What is the nature of information?

Readings for today

- Timme, N. M., & Lapish, C. (2018). A tutorial for information theory in neuroscience. eneuro, 5(3).

Topics

- Entropy
- Mutual information
- KL Divergence

Entropy

Two ways to explore

Random exploration ⟷ Directed exploration

(taking a guess) (seeking information)

Information theory

Goal: A formal theory for the transmission, processing, extraction, and utilization of information.

Approach: Quantify the *amount* of information a channel, c, can convey about a signal, s, to a receiver, r.

signal

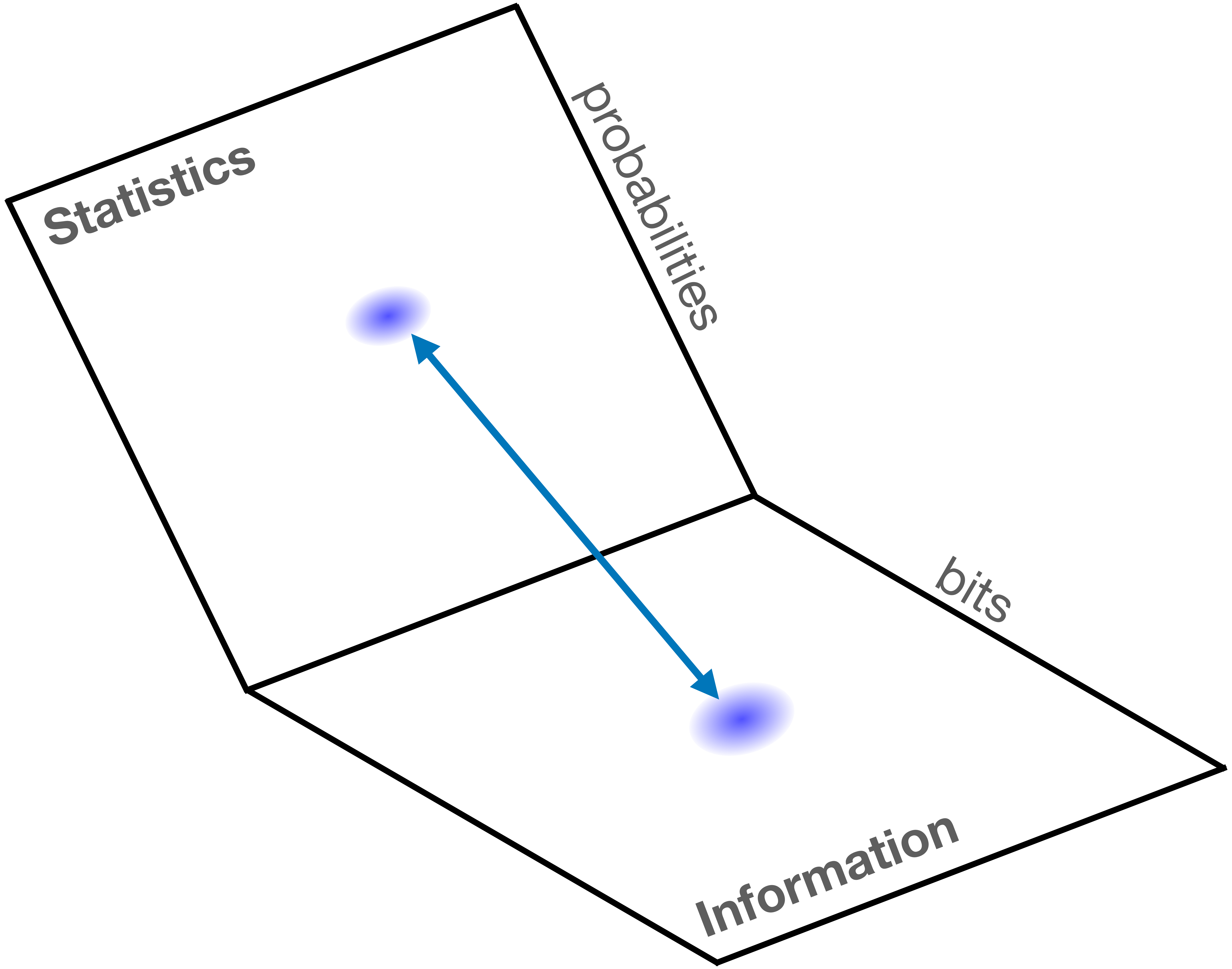
channel (c)

receiver

$$s \rightarrow r$$

(Shannon & Weaver, 1949)

