S-Information

Der Begriff bezieht sich vor allem auf die fundamentale Arbeit von Cloude E. Shannon:

A mathematical theory of communication. Bell Syst. Techn. J. 27 (Juli 1948) S. 379 - 423 und (Oktober 1948) S.623 - 656. (auch University Illonois Press 1949). Teil 2 Communication in the Presence of Noise. Proc. IRE 37 (1949) pp. 10 - 20, (*eingereicht* am 24.3.1940).

Sie enthält nicht einmal das Wort „Information“, sondern stattdessen nur communication.

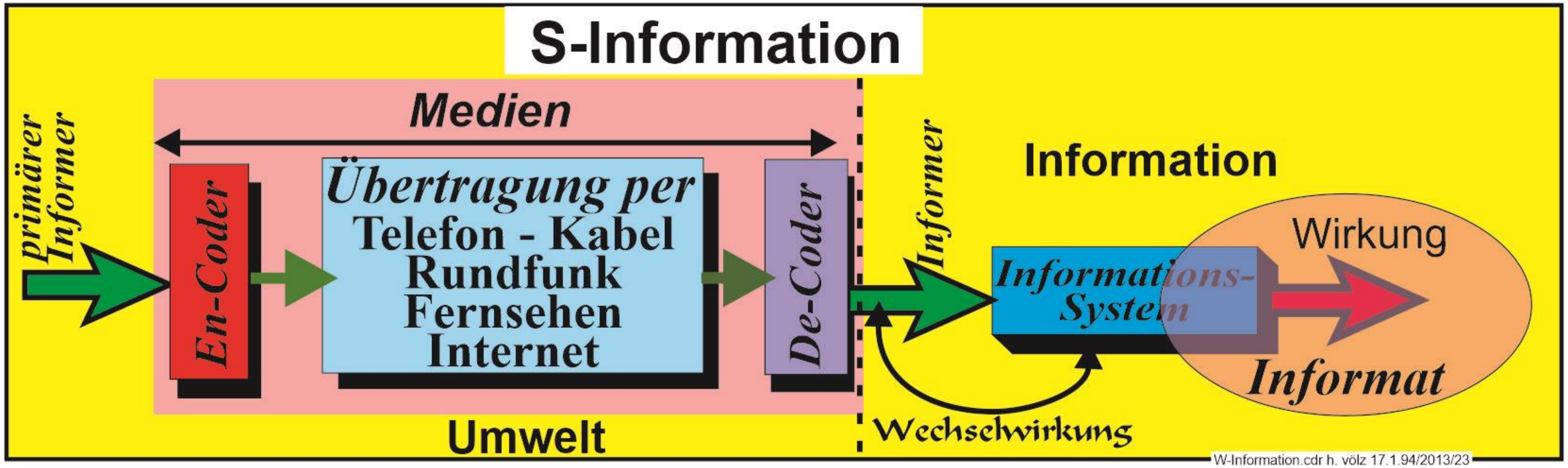
Shannon benutzt dabei bereits Inhalte der **Digitaltechnik, d**ie nicht einmal andeutungsweise vorhanden ist. Hauptziel ist die bestmögliche Übertragung bzgl. ***Energie, Zeitdauer*** sowie Überbrückung vom **Raum**

Zuweilen wird die Arbeit immer noch ***falsch interpretiert***, das gilt vor allem für den **Entropie**begriff
Auch bestehen Unklarheiten bzgl. des *Entstehungsdatum*s: 1940 oder 1948.

Deutsch übersetzt erschien die Arbeit als

Mathematische Grundlagen der Informationstheorie. R. Oldenbourg, München - Wien, 1976

Dabei erhielt sie ein teilweise inhaltlich falsches **Vorwort** von *Waren Weaver* (Einbeziehung der Semiotik).



