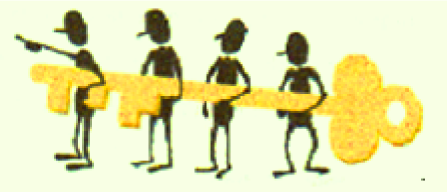
First of all we have to understand the topic term „**entropy**“.
The founder of the term was **Rudolf CLAUSIUS (1822-1888)**, a German physicist, who also found the Second Sentence of Thermodynamics. He wanted a measurement of the loss of mechanical energy. He thought especially about that amount of heat energy that couldn't be transferred into usable work (LOGAN R. (2012), p 71,72). It was known, that energy can't disappear but can only change his type. He took the word from the Old-Greek trope (trope) meaning transformation an added the first two letters of en-ergy. So he constructed entropy. The original German term was „Verwandlungsinhalt“. This entropy was constantly increasing because the type of energy permanently changed.

Another famous Viennese physician at that time – **Ludwig BOLTZMANN (1844-1906)** - developed a new formula for this entropy especial of gases. He found **S (entropy of gases) = k * log W**, k the Boltzmann constant and W the number of microstates of a gas.

And it's known that SHANNON built his entropy like the BOLTZMANN Entropy. He defined
H = - sum over m i-elements (pi * log (pi), pi probability of one element.

Today this formula is a possible measurement but has no practical importance in Information Science.

For our aim: Why SHANNON used the algorithm? - we have here a clear answer: He took it from BOLTZMANN.



6.2 Bit as new unit

In (SHANNON C. (1948), p 1) SHANNON wrote:

„The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly bits, a word suggested by **J. W. TUKEY**. A device with two stable positions, such as a relay or a flip-flop circuit, can store one bit of information. N such devices can store N bits, since the total number of possible states is 2 exp N and log with base 2 of (2 exp N) = N.“

He added: „If the base 10 is used the units may be called decimal digits.“

… This unit for base 10 was called later 1 Hartley.

Seen out of actual Information Science this definition was his most important one.

Here the genuine mathematical use of logarithm is very useful.

6.3 SHANNON's own answer to our question

In (SHANNON C. (1948) p. 1) he wrote too:

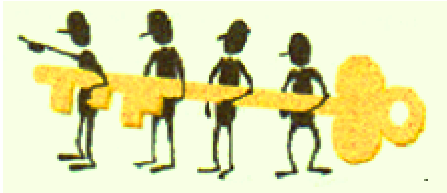
„The logarithmic measure is more convenient for various reasons:

It is practically more useful. Parameters of engineering importance such as time, bandwidth, number of relays, etc., tend to vary linearly with the logarithm of the number of possibilities. For example, adding one relay to a group doubles the number of possible states of the relays. It adds 1 to the base 2 logarithm of this number. Doubling the time roughly squares the number of possible messages, or doubles the logarithm, etc.

It is nearer to our intuitive feeling as to the proper measure. This is closely related to (1) since we intuitively measures entities by linear comparison with common standards. One feels, for example, that two punched cards should have twice the capacity of one for information storage, and two identical channels twice the capacity of one for transmitting information.

It is mathematically more suitable. Many of the limiting operations are simple in terms of the logarithm but would require clumsy restatement in terms of the number of possibilities. “

So we get a real practical explanation, why physics are better measurable by logarithm.



7 A later term: Information Content

As I was studying „Nachrichtentechnik“ at TH Wien beginning 1968 I learned this very important term which can be direct related to SHANNONS Information Theory. I found this very explaining term in my official scriptum at the Technical Highshool of Vienna in about 1968.

For actual Information Science the term Information Content is important (KRAUS G. (1968), p 60 ff) and defined as:

Information Content I = log (1/p) = - log p

It is correlated to SHANNONS Theory but I couldn’t find it in his (SHANNON C. (1948)) scriptum.

Never the less it is immense important! So everybody gets a better understanding of *word* and *term* information for all times.

