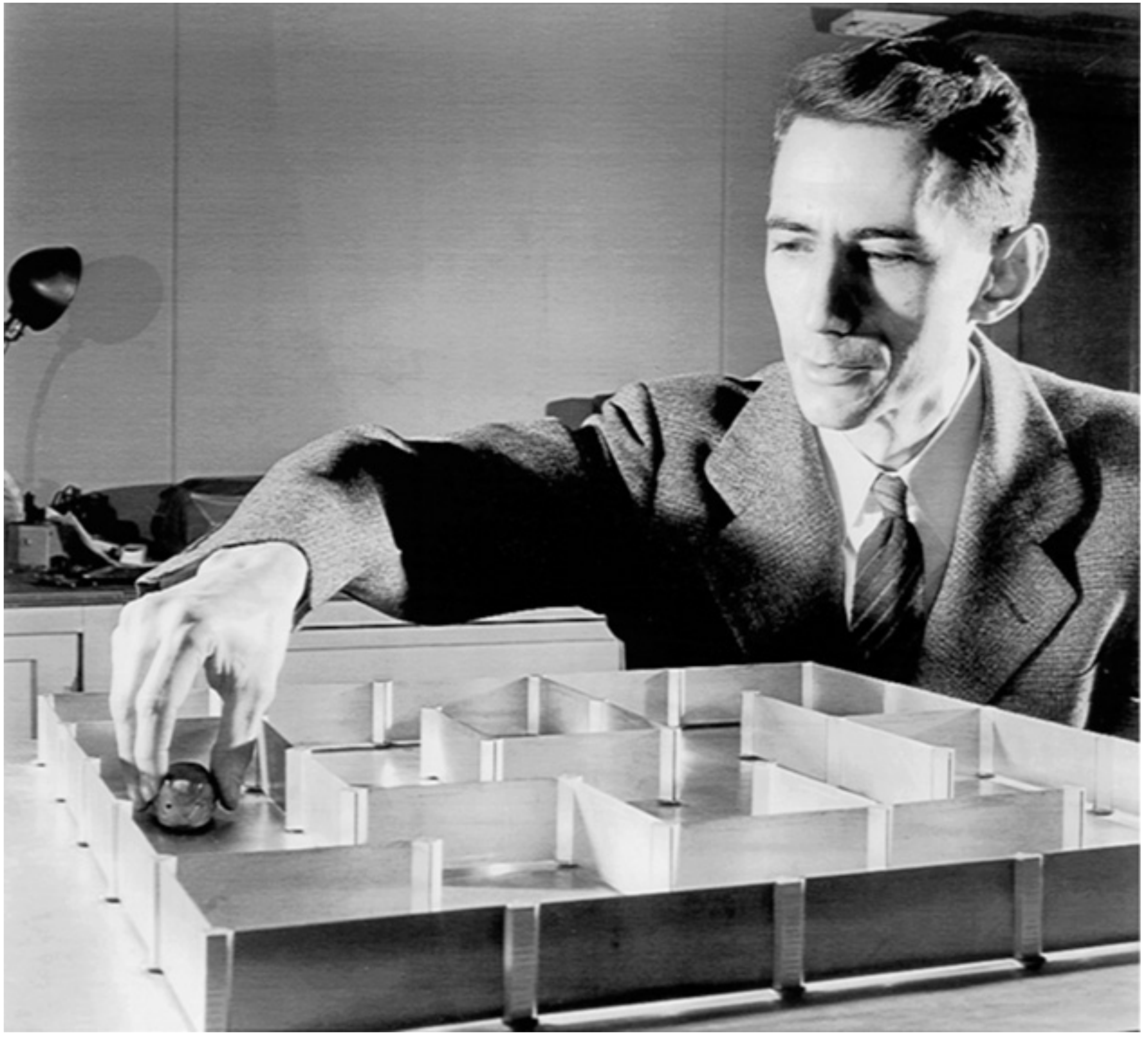
An introduction to Information Theory and its utility in neuroscience

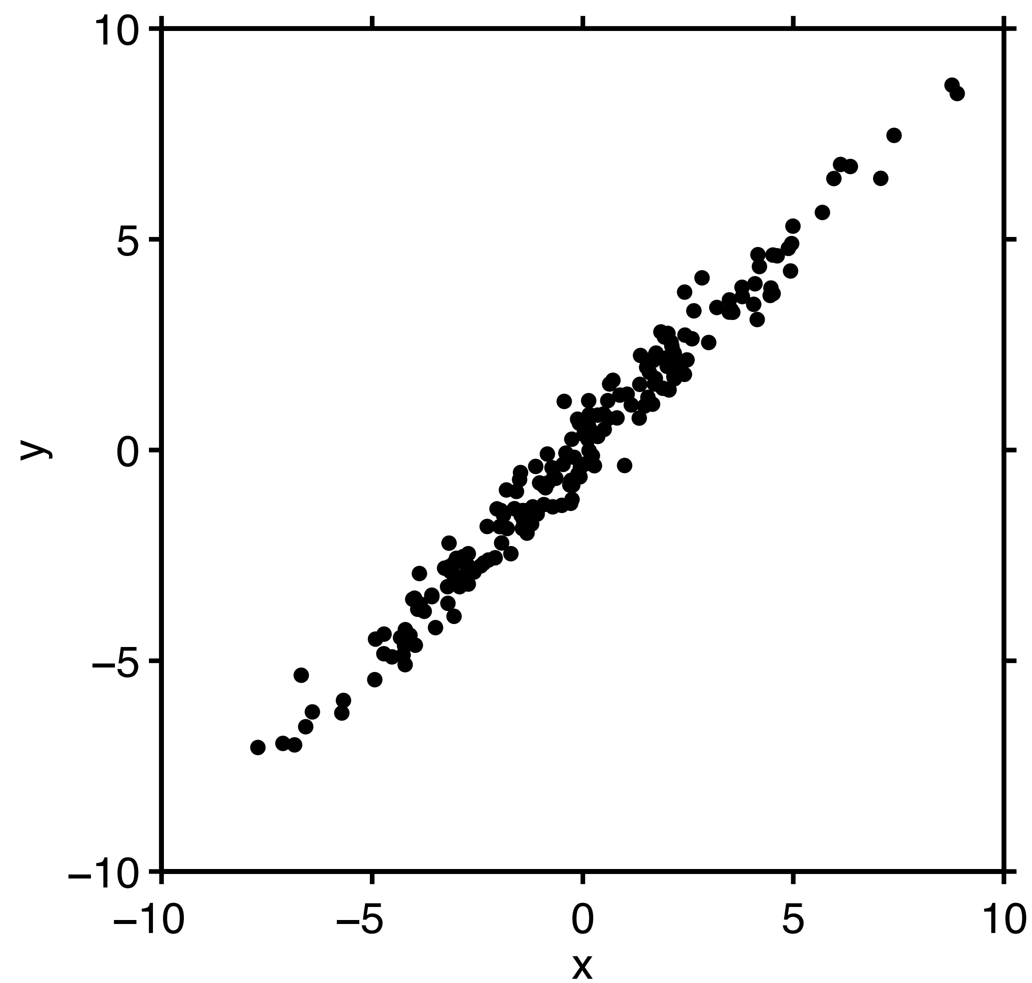


RECALL: When is information theory useful?

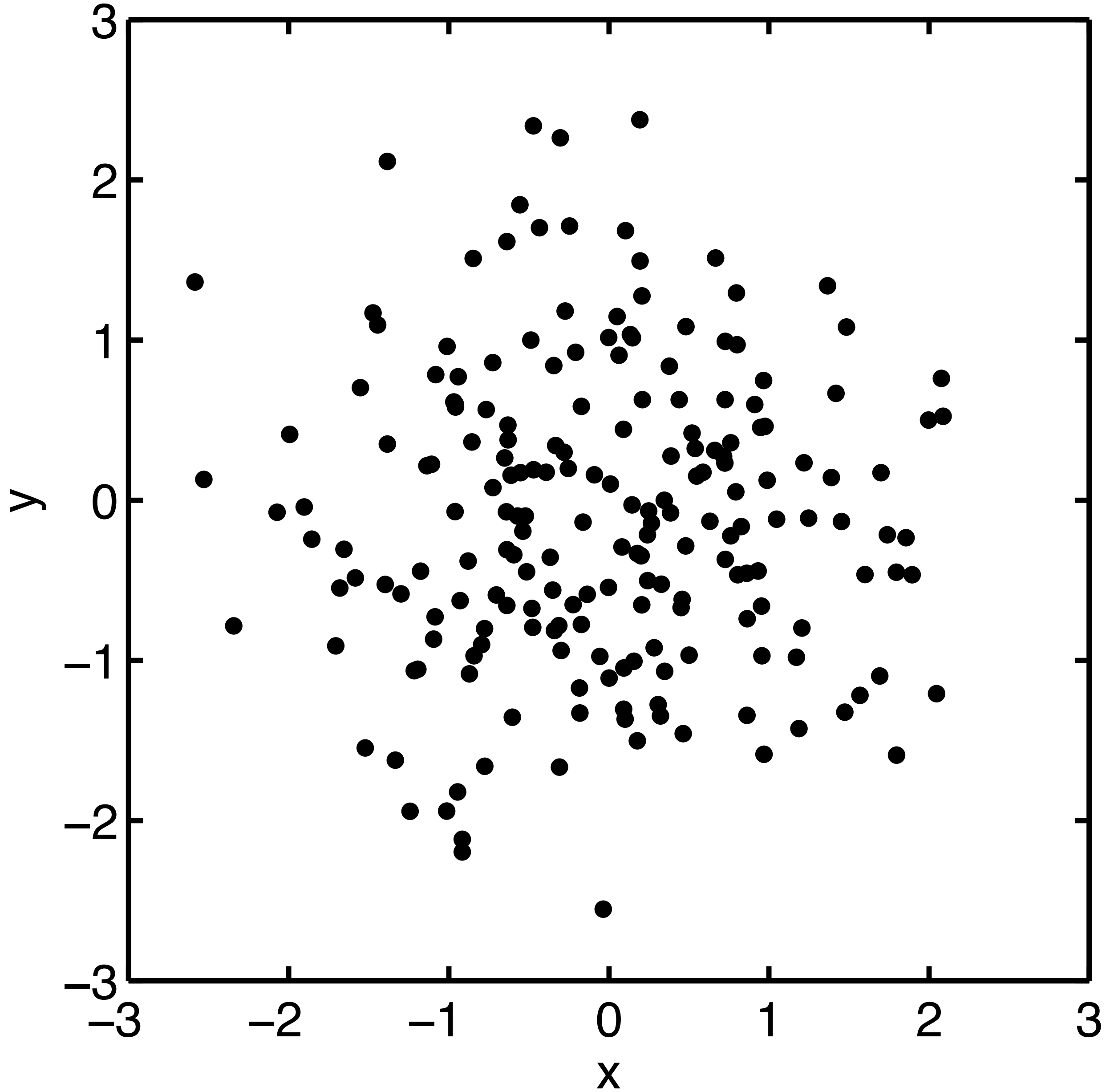
want to go beyond linear correlation

have enough data to sample P(x,y)

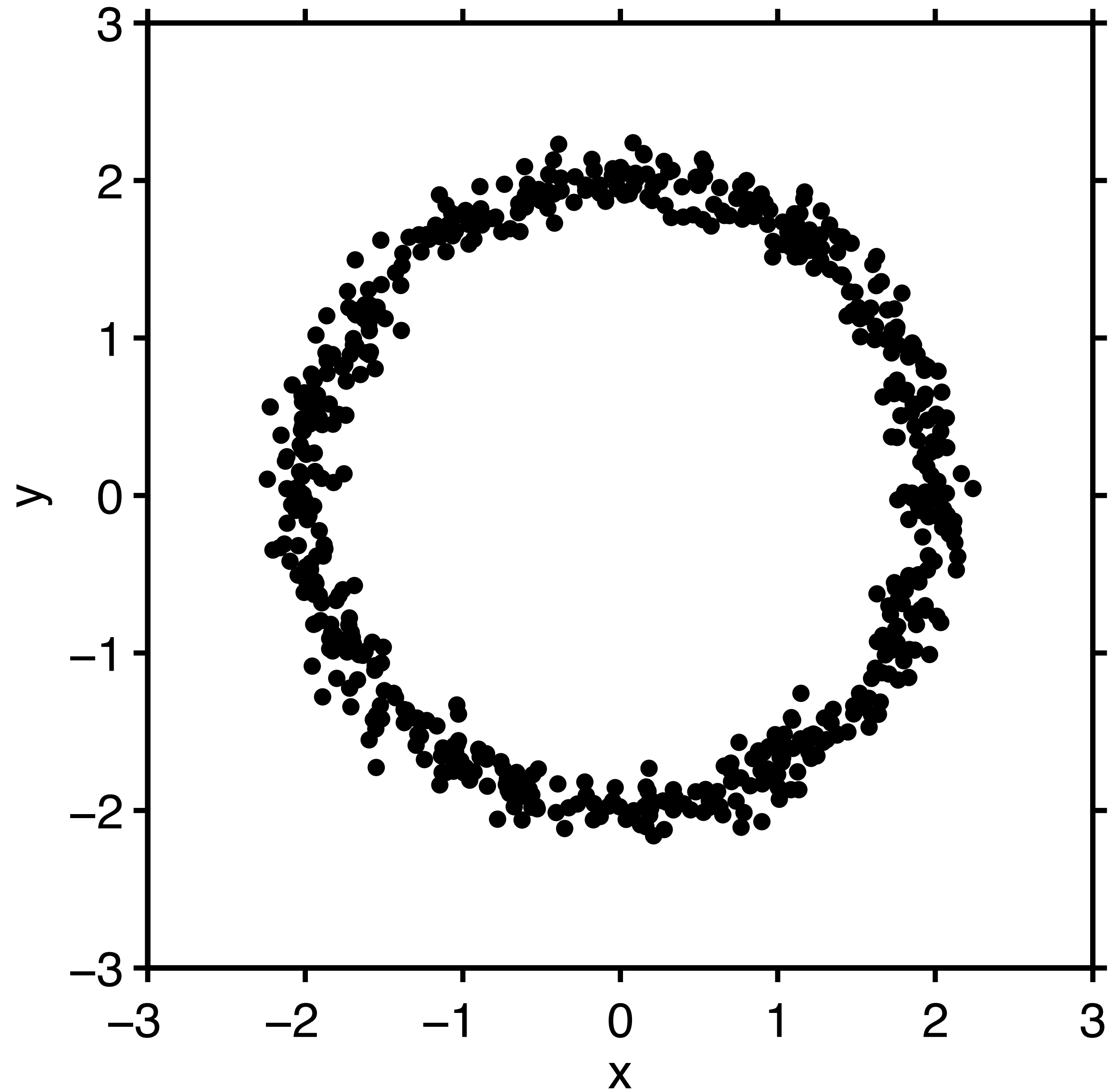
not sure what your 'code' is



Zero correlation, no information:

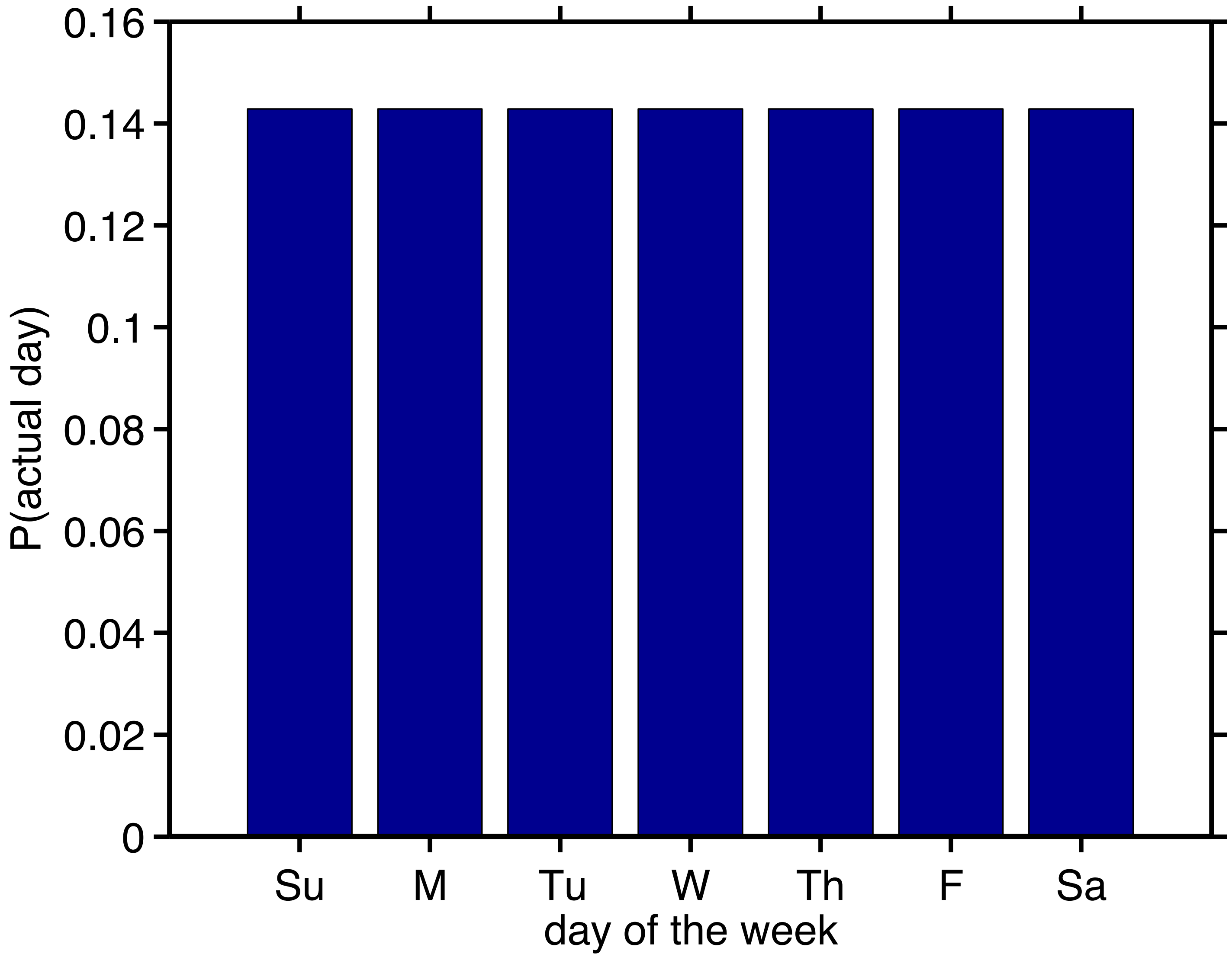


Zero correlation, but obvious information:



Example problem: student with a very strange form of amnesia...

What is his uncertainty about what day it is?



Entropy as a measure of uncertainty:

$$\begin{aligned} \text{uncertainty} &= \log\left(n\right) \\ &= \log\left(1/p\right) \\ &= -\log\left(p\right) \end{aligned}$$

$$u_i = -\log\left(p_i\right)$$

$$\langle u_i \rangle = -\sum_i p_i \log\left(p_i\right)$$

$$S(X) = -\sum_x p(x)\log_2(p(x))$$

Information as reduction in uncertainty:

$$I(X;Y) = S(X) - \langle S(X|Y) \rangle_y$$

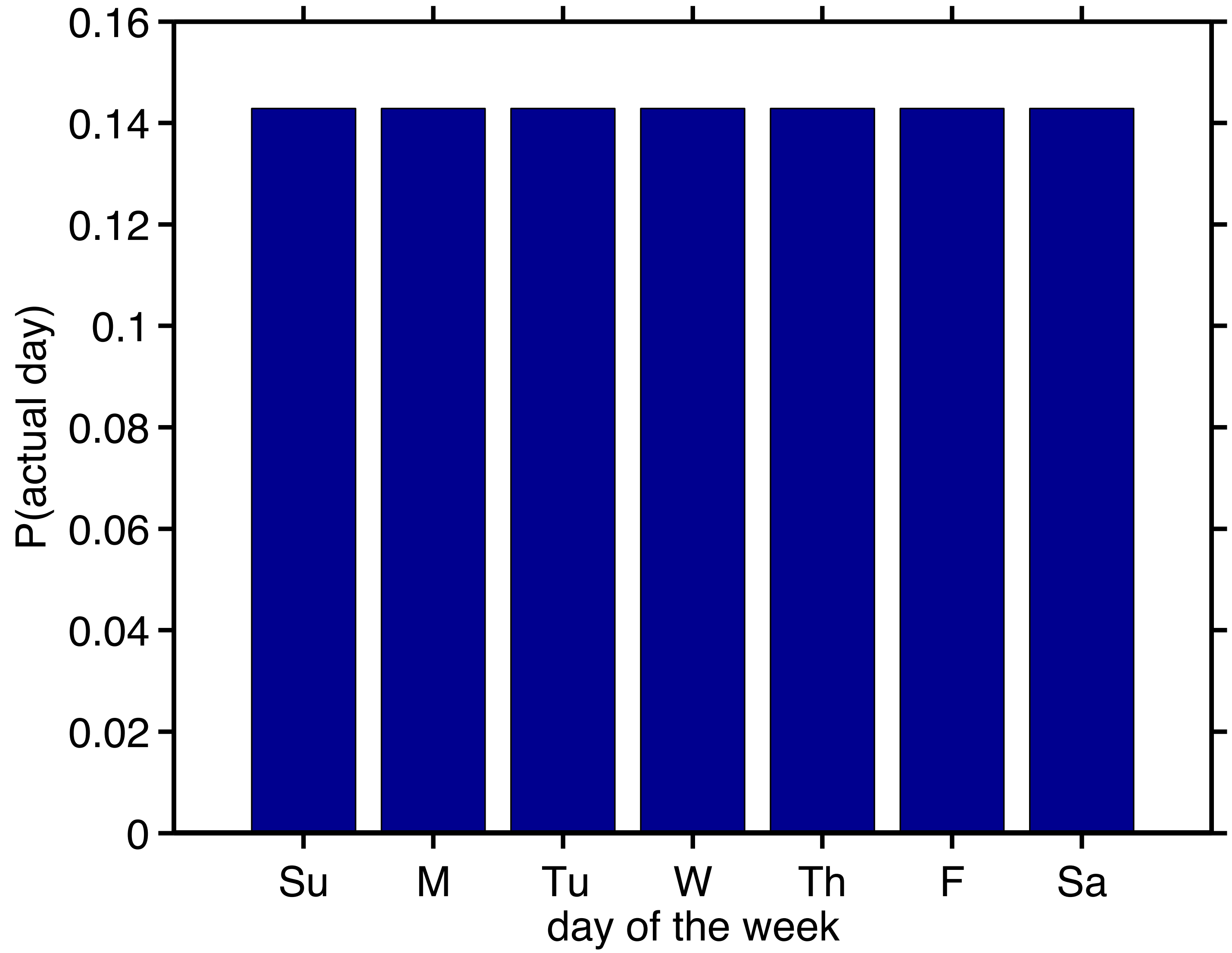
$$I(X;Y) = S(Y) - \langle S(Y|X) \rangle_x$$

$$I(X;Y) = \sum_{x,y} P(X,Y) \log_2 \left(\frac{P(X,Y)}{P(X)P(Y)} \right)$$

$$P(X,Y) = P(X|Y)P(Y)$$

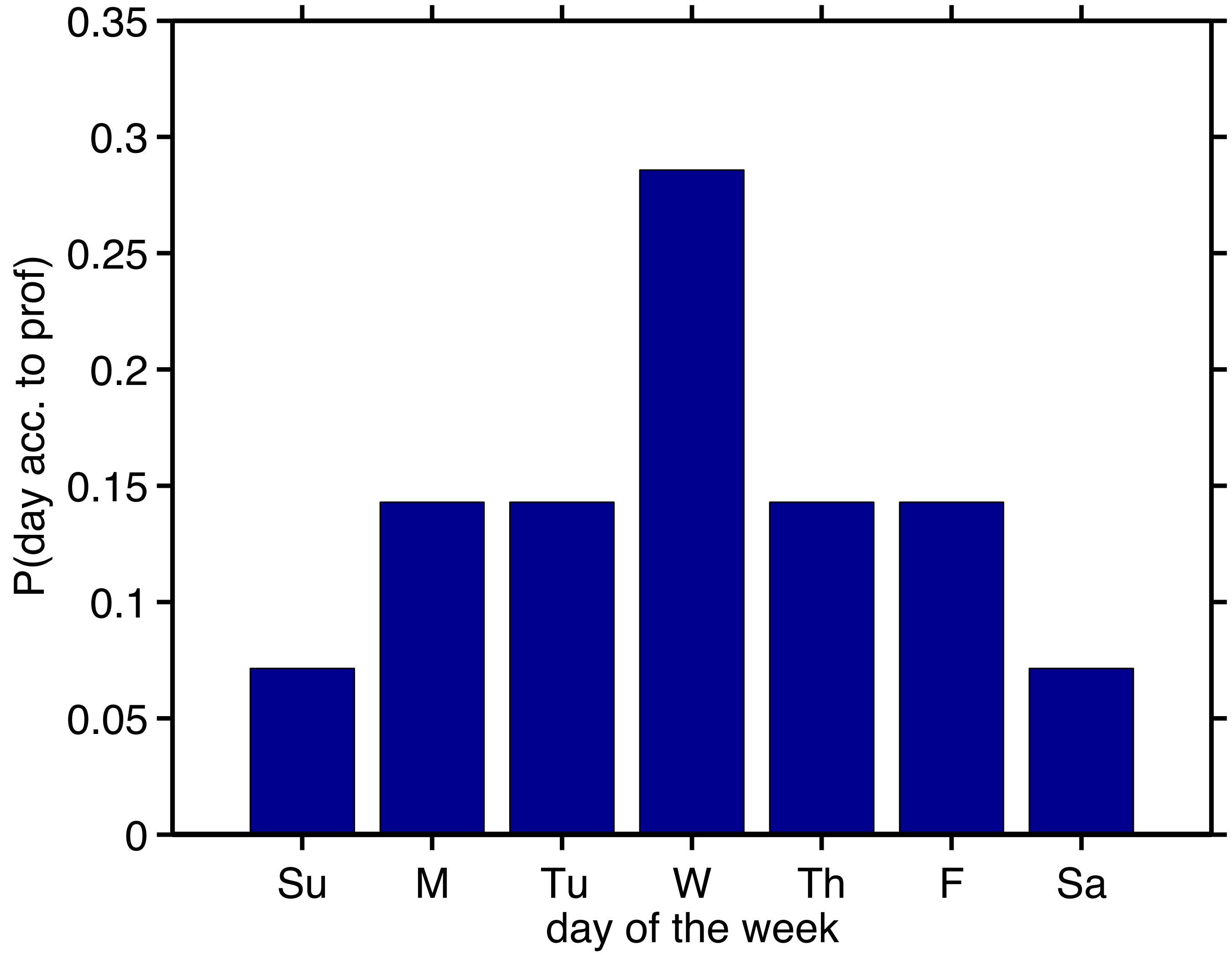
$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

What is our prior entropy?