S-Information
Medien
primärer Informer
En-Coder
Übertragung per
Telefon - Kabel
Rundfunk
Fernsehen
Internet
De-Coder
Informer
Information
Wirkung
Informations-System
Informat
Umwelt
Wechselwirkung
W-Information.cdr h. völz 17.1.94/2013/23

Stark vereinfachtes Problem

mit dem Satz: „**Dies ist ein Text**"

Er enthält: ***1-mal****:* D, T, n, x. ***2-mal***: s, t. ***3-mal***: e, i, Leerzeichen

Die Übertragung soll mit kurzen Tönen „**#**" erfolgen, die durch kurze Pausen „**°"** getrennt werden.
Zur Vereinfachung werden sie aber im Folgenden weggelassen.

lange Pause bewirken, dass ein Zeichen beendet ist und werden durch „**O**" gekennzeichnet
Es ist dann die folgende Codierung sinnvoll:

e = #	**i** = ##	*Leerzeichen*= ###	**t** = ####	**s** = #####
x = ######	**n** = #######	**T** = ########	**D** = #########	

Dann muss übertragen werden:

#########O##O#O#####O###O##O#####O####O###O#O##O#######O###O########O#O######O####O

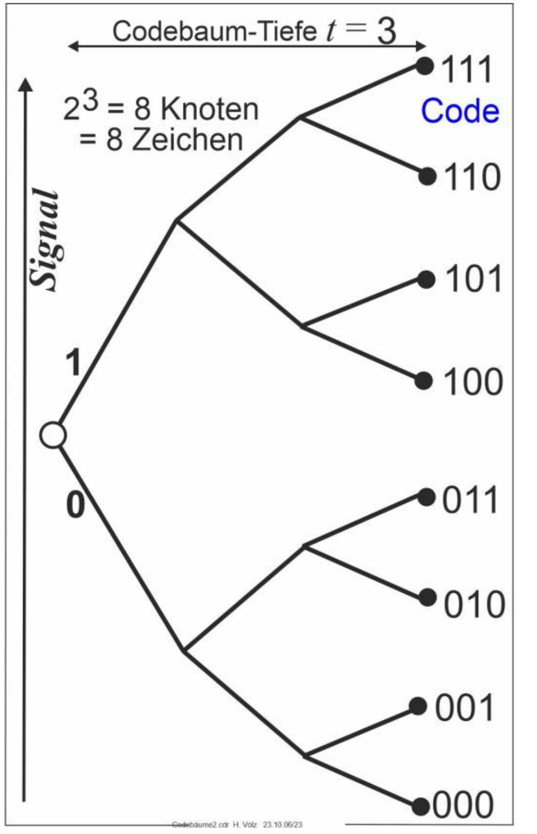
D i e s -- i s t ---- e i n --- T e x t

Für den Satz kann keine kürzere (streng binäre) Kodierung gefunden werden, jedoch mehrere gleichlange.

Es besteht nur eine gewisse Analogie zum Morsecode,
Der jedoch 2 Zeichen (*lang* und *kurz*) sowie 3 (4) unterschiedliche Pausenzeichen benutzt

Shannonsche Entropieformel

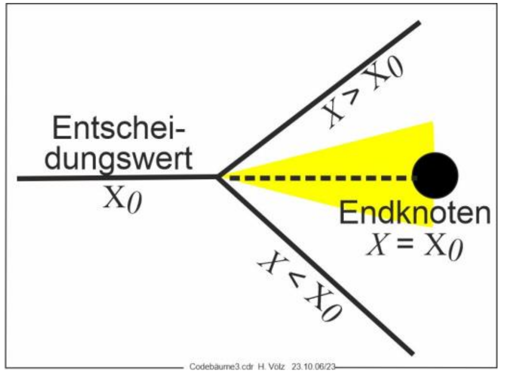
Bei einer binären Codierung sind für die Zeichen zu unterscheiden



- Der **Code** aus einer Abfolge von **0** und **1**, deren Anzahl der ***Tiefe t des Codebaumes*** entspricht
- Die **Wahrscheinlichkeit** (Häufigkeit) p_i des *i*-ten Zeichens **im Gesamtsignal** *n*

