It from bit: the science of information

Richard Harvey
IT Livery Company Professor of Information Technology, Gresham College

Professor of Computer Science, School of Computing Sciences,
University of East Anglia

#richardwharvey

information | ɪnfəˈmeɪʃ(ə)n |

noun *[mass noun]*

1 facts provided or learned about something or someone: *a vital piece of information*.

[count noun] Law a charge lodged with a magistrates' court: *the tenant may* ***lay an information against*** *his landlord*.

2 what is conveyed or represented by a particular arrangement or sequence of things: *genetically transmitted information*.

Computing data as processed, stored, or transmitted by a computer

(in information theory) a mathematical quantity expressing the probability of occurrence of a particular sequence of symbols, impulses, etc., as against that of alternative sequences.

ORIGIN late Middle English (also in the sense ‘formation of the mind, teaching’), via Old French from Latin ***informatio(n-***), from the verb ***informare***

information | ɪnfəˈmeɪʃ(ə)n |

noun *[mass noun]*

1 facts provided or learned about something or someone: *a vital piece of information*.

[count noun] Law a charge lodged with a magistrates' court: *the tenant may **lay an information against** his landlord*.

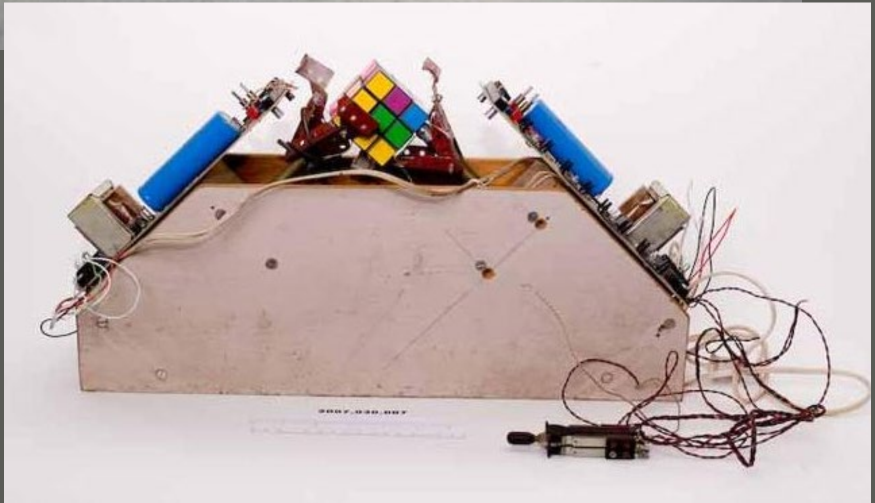
2 what is conveyed or represented by a particular arrangement or sequence of things: *genetically transmitted information*.

Computing data as processed, stored, or transmitted by a computer

(in information theory) a mathematical quantity expressing the probability of occurrence of a particular sequence of symbols, impulses, etc., as against that of alternative sequences.

ORIGIN late Middle English (also in the sense ‘formation of the mind, teaching’), via Old French from Latin ***informatio(n-***), from the verb ***informare***

Shannon information



Artefacts from the MIT Museum
https://mitmuseum.mit.edu

Shannon information

The Bell System Technical Journal

Vol. XXVII *July, 1948* *No. 3*

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.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance

[1] Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A. I. E. E. Trans.*, v. 47, April 1928, p. 617.

[2] Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal*, July 1928, p. 535.

A big idea from Claude Shannon

would produce "messages" consisting of a number of functions of three variables; (f) Various combinations also occur, for example in television with an associated audio channel.

2. A *transmitter* which operates on the message in some way to produce a signal suitable for transmission over the channel. In telephony this operation consists merely of changing sound pressure into a proportional electrical current. In telegraphy we have an encoding operation which produces a sequence of dots, dashes and spaces on the channel corresponding to the message. In a multiplex PCM system the different speech functions must be sampled, compressed, quantized and encoded, and finally interleaved

C E Shannon, "A Mathematical Theory of Communication", Vol XXVII, July 1948, No 3, Bell System Technical Journal.

Encoding text

In economics, Gresham's law is a monetary principle stating that "bad money drives out good". For example, if there are two forms of commodity money in circulation, which are accepted by law as having similar face value, the more valuable commodity will gradually disappear from circulation.

The law was named in 1860 by Henry Dunning Macleod, after Sir Thomas Gresham (1519–1579), who was an English financier during the Tudor dynasty. However, there are numerous predecessors. The law had been state…