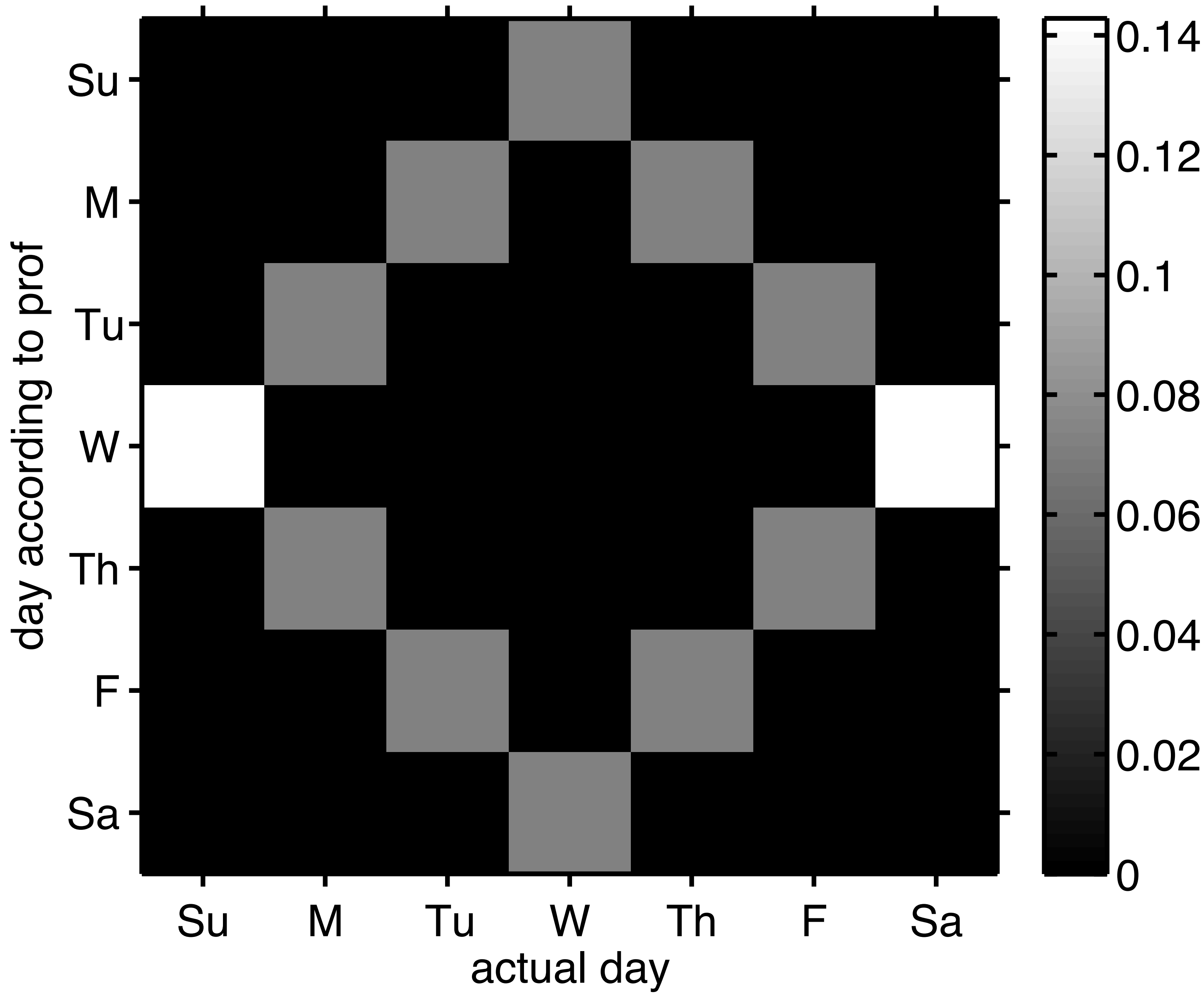
S(actual days) = 2.807 bits

How much can we possibly gain from the prof?

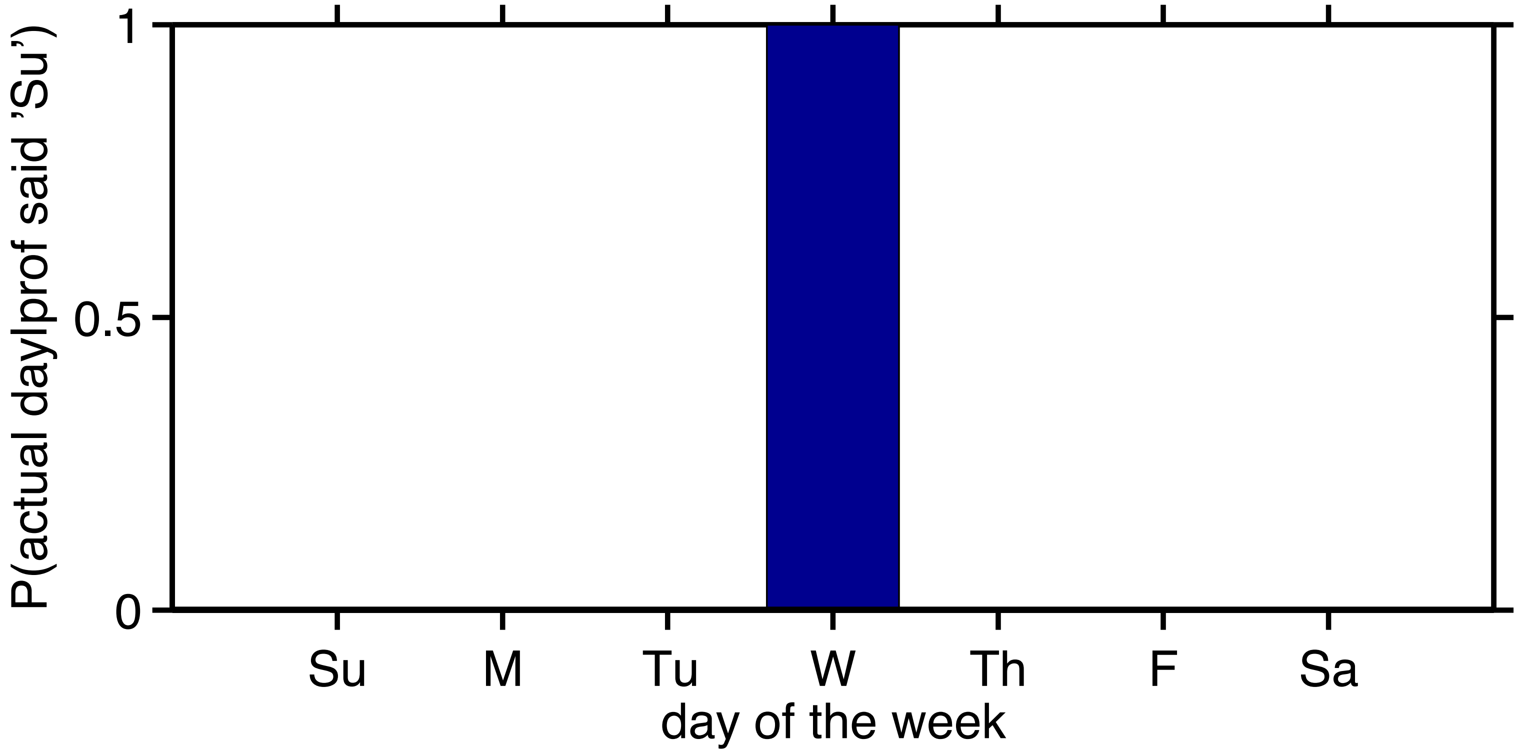


S(prof) = 2.665 bits

...given that this prof is evil:

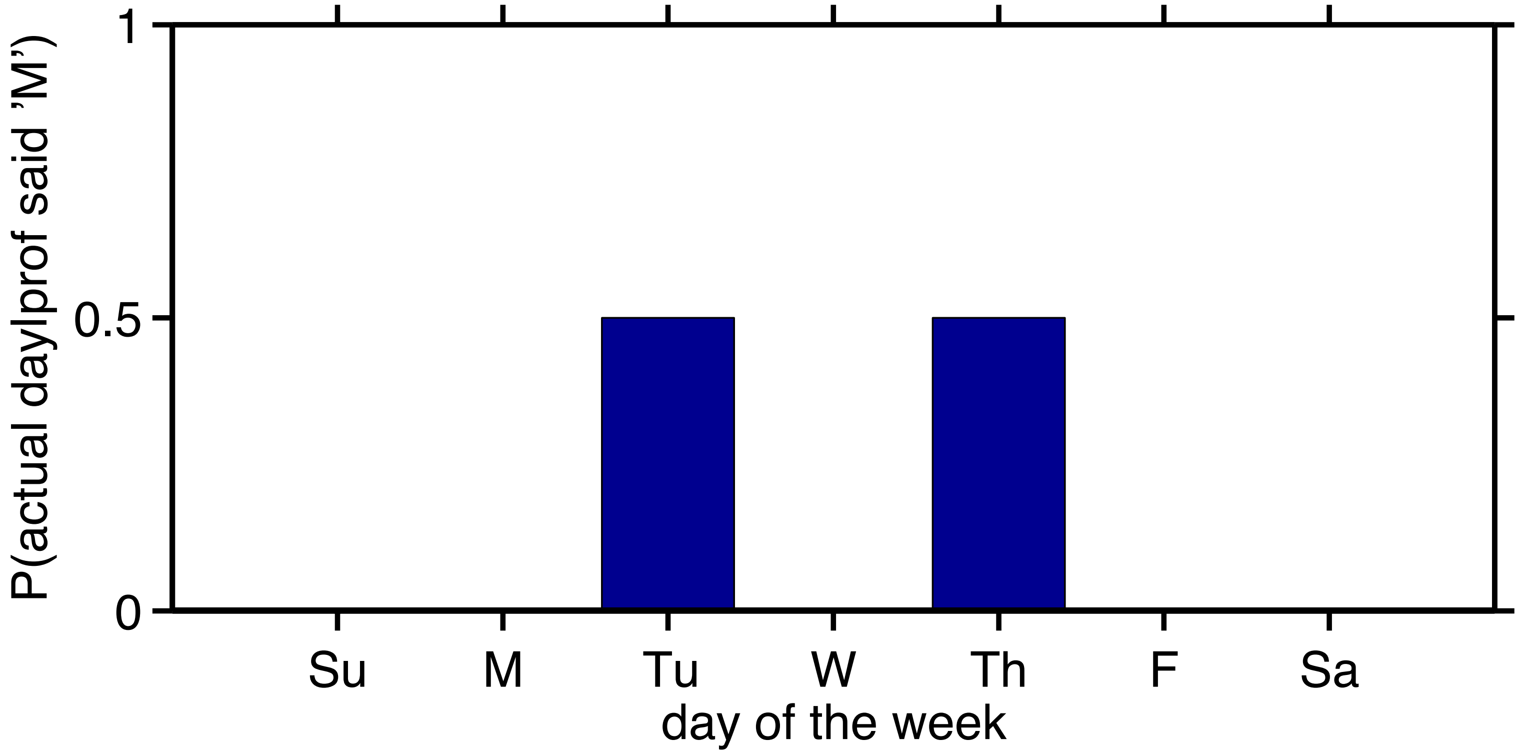


What are our conditional distributions?