Shannon bestimmte dafür den theoretisch kleinstmöglichen Codeaufwand für die Übertragung. Leider gibt es für die sehr wichtige Formel keinen unmittelbar verständlichen Beweis. Entscheidend ist jedoch immer ein binärer Codebaum. (Bild rechts oben) An jeder Verzeigung führt eine **1** nach oben eine **0** nach unten,

Im Bereich von **000** bis **111** müssen alle Signalwerte liegen, z. B. die Zahlenwerte 0 bis 1. Zum Ursprung des Codebaumes gehört dann ½, für die ersten Verzweigungen ¼ bzw. $^3/_4$.



Die Endknoten können nur mit Näherung den Entscheidungswerten entsprechen, Streng genommen müsste sie lauten $X \approx X_0$ (links). Genähert kann auch der Bereich zwischen den Verzeigungen angenommen werden (gelb). Dann ist die Codierung aber etwas unscharf.

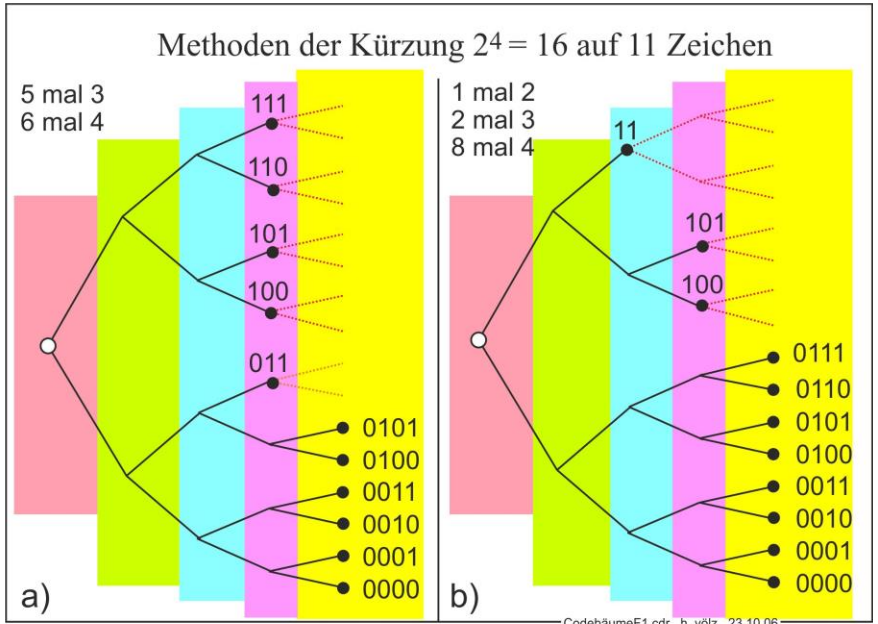
Aus diesen Betrachtungen lässt sich der Entropiesatz gemäß [Völ17] erstmalig anschaulich erklären (w. u.).

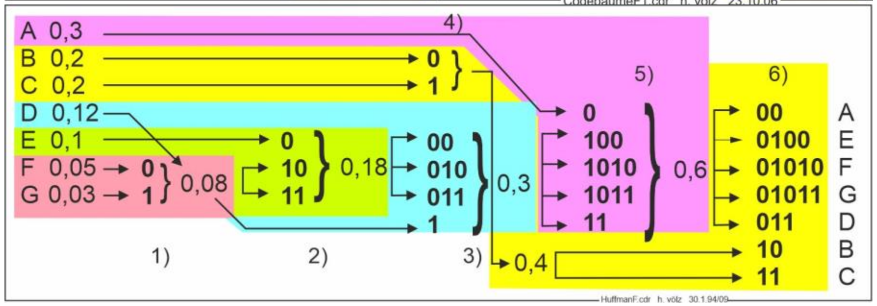
Wenn alle Endknoten gleich *t* weit vom Ursprung entfernt sind, dann ist die Wahrscheinlichkeit jedes Zeichens $p_i = 1/2h_i = p_i \cdot \mathrm{ld}(1/p_i) = -p_i \cdot \mathrm{ld}(p_i)$. Durch Summierung folgt die Shannonsche Entropieformel:

$$\boxed{H = -\sum_{1=1}^{n} p_i \cdot ld(p_i)\,.}$$

Das vielfach hervorgehobene negative Vorzeichen folgt also zwangsläufig

Die bisherigen Betrachtungen gelten nur einen ***vollständigen Codebaum.*** Dafür muss also die Länge aller Codewörter gleich sein. Dabei kann auch formal (theoretisch) die Tiefe *nicht ganzzahlig* sein. Infolge der verkürzten Längen t_{extra} sind aber zuweilen nicht alle Tiefen verwendbar. Für die richtige Anzahl der Codewörter müssen einige an den Enden zusammengefasst (abgeschnitten) werden. Drei Beispiele zeigt das nebenstehende Bild. Für die meisten Verkürzungen gibt es meist mehrere Varianten. In jedem Fall ist aber nicht mehr der ideale Wert der Entropieformel erreichbar. Es tritt eine ***Redundanz*** auf. An die Stelle der Entropie tritt dann mit den Codelängen C_{Li} der Code-Aufwand





$$C_A = \sum_{i=1}^{n} p_i \cdot c_{Li} \ .$$

Die hierbei notwendigen Berechnungen für die bestmögliche Optimierung sind oft recht komplex und so sind verschiedene Codierverfahren von Shannon über Fano 1949 bis Huffman 1952 entstanden. Danach wurde (bisher) kein neuer Weg gefunden.

Zum Teil entstehen spezielle Einschränkungen für die Codewörter, keines darf dann den Anfang eines anderen Codewortes enthalten. Existiert z. B. ***te*** so kann es nicht mehr geben: *tee*, *ter*, *test*, *terror* usw.. Dadurch entfällt das Sonderzeichen am Ende jedes Codeworts (s. oben **O**. Es wird automatisch die richtige Trennung erzeugt. Das bewirkt eine zusätzliche Kürzung des zu übertragenden Signals. Siehe auch hierzu

[Völ17] Völz, H.: Das ist Information. Shaker Verlag Aachen 2017.

Senkung der Redundanz

Da allgemein gilt $H \leq C_A$ wird die Redundanz einer Codierung eingeführt

$$R = C_A - H$$

bzw. die relative Redundanz

$$r = \frac{C_a - H}{H}\,.$$

Doch sie Bereits Shannon bewies: theoretisch ist immer $r \to 0$ möglich.
Dazu dienen Codekombination mit Gedächtnis. Bei den Zeichen A und B
Sind es die Kombinationen AA, AB, BA und BB
Das lässt sich fortsetze zu AAA, AAB, ABA, BAA, BBA, BAB, ABB, BBB und mehrstelliger weiter

Die Codewörter werden dabei immer länger (theoretisch ∞)
Das erfordert bei der Codierung und Decodierung zusätzlichen Speicherplatz und Verzögerung

Deshalb ist dies nur selten vorteilhaft

Der Name Entropie

geht auf den Energie-Erhaltungssatz der Thermodynamik zurück.
Mechanische Energie ist immer vollständig in Wärme umwandelbar, aber nicht umgekehrt (***Zeitpfeil***)

Es gibt kein **Perpetuum Mobile 1. Art**, das dauernd Energie erzeugt, ohne sie der Umgebung zu entziehen.
Thermodynamische ***Entropie*** existiert in zwei Varianten:

1854 Clausius: $\Delta S = \Delta Q/T$ gemäß dem Kreisprozess

1857 Boltzmann: $S = k_B \ln(W)$, auch als H-Theorem bezeichnet. Gemäß der Wahrscheinlichkeit des Zustandes
Temperatur T, **Wahrscheinlichkeit *W* d**es Zustandes und k_B = Boltzmann-Konstante.

