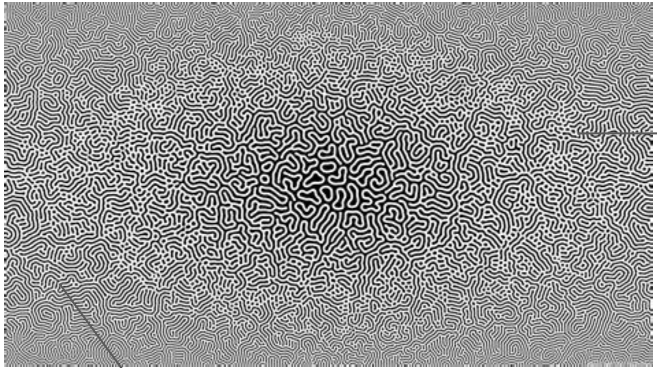
Measuring the Organization in Self-Organizing Perceptual Maps

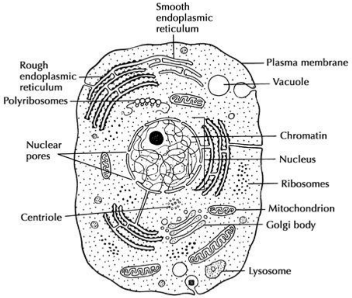
Jules Litman-Cleper Physics 256B, Spring 2022

Broad Inspiration: How do levels of complexity emerge in nature

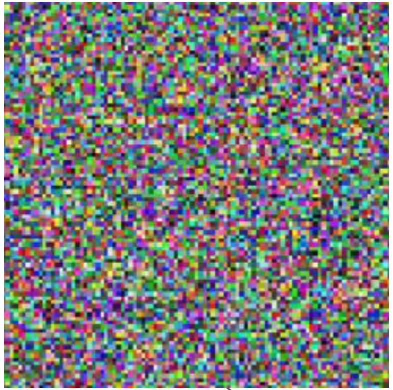
Evolution: Emergence of DNA from soup

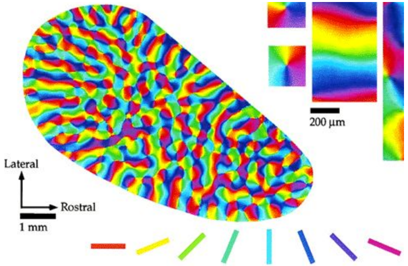


Encoding of information

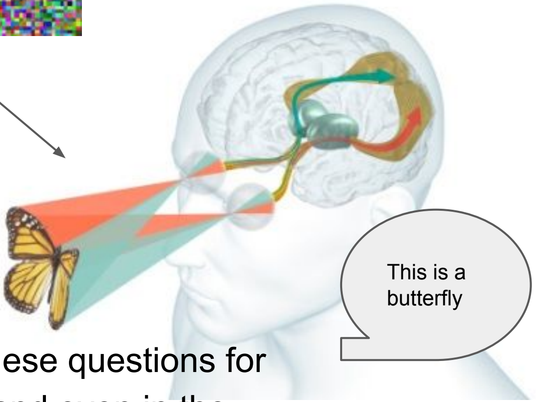


Cognition: Emergence of cognition(inference, learning, language) from perception





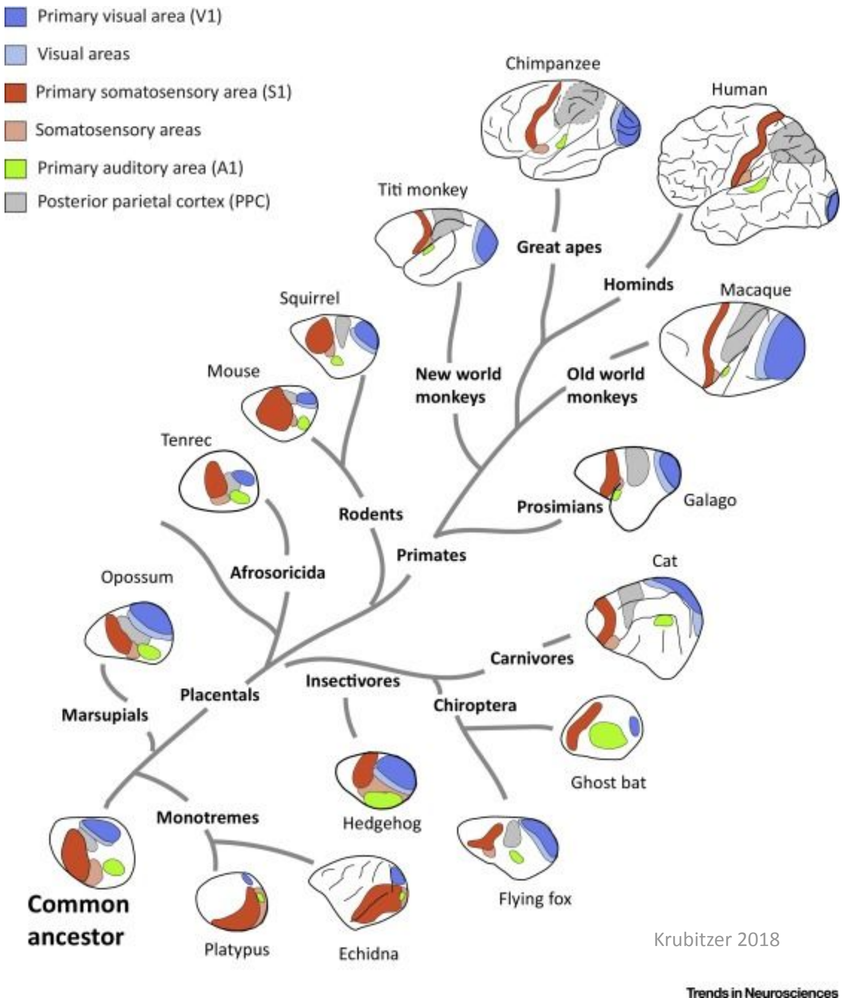
Encoding of information



This is a butterfly

As far as the **architecture of information** processing is concerned, these questions for **natural language** have **direct analogues in adaptation, evolution**, and even in the development of scientific theories [70]. - Crutchfield, *Calculi of Emergence*