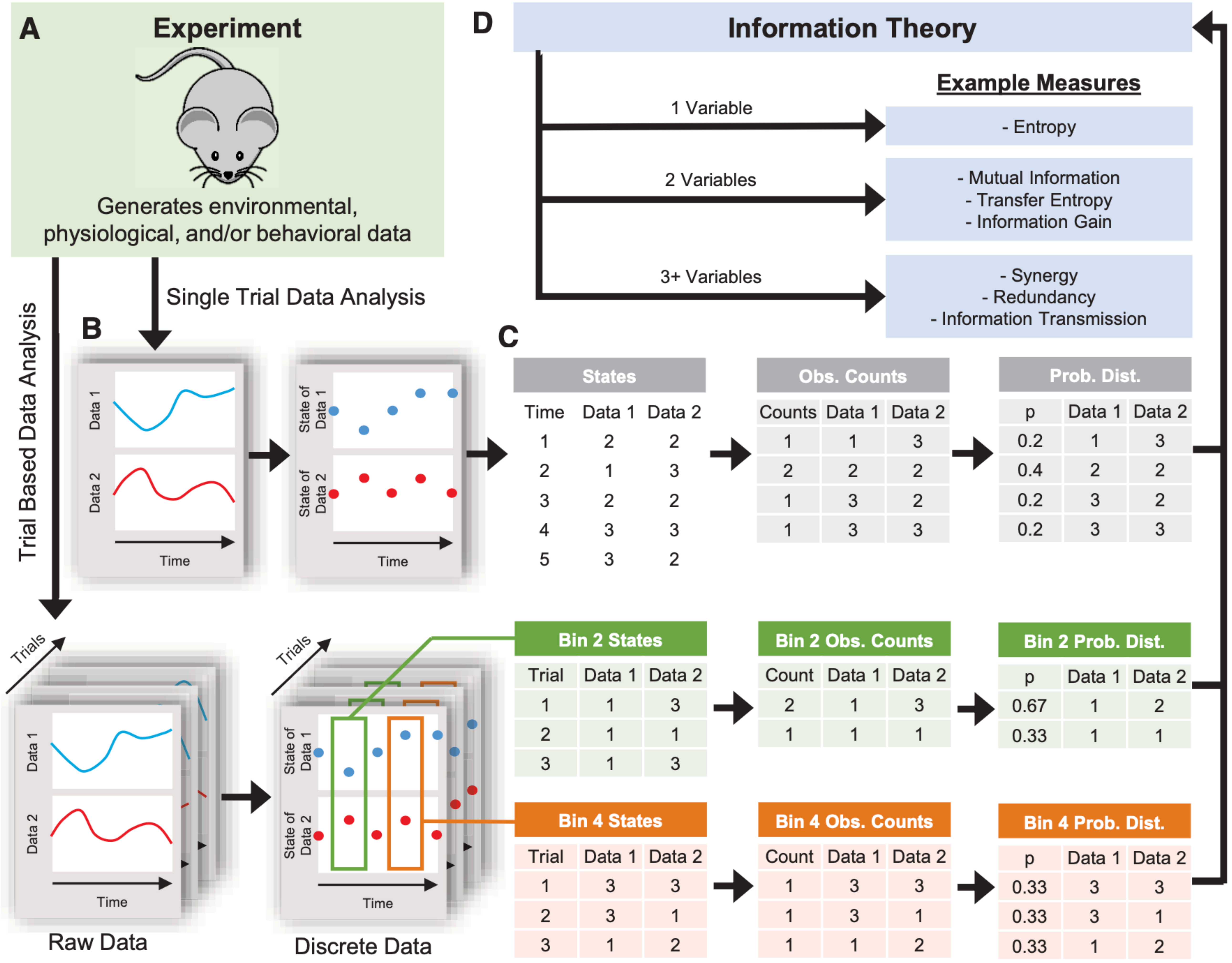
Information theory



Information theory is largely a projection of statistics, converting probabilities to bits (~$\log$ probabilities)

(Shannon & Weaver, 1949; Kullback 1959)

Information theory in neuroscience



Role in neuroscience & psychology:

- Data analysis
- Generative modeling

(Timme & Lapish, 2018)

Amount of information in s

Question: What is the *average* amount of information conveyed by s?

$$H(S) = \sum_{s \in S} p(s_i) I(s_i)$$

$$H(s) = -\sum_{s \in S} p(s) \log_2 p(s)$$

Entropy of signal s

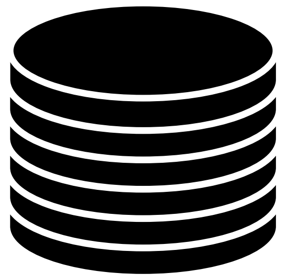
probability that i^{th} state is observed

information available from i^{th} state

$$I(s_i) = log_2(\frac{1}{p(s_i)}) \approx 0$$

$$p(s_i) = 0$$

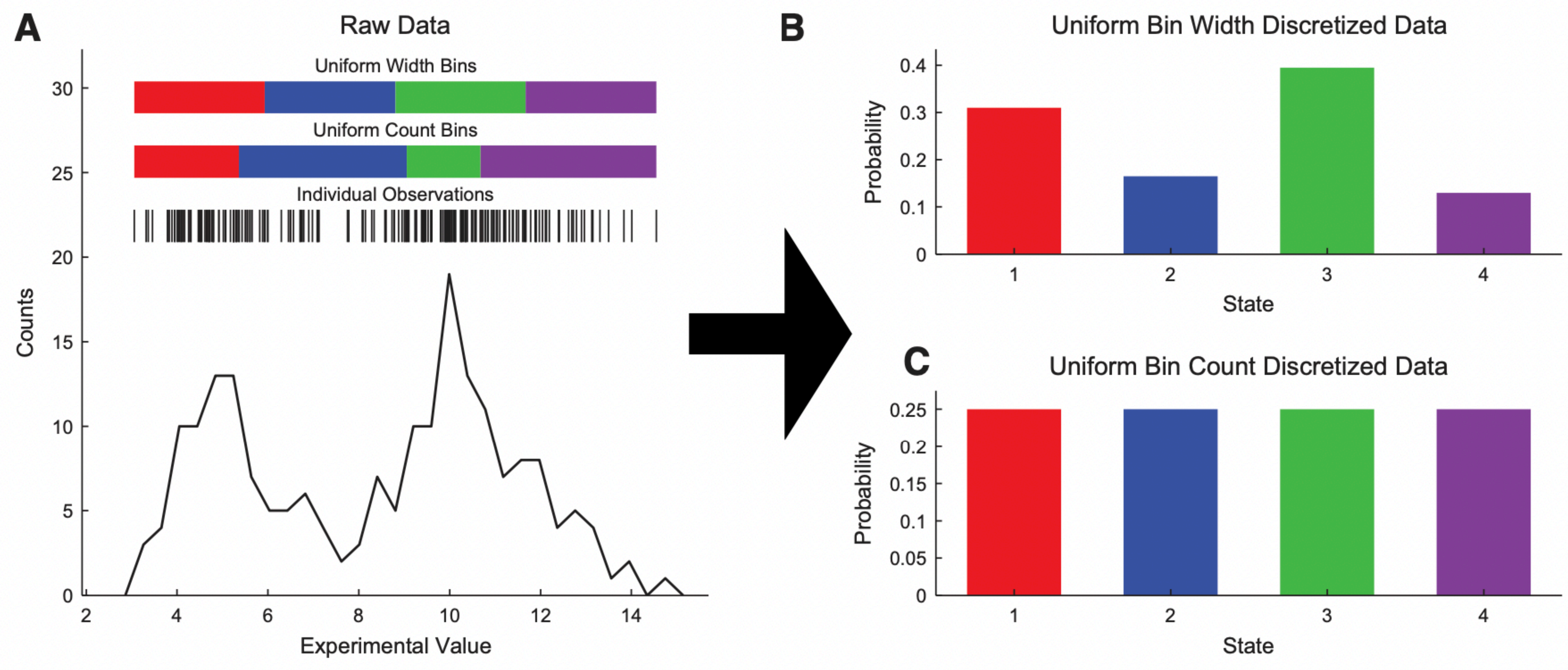
Example:



$$H(s) = -\frac{1}{2}log_2(\frac{1}{2}) - \frac{1}{2}log(\frac{1}{2})$$

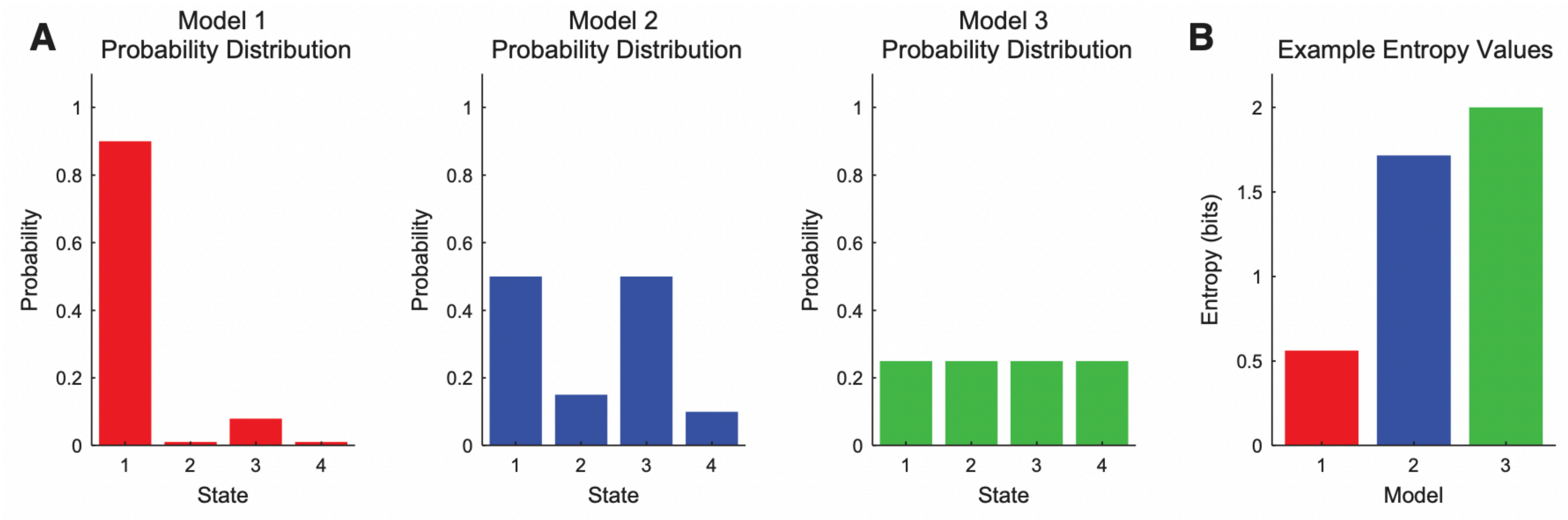
(Shannon & Weaver, 1949)

Discretizing your signal



(Timme & Lapish, 2018)

Entropy and structure



(Timme & Lapish, 2018)

Mutual information

Joint entropy

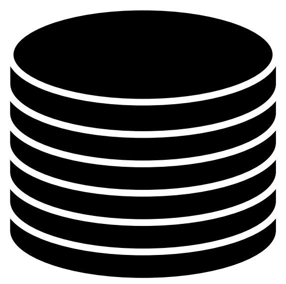
<u>Question</u>: What is the <u>*average*</u> amount of uncertainty in s and r?

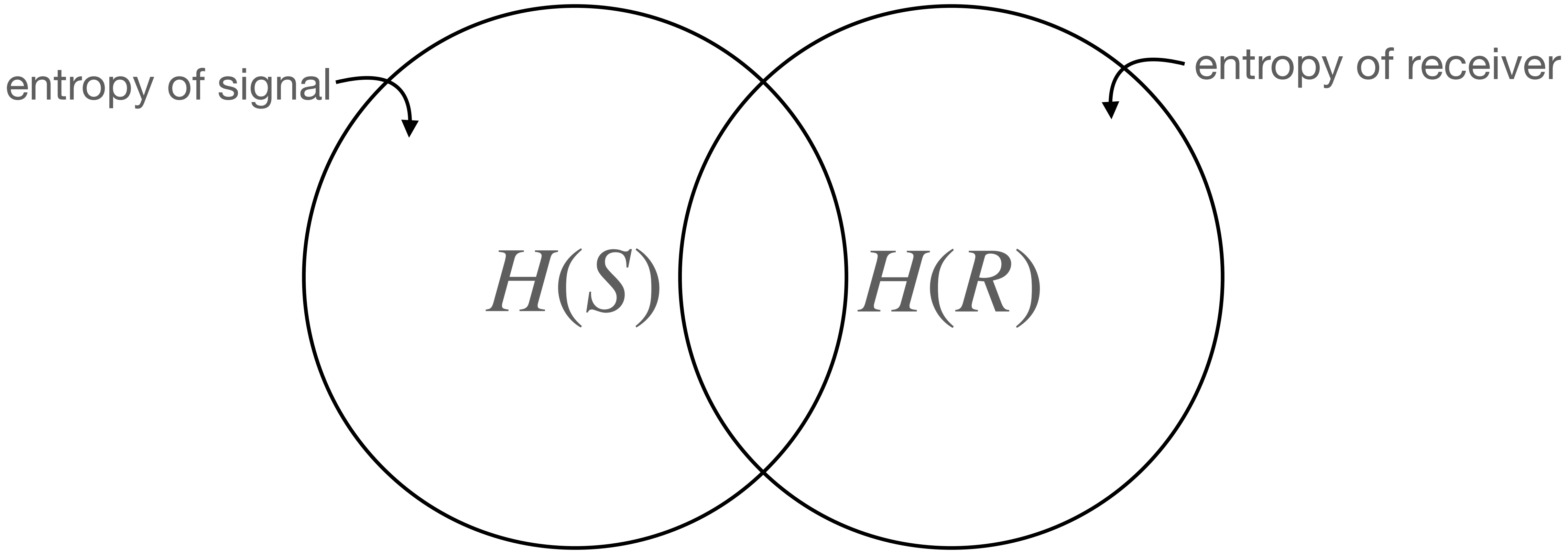
$$H(S,R) = -\sum_{s \in S, r \in R} p(s,r) \log_2 p(s,r)$$

