$$H(X) = \sum_{i=1}^{n} p(x_i)\, log_2({}^{1}/_{p(x_i)})$$

Information Theory and Visualization

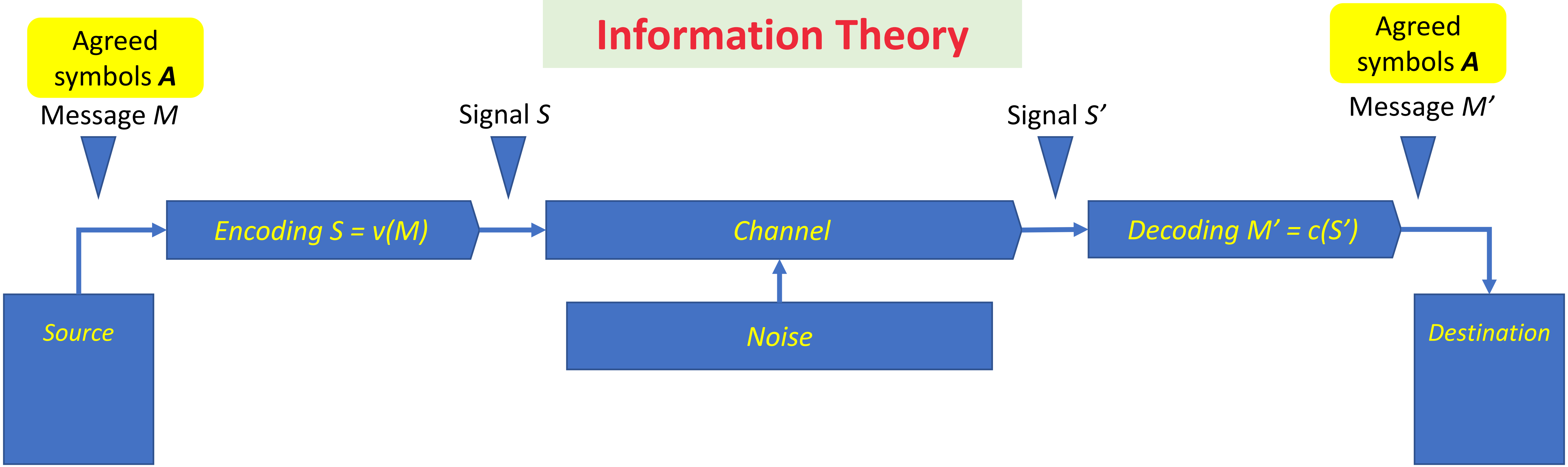
Information Theory

In 1948 Claude Shannon wrote, what has become, the definitive paper on Information Theory.

"A Mathematical Theory of Communication",
The Bell System Technical Journal, 1948.

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.

Information Theory

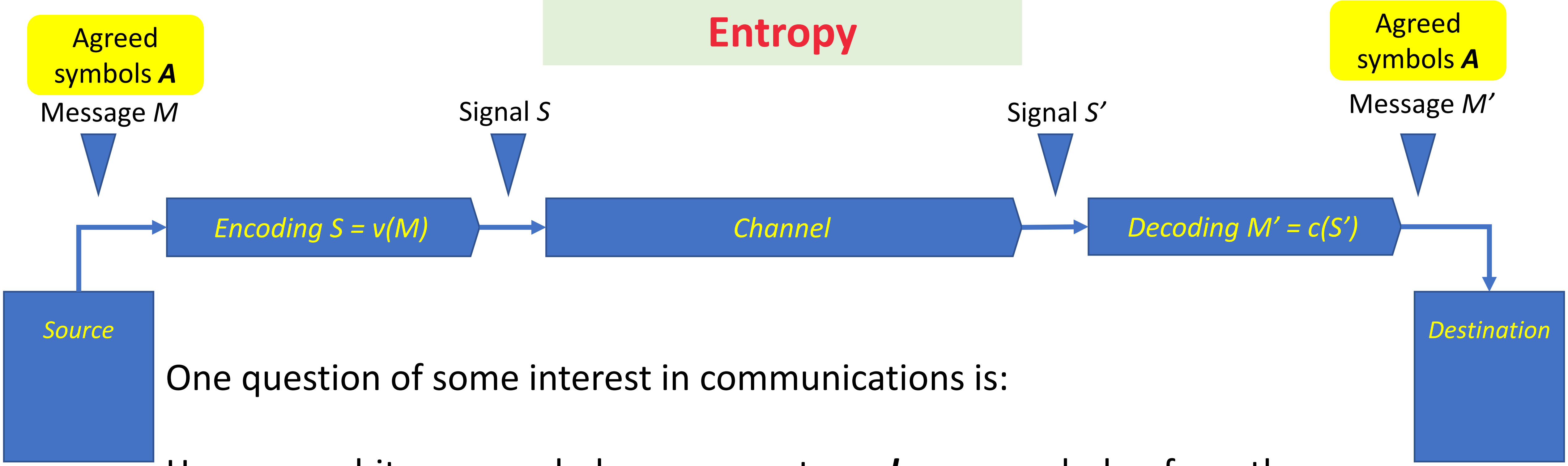


Shannon abstracts the content of a signal entirely away from the semantics of the message.

