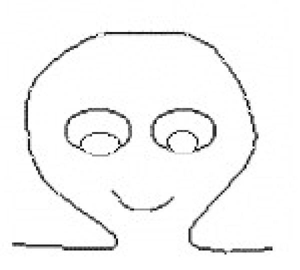
From brain to mind - challenges for mathematicians (or problems where I could use help …)



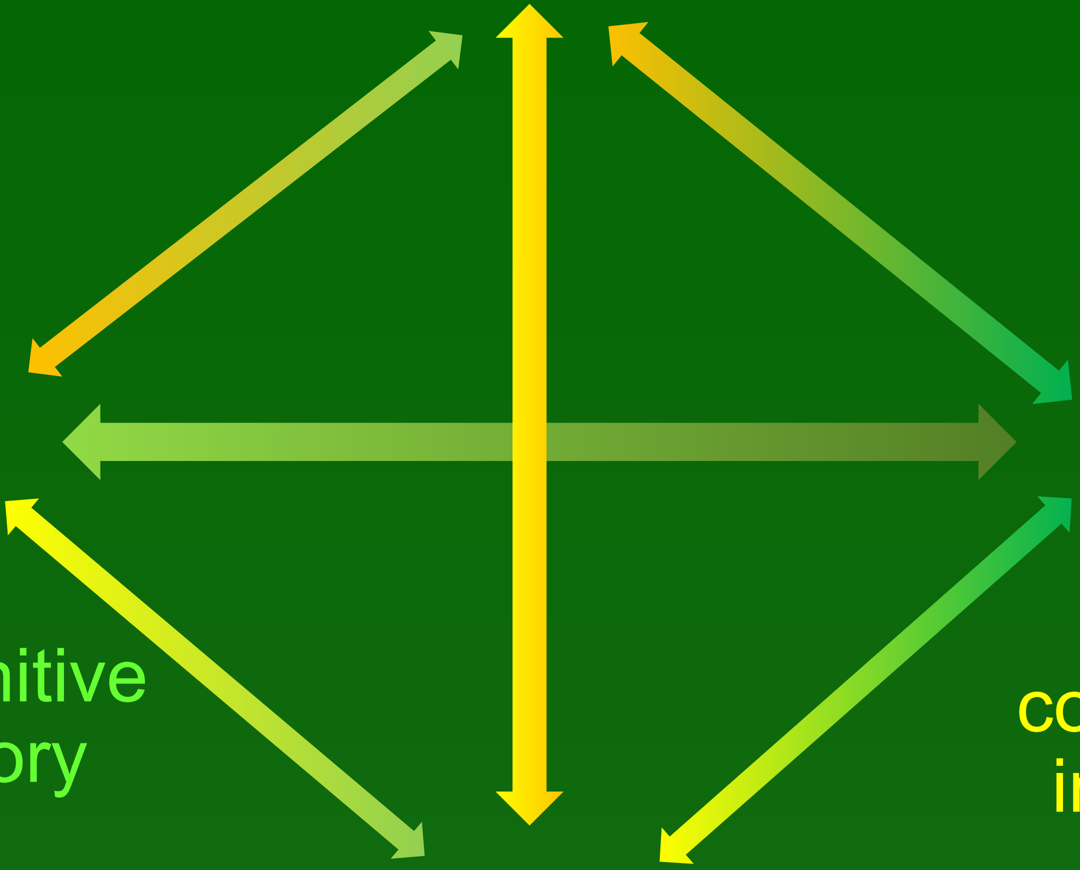
Włodzisław Duch

Department of Informatics.
Neurocognitive Laboratory.
Nicolaus Copernicus University, Toruń, Poland

Google: Wlodzislaw Duch

BioFizMat Instytut Banacha, Warszawa 19-20.05.2016.