Joint entropy of signal S & receiver R

Joint probability distribution

<u>Example</u>: $H(S,R) = 4[\frac{1}{4}\log_2(\frac{1}{4})]$





(Shannon & Weaver, 1949)

Conditional entropy

<u>Question</u>: What is the <u>*average*</u> amount of uncertainty in r after seeing s?

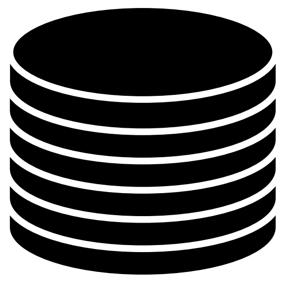
$$H(R\,|\,S) = -\sum_{s \in S, r \in R} p(s, r)\log_2 p(r\,|\,s)$$

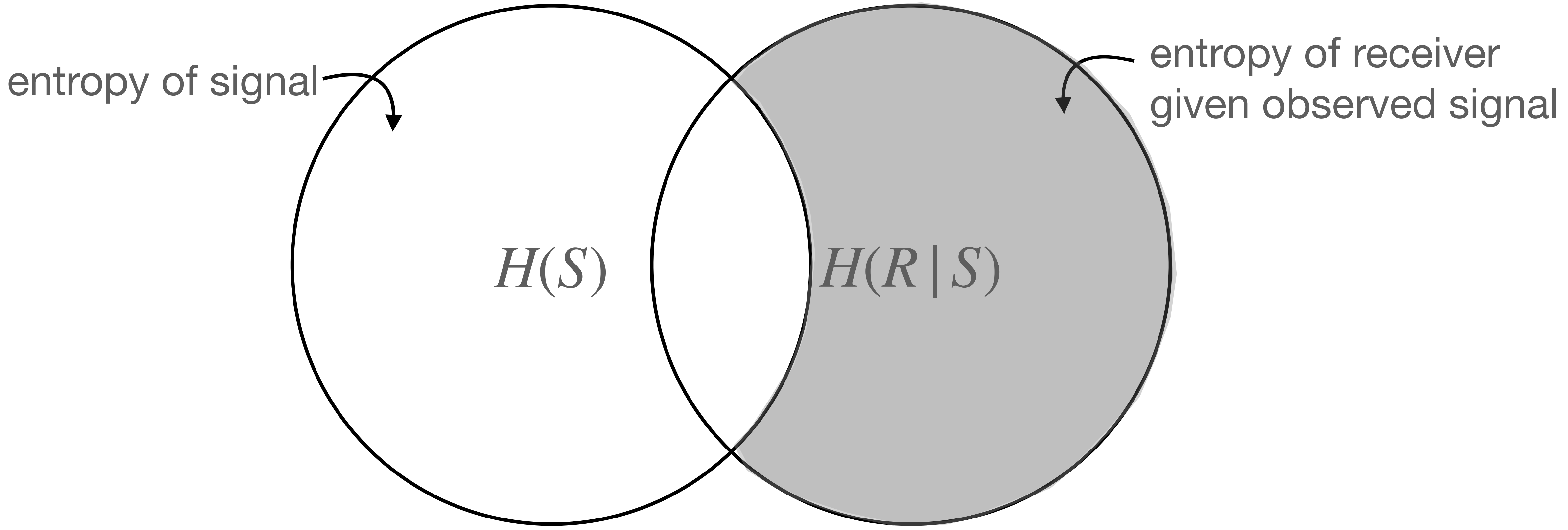
Joint entropy of signal S & receiver R

Joint probability distribution

Conditional probability distribution

<u>Example</u>: $H(R\,|\,S) = 2[0.4(\frac{1}{0.8})] + 2[0.1\,\log_2(\frac{1}{0.2})]$





$$H(S, R) = H(S) + H(R\,|\,S)$$

(Shannon & Weaver, 1949)

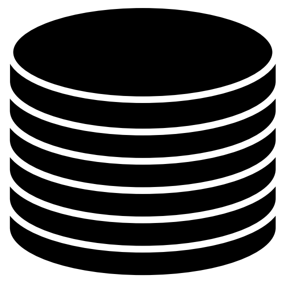
Conditional entropy

Example:

Table 1. Marginal and joint probability distributions for an example system of two dependent coins.