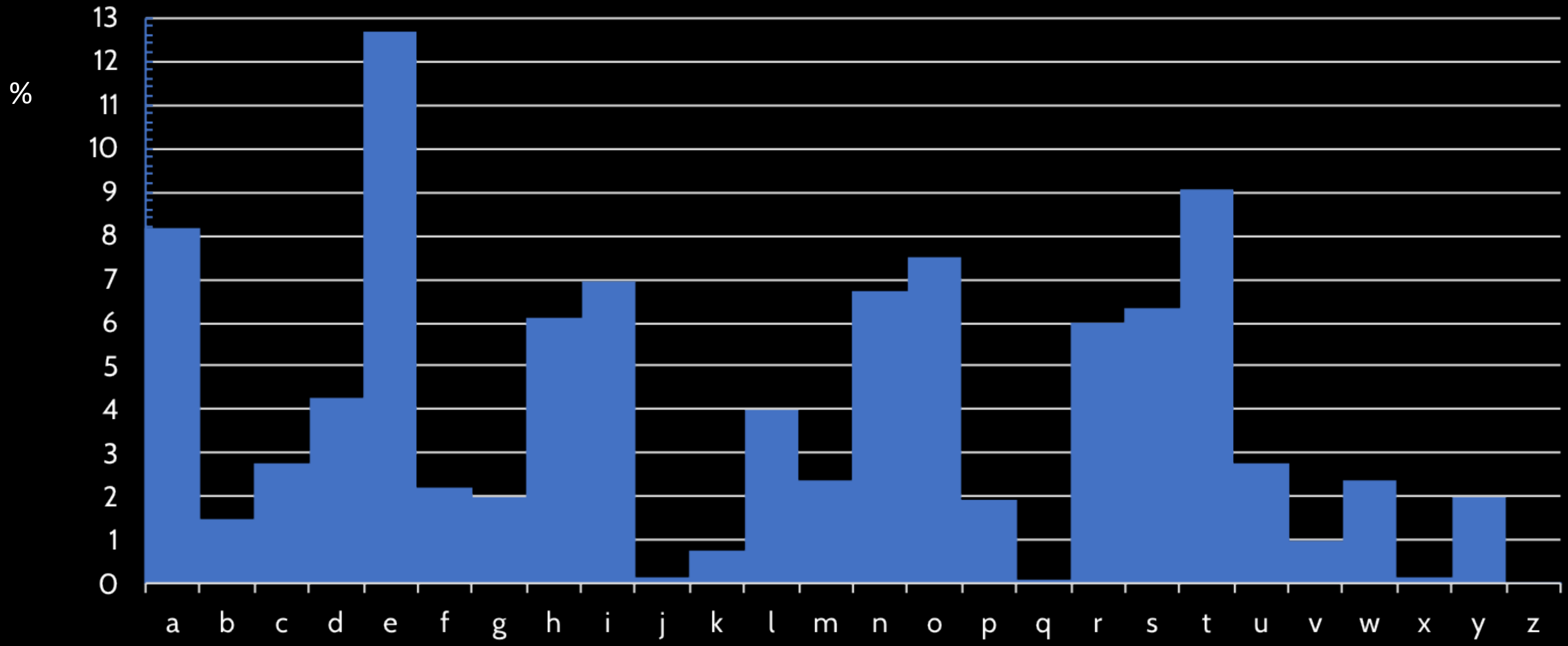
Wiikipedia entry on “Gresham’s Law”

How to code this text?

⋮								
A	0	1	0	0	0	0	0	1
B	0	1	0	0	0	0	1	0
C	0	1	0	0	0	0	1	1
D	0	1	0	0	0	1	0	0
E	0	1	0	0	0	1	0	1
F	0	1	0	0	0	1	1	0
⋮								

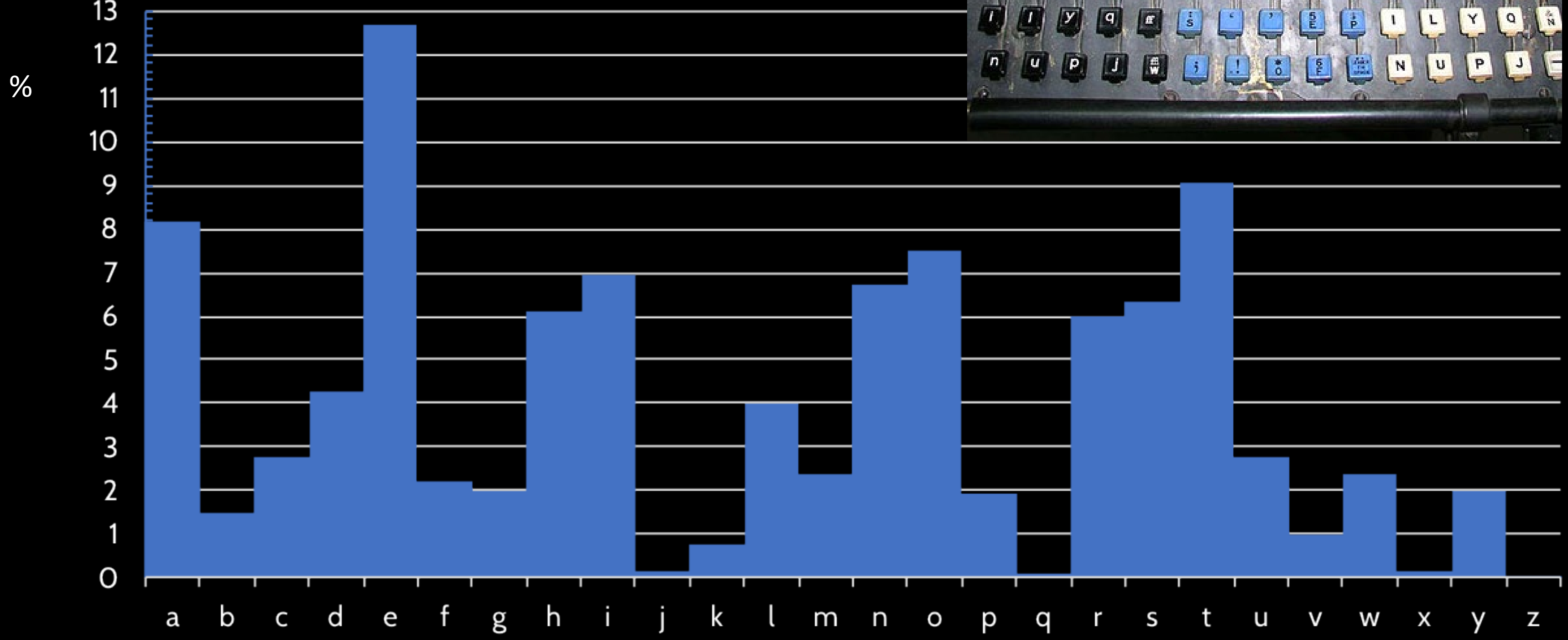
500 characters $\times$ 8 bits per character = 4000 bits

But some letters are more common than others



If Youth, throughout all history, had had a champion to stand up for it; to show a doubting world that a child can think; and, possibly, do it practically; you wouldn't constantly run across folks today who claim that "a child don't know anything." A child's brain starts functioning at birth; and has, amongst its many infant convolutions, thousands of dormant atoms, into which God has put a mystic possibility for noticing an adult's act, and figuring out its purport. …

Gadsby by Ernest Vincent Wright


%
13
12
11
10
9
8
7
6
5
4
3
2
1
0
a b c d e f g h i j k l m n o p q r s t u v w x y z

So what about a different code?