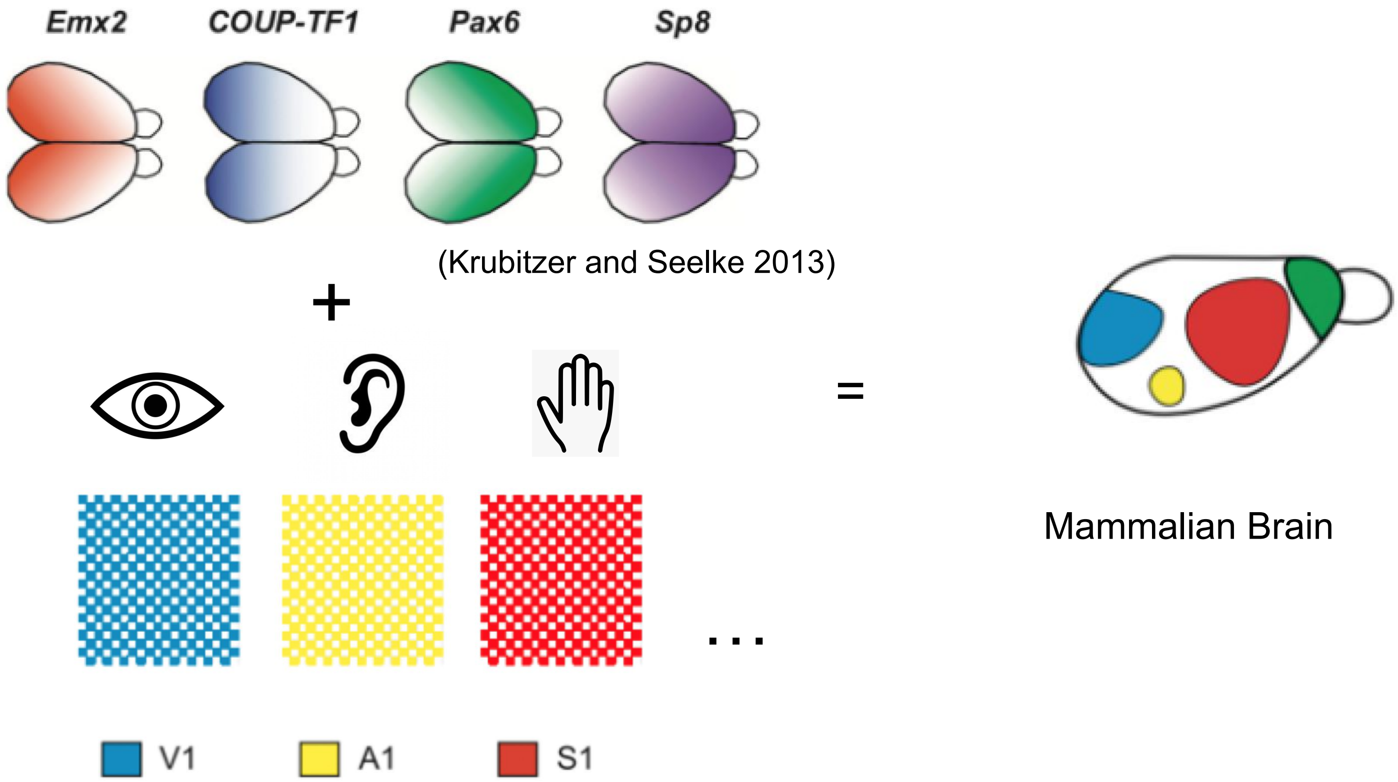
All mammals have a partitioned neocortex, partitions are called cortical fields



How do diverse cortical regions develop?

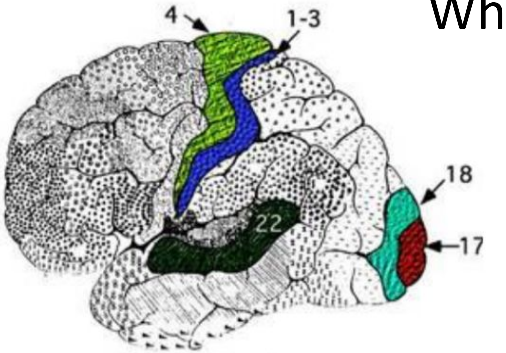
Graded Expression of transcription factors

Activity-dependent Sensory experience define feature maps

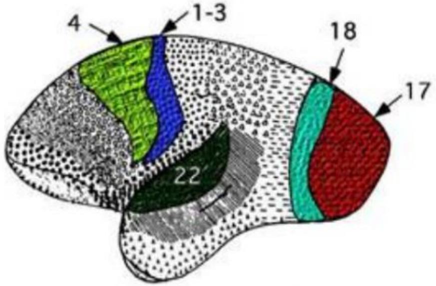


Both inherited genetic factors and activity-dependent sensory factors organize information in such a way as to produce adult brain phenotypes.

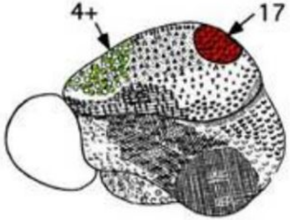
What is a cortical field, how are they defined?



