



observer A has perfect knowledge of alphabet A and symbol probabilities P_A and the observer B has perfect knowledge of code alphabet B and codeword probabilities P_B . Neither observer, however has any knowledge whatsoever of the other observer's block box. Now suppose each time observer B observe a codeword he asks observer A what symbol had been sent by the information source.

How much information does observer B obtained from the answer of observer A?

If the answer to this is "none" then all of the information presented to the encoder passed through it to reach observer B and the encoder was information lossless \rightarrow observer B knows A as much as he knows B.

On the other hand, if observer A 's report occasionally surprises observer B, then some information was lost in the encoding process.

\leftarrow errors are introduced!

A's report then serves to decrease the uncertainty observer B has concerning the symbols being emitted by the encoder.

The reduction in uncertainty about B conveyed by the observation A is called mutual information, $I(B;A)$. The information presented to observer B by his observation is merely the entropy $H(B)$. If the observer observes symbol b and then learn from his partner that the source symbol was a, observer A's report conveys information.

$$H(B | A = a) = \sum_{b \in B} P_{b|a} \log \left(\frac{1}{P_{b|a}} \right)$$

The amount by which B's uncertainty is therefore reduced is

$$H(B | A) = \sum_{a \in A} P_a H(B | A = a) = \sum_{a \in A} \sum_{b \in B} P_a P_{b|a} \log \left(\frac{1}{P_{b|a}} \right)$$

and averaged over the course of all observation, the average information conveyed by A's report will be

$$I(B; A) = H(B) - H(B | A) = \sum_{b \in B} \sum_{a \in A} P_{b,a} \log \frac{P_{b,a}}{P_b P_a}$$