Maybe we could use fewer bits for the most common letters?

⋮							
A	1	0	0	0			L_A =4
B	1	1	1	1	0	0	L_B =6
C	0	1	1	0	1		L_A =5
D	0	0	0	0	0		L_A =5
E	1	1	0				L_A =3
F	0	1	0	1	1		L_A =5
⋮							

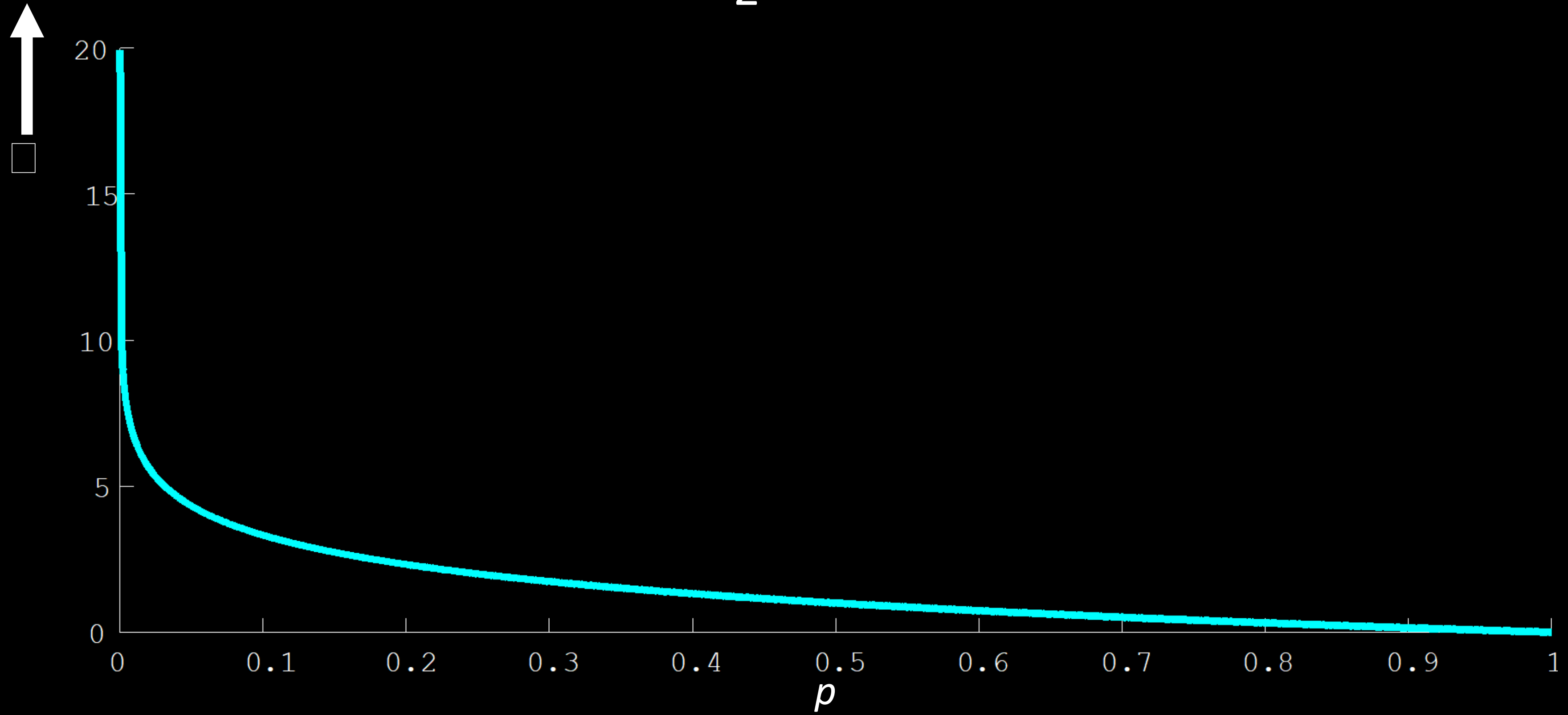
Average length of this code is

$p_A L_A + p_B L_B + \ldots + p_Z L_z$

= 4.2 bits

Roughly half the size of ASCII

Information $I = -\log_2 p$



So what's the best we could do?

Shannon Entropy

$H = -\sum_{n=1}^{N} p_n \log_2 p_n$ = Average($-\log_2 p$)

For our probabilities:

H = 4.176 bits per character.