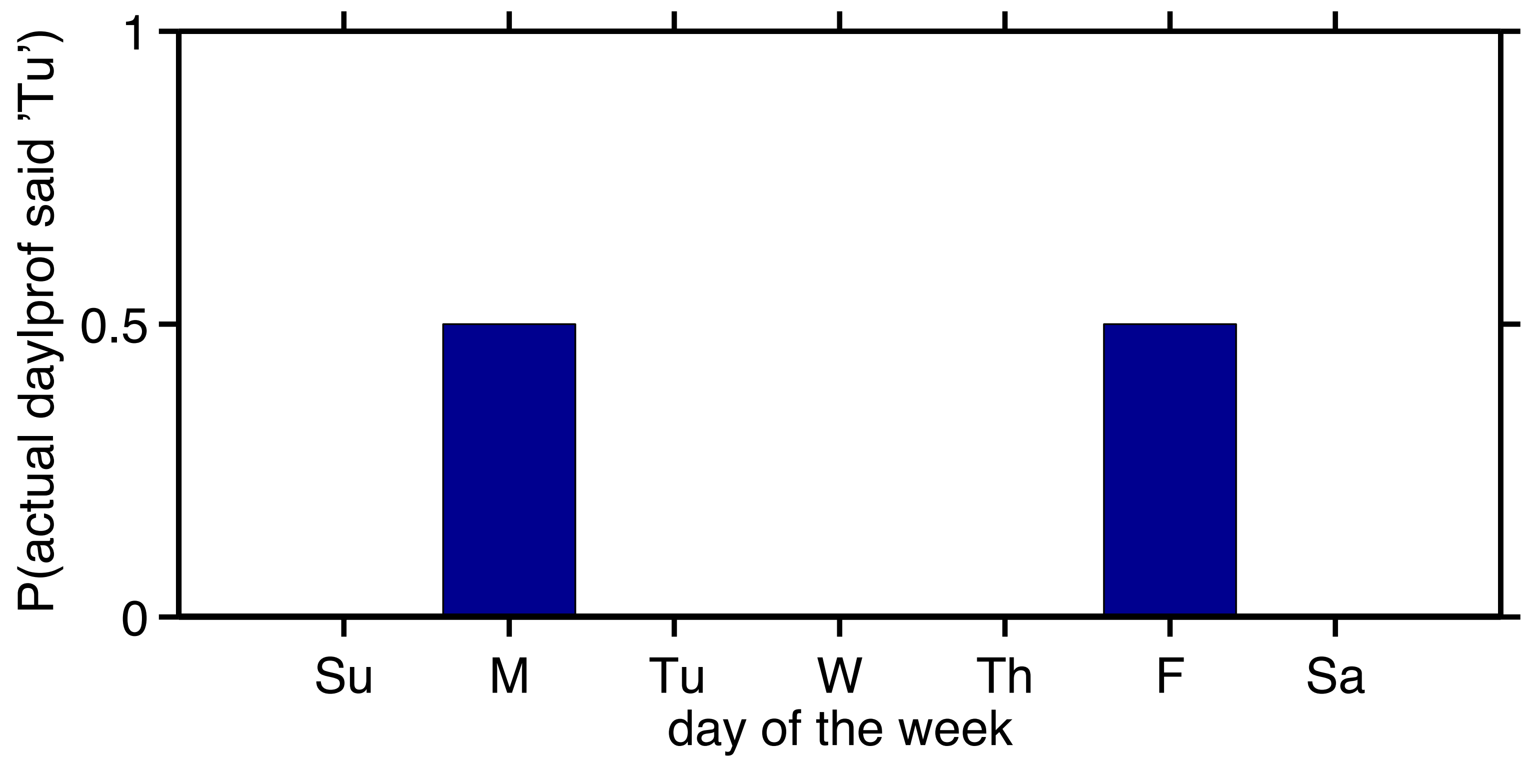
$$
\begin{aligned}
S(\text{days}|\text{prof} = \text{`Sunday'}) = & \\
& - \sum_{\text{days}} p(\text{day}|\text{prof} = \text{`Sunday'}) \log_2 \left( p(\text{day}|\text{prof} = \text{`Sunday'}) \right) \\
= & - p(\text{`Sunday'}|\text{prof} = \text{`Sunday'}) \log_2 \left(p(\text{`Sunday'}|\text{prof} = \text{`Sunday'}) \right) \\
& - p(\text{`Monday'}|\text{prof} = \text{`Sunday'}) \log_2 \left(p(\text{`Monday'}|\text{prof} = \text{`Sunday'}) \right) \\
& - \ldots \\
& - p(\text{`Saturday'}|\text{prof} = \text{`Sunday'}) \log_2 \left(p(\text{`Saturday'}|\text{prof} = \text{`Sunday'}) \right) \\
= & 0 + 0 + 0 + 1 \cdot 0 + 0 + 0 + 0 \\
= & 0
\end{aligned}
$$

What are our conditional distributions?



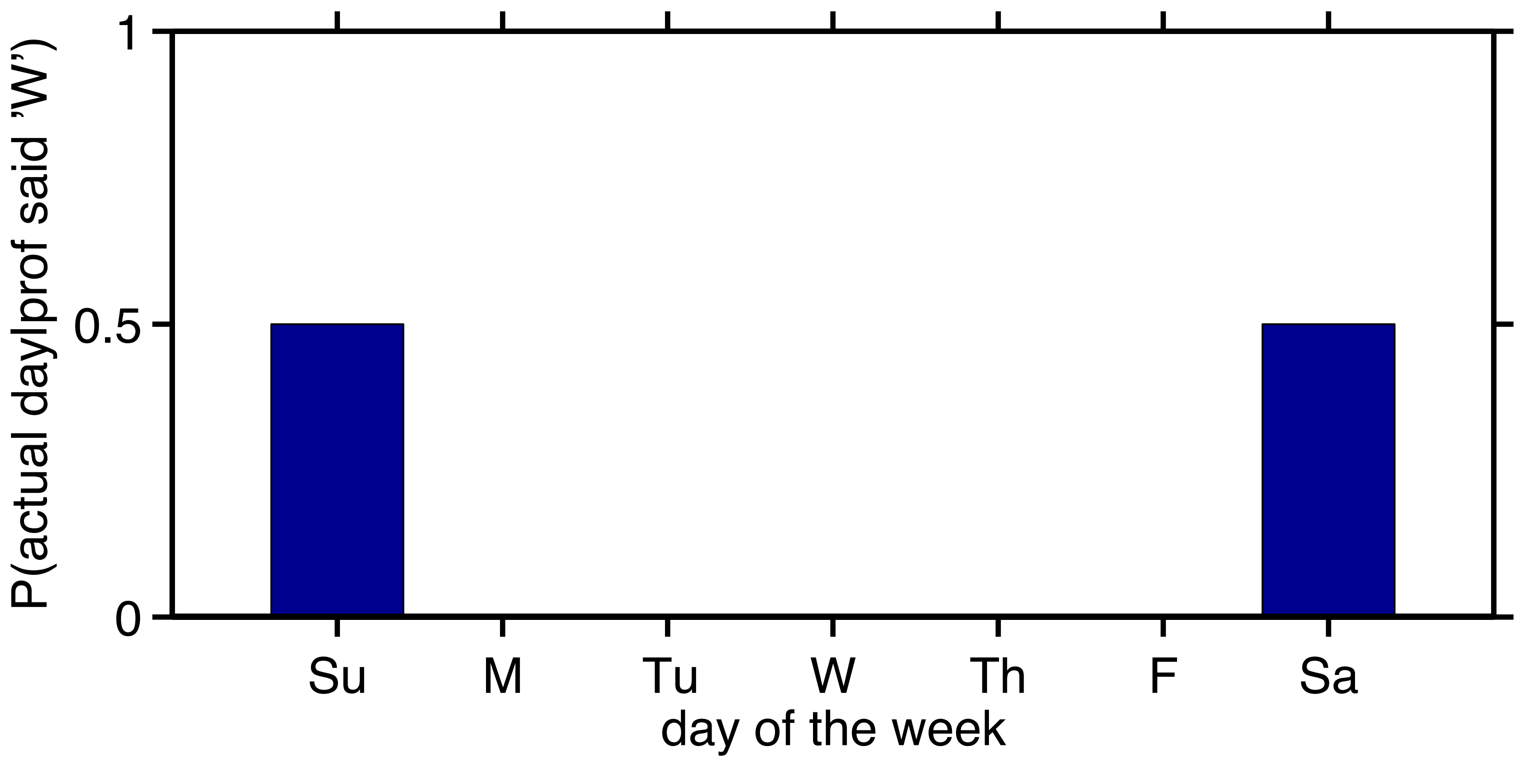
$$S(\text{days}|\text{prof} = \text{`Monday'}) = 1\text{bit}$$

What are our conditional distributions?



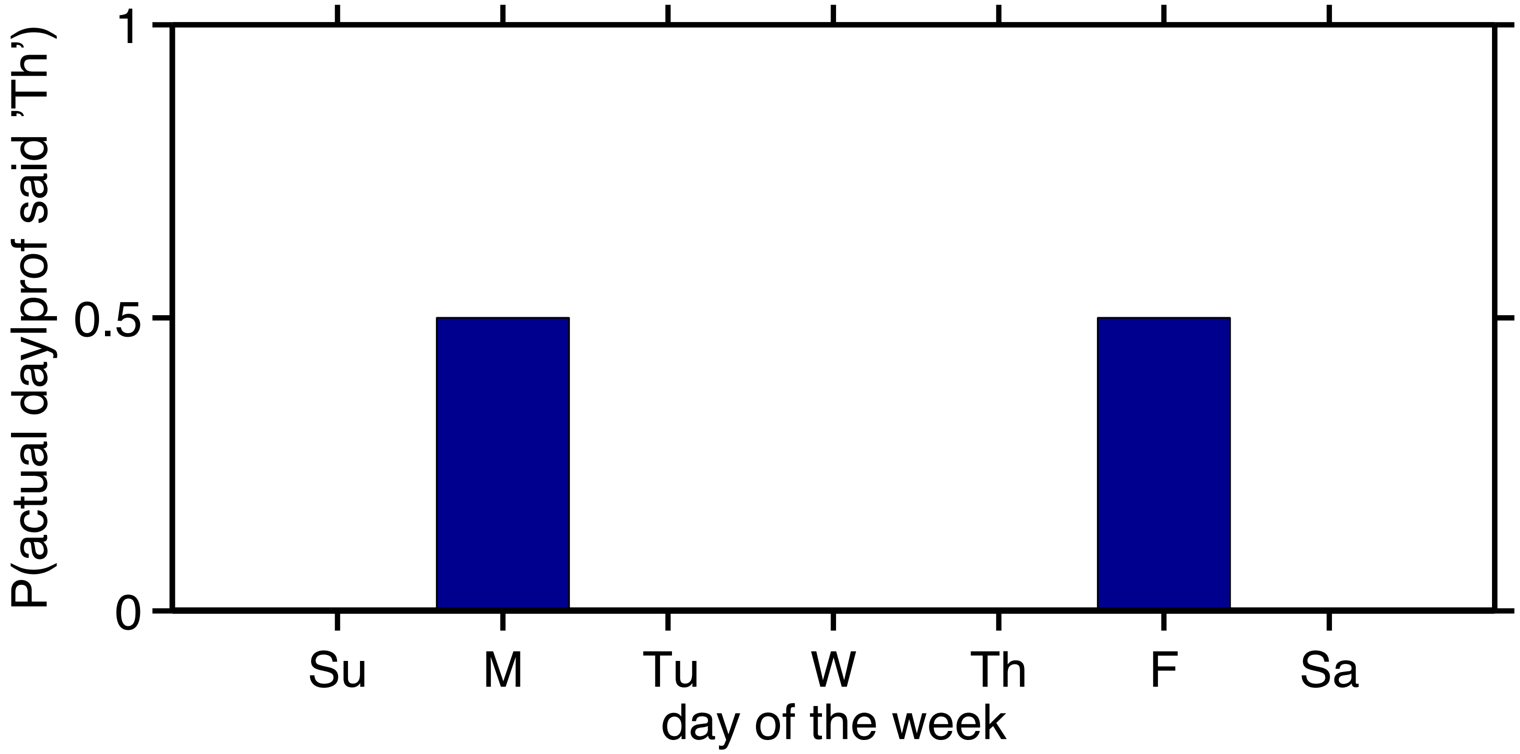
$$S(\text{days}|\text{prof} = \text{`Tuesday'}) = 1\text{bit}$$

What are our conditional distributions?



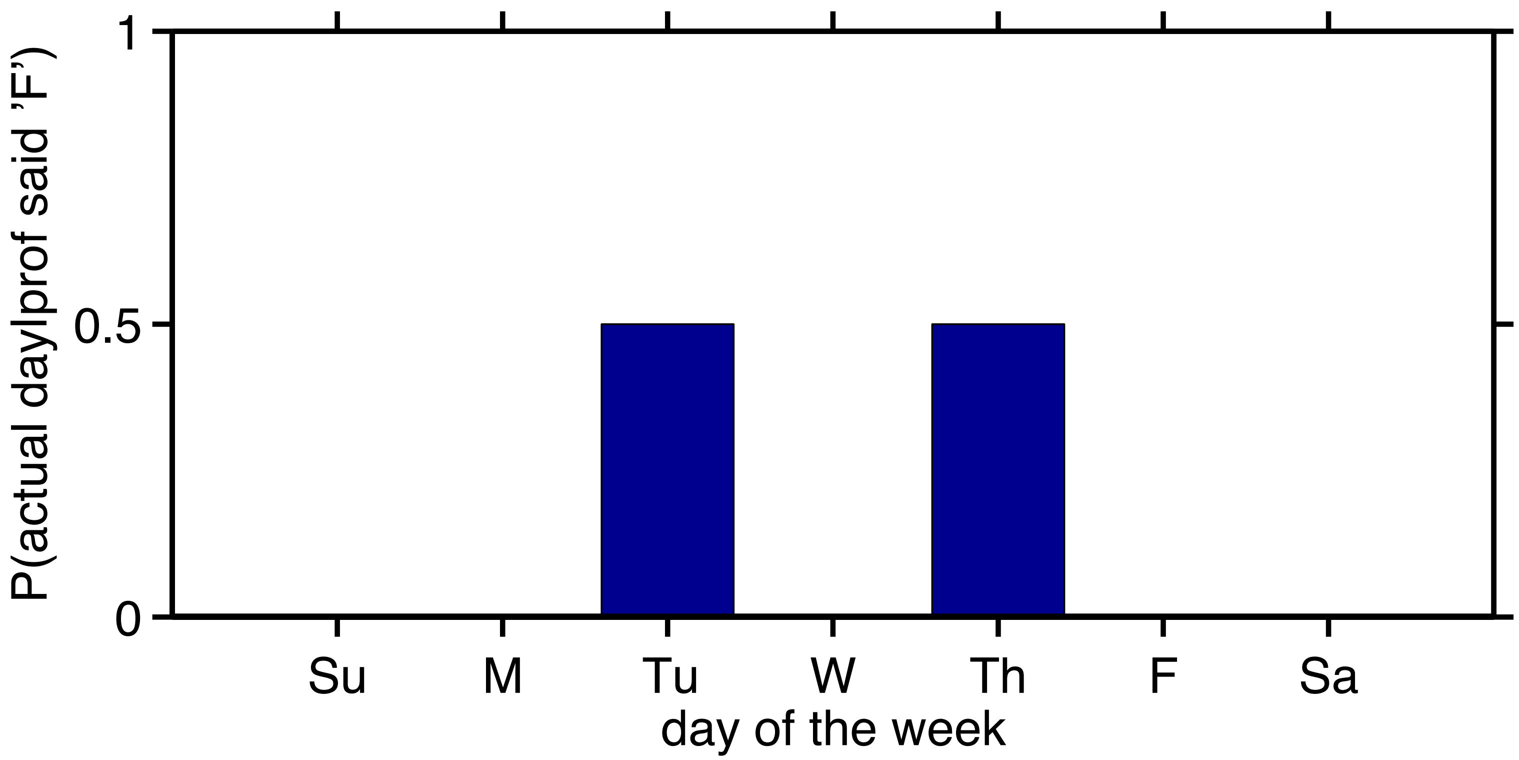
$$S(\text{days}|\text{prof} = \text{`Wednesday'}) = 1\text{bit}$$

What are our conditional distributions?