Hierzu gehört auch der **a*bsoluten Nullpunkt*** $T = 0$, der 1848 von Kelvin vorgeschlagen wurde.

Aus diesen Gründen beschrieb Clausius 1865 den ursprünglichen Entropie-Begriff:

> *„Sucht man für S (die Entropie) einen bezeichnenden Namen, so könnte man, ähnlich wie von der Größe U (der inneren Energie) gesagt ist, sie sey der Wärme- und Werkinhalt des Körpers, von der Größe S sagen, sie sey der Verwandlungsinhalt des Körpers. Da ich es aber für besser halte, die Namen derartiger für die Wissenschaft wichtiger Größen aus den alten Sprachen zu entnehmen, damit sie unverändert in allen neuen Sprachen angewandt werden können, so schlage ich vor, die Größe S nach dem griechischen Worte »tropae«, die Verwandlung, die Entropie des Körpers zu nennen. Das Wort Entropie habe ich absichtlich dem Wort Energie möglichst ähnlich gebildet, denn die beiden Größen, welche durch diese Worte benannt werden sollen, sind ihren physikalischen Bedeutungen nach einander so nahe verwandt, daß eine gewisse Gleichartigkeit in der Benennung mir zweckmäßig zu seyn scheint.“*

Die Struktur der Boltzmann-Formel mit ln(W) war für Wiener die Begründung der Benennung bei Shannon

Die Vielzahl der Entropien

Der Begriff Entropie ist heute für über zehn mathematische Formeln und deren Kontext gebräuchlich.
Dabei gibt es nur selten ähnliche oder gar übereinstimmende Inhalte.
Details Siehe: Völz, H.; Ichrealität – meine Weltsicht. Shaker Verlag, Düren 2021

1854 Clausius: Thermodynamik
1865 Boltzmann: statistische Thermodynamik
1948 Shannon: Optimal-Codierung.
1954 Carnap-Entropie (Semiotik, ähnlich Quantität Qualität).
1962 Renyi-Entropie (rein formal mathematische Verallgemeinerung von Shannon).
1963 Bongard-Weiß-Entropie (auch subjektive Wahrscheinlichkeit).
1965 Marko: bidirektionale Information, weiterentwickelt für viele Sender und Empfänger.
1987 Hilberg: deterministische Informationstheorie (Statistik entsteht über zufällige Schalter).

Relativ oft wird aber der Gehalt der Formeln nicht richtig verstanden oder vieles miteinander vermengt
Sogar bei Experten behaupten dann falsche Kombinationen und Folgerungen.
Mehrfach werden sowohl Boltzmann- und Shannon-Entropie gleich oder sich summarisch ergänzend angesehen
Neg-Entropie wegen des Vorzeichens, ist dann die Begründung

Freitod von Boltzmann

Heute gelten die Ergebnisse von Boltzmann als eine besonders große wissenschaftliche Leistung.
Doch zu seiner Zeit war das deutlich anders.
Von den meisten Physikern wurde sie aus zwei Gründen stark kritisiert.

Zunächst galt die ***Atom-Annahme*** als unbegründet, teilweise sogar als falsch:

„Hat denn schon jemand ein Atom gesehen?"
Das galt sogar, obwohl sie in der Chemie durch Dalton für die Stöchometrie eingeführt
und generell sehr erfolgreich benutzt wurden.

Der zweite Widerspruch kam vor allem von Mach, Ostwald, Poincaré und Zermelo.
Er betraf die ***statistische Irreversibilität*** sowie die Ableitung des 2. Hauptsatzes aus reversiblen Mikrozuständen.