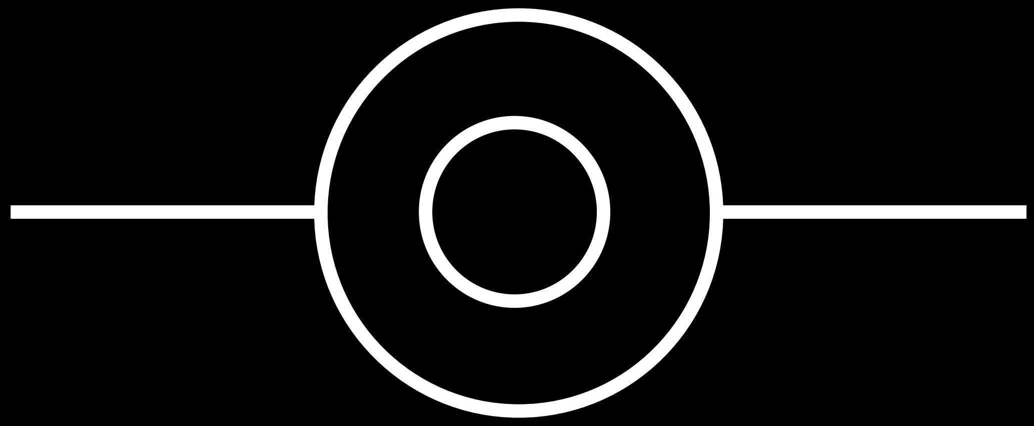
Improbability =
surprise = information = bits

Information (and surprise) is conditional

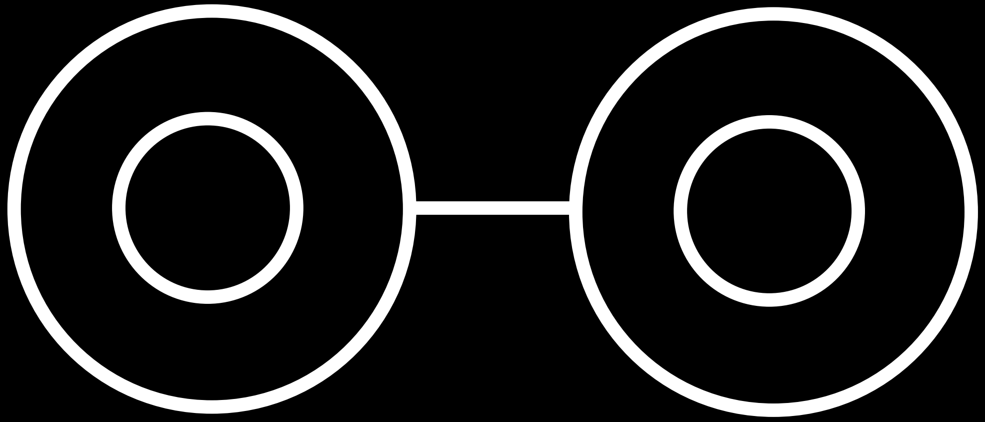
Given that I know X, how much extra information is needed to describe Y?

Information is conditional

Surprise is conditional



Surprise is conditional



Magic compression

A *magic* or universal compressor would be able to compress every file…

If such a thing existed then it would be able to compress files that themselves were compressed files

So the universal compressor does not exist!

Lossless compression

http://squeezechart.com

Video data

1080 lines × 1920 columns = 2M pixels

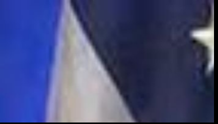
2M pixels × 3 colours × 8 bits = 49.8 Mbits per frame

49.8 Mbits per frame × 25 frames per second = 1.2 Gbits per second

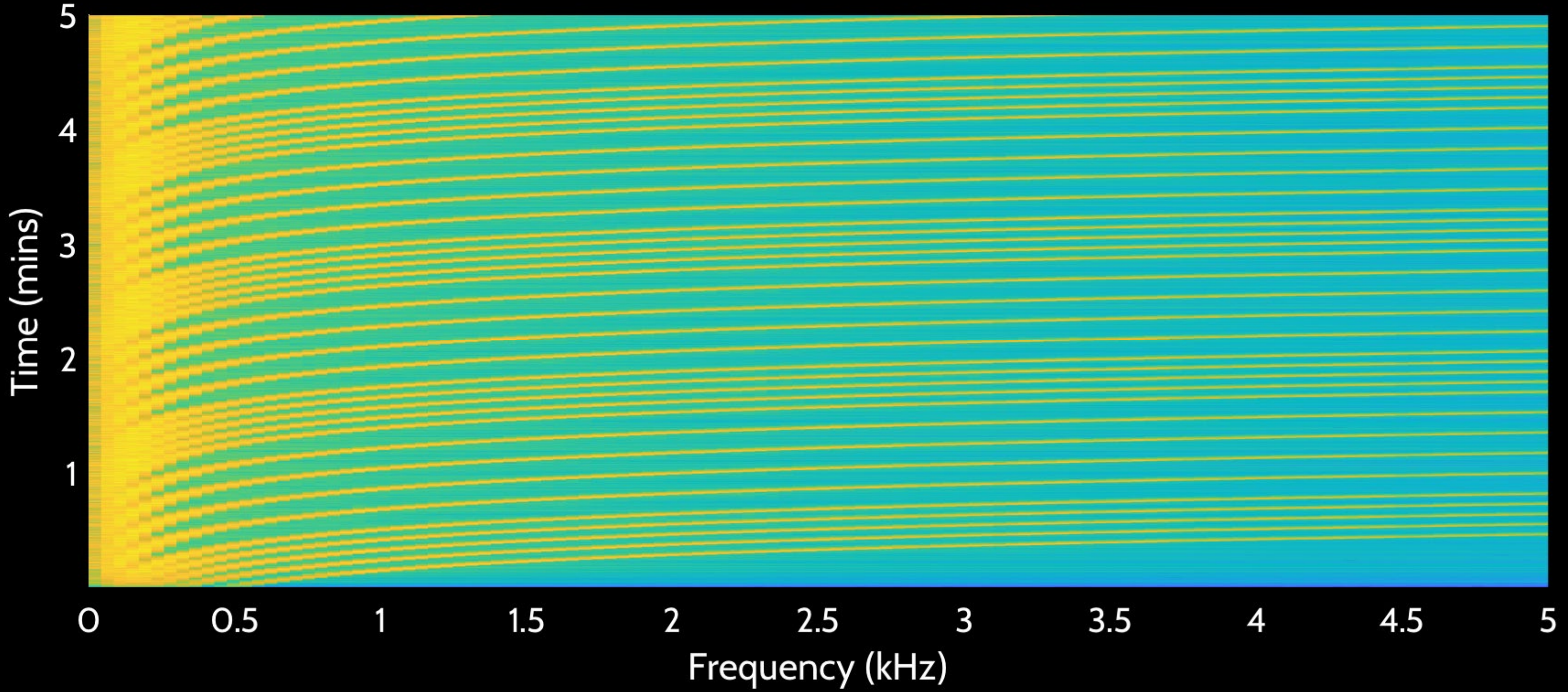
But DVB channels range from 5 Mbits per second to 50 Mbits per second.