Cognitive sciences
Cog



Bio

Neurocognitive
Laboratory

Info
Neural
networks
computational
intelligence,
machine
learning

Nano
Fizyka kwantowa

Computational physics

Graphical representation of model spaces. Vol. I Basics.
Springer Verlag, Berlin Lecture Notes in Chemistry
Vol. 42 (1986); Vol. II has never been written …

Lecture Notes in Chemistry

[illegible]

42

W. Duch

GRMS or Graphical Representation of Model Spaces

Vol. I Basics

Springer-Verlag

Idea:

Differential equations, such as Schrodinger equations in quantum mechanics are approximated by algebraic equations defined in tensor spaces with proper symmetry.

$$\hat{A}\Psi\left(x_1, x_2 \ldots x_N\right) = E\Psi\left(x_1, x_2 \ldots x_N\right) \rightarrow \mathbf{AC} = E\mathbf{C}$$

In finite dimensional N-particle Hilbert spaces eigenfunctions Ψ become linear combinations of basis functions (usually a large number):

$$\Psi\left(x_1, x_2 \ldots x_N\right) = \sum_I C_I \Phi_I\left(x_1, x_2 \ldots x_N\right)$$

Tensor spaces

Matrix elements can be efficiently calculated if N-D functions are constructed from tensor products based on subsets of 1-D functions.

$$\Phi_I\left(x_1,x_2\ldots x_N\right)=\sum_I a_{I1..IN}\phi_{I2}\left(x_2\right)\phi_{I2}\left(x_1\right)\ldots\phi_{IN}\left(x_N\right)$$

$$Ik=1\ldots n$$

Operators in N-particle Hilbert spaces become matrices:

$$\hat{\mathbf{A}}_M=\mathbf{1}_M\hat{\mathbf{A}}\mathbf{1}_M=\sum_{L=1}^{M}|L\rangle\langle L|\;\hat{\mathbf{A}}\sum_{L=1}^{M}|R\rangle\langle R|=\sum_{L,R=1}^{M}|L\rangle\langle L|\hat{\mathbf{A}}|R\rangle\langle R|$$

Problem: number of N-D tensor products constructed from *n* 1-D functions may become very large, ex: 10-D for *n*=28 is >2x10^{16}

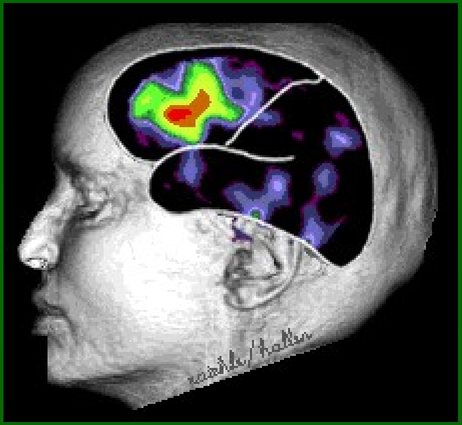
Solution: represent tensor spaces graphically, use symmetries, learn to solve equations directly from graphs with constructing matrices.

Geometric model of mind/brain

Brain ⇔ Mind

Objective ⇔ Subjective

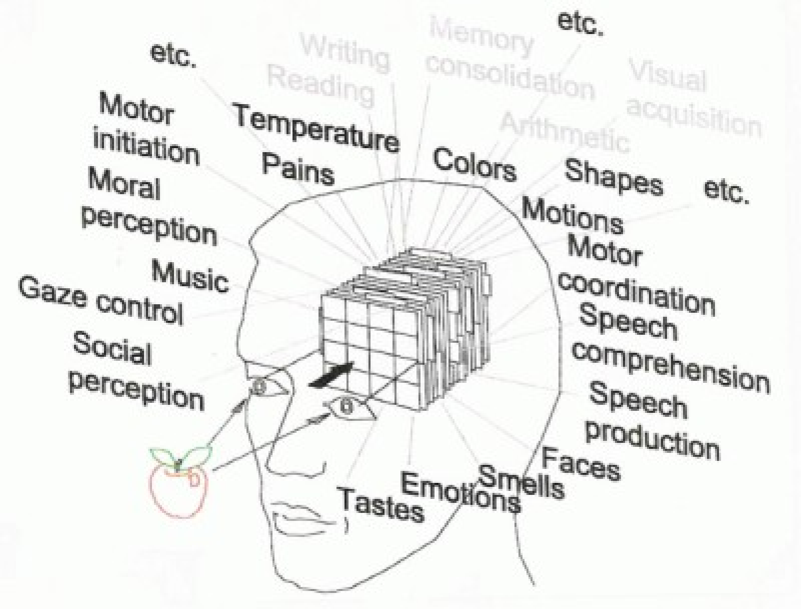
Brain states are approximated fairy well by neurodynamics, with neuron activity estimated by EEG, MEG, NIRS-OT, PET, fMRI …



How to describe mental states?