He defines the role of a communication system simply as
reproducing a message exactly, sent from the source at the destination.

An alphabet ***A***, is the pre-agreed set of discrete symbols that may be in the message.
A random variable ***X*** selects symbols x_i from the alphabet to form the current message ***M***

Entropy



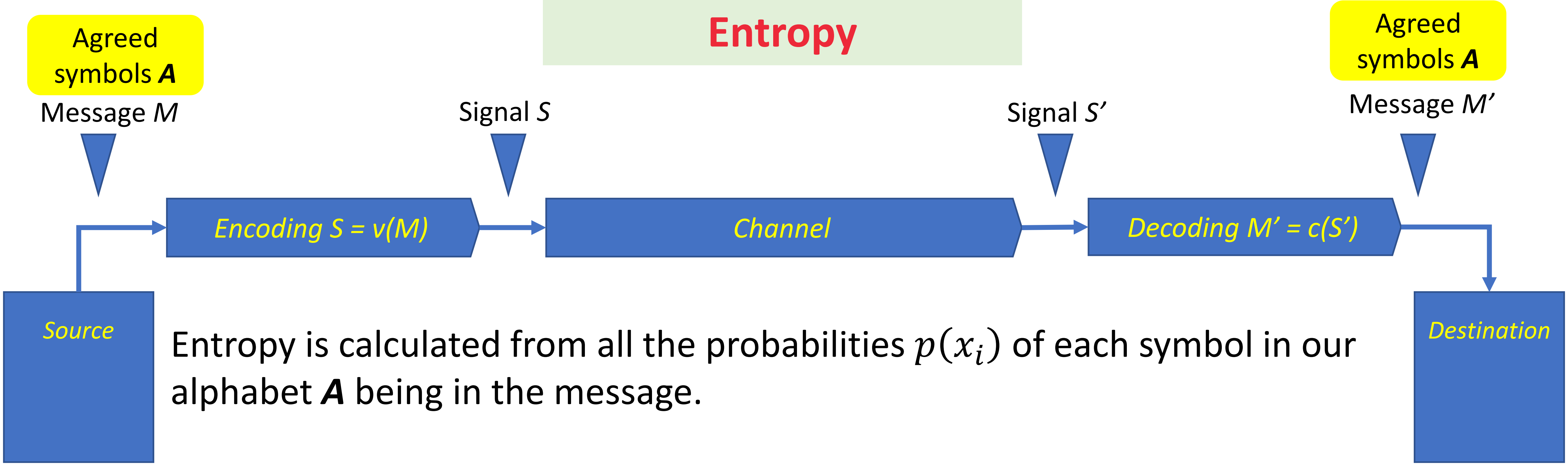
One question of some interest in communications is:

How many bits are needed on average to ***code*** one symbol x_i from the message ***M*** in the signal ***S ?***

Entropy $\boldsymbol{H}$ gives an answer, among other things it can be thought of as a measure of how *incompressible* a message is.

See John Stone's book, Chapters 1 and 2 for a very clear introduction.

Entropy

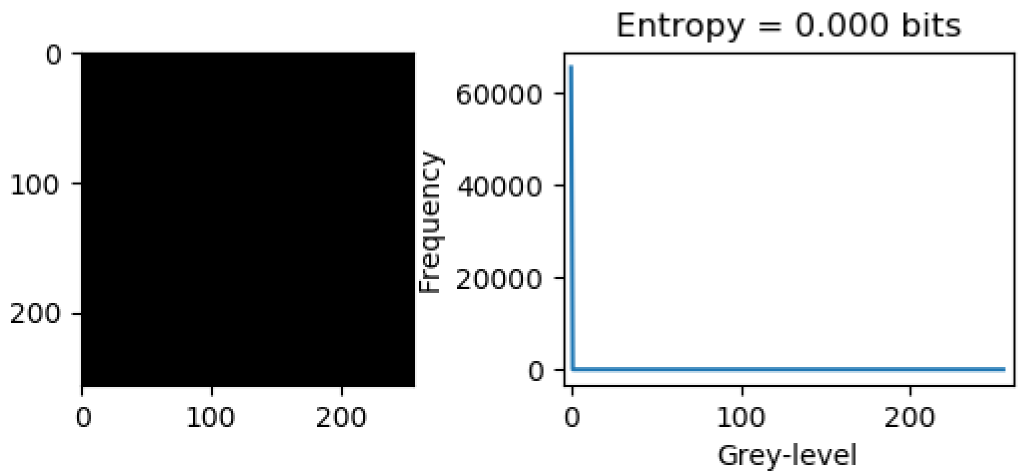


Entropy is calculated from all the probabilities $p(x_i)$ of each symbol in our alphabet ***A*** being in the message.

$$\boldsymbol{H(X) = \sum_{i=1}^{n} p(x_i)\, log_2 ({}^{1}/_{p(x_i)})}$$

: $\boldsymbol{H(X)}$ measures uncertainty (which is also information content)
smooth messages have low uncertainty & can be coded efficiently,
it also helps capture the idea of localization or coherence that many efficient algorithms rely on.