Compression ratios of 250 to 25 needed.


USA TODAY
NETWORK



Sound is not what meets the ear



For example: https://www.youtube.com/watch?v=BzNzgsAE4F0

I wonder if this is encoded for humans?

By Francis Barraud

1415926535…

$I = -\log_2 (1/10) = 3.3$ bits per symbol

10 symbols = 33 bits

1415926535897932384626433832795028841971693993751058209749445923078164062862089986280348253421170 67…

100 symbols = 330 bits

14159265358979323846264338327950288419716939937510582097494459230781640628
62089986280348253421170679821480865132823066470938446095505822317253594081
28481117450284102701938521105559644622948954930381964428810975665933446128
47564823378678316527120190914564856692346034861045432664821339360726024914
12737245870066063155881748815209209628292540917153643678925903600113305305
48820466521384146951941511609433057270365759591953092186117381932611793105
11854807446237996274956735188575272489122793818301194912983367336244065664
30860213949463952247371907021798609437027705392171762931767523846748184676
69405132000568127145263560827785771342757789609173637178721468440901224953
43014654958537105079227968925892354201995611212902196086403441815981362977
47713099605187072113499999983729780499510597317328160963185950244594553469
08302642522308253344685035261931188171010003137838752886587533208381420617
17766914730359825349042875546873115956286388235378759375195778185778053217
1226…

1000 symbols = 3300 bits

…and so on…

1 million symbols = 3.3 Mbits

…but…

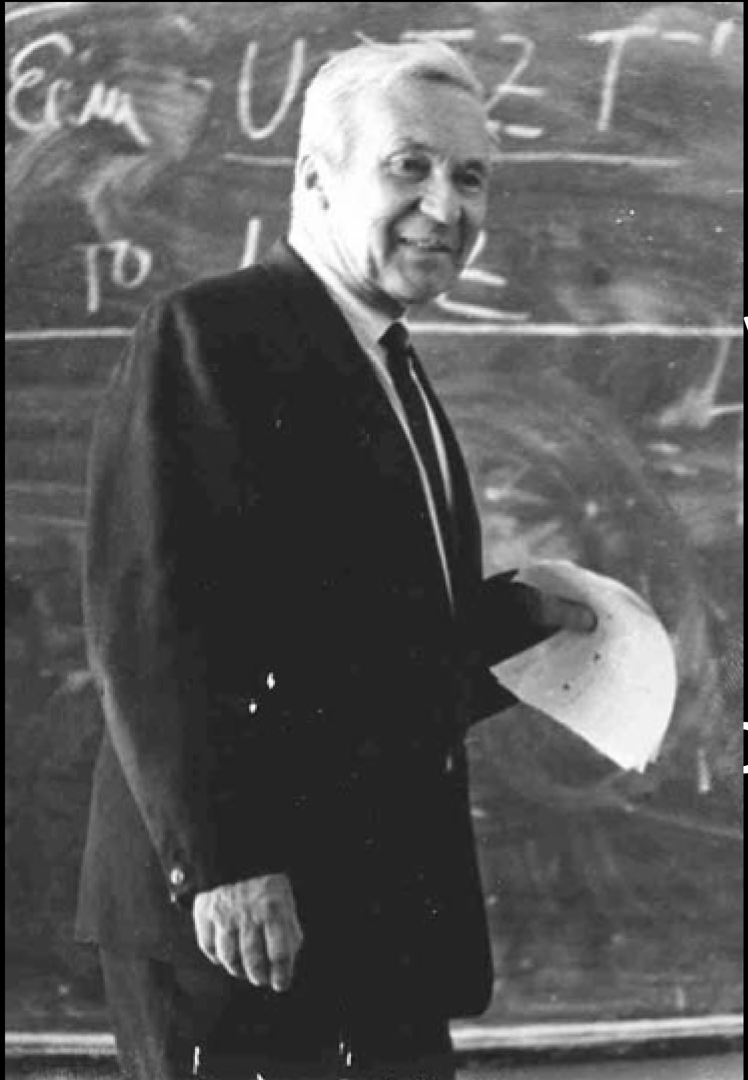
1415926535…

are the first 10 digits of…

3.1415926535… = π

KOLMOGOROV INFORMATION

The length of the shortest program that can compute the object



V INFORMATION

the shortest program
oute the object

160 character program for π :

```
int a=10000,b,c=2800,d,e,f[2801],g;
main(){for(;b-c;)f[b++]=a/5;
     for(;d=0,g=c*2;c-
=14,printf("%.4d",e+d/a),e=d%a)for(
b=c;d+=f[b]*a,
     f[b]=d%--g,d/=g--,--b;d*=b);}
```