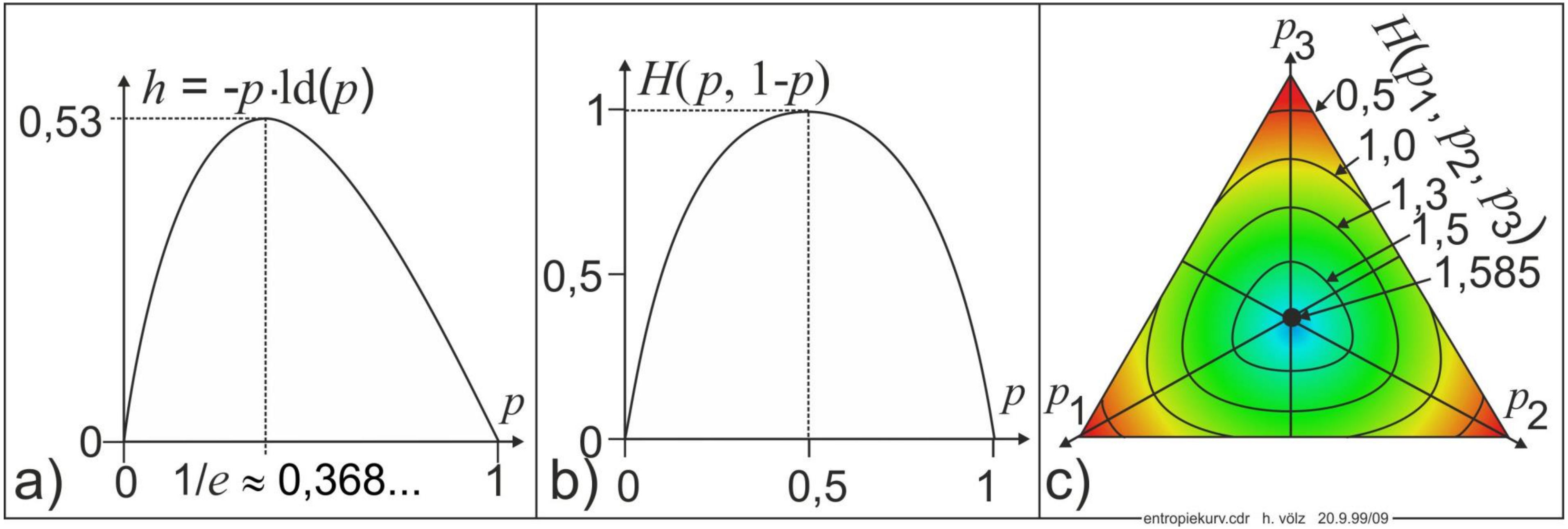
Das trieb Boltzmann schließlich am 5.9.1906 in den Freitod.

Hierzu stellte Planck später fest, *dass sich eine neue wissenschaftliche Wahrheit normalerweise*

> *„nicht in der Weise durchzusetzen pflegt, daß ihre Gegner überzeugt werden und sich, als belehrt erklären, sondern vielmehr dadurch, daß die Gegner allmählich aussterben und daß die heranwachsende Generation von vornherein mit der Wahrheit vertraut gemacht wird".*

Shannon-Entropien-Verlaufe

Sie sind bei kontinuierlich veränderlichen Wahrscheinlichkeiten u. a. für 1, 2 und 3 Zeichen darstellbar.