Entropy for an Image



Message ***M*** is the array of pixels in the image, 256x256 = 65,536 pixels.

Alphabet ***A*** is all possible values which can be stored in 8 bits per pixel, which is 0-255.

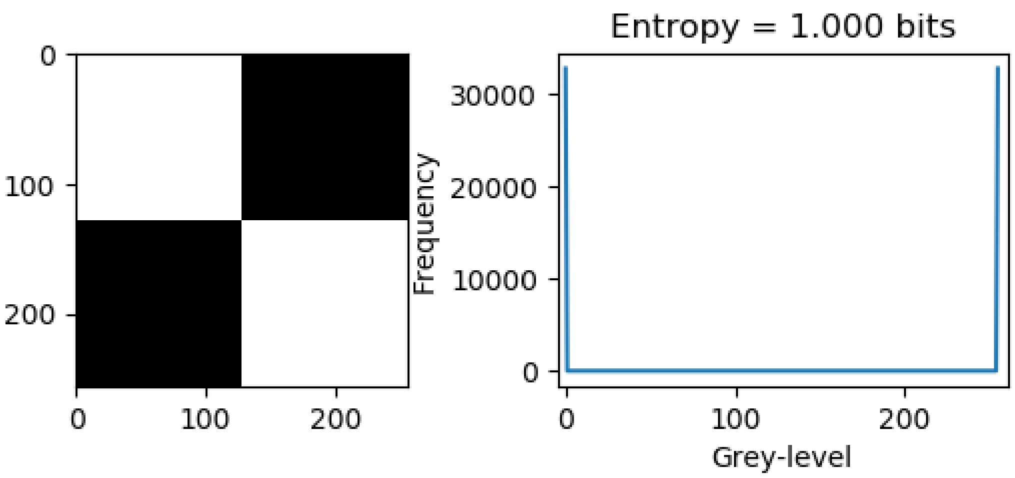
We need to know the probability of each symbol in the alphabet being in the message,

We use a “training” image to do this and calculate the frequency histogram above right.
From this histogram we find the probability of 0 (black) is 1, everything else has 0 probability.

Calculating entropy ***H*** for this image using these probabilities results in ***H =*** 0 bits.

This is telling us if we will only ever send black pixels we need not send anything as the shared alphabet (with probabilities) will tell us we will only ever receive black pixels.

Entropy for a Checkerboard Image



Message ***M*** is the array of pixels in the image, 256x256 = 65,536 pixels.

Alphabet ***A*** is all possible values which can be stored in 8 bits per pixel, which is 0-255.

Entropy is an upper limit on the average number of bits we need to code a message symbol.

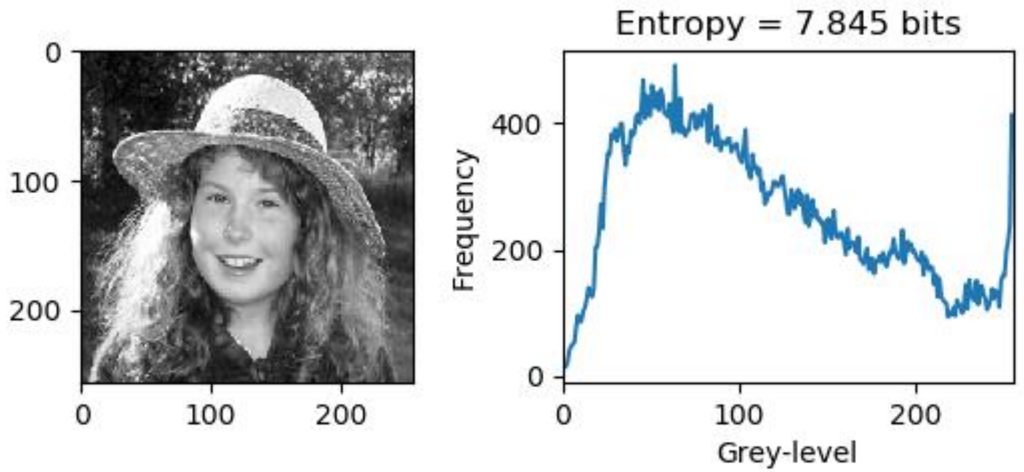
In image coding we can often do (much) better.

Again, we calculate entropy *just for this image* :

50% of the pixels are 255 (white) and 50% are 0 (black)
pixels can only have one of two values
calculating the entropy it is ***H*** = 1 bit

This would result in coding savings of at least 87.5% saving over the raw message data.

