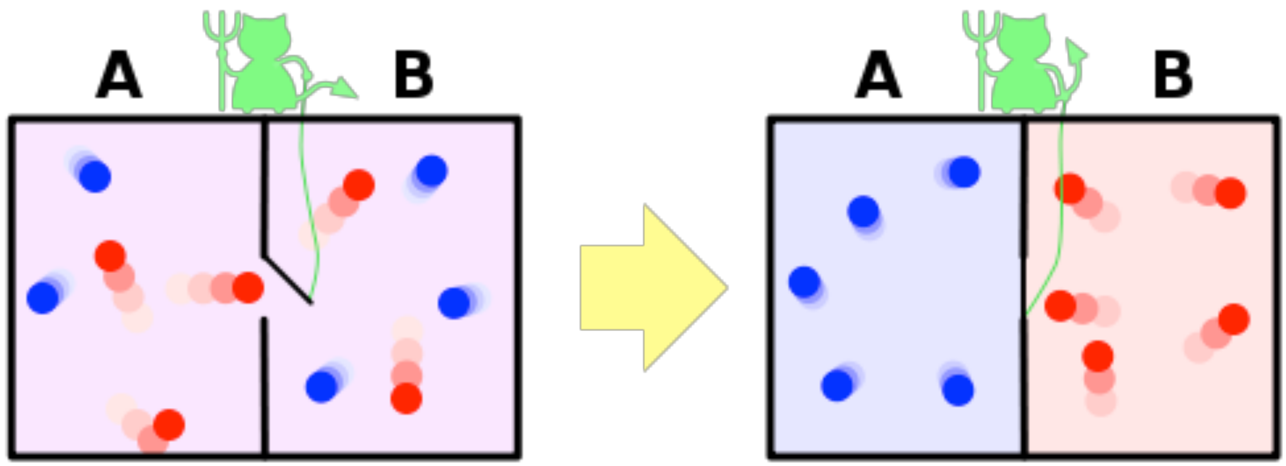
Maxwell's Demon



Shannon Information

- Shannon Entropy H to measure basic information capacity:
  - For a discrete random variable X with a probability mass function p(x), the entropy of X is defined as:

$$H(X) = -\sum p(x)\log_2 p(x)$$

  - Entropy is measured in bits.
  - H measures the average uncertainty in the random variable.
- Example 1:
  - Consider a random variable with uniform distribution over 32 outcomes.
  - To identify an outcome, we need a label that takes on 32 different values. How big does my label need to be to communicate the outcome?

$$H(X) = -\sum_{i=1}^{32} p(i)\log_2 p(i) = -\sum_{i=1}^{32} \frac{1}{32}\log_2 \frac{1}{32} =$$

Shannon Information

- Shannon Entropy H to measure basic information capacity:
  - For a random variable X with a probability mass function p(x), the entropy of X is defined as:

$$H(X) = -\sum p(x)\log_2 p(x)$$

  - Entropy is measured in bits.
  - H measures the average uncertainty in the random variable.
- Example 1:
  - Consider a random variable with uniform distribution over 32 outcomes.
  - To identify an outcome, we need a label that takes on 32 different values. How big does my label need to be to communicate the outcome?

$$H(X) = -\sum_{i=1}^{32} p(i)\log_2 p(i) = -\sum_{i=1}^{32} \frac{1}{32}\log_2 \frac{1}{32} = -32 \cdot \frac{1}{32} = 5 \text{ bits}$$

00001, 00010, 00011…11111

Random Variables

- A point in a sample space (e.g., the outcome of an experiment).
- Probability distribution of the random variable X:

$$P\{X = x_j\} = f(x_j) \qquad (\mathrm{j} = 1,2,\ldots)$$

Toss 3 fair coins.
Let X denote the number of heads appearing.
X is a random variable taking on one of the values (0,1,2,3).

P{X=0} = 1/8; P{X=1} = 3/8; P{X=2} = 3/8; P{X=3} = 1/8.

Sum of the probabilities is 1
All probabilities are >= 0

- Example 2:
  - A horse race with 8 horses competing.
  - What is the entropy of the race ?
  - The probabilities of 8 horses are:

    $$\left(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{16},\frac{1}{64},\frac{1}{64},\frac{1}{64},\frac{1}{64}\right)$$

  - Calculate the entropy H of the horse race:

    $$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{16}\log\frac{1}{16} - 4\frac{1}{64}\log\frac{1}{64} = 2 \text{ bits}$$

  - Suppose that we wish to send a (short) message to another person indicating which horse won the race.
  - Could send the index of the winning horse (3 bits).
  - Alternatively, could use the following set of labels:
    - 0, 10, 110, 1110, 111100, 111101, 111110, 111111.
    - Average description length is 2 bits (instead of 3).
    - Huffman code: variable length ‘prefix free’ code for maximum lossless compression

- 0, 10, 110, 1110, 111100, 111101, 111110, 111111
- ½, 2/4, 3/8, 4/16, 6/64*4
- 1 + 10/16 + 24/64
- 2

We would like to develop a usable measure of the information we get from observing the occurrence of an event having probability p .

- We will want our *information* measure $I(p)$ to have several properties (note that along with the axiom is motivation for choosing the axiom):

  1. Information is a non-negative quantity: $I(p) \geq 0$.

  2. If an event has probability 1, we get no information from the occurrence of the event: $I(1) = 0$.

  3. If two independent events occur (whose joint probability is the product of their individual probabilities), then the information we get from observing the events is the sum of the two informations: $I(p_1 * p_2) = I(p_1) + I(p_2)$. (This is the critical property . . .)

  4. We will want our *information* measure to be a continuous (and, in fact, monotonic) function of the probability (slight changes in probability should result in slight changes in *information*).

cogs.csustan.edu/~tom/SFI-CSSS/info-theory/info-lec.pdf

– Suppose you live in a city where ½ the time it's sunny and ½ the time it's cloudy. I say it's sunny. How much information did I give you?

– You live in Abq, 7/8 of the days its sunny, 1/8th it's cloudy. I say it's sunny. How much information did I give you?

Given: $p(W = sunny) = 7/8$ and $p(W = cloudy) = 1/8$

By observing W = sunny we get $-\log(7/8) \sim= 0.2$ bits of info

By observing W = cloudy is $-\log(1/8) = 3$ bits of info

On average, we observe W= sunny 7/8 of the time, and W = cloudy 1/8 of the time

The AVERAGE INFORMATION we receive by observing W is

$-7/8 * \log(7/8) + -1/8 \log(1/8) \sim= 0.54$

The SHANNON INFORMATION, $H(W_{abq}) \sim= 0.54$

In New York half the days are sunny, and ½ the days are cloudy

The SHANNON INFORMATION, $H(W_{NY}) = 1$