π can be computed with a 160 character (1280 bit) program

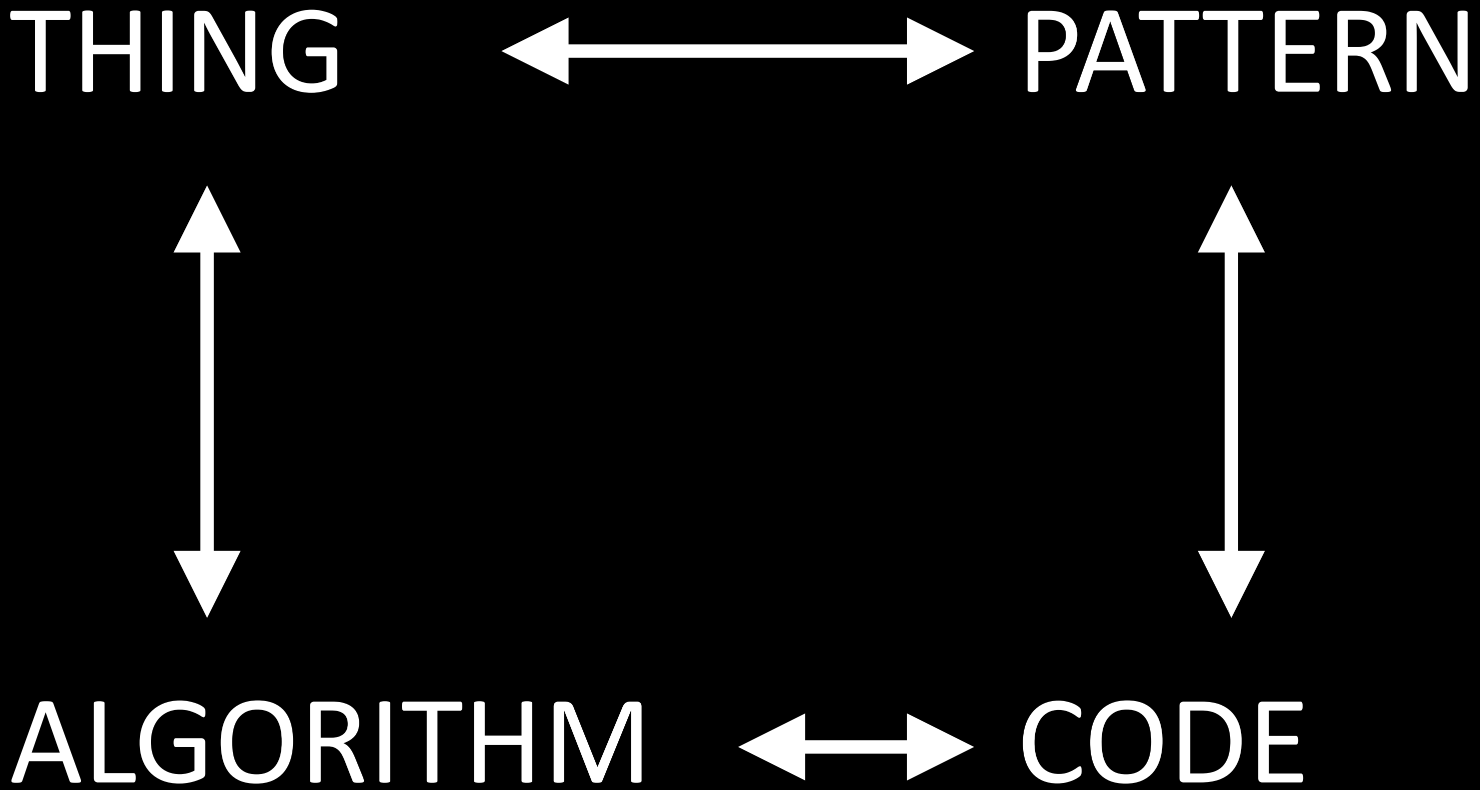
so the information in π is < 1280 bits

KOLMOGOROV INFORMATION IS CONDITIONAL TOO

$K(x|y)$ is the size of the shortest program that computes x given y as input

A string is incompressible if K(x) = Length(x)

INFORMATION = SIZE OF MACHINE THAT CREATES IT


THING
PATTERN
ALGORITHM
CODE

OLD SCIENCE = WHAT IS THAT THING?

NEW SCIENCE = WHAT CREATES THAT THING?