Entropy for a Natural Image



Message ***M*** is the array of pixels in the image, 256x256 = 65,536 pixels.

Alphabet ***A*** is all possible values which can be stored in 8 bits per pixel, which is 0-255.

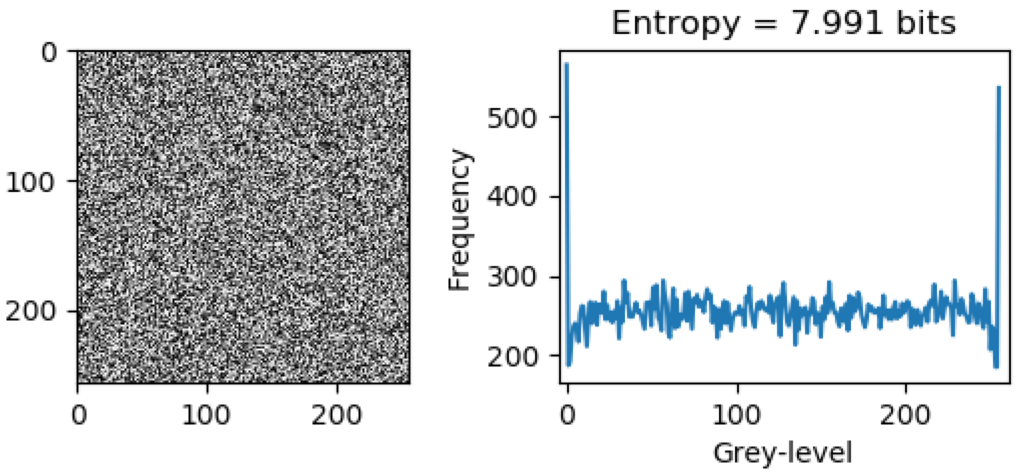
Again, we calculate entropy *just for this image* :

The frequencies of each Grey level vary and the resulting Entropy value gives us a guide to how many bits per pixel on average we need to code this image, here ***H*** = 7.845

H is just less than the 8 bits per pixel we need for 256 grey levels, and shows we could save at least 2% of the total by recoding the image grey levels in the signal.

Another way to look at this is that this image has a lot of information in it (high entropy).

Entropy for an Image of White Noise



Message ***M*** is the array of pixels in the image, 256x256 = 65,536 pixels.

Alphabet ***A*** is all possible values which can be stored in 8 bits per pixel, which is 0-255.

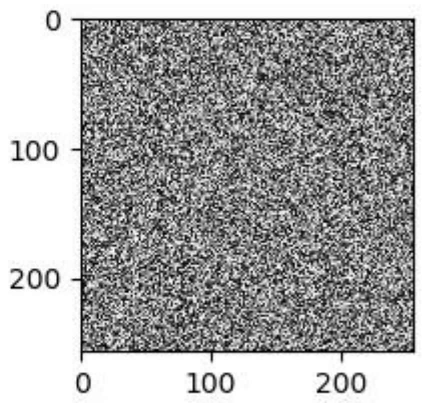
Again, we calculate entropy *just for this image* :

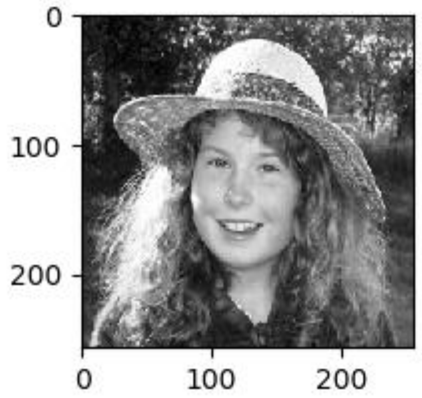
The frequencies of each Grey level vary and the Entropy value gives us a guide to how many bits per pixel we need to code this image.

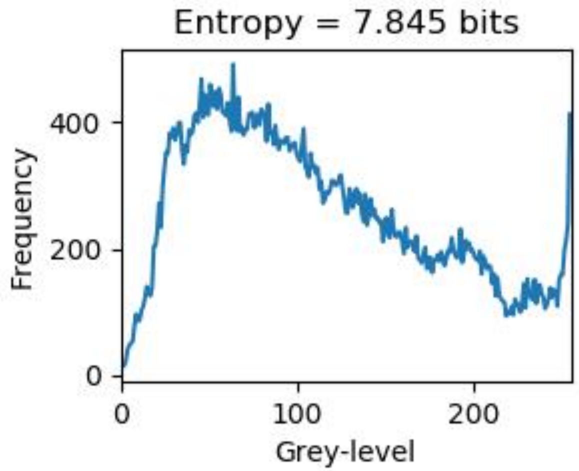
This is almost 8 bits per pixel it is uncompressible and contains almost the most Entropy (information) possible in this one image ***H*** = 7.991 bits

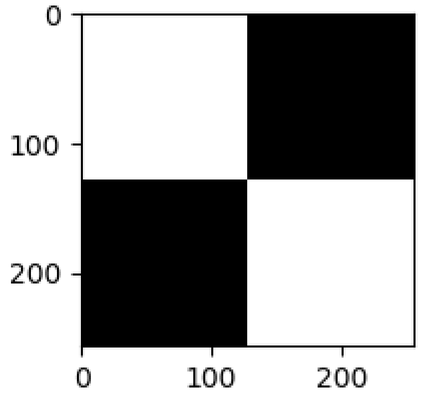
It might seem odd that this random image contains information, but imagine each of the non-zero pixels were stars, it would be very many stars represented in one image.