	c_1 = *heads*	c_1 = *tails*	Marginal Distributions for Coin 2
c_2 = *heads*	$p(c_1 = heads, c_2 = heads) = 0.4$	$p(c_1 = tails, c_2 = heads) = 0.1$	$p(c_2 = heads) = 0.5$
c_2 = *tails*	$p(c_1 = heads, c_2 = tails) = 0.1$	$p(c_1 = tails, c_2 = tails) = 0.4$	$p(c_2 = tails) = 0.5$
Marginal distributions for coin 1	$p(c_1 = heads) = 0.5$	$p(c_1 = tails) = 0.5$	

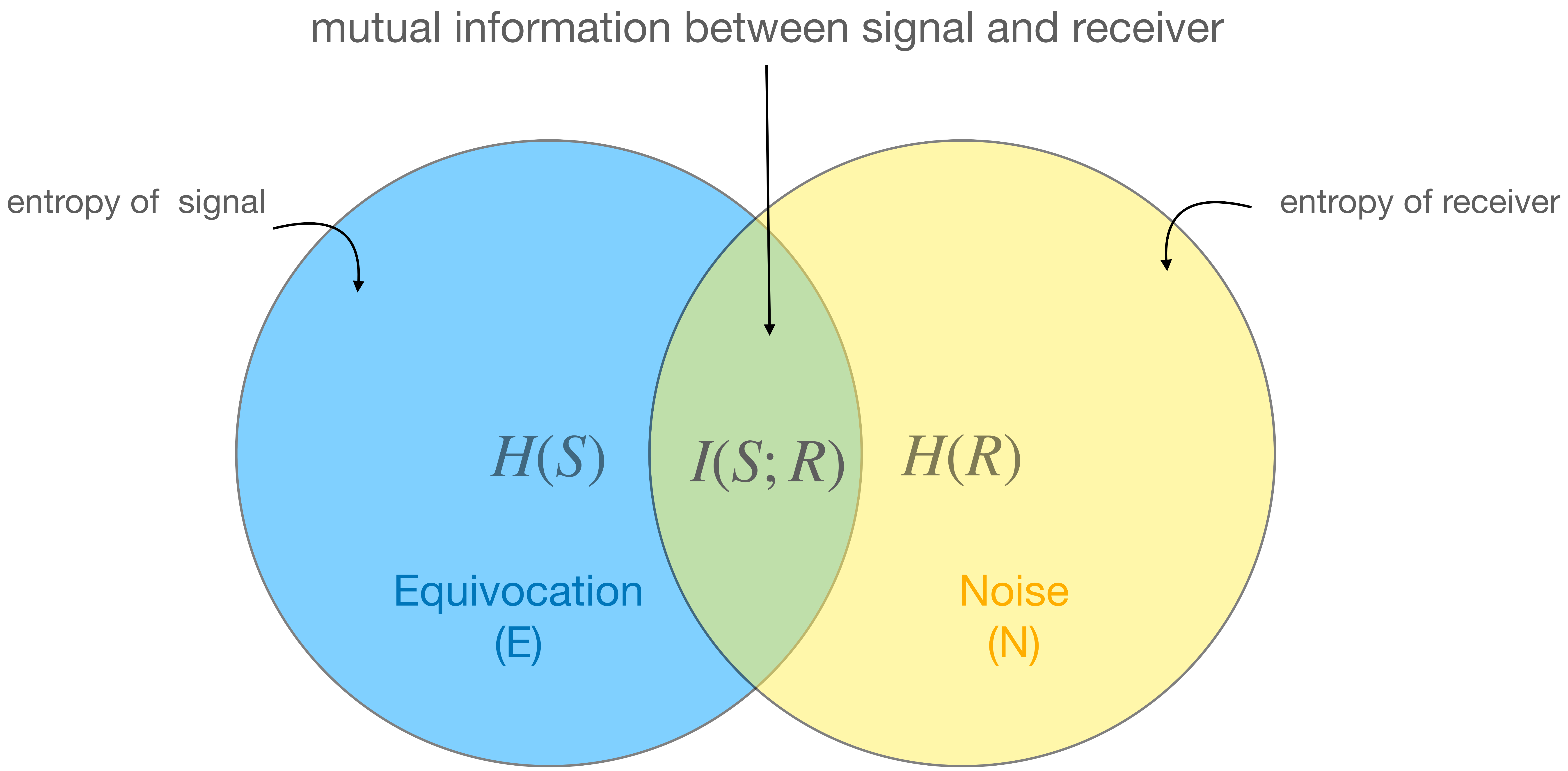
The joint distribution describe the likelihood for each possible combination of the two coins. The marginal distributions describe the likelihood for each coin alone. Marginal distributions can be found by summing across rows or columns of the joint distribution (Eqn. 1).



$$H(R\,|\,S) = 2[0.4(\frac{1}{0.8})] + 2[0.1\log_2(\frac{1}{0.2})]$$

(Timme & Lapish, 2018)