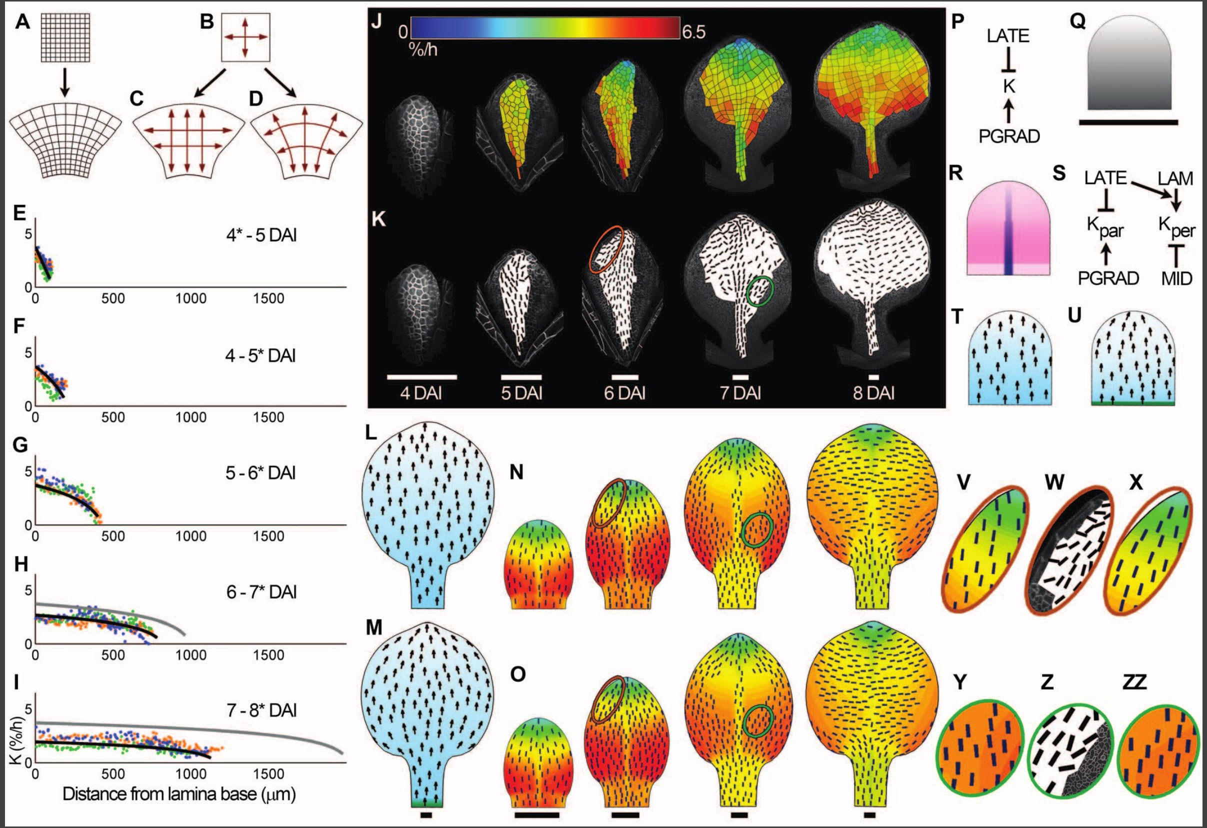
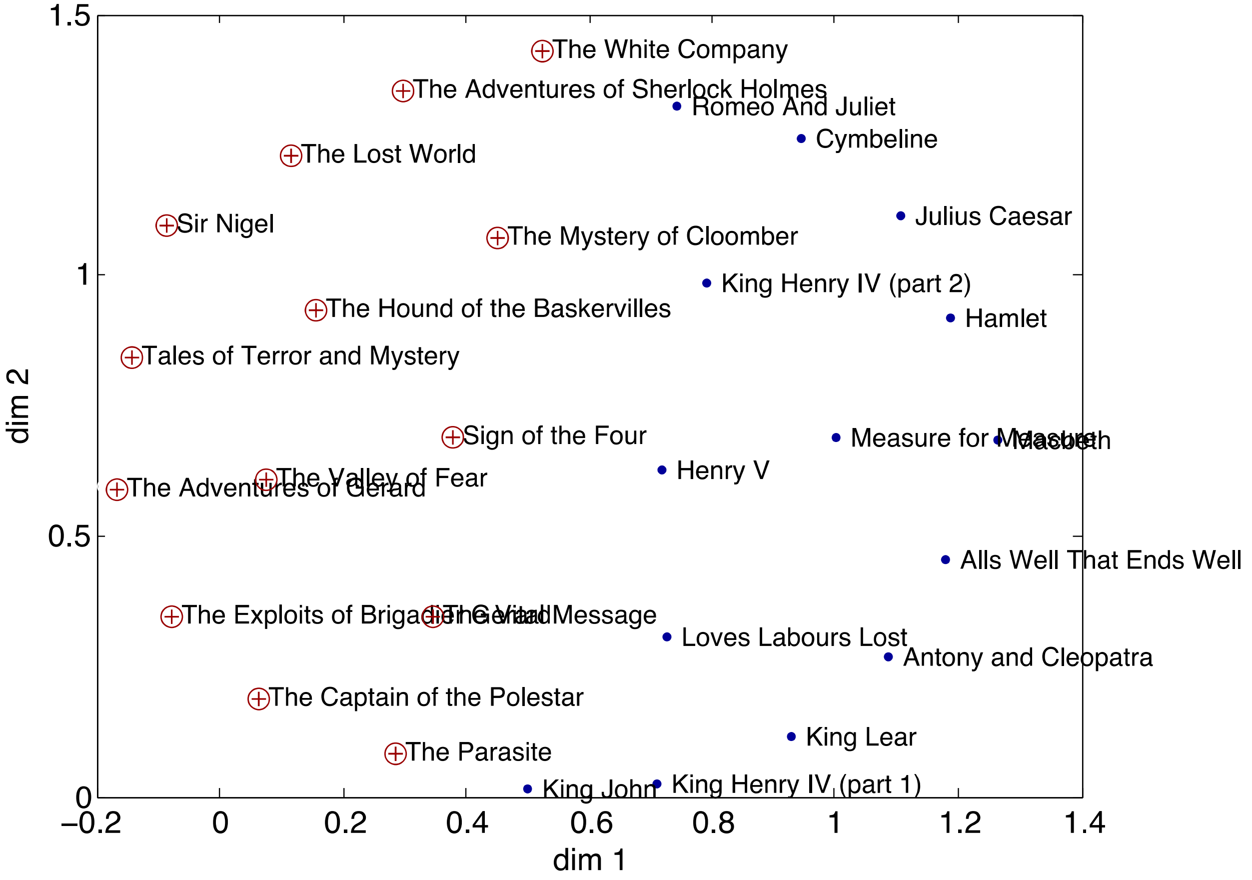


Fig. 1. Leaf growth analysis. (**A**) Tissue deforms through growth. (**B**) Orthogonal organizing system which (**C**) retains its original arrangement or (**D**) deforms during growth. (**E** to **I**) Midline proximodistal growth rates for three replicates (orange, green, and blue), and 1D models (black and gray lines). Distances from lamina base correspond to those on the day indicated by an asterisk. (**J**) Areal growth rates (heat map) and (**K**) principal directions of growth (black lines, where anisotropy > 10%) at the end of each period. (**L**) Resultant shape, POL levels and specified growth orientations (arrows) for nondeforming and (**M**) deforming (organizer-based) models. (**N**) Resultant shapes, areal growth rates, and directions of growth (black lines, where anisotropy > 5%) for 2D nondeforming and (**O**) deforming (organizer-based) models. Heat map and staging as in (J). (**P**) 1D model regulatory network. (**Q**) 2D distribution of PGRAD (gray). (**R**) MID (blue) and LAM (magenta) distributions. (**S**) 2D model regulatory network. (**T**) Initial POL (cyan) distribution for nondeforming and (**U**) deforming models. PROXORG in green. (**V** to **X**) Enlargement of brown ellipses in (N) (K), and (O), respectively. (**Y**, **Z**, and **ZZ**) Enlargement of green ellipses in (N), (K), and (O), respectively. Scale bars, 100 µm.