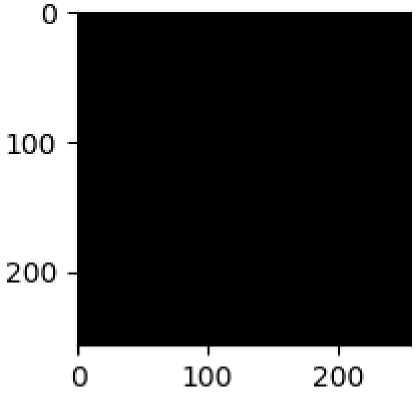
Entropy Coding for General Images











If we wanted to use lossless Entropy Coding to code images in general (not just one image) we would need to collect a *training set* of many images and average the probabilities from each one for each symbol in our alphabet ***A*** (the grey values from 0-255).

However, better coding methods have been devised e.g. difference coding that uses redundancy (coherence) to create more efficient alphabets for images.

Also remember we are showing humans the images, they have limited perceptual abilities and we can exploit that in lossy coding schemes (JPEG does this very well).

Entropy – a summary

Entropy as a concept underlies all Data Science, you can see it as:

a measure of information (uncertainty) in a signal, with no concern for semantics.

on average how many bits it will take to losslessly code a symbol into a signal for transmission, given a pre-agreed alphabet of symbols and their probabilities.

a metric for comparing two signals (using same ***A***) based on information content, *Approximate Entropy* and *Sample Entropy* – used in ECG trace comparisons.

Practically a signal might be a csv file, an image, a book, a sound, or any other data.

Entropy ***H*** is a concept central to the development of two of Shanon's key theorums:

The source coding theorem can't compress lossless-ly any better than ***H***

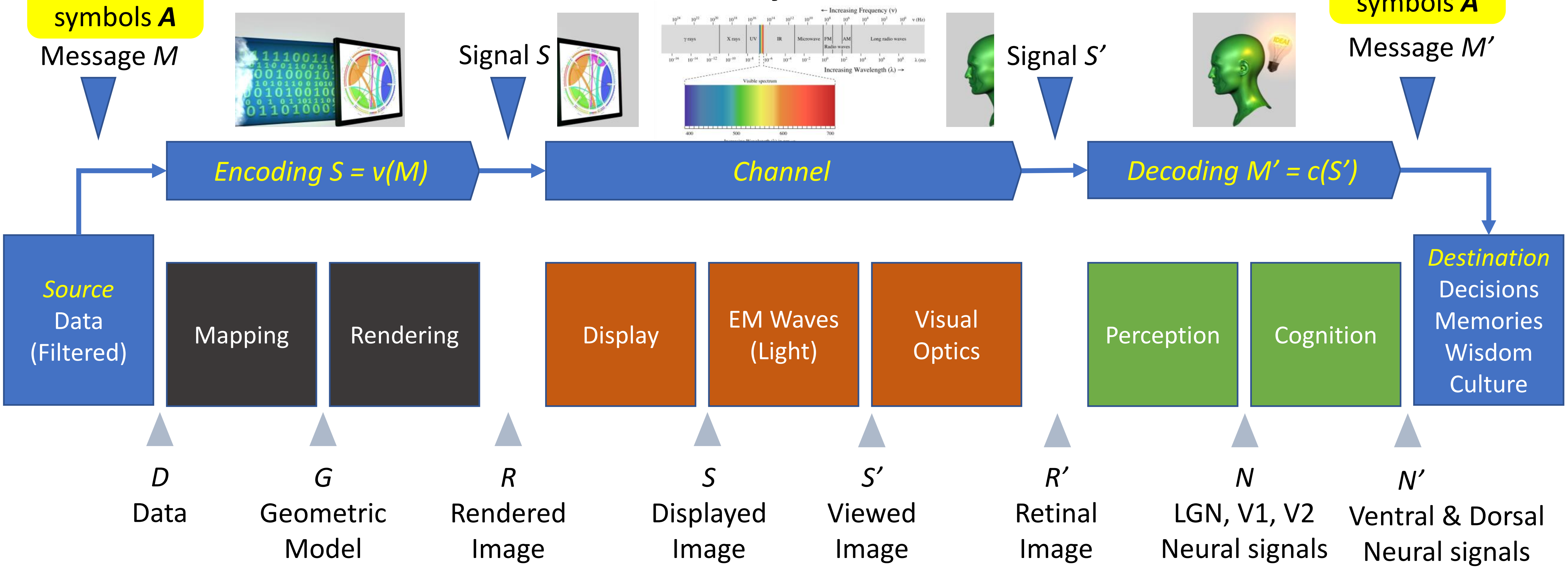
The noisy channel coding theorem defines coding bounds for a given error rate.