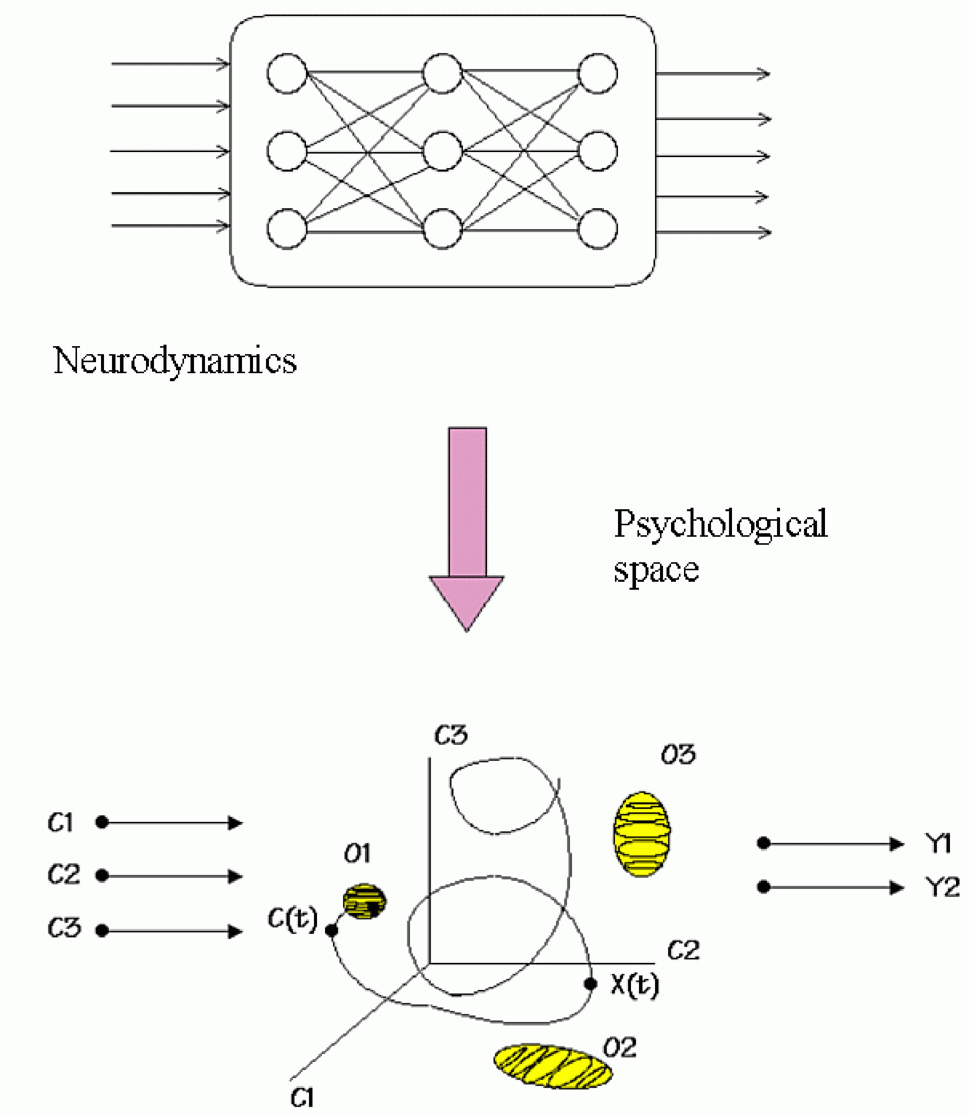
Define mental space, based on dimensions that are related to subjective experience.

Mental state is than described by trajectory in this “psychological space” (Shepard, Gardenfors, Fauconniere etc).



Problem: lack of phenomenology.