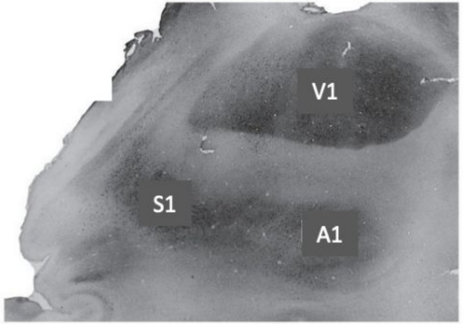
Homo sapiens, human

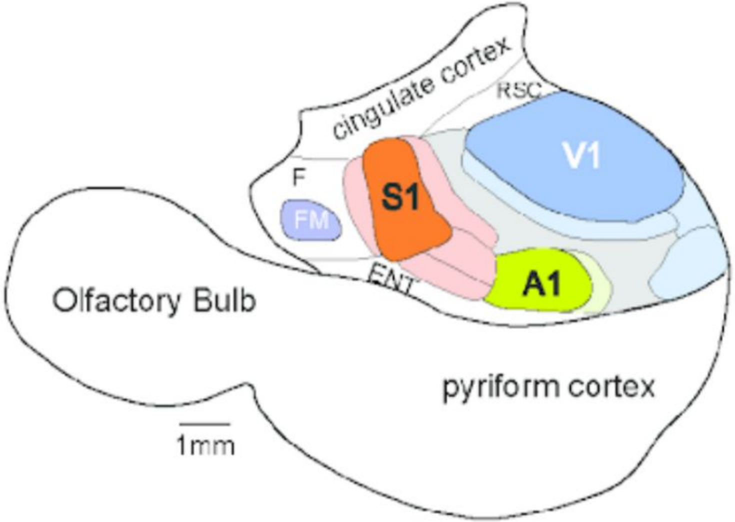


Callithrix jacchus, common marmoset

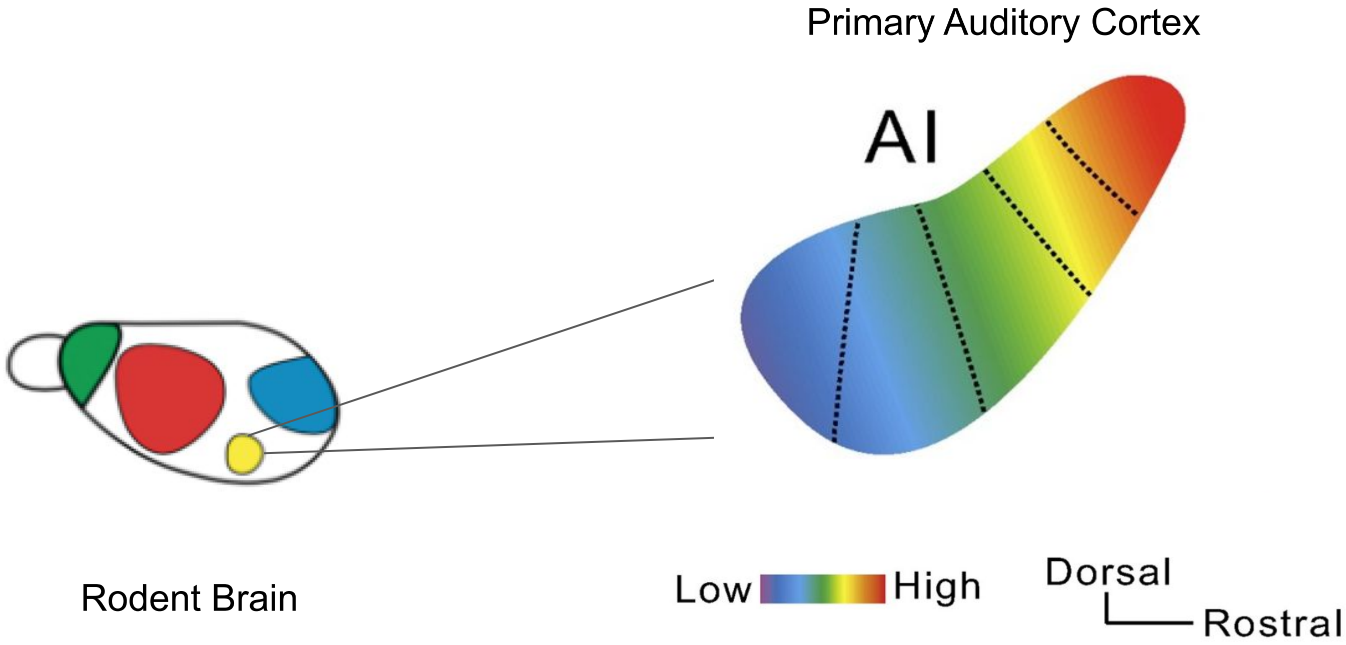


Erinaceus europaeus, European hedgehog



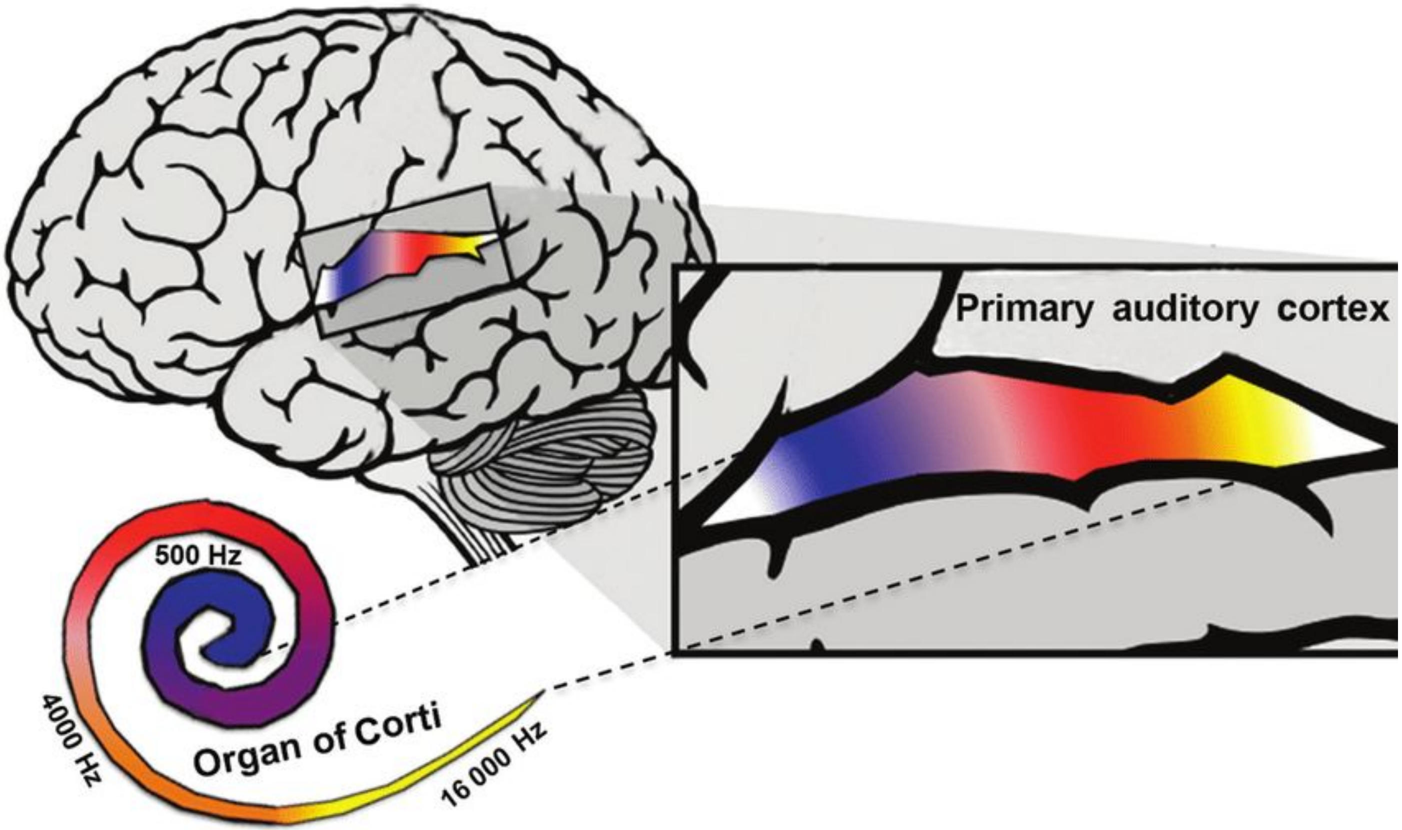


Structures within cortical fields: Topography and Topology