But we won't consider these further here.

Keep in mind: entropy does not measure ideas or semantics. https://youtu.be/2s3aJfRr9gE

What is Visualization?

Information Theory and Visualization



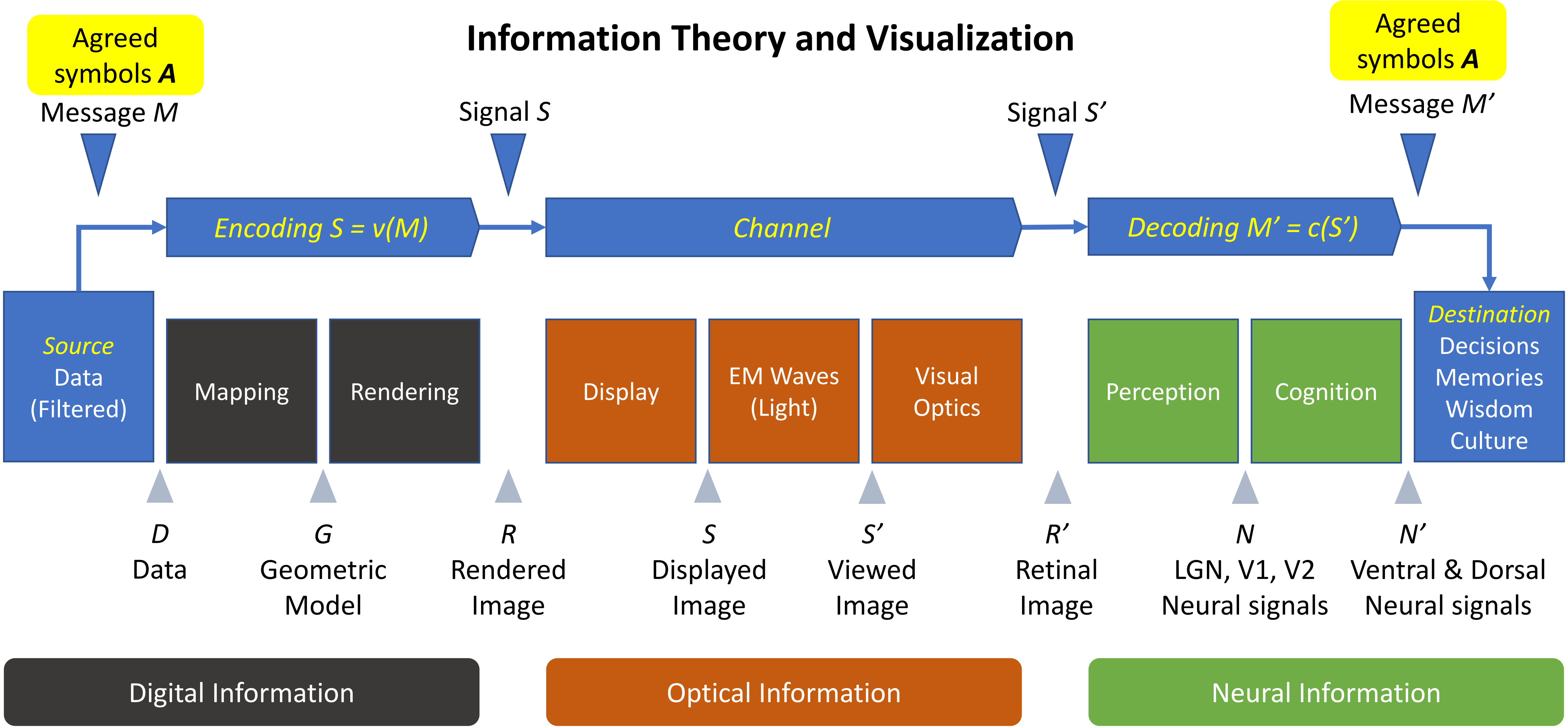
Each stage of the visualization pipeline from data to human cognition can be mapped onto the inf. theoretic model.