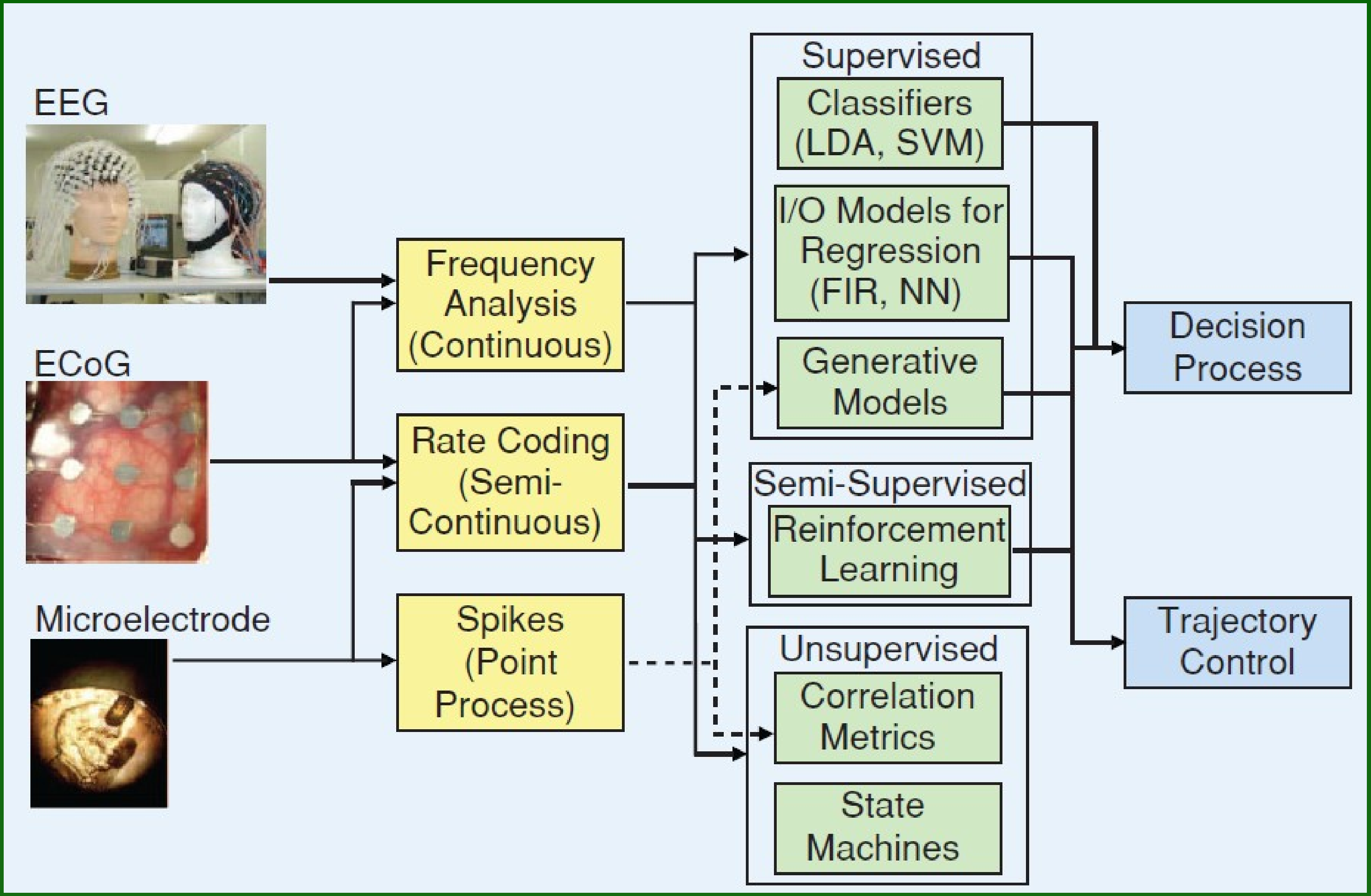
Hurlburt & Schwitzgabel, Describing Inner Experience? MIT Press 2007



From neurodynamics (simulated, observed) to geometric description of mental states in some “mind space”, or P-space: attractor networks?

BCI

To some degree this is what we do in Brain-Computer Interfaces, mapping brain states to intentions/decisions or feedback signals.