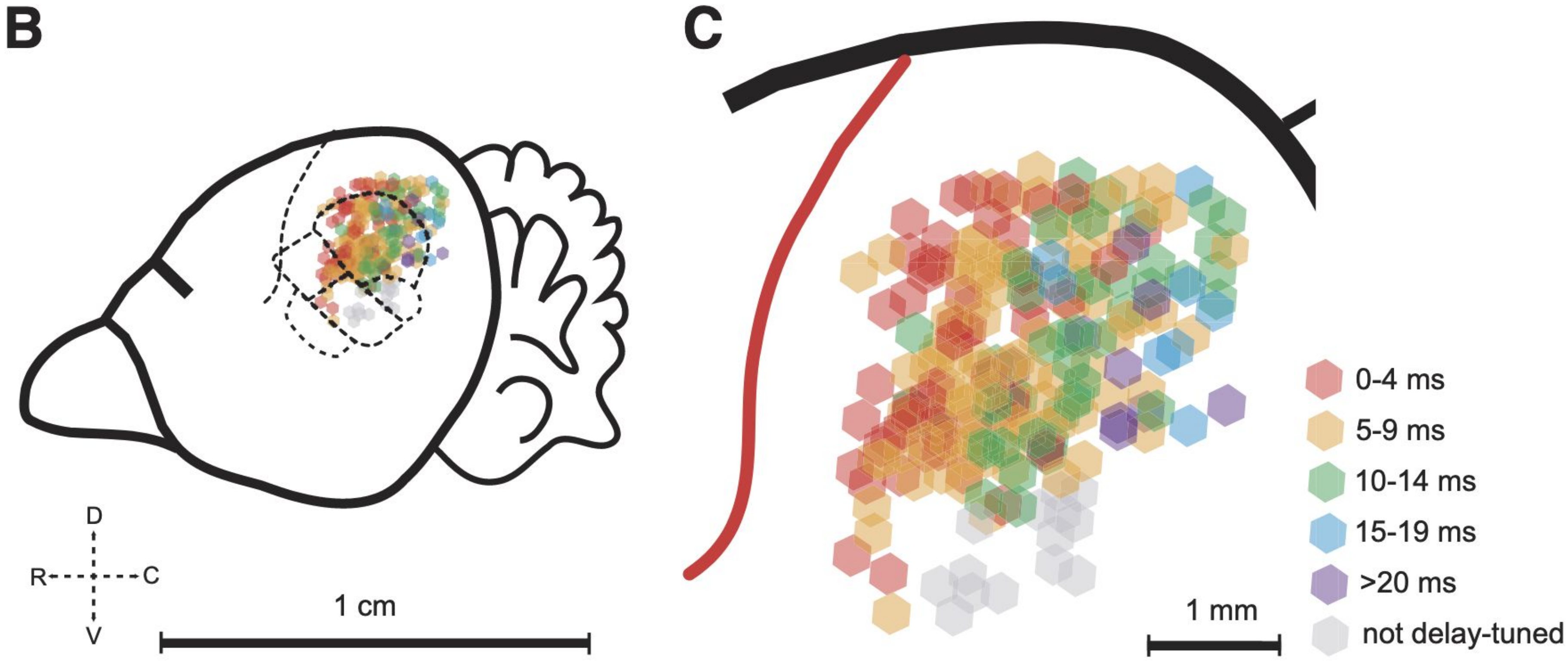
(Tsukano et al, 2016)

Within cortical fields Topography and Topology



Within cortical fields Topography and Topology

Cortical representation of echo delay time in A1

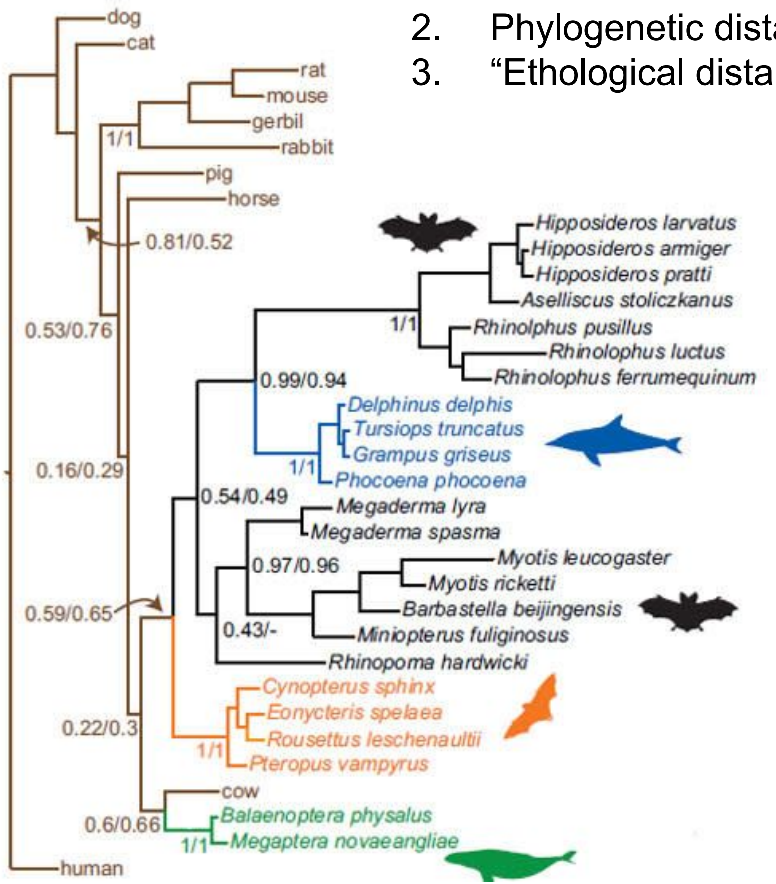


Short-tailed Bat (*C perspicillata)*

(Hagemann, 2009)

Measures for Multi-species comparison of cortical fields:

1. Cortical field areas
2. Phylogenetic distance
3. "Ethological distance"



Ethological behaviors shape encoded feature spaces

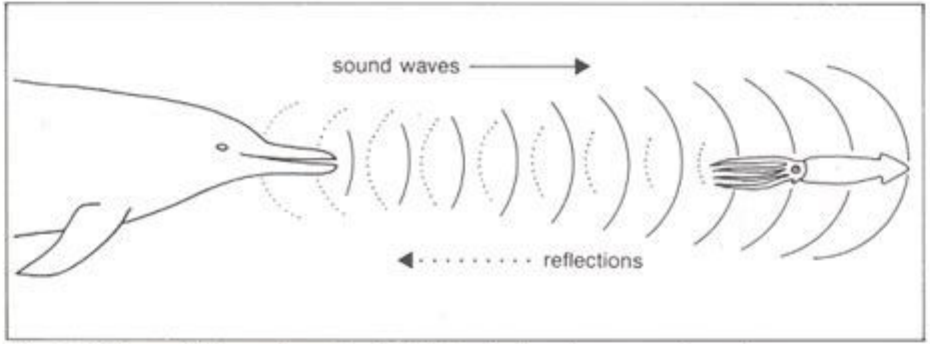
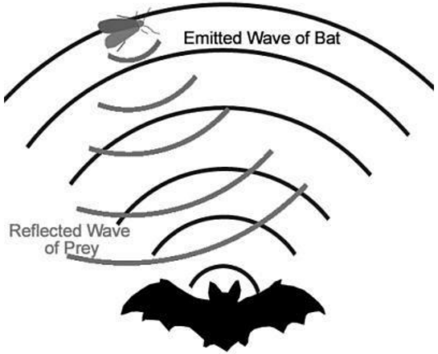
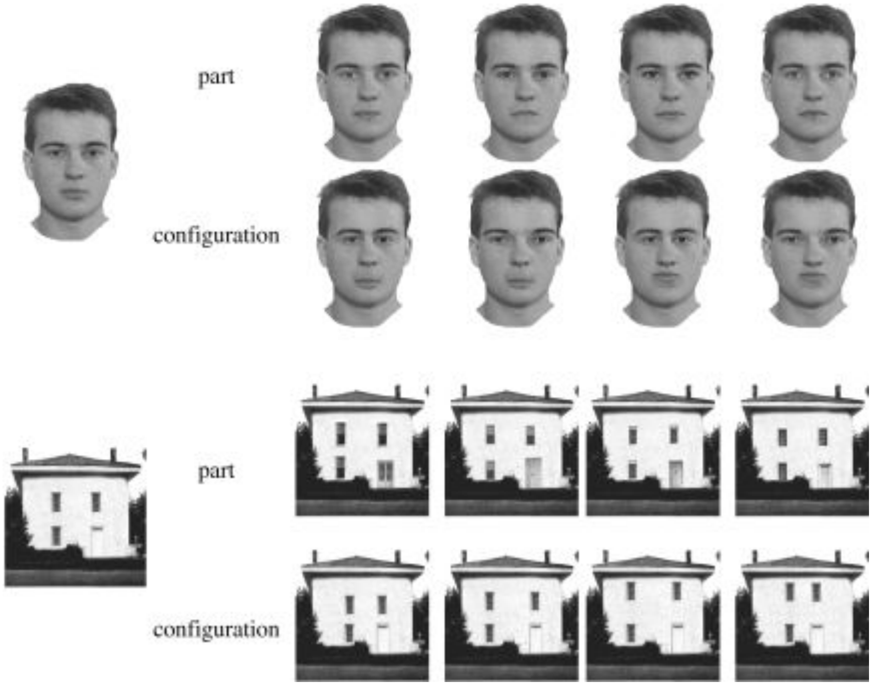


Fig. 6 Schematic illustration of a dolphin's echolocation system in action.

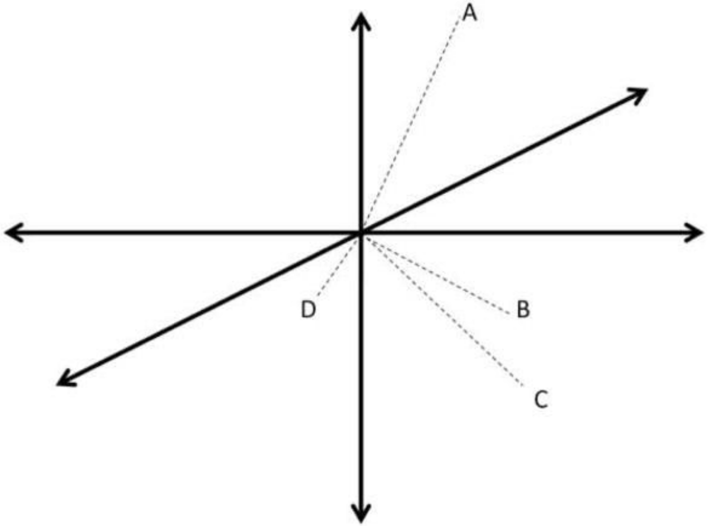


Complex features have also been discovered to have *topological* organization in multimodal regions of the cortex

The fusiform face area: a cortical region specialized for the perception of faces



Kanwisher, 2006



A hypothetical illustration of a three-dimensional face space (Valentine, 1991) from Parr, 2012.

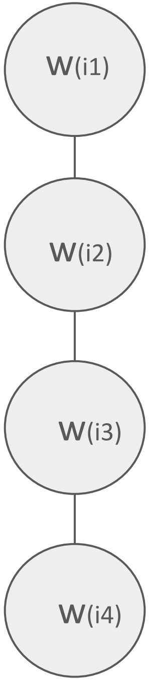
How do complex feature spaces organize?

How many feature spaces does there really need to be for cognition?

How could new feature spaces come about?