Two key functions here are:

the visualization encoding function — *v* where $S = v(M)$

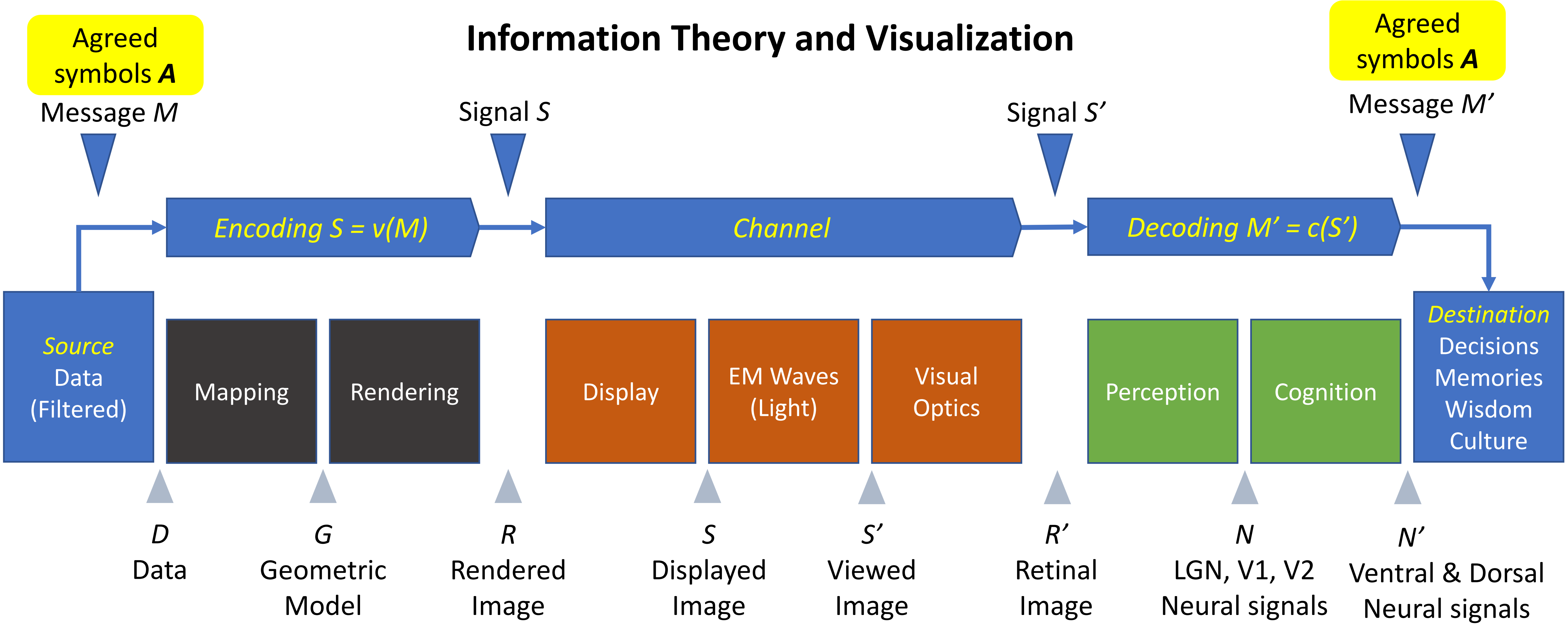
the human decoding function — *c* where $M' = c(S')$

Information Theory and Visualization



We can observe that the physical representation of information in this pipeline changes, each of these changes in representation may add unique forms of noise to the signal and change the message we intended to send.