Information Content is reciprocal to the property of experience.

With other words we can say: If the Content of Information (= the amount of news in it) is very high the property of their experience must be very low and vice versa. That's a very useful definition !

I learned to use the unit **bit** for it by using base 2 of logarithm and to use **Hartley** as unit when logarithm has base 10.

In actual Information Science this unit 1 bit is used for this Information Content and for the amount of storage (= amount of data) too. Both terms are semantically in same usage: You can store the Information Content of 1 bit on the storage length of 1 bit. Here the practical importance is founded.

The logarithm is used because of SHANNON's explanations above.

But to understand the actual term information it's not important. We even can drop here the logarithm.

8 Conclusions

- Logarithm is useful because of same structures in physics (elements for transmission of information) and mathematical definitions (f. i. exponent → multiplication → summation). This fact may be important today at computing a big amount of data.

- SHANNON explained the optimal usage of logarithm by himself (see 6.3 above).

- For actual Information Science the usage of logarithm isn't so important, because we don't use tables or slide-rules as it was usual in the time of BOLTZMANN, NYQUIST, HARTLEY and SHANNON. We use our scientific hand-held calculators and common computers.

- For undstanding SHANNONS term of Information we can substitute his term in most cases by the new term **data or structure of data** (PLOCHBERGER F. (2012), p 22 ff).

9 References

LOGAN R. (2012),

Robert K. LOGAN, What is Information? Why is it relativistic and what is its relationship to Materiality, Meaning and Organization, University of Toronto, Canada, 14 February 2012, ISSN 2078-2489, www.mdpi.com/journal/information.

BURGIN M. (2011),

Mark BURGIN, Information: concept Clarification and Theoretical Representation, University of California, LosAngeles, ISSN 1726-670X, tripleC 9(2): 347-357, 2011, http://www.triple-c.at

BRENNER J. (2012),

Joseph E. BRENNER, Mark BURGIN's Theory of Information, International Centre for Transdisciplinary Research, 75019 Paris, France, 1 June 2012, Book Review, ISSN 2078-2489, www.mdpi.com/journal/information

BURGIN M. (2010),

Mark BURGIN, Theory of Information, Fundamentality, Diversity and Unification, University of California, Los Angeles, USA, 2010, World Scientific, ISBN-13 978-981-283-548-2

FISHER R. (1925),

Ronald A. FISHER, Theory of Statistical Estimation, 1925, Proceedings of the Cambridge Philosophical Society, 22:700-725, 25 pages,

http://digital.library.adelaide.edu.au/dspace/bitstream/2440/15186/1/42.pdf

HARTLEY R. (1928),

Ralph V. L. HARTLEY, Transmission of Information, published first time 1927, printed in Bell Systems Technical Journal, 1928, VII, p 535-563

KRAUS G. (1968),

Günther KRAUS, Telegraphie, Offizielles Vorlesungsskriptum der TU (TH) Wien, Institut für Niederfrequenztechnik, 1968

PLOCHBERGER F. (2011),

Franz PLOCHBERGER, H.NYQUIST, R.V.HARTLEY und C.E.SHANNON aus der Sicht der heutigen Informationswissenschaft, Eigenverlag, http://www.plbg.at/Werke/deutsch/NYQUIST,HARTLEY,SHANNON.pdf

PLOCHBERGER F. (2012),

Franz PLOCHBERGER, Axioms around the term Information (2012), Part of Dissertation at University Vienna and Technical University Vienna, 2012,

http://www.plbg.at/Werke/english/Axioms%20around%20the%20term%20Information.pdf

SHANNON C. (1948),

Claude E. SHANNON, A Mathematical Theory of Communication, The Bell System Technical Journal, Vol.27, pp 379-656, July, October, 1948 (55 pages)

NYQUIST H. (1924),

Harry NYQUIST, Certain factors affecting telegraph speed, Bell System Technical Journal*, 3, 324–346*