Self Organizing maps

1. What are self-organizing maps: Teuvo Kohonen, 1982



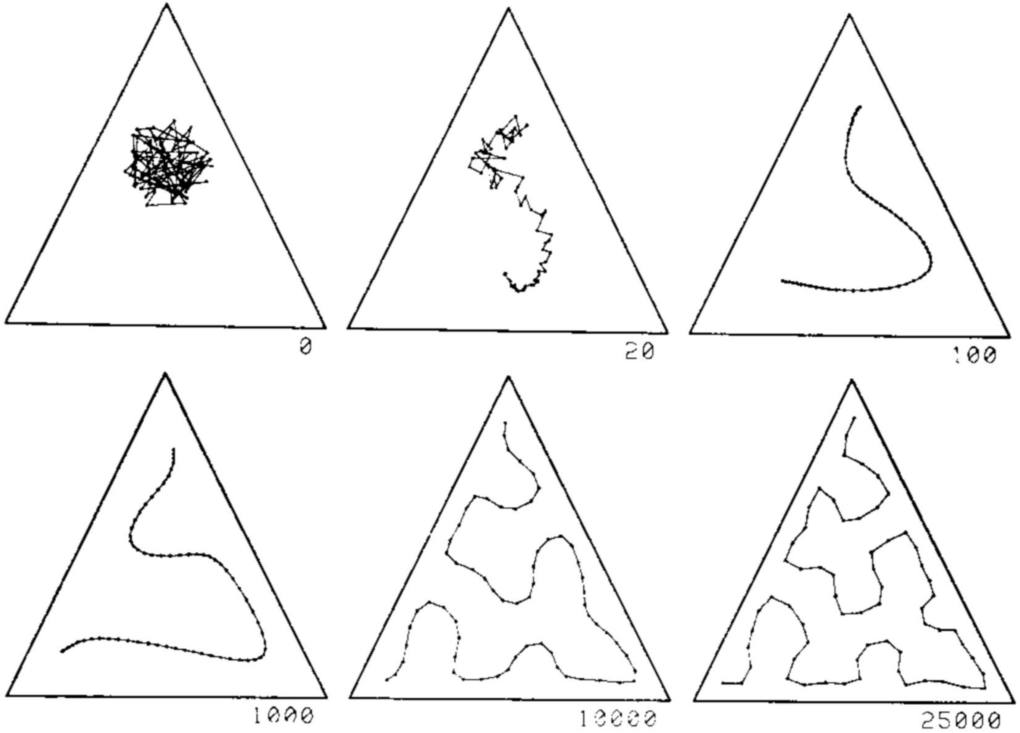


Fig. 4. Weight vectors during the ordering process, one-dimensional array.

Self Organizing maps

1. What are self-organizing maps

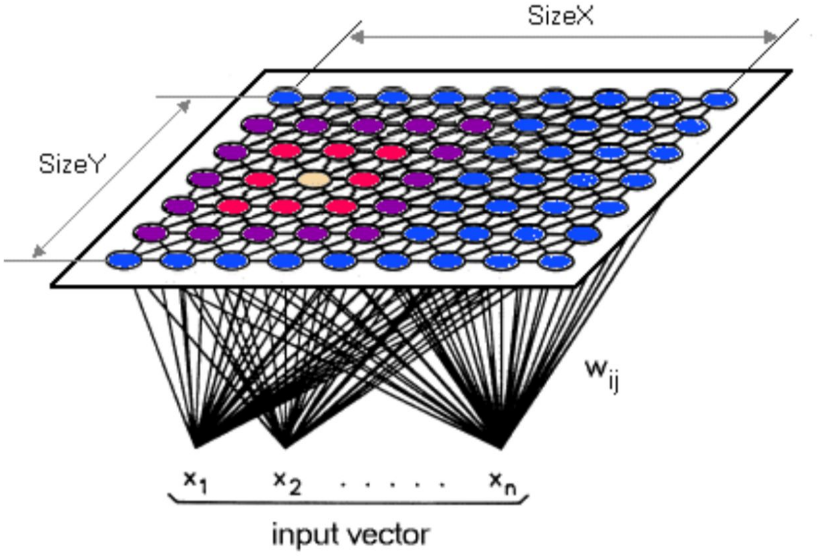
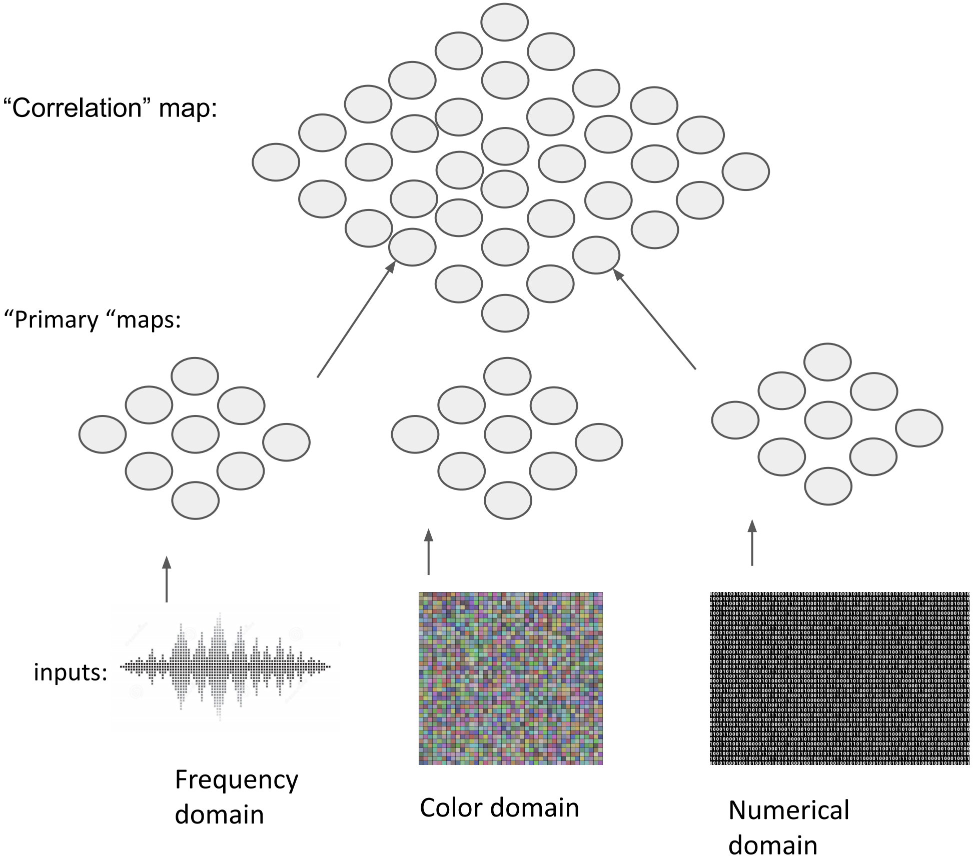