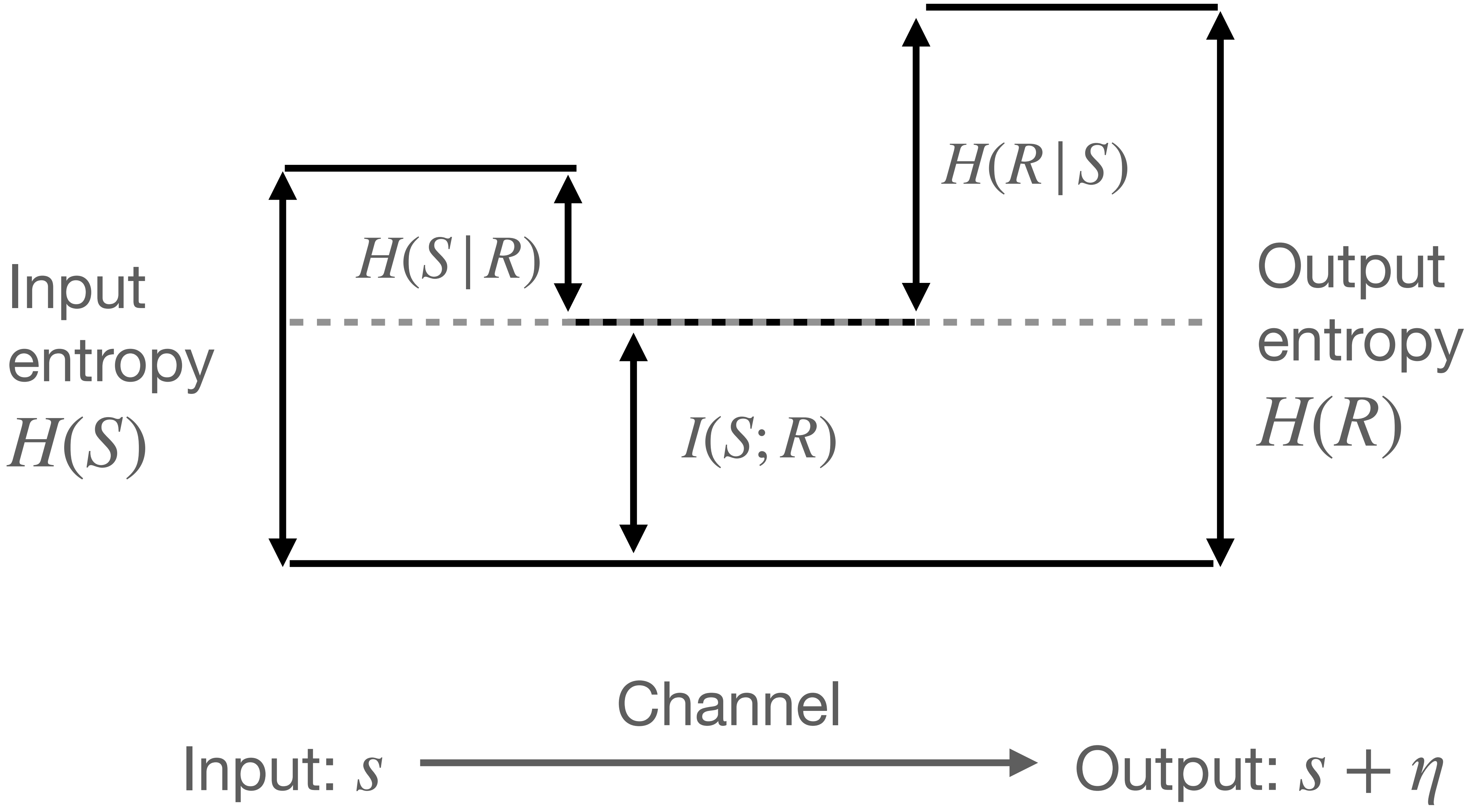
$I(S;R)$ mutual information



$I(S;R)$: The information transmitted from s to r is the total amount of information available at r, $H(R)$, minus noise.

(Timme & Lapish, 2018)

$I(S; R)$ mutual information

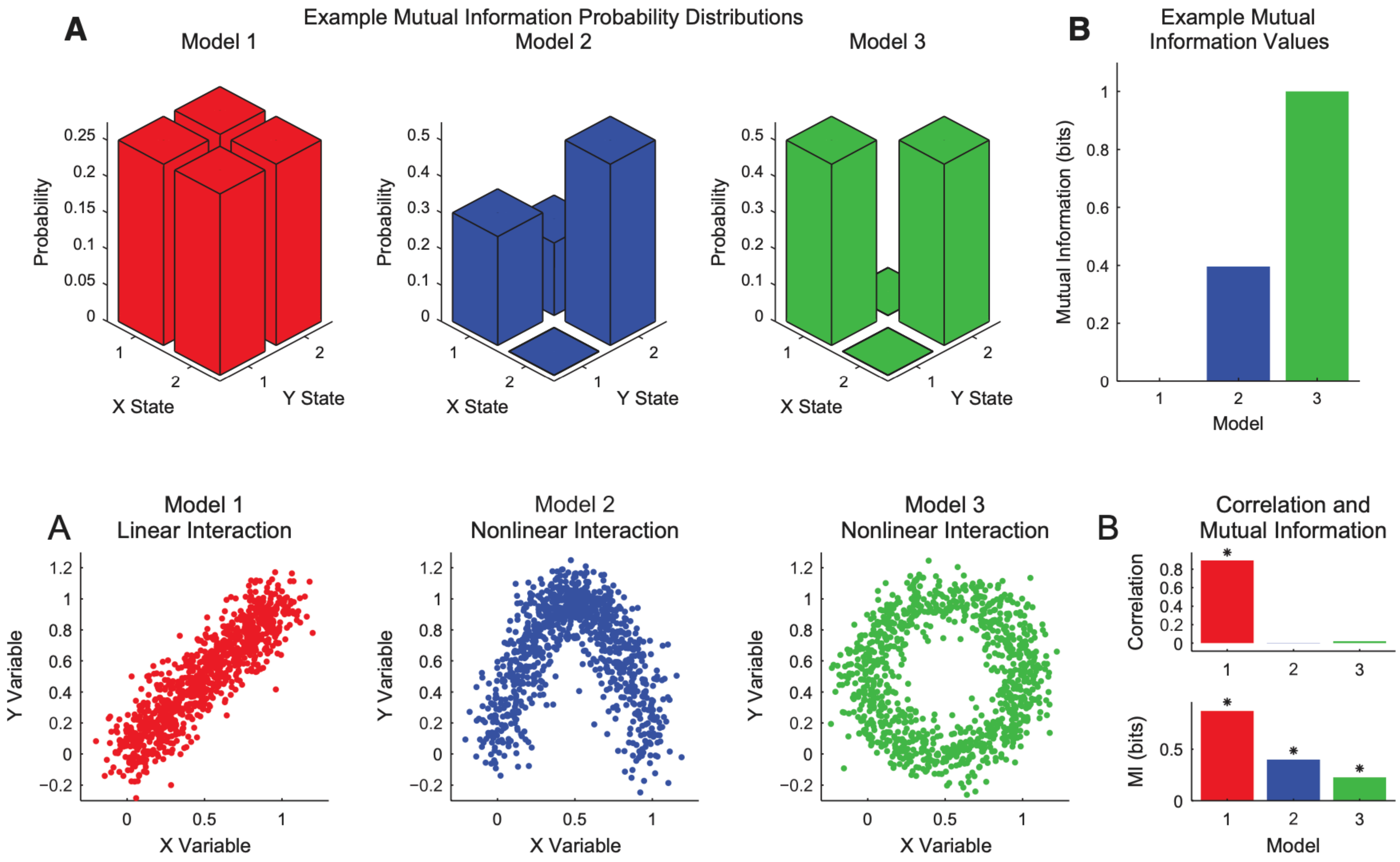


Mutual information:

$$I(S; R) = H(S) - H(S \mid R)$$
$$= H(R) - H(R \mid S)$$
$$= \sum_{s \in S, r \in R} p(s, r) \log_2\left(\frac{p(s, r)}{p(s)p(r)}\right)$$

(Stone, 2022; Timme & Lapish, 2018)

Examples mutual information



(Timme & Lapish, 2018)

KL Divergence

D_{KL} Kullback-Leibler divergence

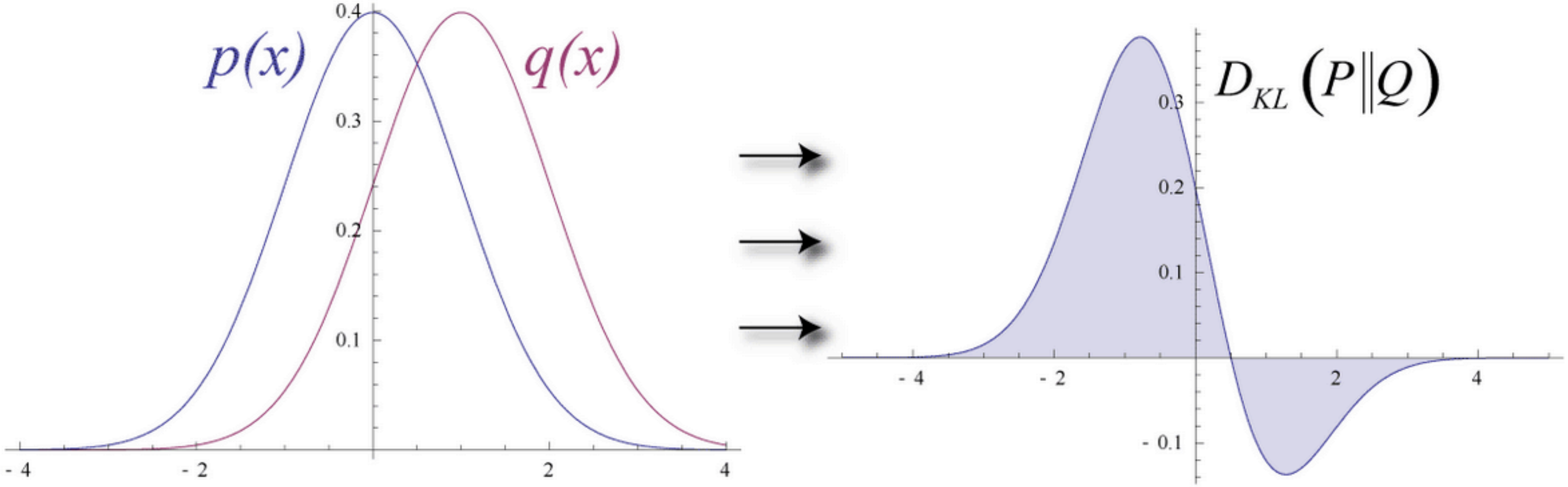
<u>Question</u>: What is the difference, in relatively entropy, between two distributions?

$$D_{KL}(p(X)||q(X)) = \Sigma_{x \in X} p(x) log \frac{p(x)}{q(x)} dx$$

KL divergence between $p(x)$ and $q(X)$

probability of x from distribution p

probability of x from distribution q



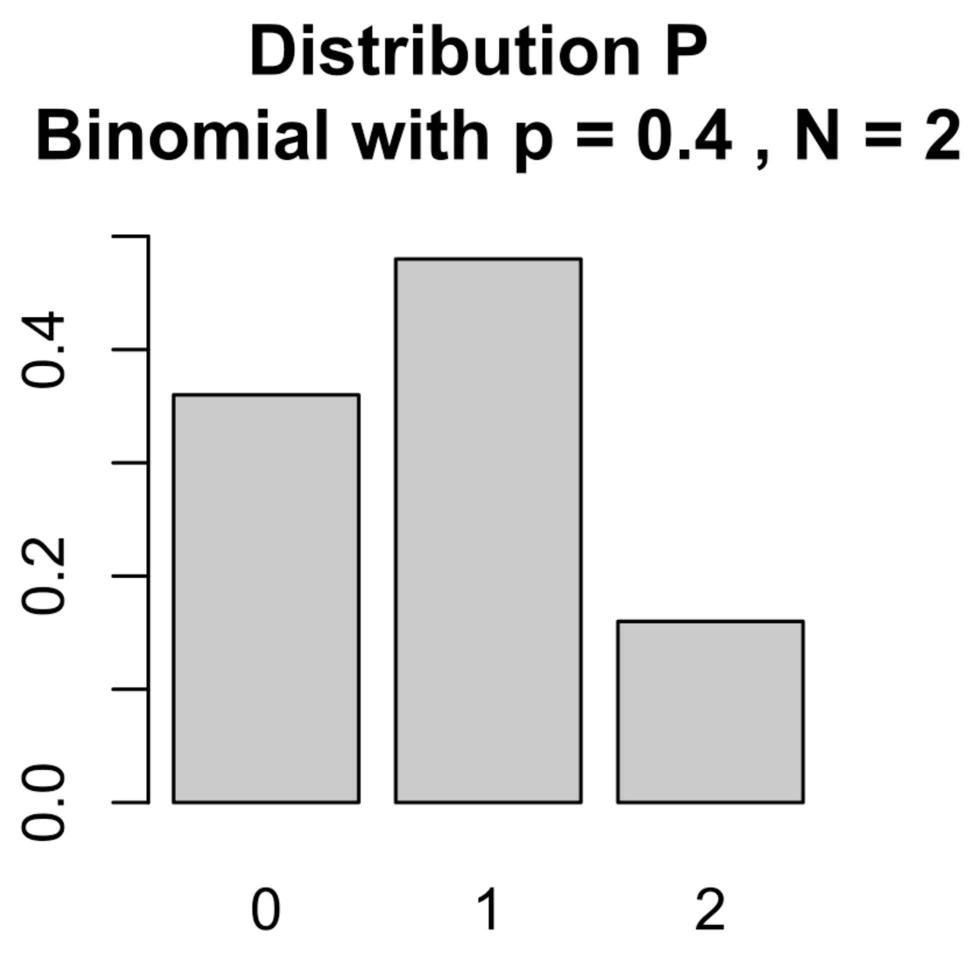
The *statistical distance* between two distributions, not Euclidian distance.