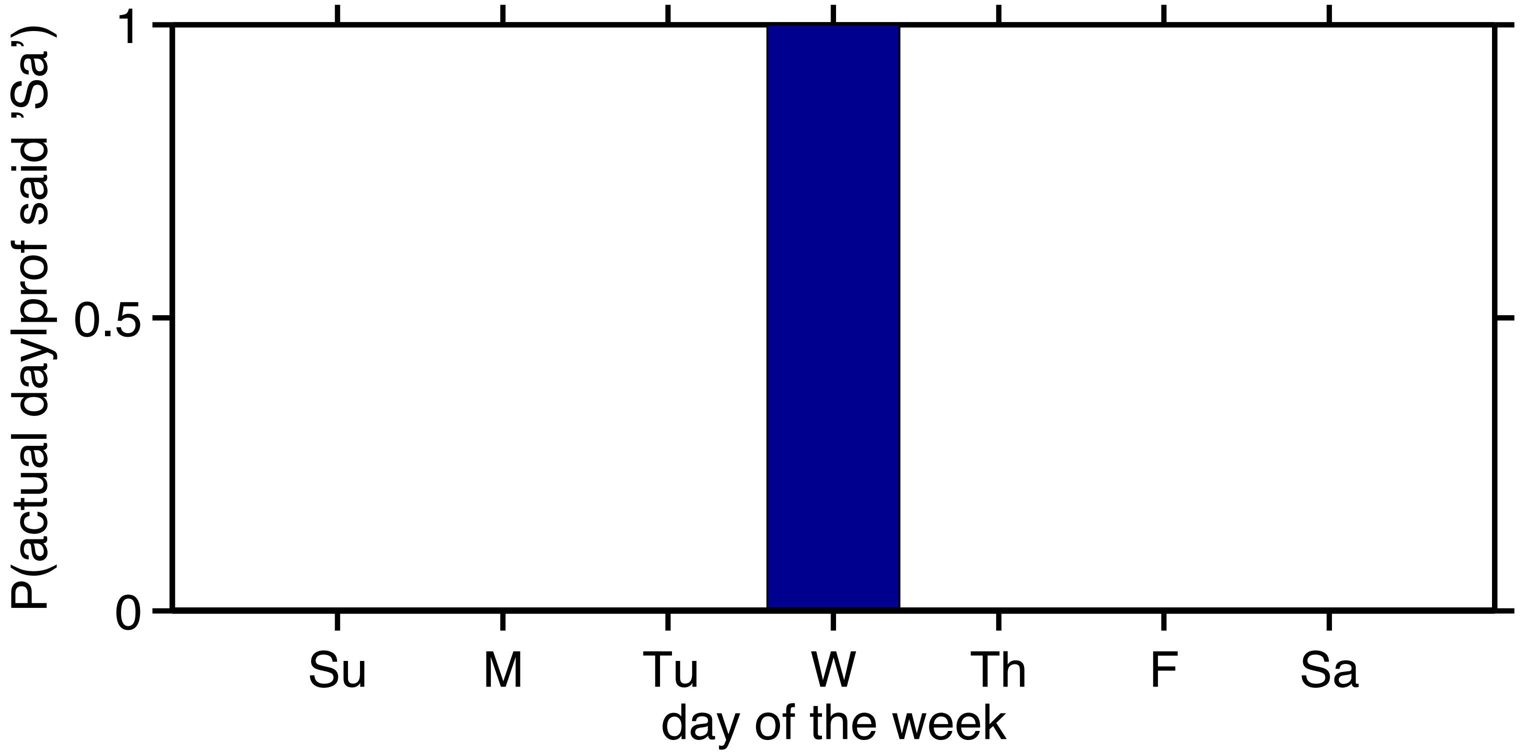
$$S(\text{days}|\text{prof} = \text{`Thursday'}) = 1\text{bit}$$

What are our conditional distributions?



$$S(\text{days}|\text{prof} = \text{'Friday'}) = 1\text{bit}$$

What are our conditional distributions?

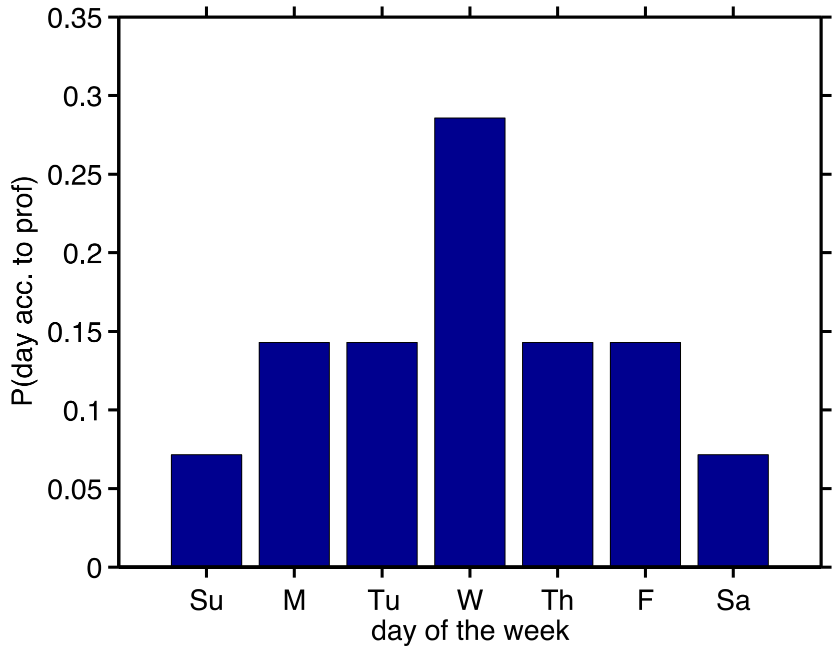


$$S(\text{days}|\text{prof} = \text{`Saturday'}) = 0\text{bits}$$

Total information prof can give about the days of the week:

$$I(X;Y) = S(X) - \langle S(X|Y) \rangle_y$$

$$\begin{aligned} I(\text{days};\text{prof}) =& S(\text{days}) + \\ & - S(\text{days}|\text{prof} = \text{`Sunday'}) p(\text{prof} = \text{`Sunday'}) \\ & - S(\text{days}|\text{prof} = \text{`Monday'}) p(\text{prof} = \text{`Monday'}) \\ & - \ldots \\ & - S(\text{days}|\text{prof} = \text{`Saturday'}) p(\text{prof} = \text{`Saturday'}) \end{aligned}$$

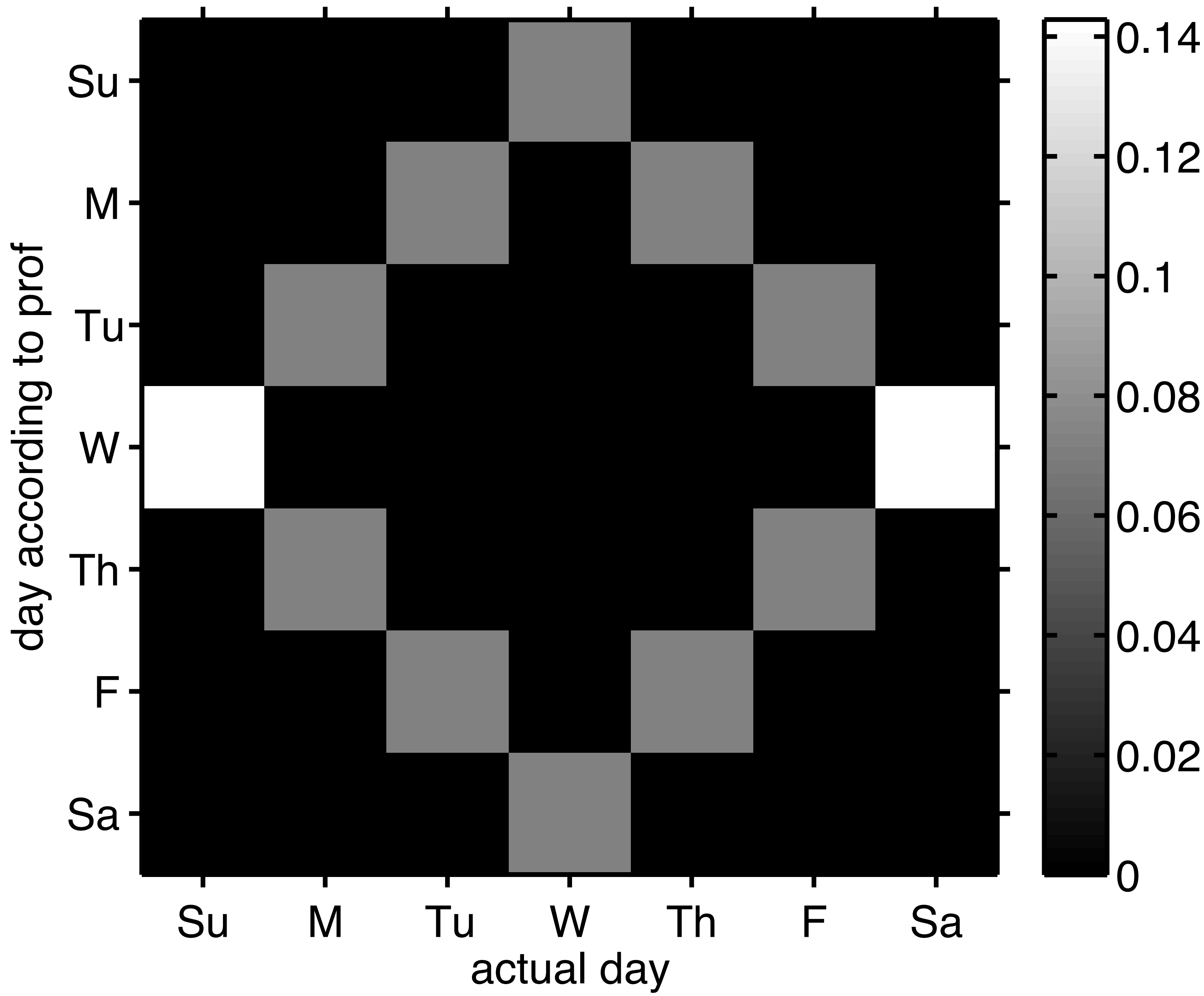


$$\begin{aligned} I(\text{days};\text{prof}) =& 2.807 + \\ & - 0 \cdot 0.07 \\ & - 1 \cdot 0.14 \\ & - \ldots \\ & - 0 \cdot 0.07 \\ & = 1.950 \text{bits} \end{aligned}$$

$$I(X;Y) = \sum_{x,y} P(X,Y) \log_2 \left(\frac{P(X,Y)}{P(X)P(Y)} \right)$$

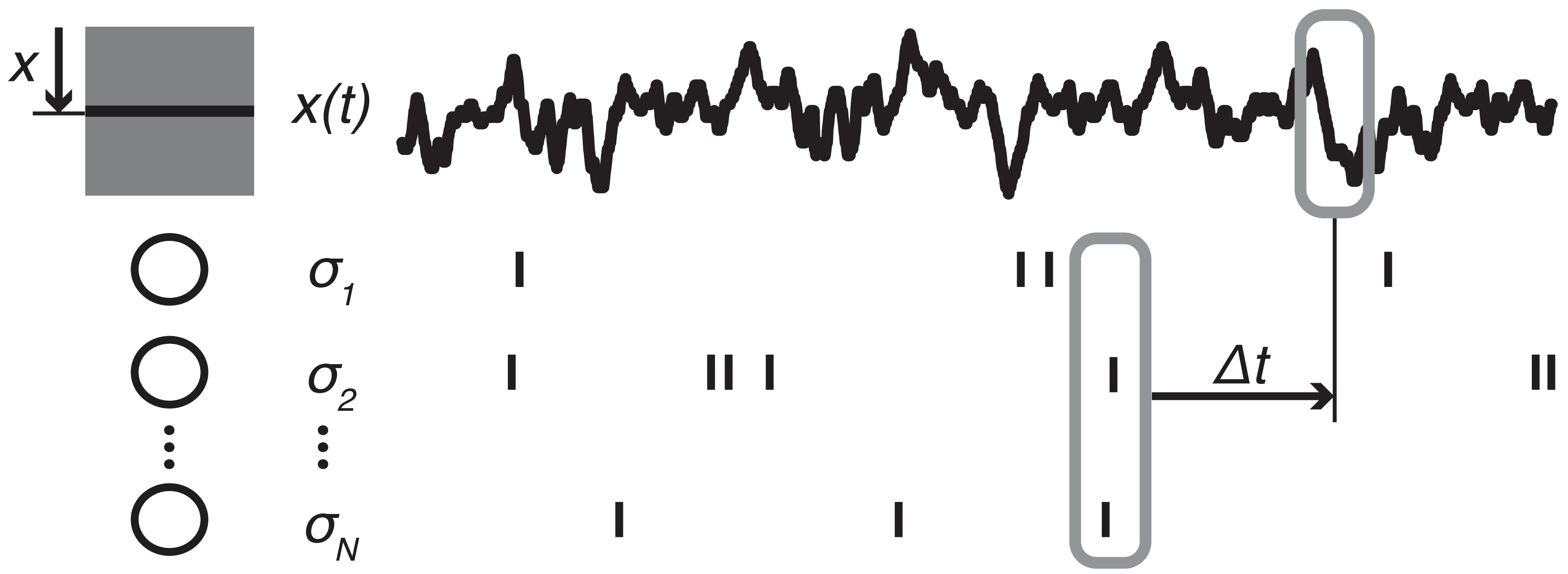
Total info:

$$I(X;Y) = \sum_{x,y} P(X,Y) \log_2 \left(\frac{P(X,Y)}{P(X)P(Y)} \right)$$



$$I(\text{days}; \text{prof}) = 1.950\text{bits}$$

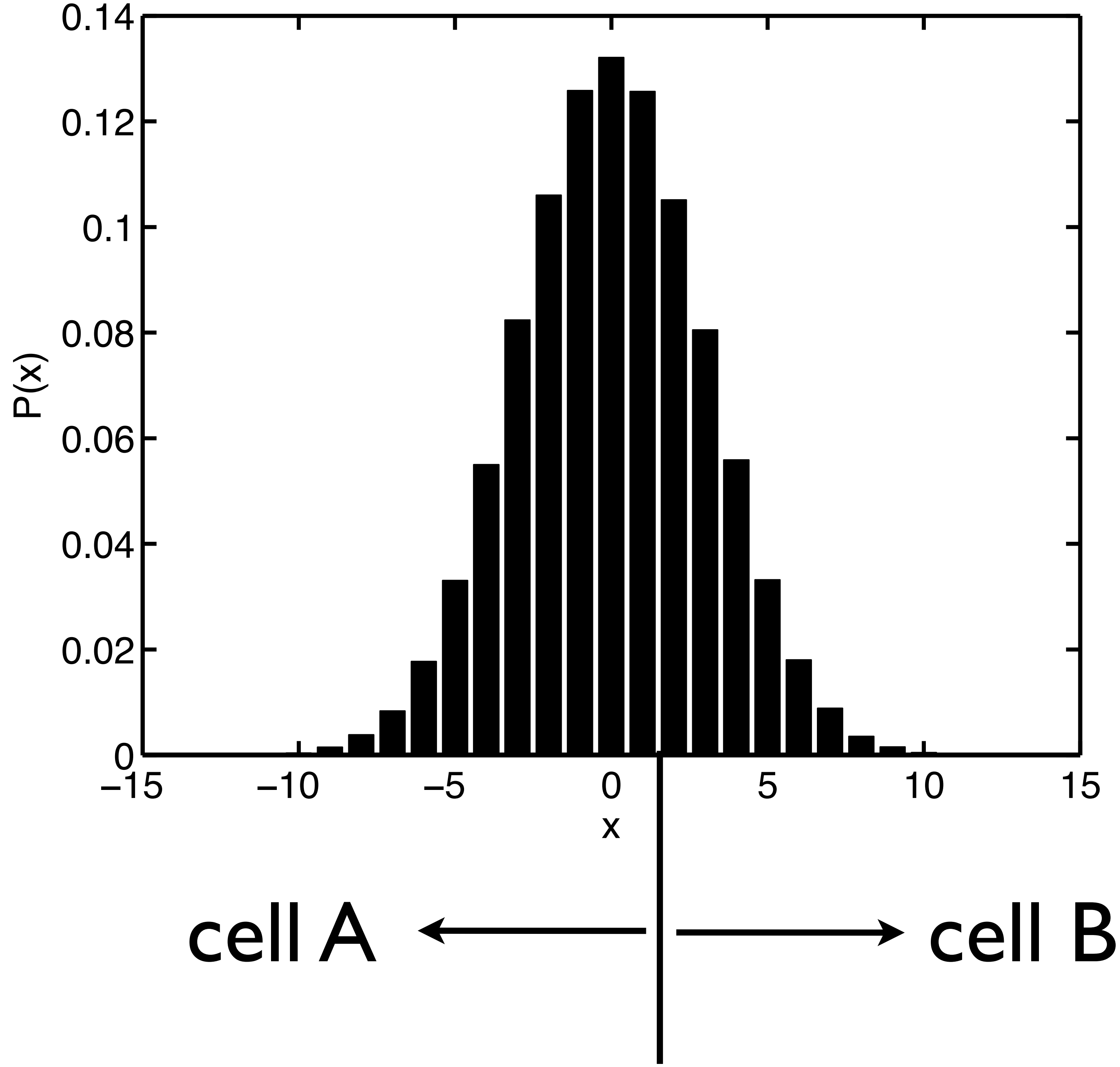
Example from a neural recording:



Let's examine $N = 2$ and $\Delta t = 20\text{ms}$

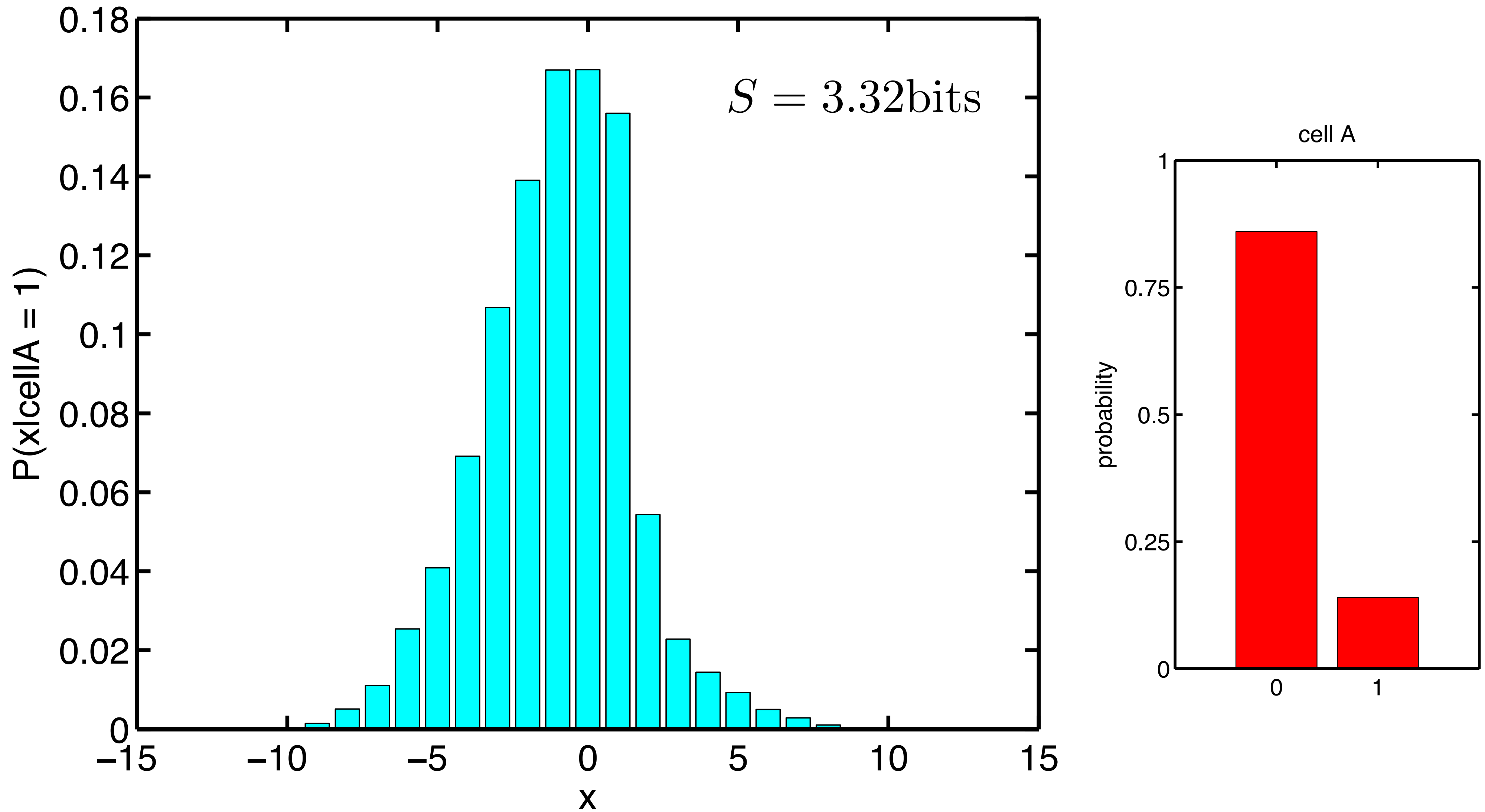
$$2^N = 4 \qquad \{(00), (01), (10), (11)\}$$

Prior distribution on the stimulus:



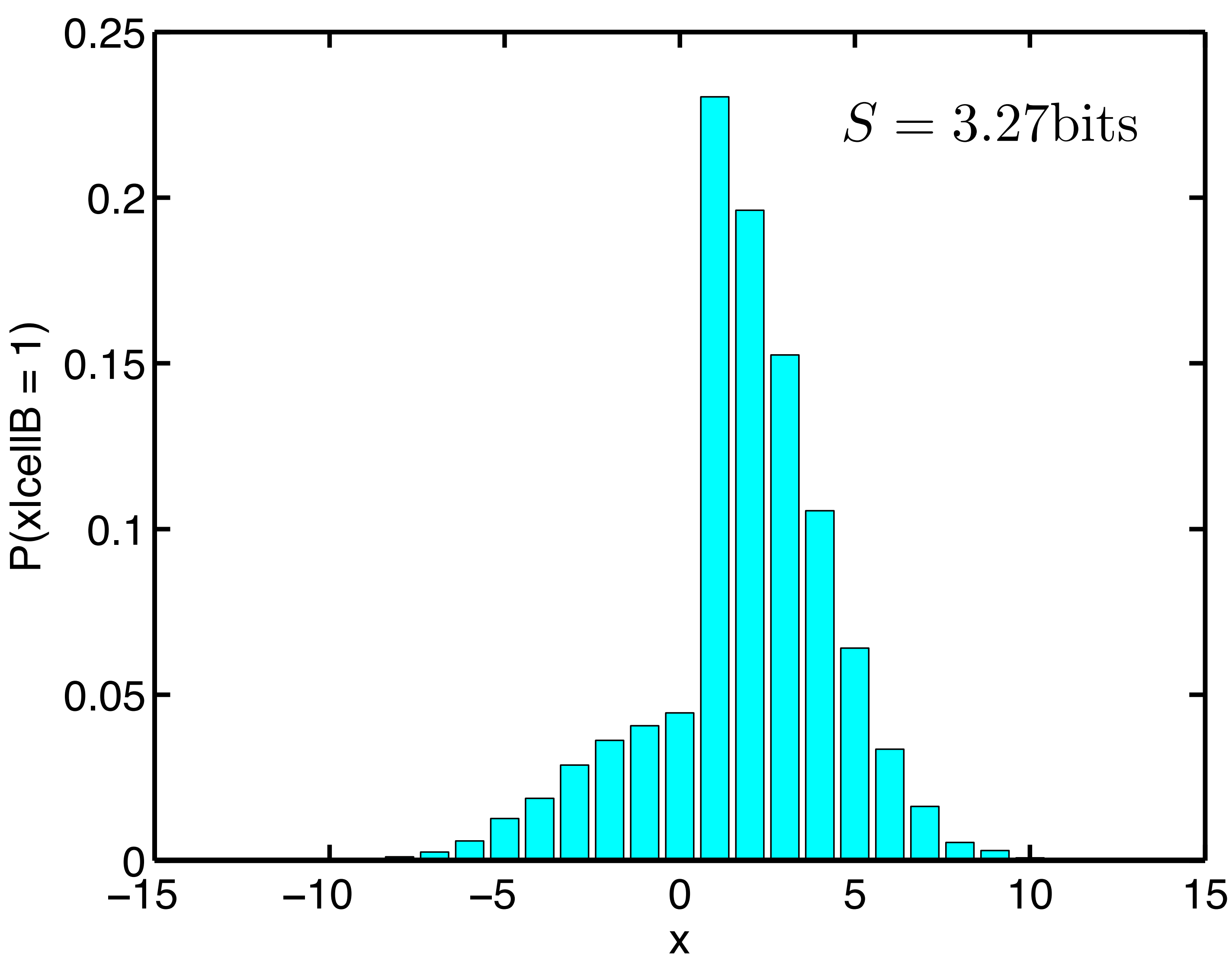
$S(x) = 3.64$ bits

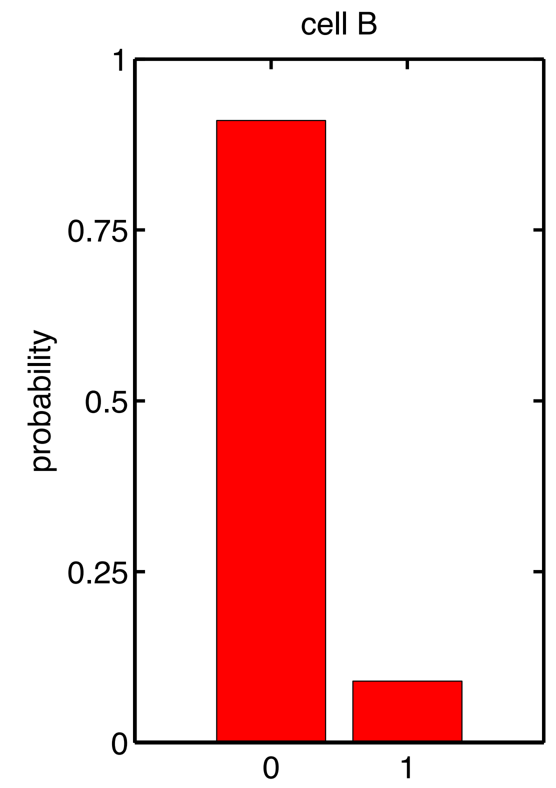
Conditional distribution for spikes in cell A:



$$\begin{aligned} I(\text{cell A}; x) =& S(x) - S(x|\text{cell A} = 1)p(\text{cell A} = 1) - S(x|\text{cell A} = 0)p(\text{cell A} = 0) \\ =& 3.64 - 3.32 \cdot 0.14 - 3.62 \cdot 0.86 \\ =& 0.027\text{bits} \end{aligned}$$