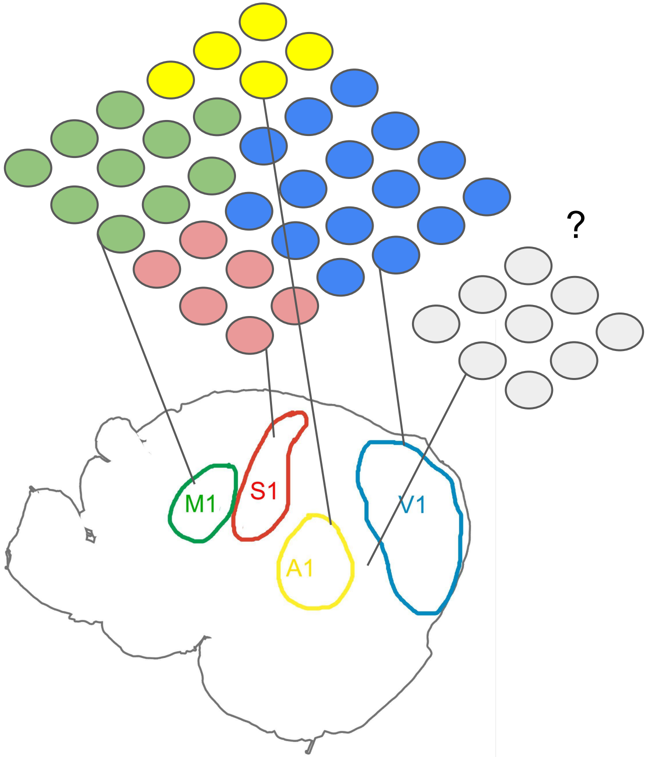
FIGURE 3.3: A Kohonen model with the BMU in yellow, the layers inside the neighbourhood radius in pink and purple, and the nodes outside in blue.

(Sarkar, 2018)

Hopes for modelling with an SOM



Space of causal correlations between maps with a 'saliency' rule?



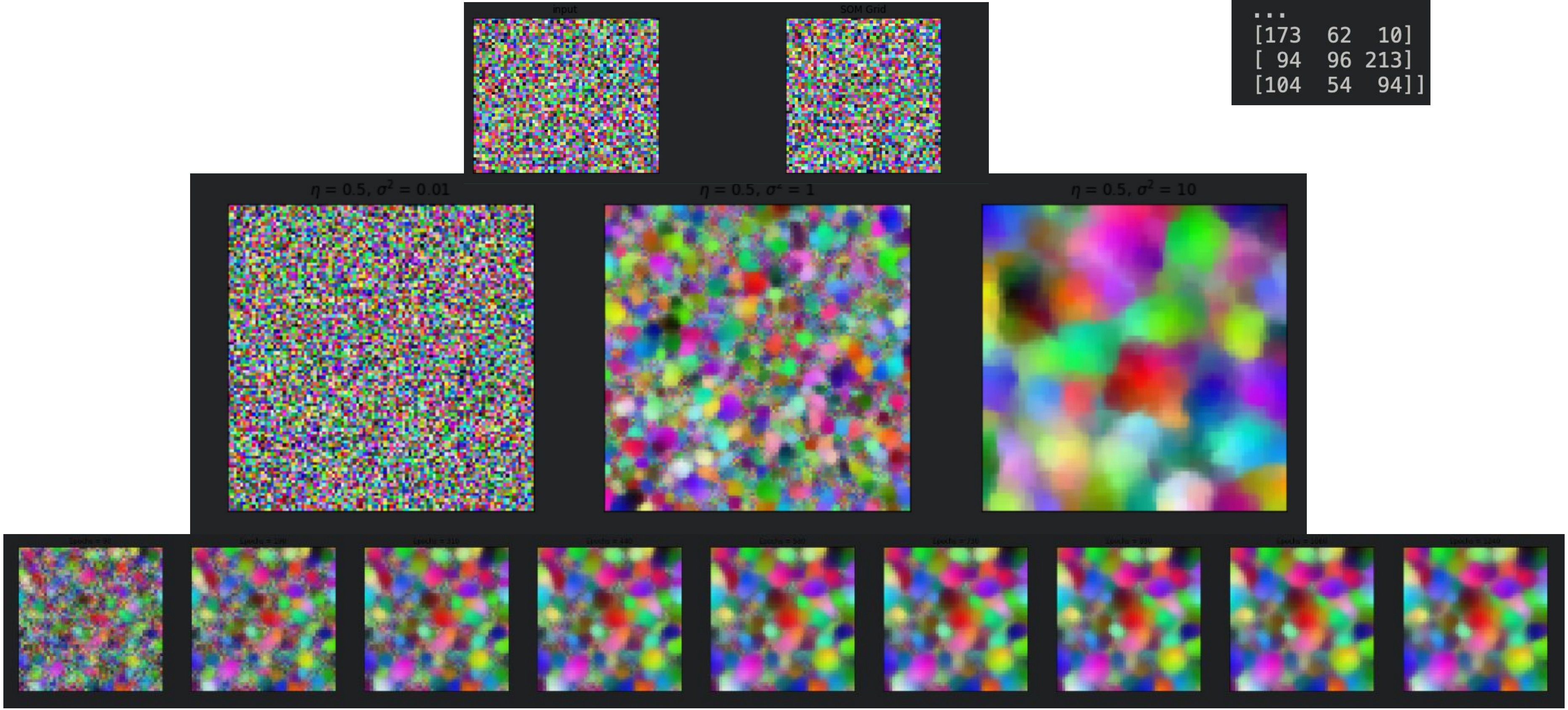
Geometric constraint to predict complex feature spaces between primary fields

Color version of self-organized map

Each node location (x, y) has a color value (r, g, b) from 0-255

```
[[172  47 117]
 [192  67 251]
 [195 103   9]
 ...
 [173  62  10]
 [ 94  96 213]
 [104  54  94]]
```

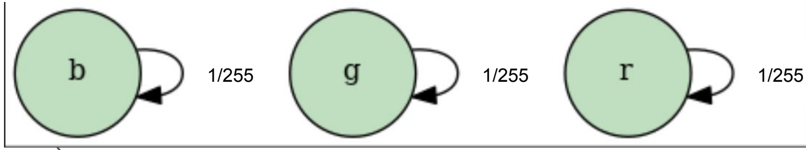
Input SOM

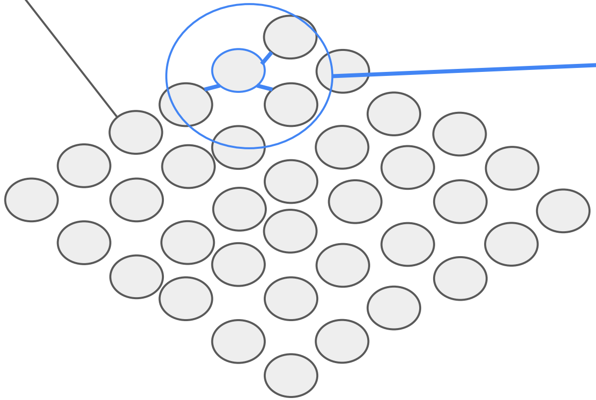


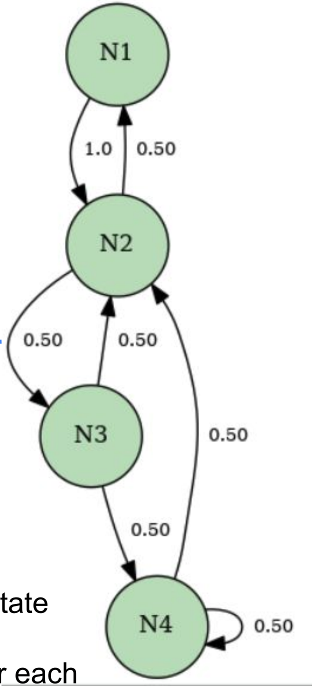
Various Stumbling blocks: probability distribution while updating?