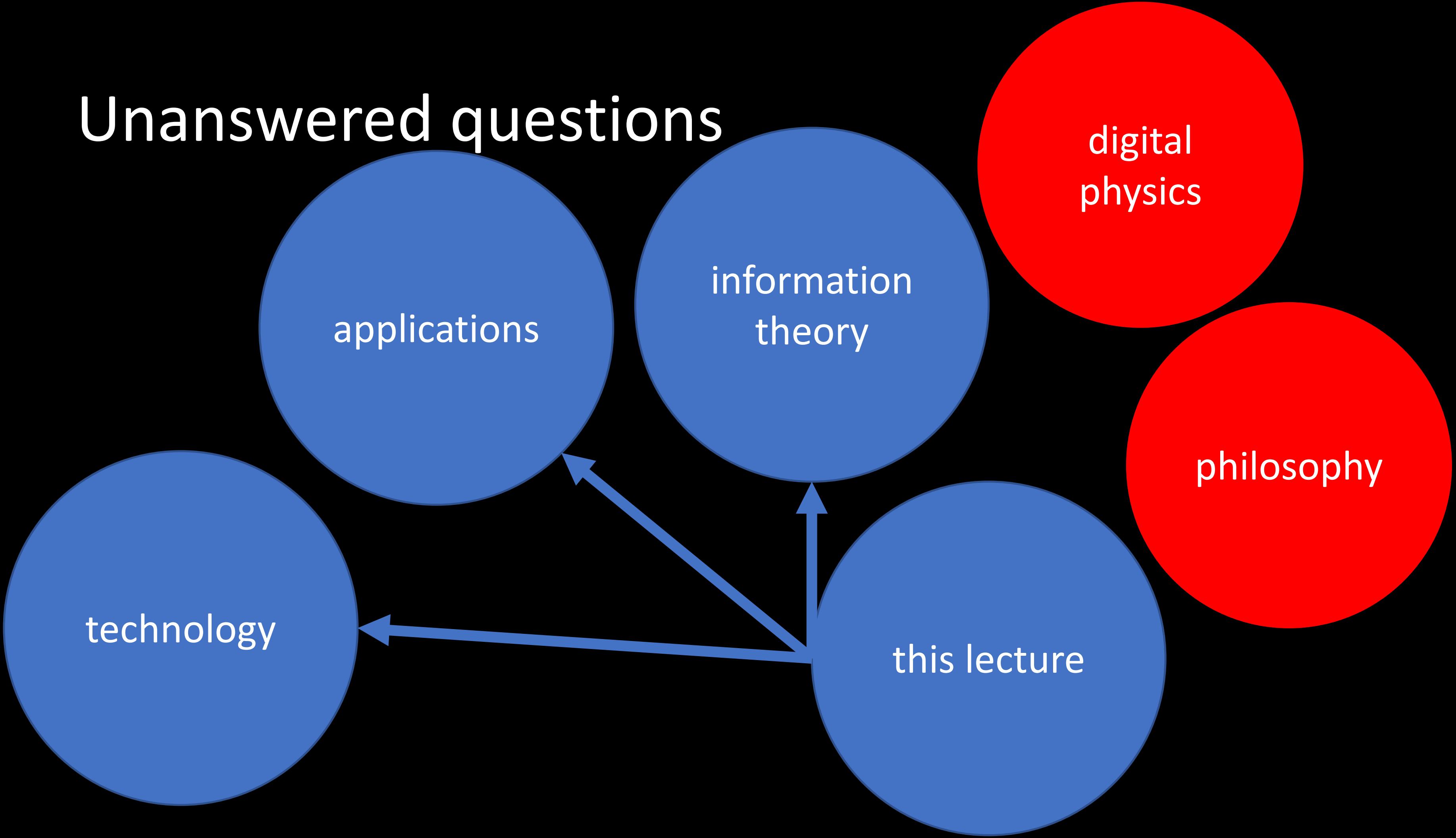
"Generation of Leaf Shape Through Early Patterns of Growth and Tissue Polarity",

Erika E. Kuchen, Samantha Fox, Pierre Barbier de Reuille, Richard Kennaway, Sandra Bensmihen, Jerome Avondo, Grant M. Calder, Paul Southam, Sarah Robinson, Andrew Bangham, Enrico Coen, Science, 335, 2 March 2012


The White Company
The Adventures of Sherlock Holmes
Romeo And Juliet
Cymbeline
The Lost World
Sir Nigel
Julius Caesar
The Mystery of Cloomber
King Henry IV (part 2)
The Hound of the Baskervilles
Hamlet
Tales of Terror and Mystery
Sign of the Four
Measure for Measure
Macbeth
Henry V
The Valley of Fear
The Adventures of Gerard
Alls Well That Ends Well
The Exploits of Brigadier Gerard
The Vital Message
Loves Labours Lost
Antony and Cleopatra
The Captain of the Polestar
King Lear
The Parasite
King John
King Henry IV (part 1)
dim 1
dim 2
1.5
1
0.5
0
−0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4

IT FROM BIT?

Unanswered questions



Moving onwards:

- speech (27th Nov)
- vision (12th Feb)
- learning (19th March)
- text (16th April)
- creativity (28th May)