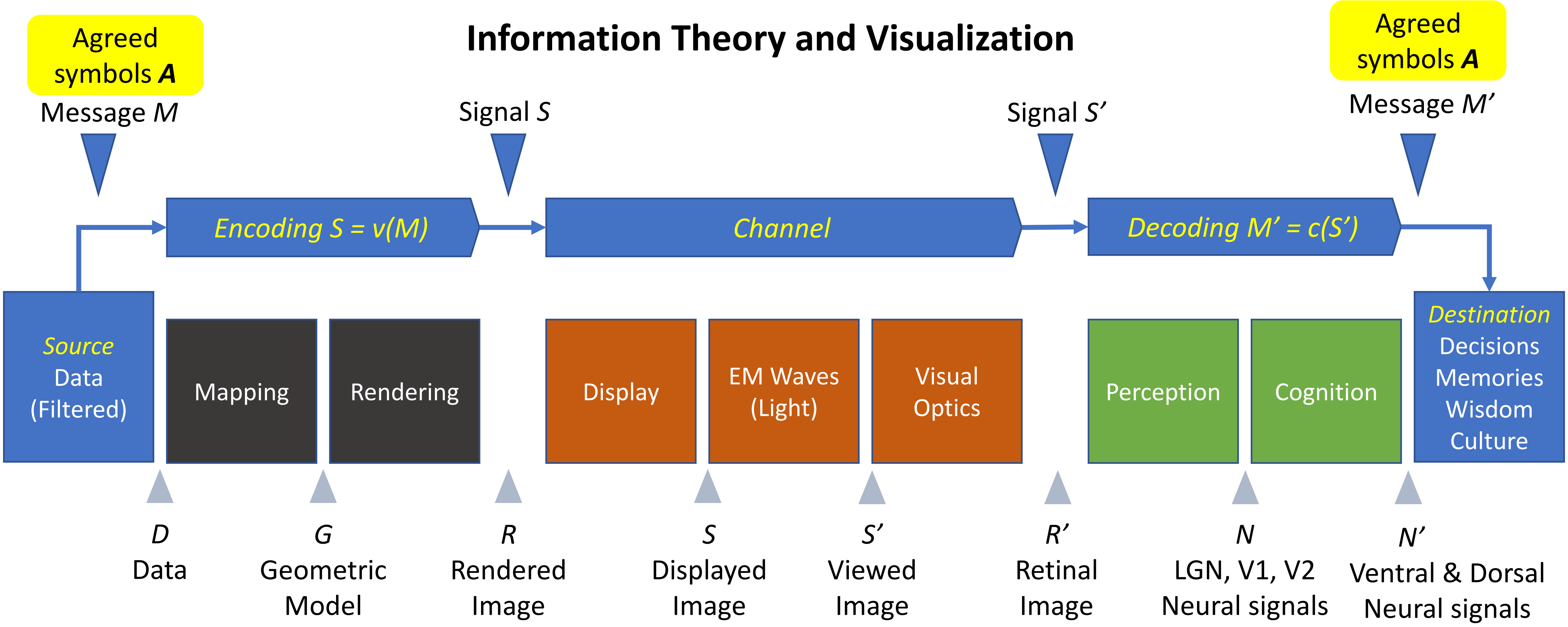
Information Theory and Visualization



Shannon assumed the destination was mechanically decoding and relaying a message from the source.

Humans are ***far*** from ideal receivers of messages, and it is humans that we need to be the decoders.

Information Theory and Visualization



Visualization lends itself to an empirical psychometric approach to determining what works:

Evaluation Methodologies: Measure Response Time, Accuracy, Memorability, Cost.

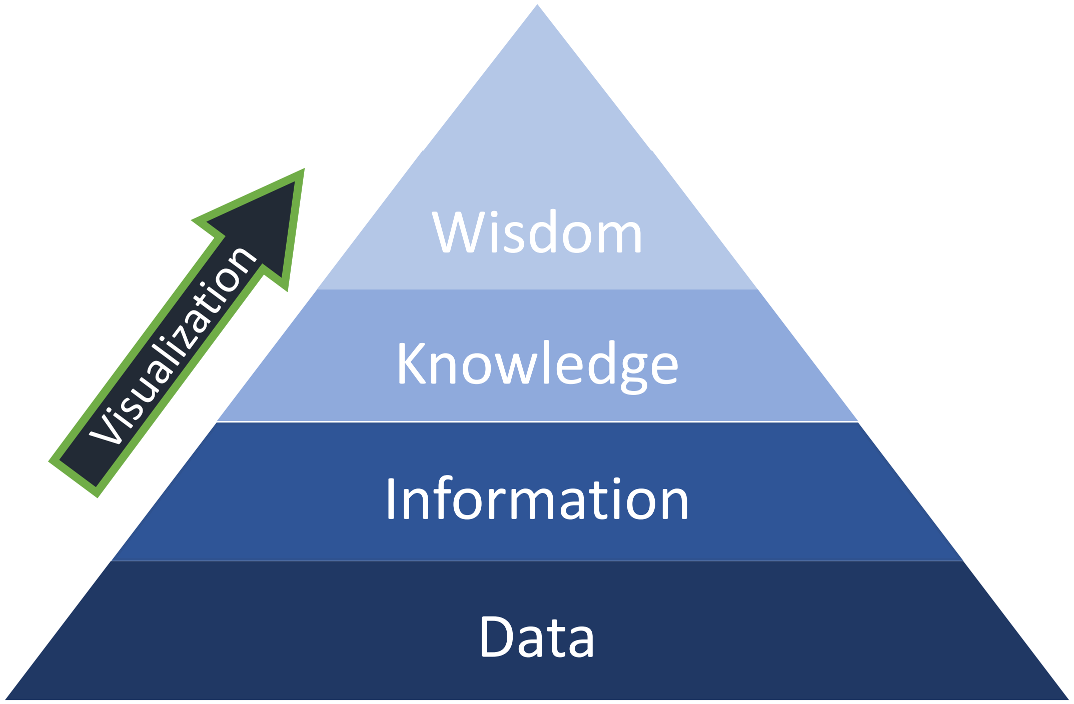
Quality Metrics: Predict Response Time, Accuracy, Memorability, Cost.

What is Visualization?

Information theory = bits

Visualization = ideas

From data to wisdom



Computer Science and Engineering

Wisdom: principles and values that inform decisions.

Knowledge: information that is understood as ideas by a human being.

Information: structured data with clear meaning or purpose.

Data: symbols and signs often representing something in the world.

Human Psychology & Psychophysics

Information Theory

References and sources of data/code

http://jim-stone.staff.shef.ac.uk/BookInfoTheory/InfoTheoryBookMain.html

https://youtu.be/2s3aJfRr9gE