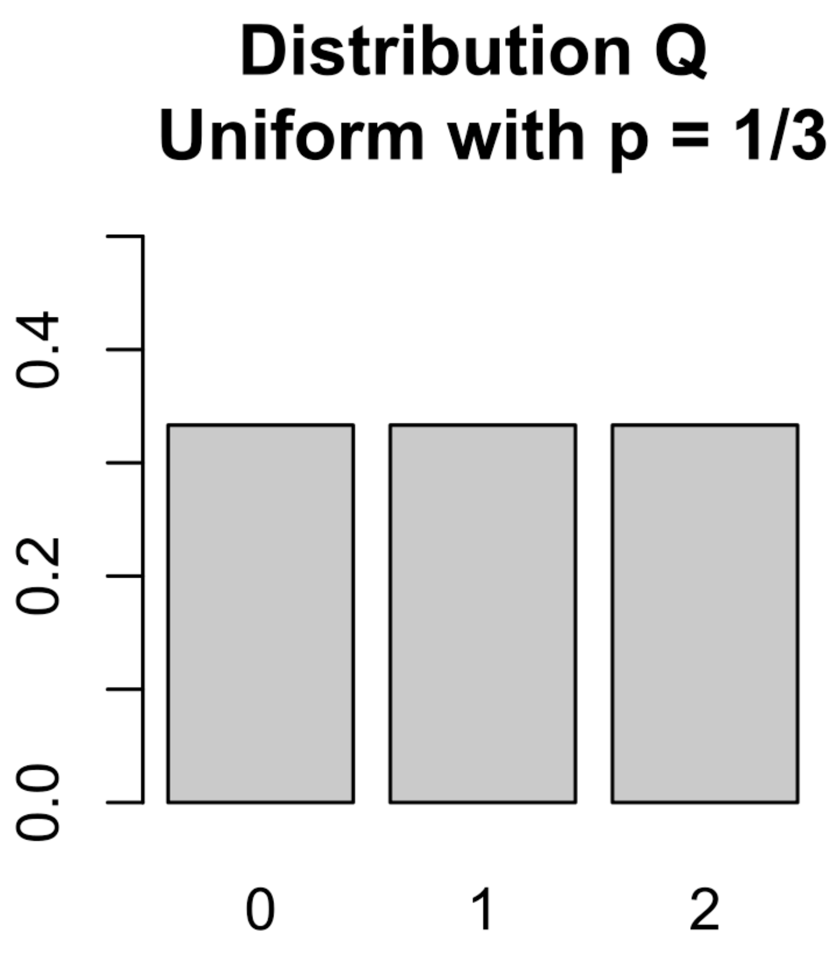
$$D_{KL}(p(X)||q(X)) \neq D_{KL}(q(X)||p(X))$$

https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence

(Kullback & Leibler, 1951)

Differences in entropy





x	0	1	2
Distribution $P(x)$	$\frac{9}{25}$	$\frac{12}{25}$	$\frac{4}{25}$
Distribution $Q(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$$\begin{aligned} D_{\mathrm{KL}}(P \| Q) &= \sum_{x \in \mathcal{X}} P(x) \ln\left(\frac{P(x)}{Q(x)}\right) \\ &= \frac{9}{25} \ln\left(\frac{9/25}{1/3}\right) + \frac{12}{25} \ln\left(\frac{12/25}{1/3}\right) + \frac{4}{25} \ln\left(\frac{4/25}{1/3}\right) \\ &= \frac{1}{25}(32 \ln(2) + 55 \ln(3) - 50 \ln(5)) \approx 0.0852996 \end{aligned}$$

$$\begin{aligned} D_{\mathrm{KL}}(Q \| P) &= \sum_{x \in \mathcal{X}} Q(x) \ln\left(\frac{Q(x)}{P(x)}\right) \\ &= \frac{1}{3} \ln\left(\frac{1/3}{9/25}\right) + \frac{1}{3} \ln\left(\frac{1/3}{12/25}\right) + \frac{1}{3} \ln\left(\frac{1/3}{4/25}\right) \\ &= \frac{1}{3}(-4 \ln(2) - 6 \ln(3) + 6 \ln(5)) \approx 0.097455. \end{aligned}$$

https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence

Take home message

- Information is defined by its uncertainty. We call this entropy.
- Mutual information reflects the degree of association between signals.
- The KL divergence reflects the relative difference in information contained in two distributions of signal.

Lab time!

https://coaxlab.github.io/BIX-book/notebooks/lab6-information.html