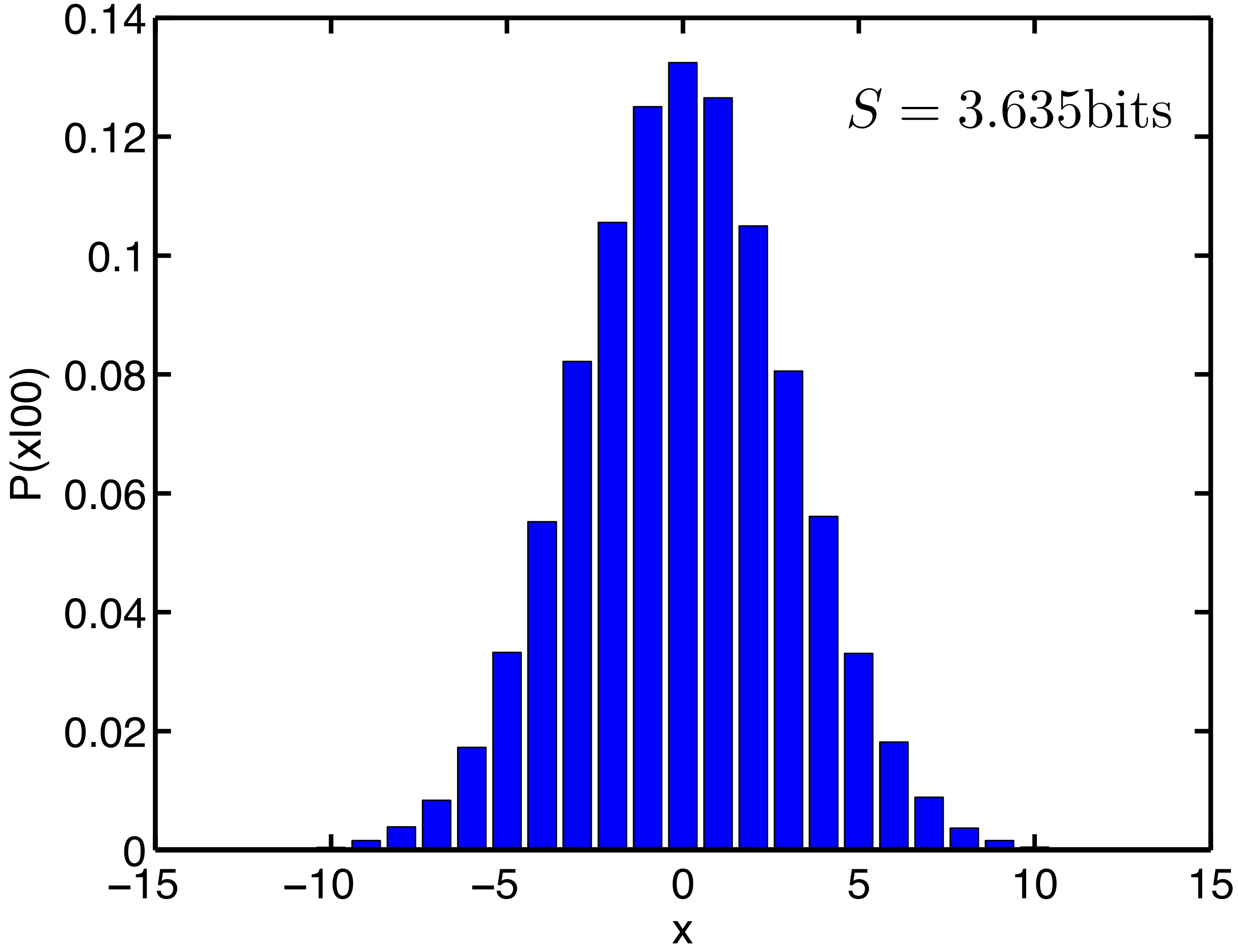
Conditional distribution for spikes in cell B:

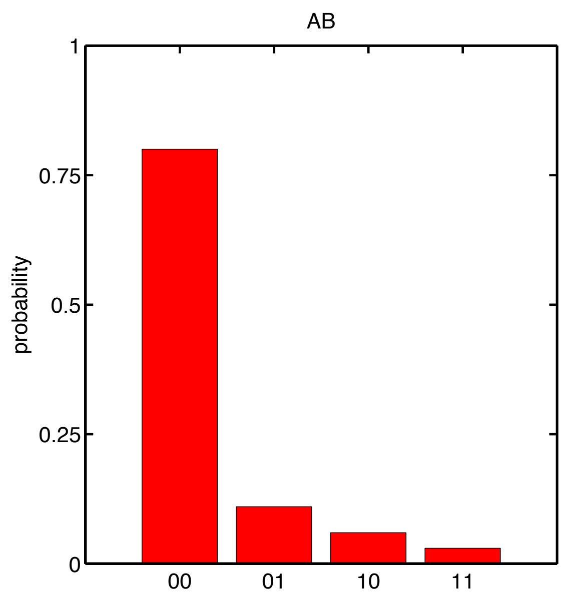




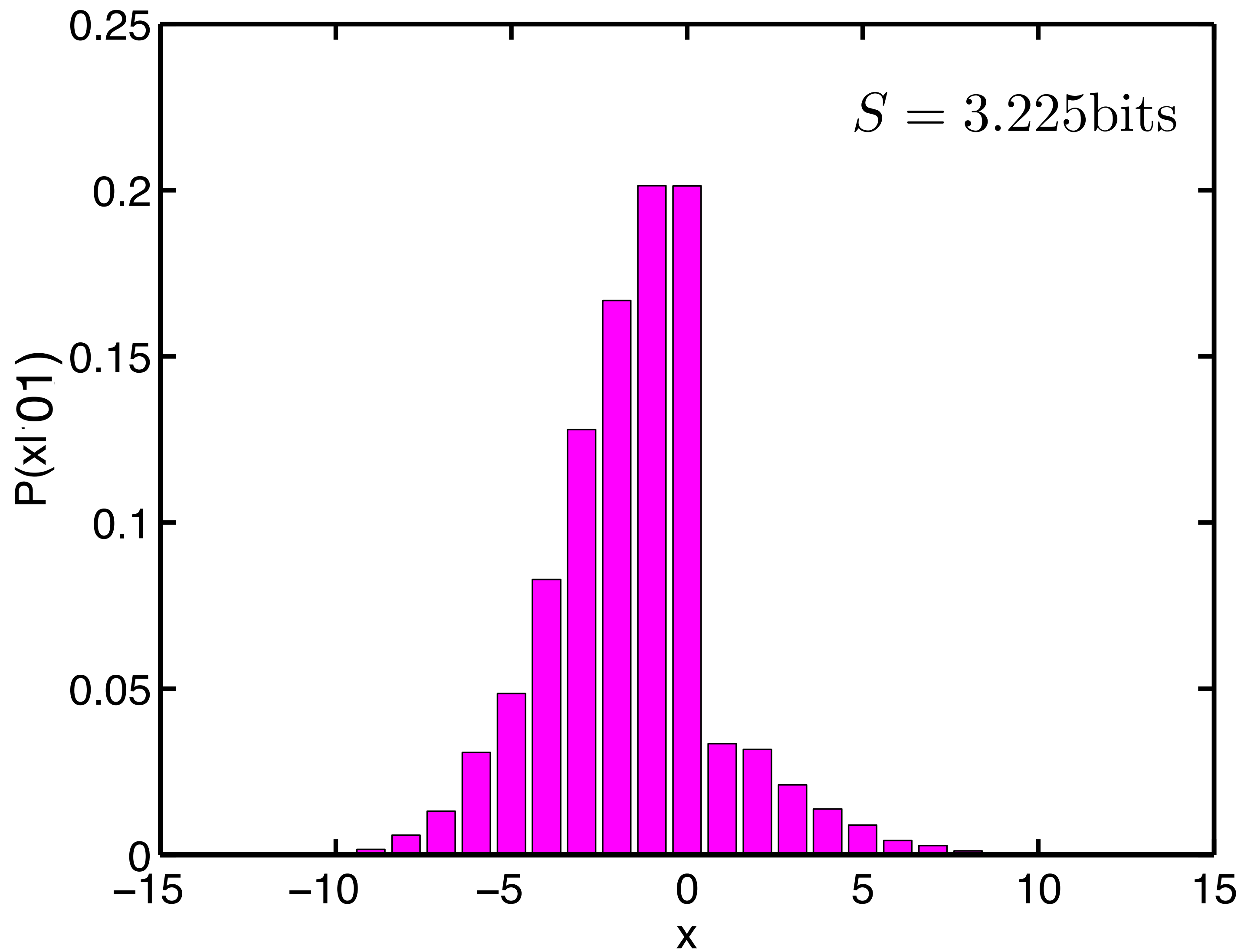
$$
\begin{aligned}
I(\text{cell B}; x) =& S(x) - S(x|\text{cell B} = 1)p(\text{cell B} = 1) - S(x|\text{cell B} = 0)p(\text{cell B} = 0) \\
=& 3.64 - 3.27 \cdot 0.09 - 3.62 \cdot 0.91 \\
=& 0.043\text{bits}
\end{aligned}
$$

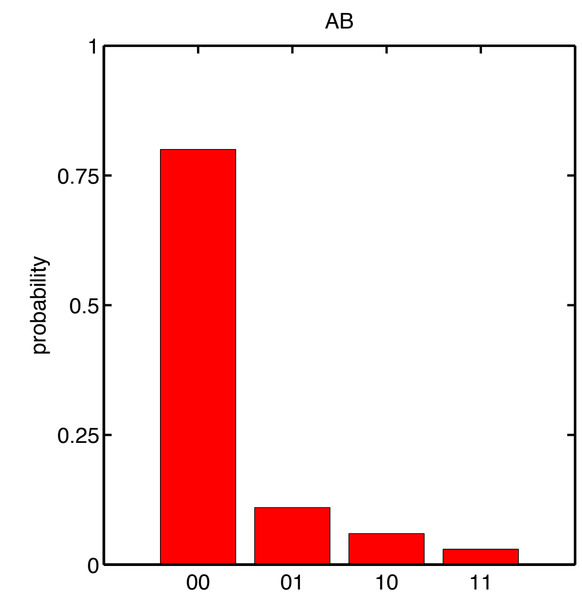
Patterns across both cells: both silent (00)