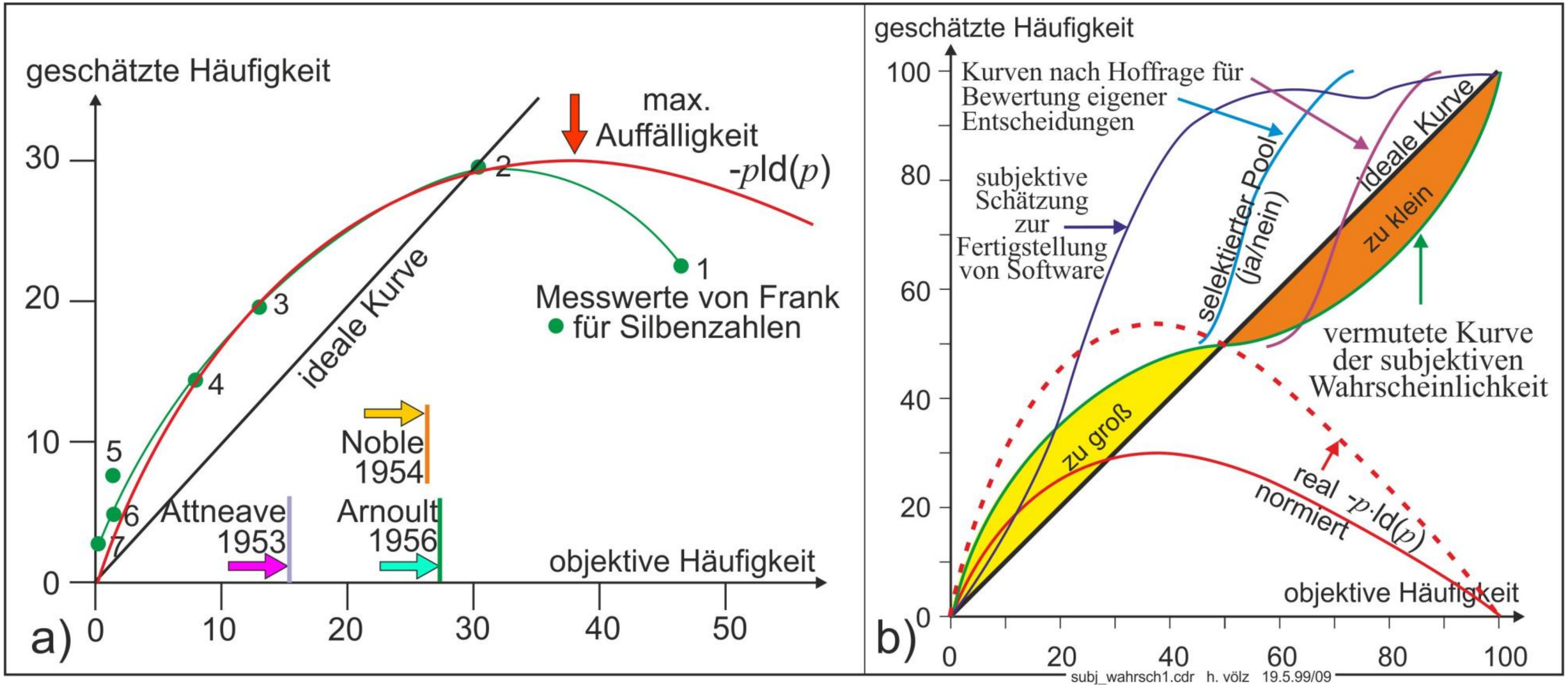


Auffällig ist der Verlauf für den Einzelwert $h = -p \cdot \mathrm{ld}(p)$ mit ausgeprägtem Maximum bei $p = 1/e \approx 0{,}368$.
Sie fallen subjektiv auf. Das entdeckte Helmar Frank und benannte den Wert „Auffälligkeit“ [Fra69]

Ansonsten ergibt sich immer ein Maximum für H bei Gleichverteilung: $p_i = 1/n$ *mit* $H_{max} = \mathrm{ld}(n) \geq C_{A.}$
Das entspricht zugleich der Obergrenze für *H*. Hierfür gilt die Tabelle

n	2	3	4	8	1024	1048576
$\boldsymbol{H_{max}}$	1	1,585	2	3	10	20

Subjektive Wahrscheinlichkeit



Hinweis Kaufhaus usw. Übergang zur Kunst

.

Superzeichen

Die Kodierung mit Kombinationszeichen (s. o.) besitzt auch für unsere Wahrnehmung Vorteile.
Leider gilt aber für die Ziffernhäufigkeiten das Newcombschen Gesetz (gemäß Intervallänge!):

Ziffern	1	2	3	4	5	6	7	8	9
Häufigkeit	30,1	17,6	12,5	9,7	7,9	6,7	5,8	5,1	4,6

Außerdem existiert bei Preisen noch eine die Überhöhung für $p(9)$.