1. *equivocation* H[Y|X]
2. *ambiguity*: H[X|Y]
3. *organization/structure* ? increase in organization of information might not change entropy of underlying information?

How to get conditional distribution Pr(Y|X) between input and output processes?







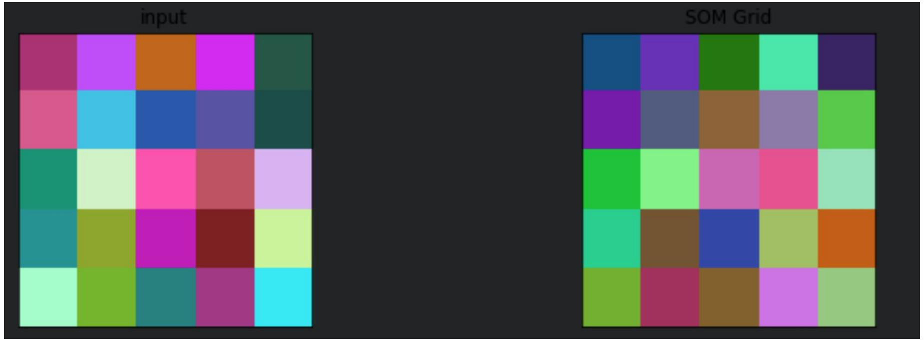
How to find probability of a node being a BMU at initialization versus given neighboring nodes?

There are 3000 random colors in the input source, so there are 3000 x (255, 255, 255) = 49744125000, so the initially its a uniform probability distribution of 1/49744125000

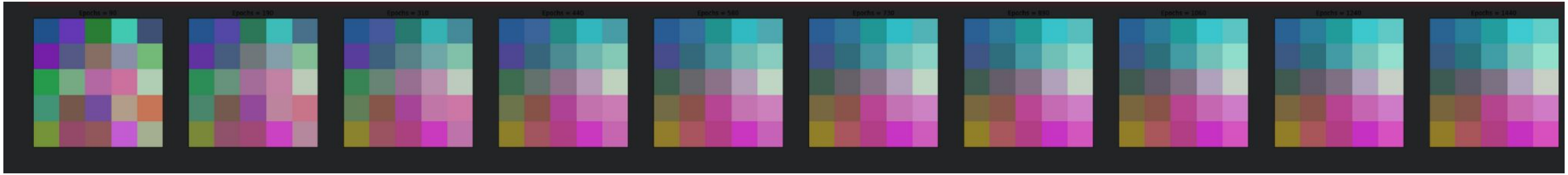
…a mixed state distribution changing for each epoch?

Color version of self-organized map

Input SOM



Epochs



Defining channel capacity for a self-organized map

Communication channel from Input “Information Source” to “Output” SOM Process, BMUs as output (Y)
Build a Distribution that roughly corresponds

Source Entropy: H(X)

For x in X
Each initial input H(x) = 1.58

There are 25 random colors in the input source, so there are 25 x (255, 255, 255) = 414534375 possible states, so the probability is 1/414534375

Class:	Distribution	X	p(x)
Alphabet:	('b', 'g', 'r') for all rvs	b	1/3
Base:	linear	g	1/3
Outcome Class:	str	r	1/3
Outcome Lenght:	1		

Initial Output Entropy: H(Y)

overall output H(Y) = 4.64

Uncertainty about which location will be the Best matching unit is initially 4.64 (?)

Class:	Distribution	a	b	x	y	p(x)
Alphabet:	(('A',), ('1', '2', '3', '4', '5'), ('B',), ('1', '2', '3', '4', '5'))	A	1	B	1	1/25
Base:	linear	A	1	B	2	1/25
Outcome Class:	str	A	1	B	3	1/25
Outcome Lenght:	4	A	1	B	4	1/25
		A	1	B	5	1/25

Defining channel capacity for a self-organized map

Communication channel from Input “Information Source” to “Output” SOM Process, BMUs as output (Y)
Build a Distribution that roughly corresponds

Source Entropy: H(X)

For x in X
Each initial input H(x) = 1.58

There are 25 random colors in the input source, so there are 25 x (255, 255, 255) = 414534375 possible states, so the probability is 1/414534375

		X	p(x)
Class:	Distribution	**X**	**p(x)**
Alphabet:	('b', 'g', 'r') for all rvs	b	1/3
Base:	linear	g	1/3
Outcome Class:	str	r	1/3
Outcome Lenght:	1		

Initial Output Entropy: H(Y)

overall output H(Y) = 3.32

Uncertainty about which location will be the Best matching unit is initially 3.32?)

Class:	Distribution	**m**	**n**	**p(x)**
Alphabet:	(('A', 'B'), ('1', '2', '3', '4', '5'))	A	1	1/10
Base:	linear	A	2	1/10
Outcome Class:	str	A	3	1/10
Outcome Lenght:	2	A	4	1/10
		A	5	1/10
		B	1	1/10

Defining channel capacity for a self-organized map: Next steps

How to get correct probability distribution of the SOM as it is being trained, and after training:

1. Get the correct probability distribution empirically:
   a. Keep track of nodes
   b. count how many times the node becomes a BMU
   c. counts are a sampling probability distribution, use to calculate H(Y)

2. There should be obvious dimensionality reduction properties in the change in conditional entropy
   a. How many input vectors map to a 2D output: H(Input vectors | x,y) and H(x,y | Input vectors)

3. Try not using colors, and instead use eMs to generate binary strings that have known properties, then test classification with an SOM.

4. Is there a way to jump model classes with SOMs, such as by iterating an SOM so that outputs become inputs (correlations of correlations)?

Understand information architectures of feature selectivity:

How do topological feature maps emerge (evolutionarily and statistically)

Are there constraints to the amount or type of feature spaces that can exist (dimensionality)?

How do completely new feature spaces come about?

How do feature maps relate to generalization capacities in cognition (inference, learning, flexible behaviors)?

Thank you James Crutchfield, Mikhael Seeman and the class of Physics 256B, Spring 2022

jlitmancleper@ucdavis.edu