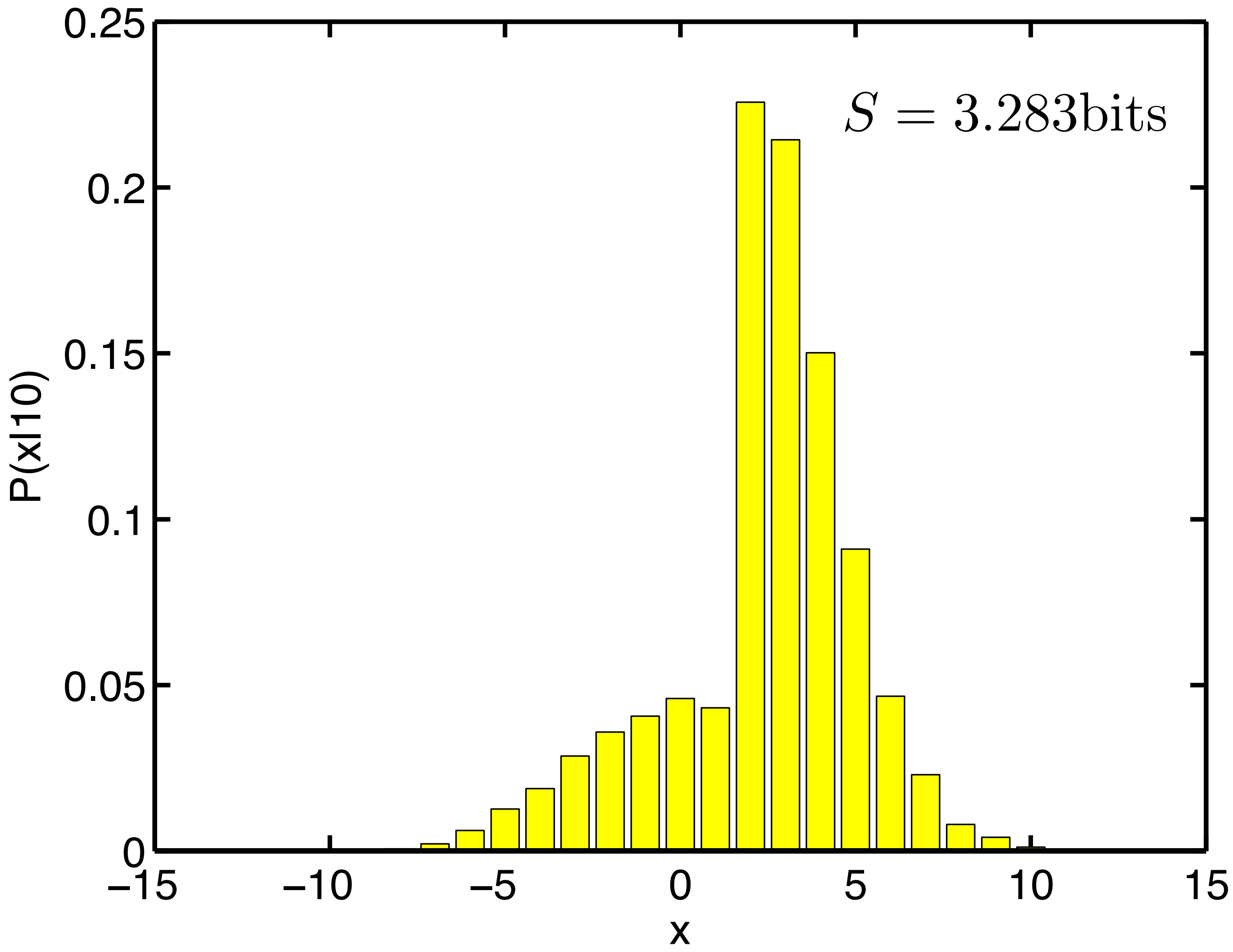


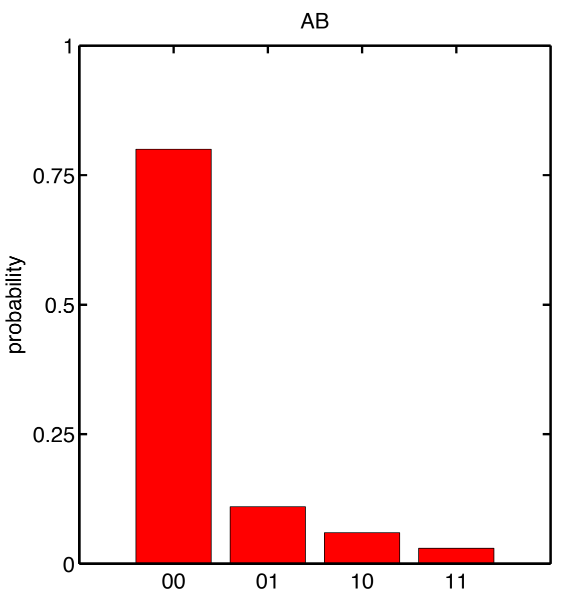
Patterns across both cells: A only (01)



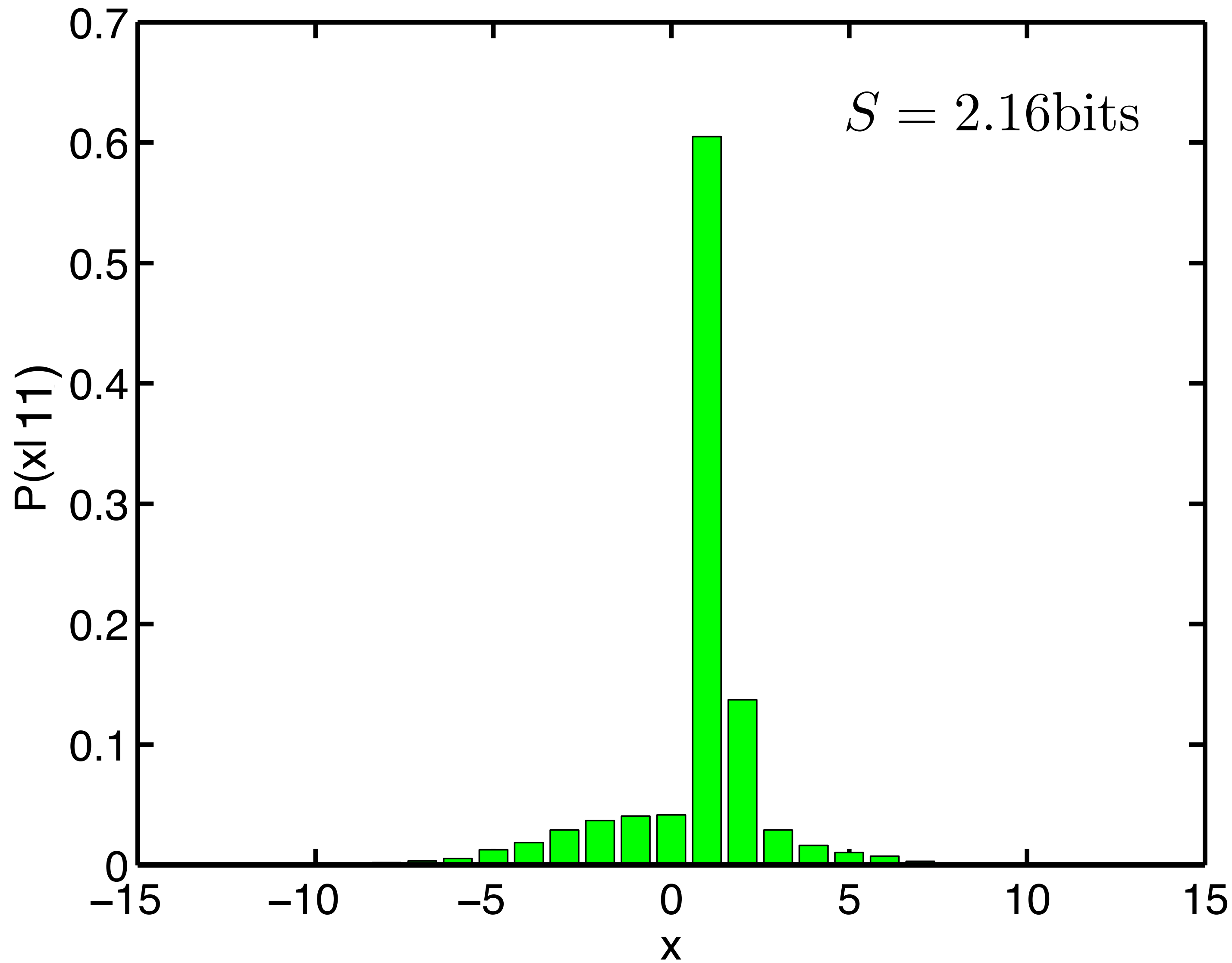


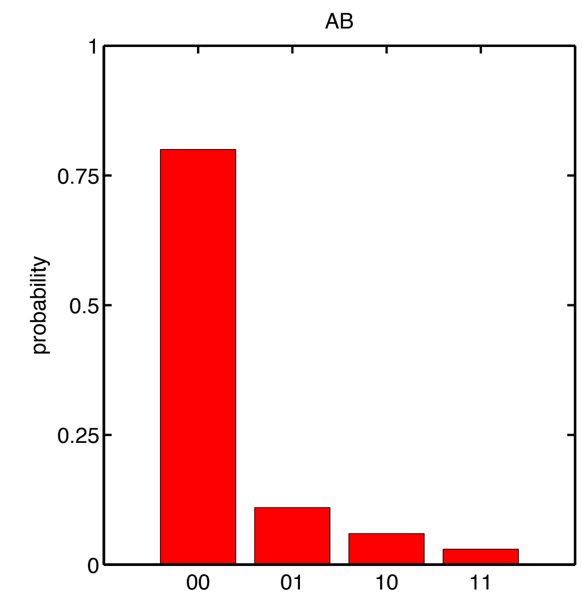
Patterns across both cells: B only (10)





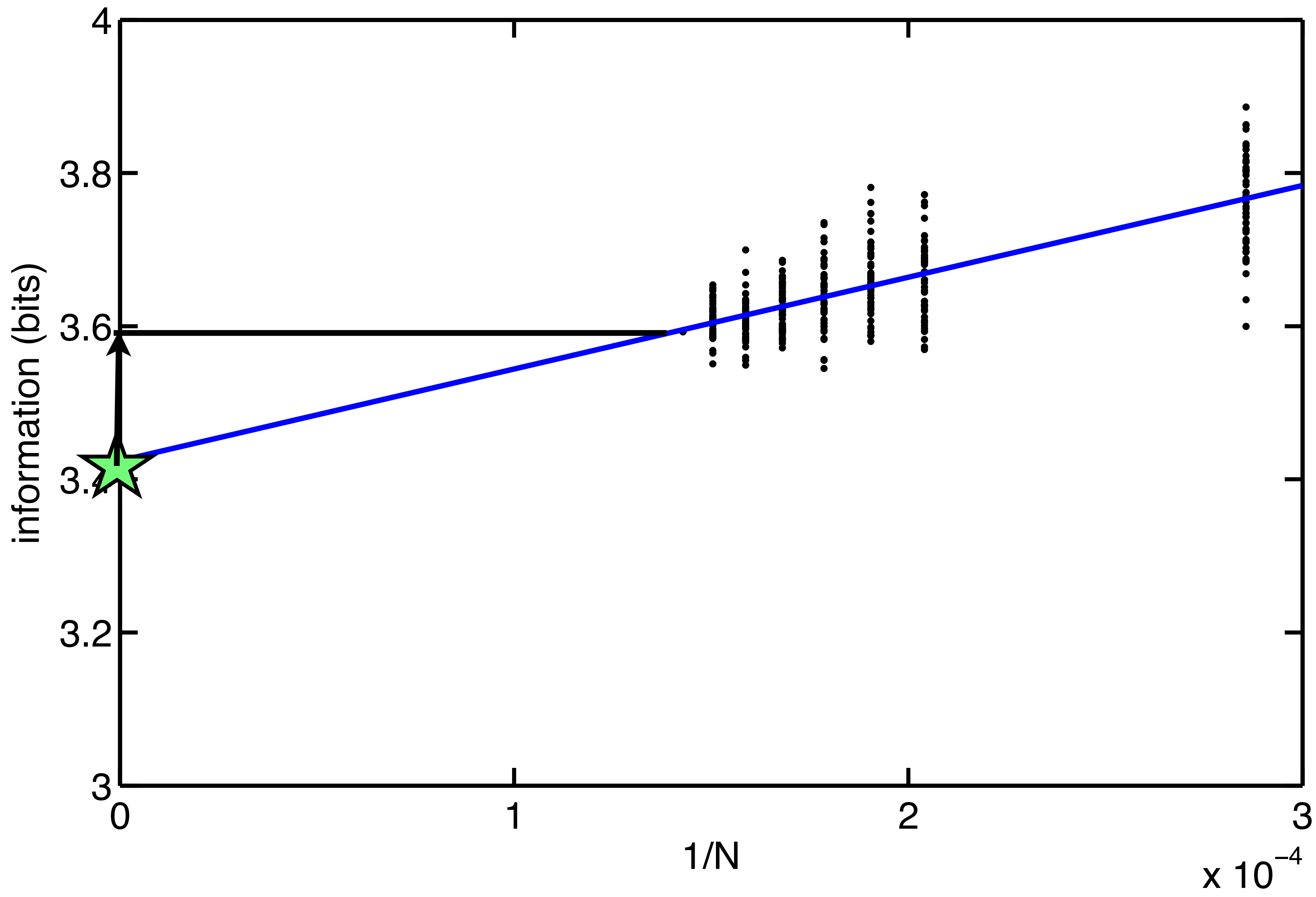
Patterns across both cells: both spiking (11)





$$
\begin{aligned}
I(AB;x) =& S(x) - S(x|00)p(00) - S(x|01)p(01) - S(x|10)p(10) - S(x|11)p(11) \\
=& 3.64 - 3.635 \cdot 0.8 - 3.225 \cdot 0.11 - 3.283 \cdot 0.06 - 2.16 \cdot 0.03 \\
=& 0.11\text{bits} \quad > 0.043 \text{ bits} + 0.027 \text{ bits}
\end{aligned}
$$

It's important to estimate your sampling bias:



J Neurophysiol. 2007 Sep;98(3):1064-72. Epub 2007 Jul 5.
Correcting for the sampling bias problem in spike train information measures.
Panzeri S, Senatore R, Montemurro MA, Petersen RS.

The tip of the iceberg:

The main results of information theory and coding are more substantial:

source coding Th'm: gives bounds on compressibility of signals given their entropy

channel coding Th'm: describes the limit of the information rate of a noisy channel, and the maximal efficiency of error-correcting codes