Hiervon abgesehen, hat Cube die Wahrnehmung von Superzeichen (als Kombinationen) untersucht.
Ein Text bestehe dazu aus k Wörtern mit je n Zeichen, also insgesamt $m = k \cdot n$ Zeichen.
Für die Einzelentropie eines Wortes gilt dann $H_1 \approx n \cdot \mathrm{ld}(n)$.
Folglich liefern alle Wörter $H_2 \approx m \cdot \mathrm{ld}(n)$.
Zusammengefasst gilt $H_3 \approx k \cdot \mathrm{ld}(k)$. Wegen $k = m/n$ gilt folgt schließlich

$$H_{total} \approx m \cdot \mathrm{ld}(n) + \frac{m}{n} \cdot ld\left(\frac{m}{n}\right).$$

Ihr Minimum liegt bei $m = n \cdot e^{n-1}$. Daraus leiten sich die folgenden Werte ab:

bei maximaler Wortlänge m	5,44≈6	22,2≈23	80,3≈81	273	891	2824	8773
optimale Gruppengröße k	2	3	4	5	6	7	8

So sind die typischen Trennungen bei Telefonnummern z. B. 289 517 091, bei Bankkonten als
IBAN z. B. DE77 1018 0000 0916 6311 03 usw. gut erklärbar.
Ähnliches gilt auch für Wörter
Auch deshalb sind Silbenwiederholungen (und Ohrwürmer) oft so einprägsam.

Thermodynamische kontra Shannon-Entropie

In der Literatur werden häufig – selbst von hervorragenden Wissenschaftlern – die thermodynamische und Shannon-Entropie fälschlich in (sehr) enge Beziehung zueinander gebracht. Dazu die Formel-Varianten

Daten-Art	Shannon-Entropien	Thermodynamische Entropien
diskret/digital	$H = -\sum_{1=1}^{n} p_i \cdot ld(p_i)$	$S = k \cdot \ln(W)$
kontinuierlich	$H_K = \mathrm{ld}(n_{AS})$ bzw. $h_{Nutz}(x) = -\int_{-\infty}^{+\infty} p(x) \cdot \mathrm{ld}(p(x)) \cdot dx$	$\Delta S = \Delta Q/T$

Die Zeilen – kontinuierlich und diskret/digital – sind dabei allerdings nur z. T. korrekt.
Bei der Boltzmann-Entropie ist die Wahrscheinlichkeit *W* kontinuierlich.
Sie gilt immer nur für einen ***einzelnen Zustand*** bestimmt, bei Shannon aber für eine **Gesamtheit**.
Die Clausius-Entropie ist bei allen bekannten Vergleichen nicht einbezogen
Sie ermöglicht offensichtlich auf keine Weise einen sinnvollen Vergleich mit der Shannon-Entropie.

Gemeinsamkeiten

Trotz aller Unterschiede gibt es fünf Gemeinsamkeiten:

1. Beide sind Grundlage und Ursache von wissenschaftlich-technischen Revolutionen.
   Die Thermodynamik (Dampfmaschine) für die erste industrielle Revolution.
   Die Shannon-Entropie für die Informations-, Nachrichten- und Computer-Technik.
2. Beide nutzen den Logarithmus von Wahrscheinlichkeiten.
3. Beide verweisen auf Austauschmöglichkeiten von einigen Parametern.
4. Es existieren teilweise mehrere Beschreibungen (Formeln) und Inhalte gemäß den beiden obigen Zeilen.
5. Beide Entropien betreffen Theoriegrenzen, die verallgemeinert als Wirkungsgrad interpretiert werden können.
   Dabei ist zu beachten, dass Wirkungsgrade oft recht schwierig zu bestimmen sind, denn sie erfordern:
   a) die entscheidenden Parameter zu finden (z. B. Temperaturen. Wahrscheinlichkeiten),
   b) die genaue Kenntnis der typischen Grundlagen (Kreisprozess).
   c) eine Beschreibung (Formel) der Zusammenhänge so, dass eine Berechnung möglich ist.