P-spaces

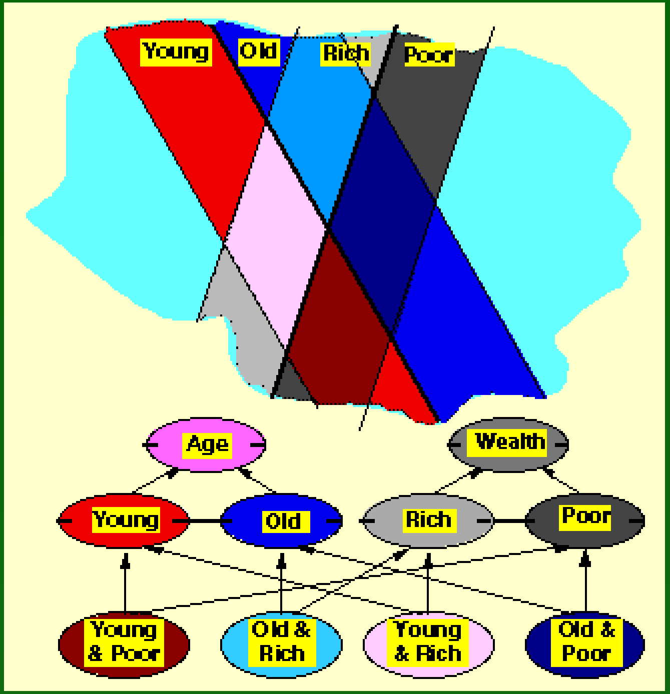
Psychological spaces: how to visualize inner life?

K. Lewin, The conceptual representation and the measurement of psychological forces (1938), cognitive dynamic movement in phenomenological space.

George Kelly (1955):
personal construct psychology (PCP), geometry of psychological spaces as alternative to logic.

A complete theory of cognition, action, learning and intention.

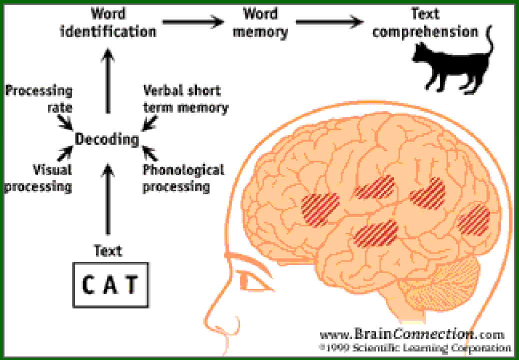
PCP network, society, journal, software …
quite active group. No connection to brain.



Many things in philosophy, dynamics, neuroscience and psychology, searching for new ways of understanding cognition, are relevant here.

P-space definition

P-space: region in which we may place and classify elements of our experience, constructed and evolving,
„a space without distance", divided by dichotomies.



P-spaces should have (R. Shepard 1957-2001):

- minimal dimensionality;
- distances that monotonically decrease with increasing similarity.

This may be achieved using **multi-dimensional non-metric scaling** (MDS), reproducing in low-dimensional spaces original similarity relations estimated from empirical data.

Many object recognition and perceptual categorization models assume that objects are represented in a multidimensional psychological space; similarity between objects ~ 1/distance in this space.

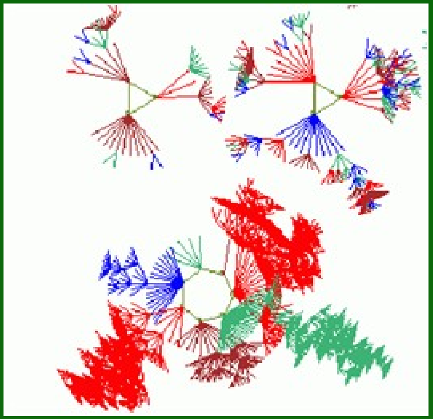
Can one describe the state of mind in similar way?

Static Platonic model

Newton introduced space-time, arena for physical events.

Mind events need psychological spaces.

Goal: integrate neural and behavioral information in one model, create model of mental processes at intermediate level between psychology and neuroscience.



Static version: short-term response properties of the brain, behavioral (sensomotoric) or memory-based (cognitive). Dynamic: look at brain/mental trajectories.

Approach:

- simplify neural dynamics, find invariants (attractors), characterize them in psychological spaces;
- use behavioral data, represent them in psychological space.

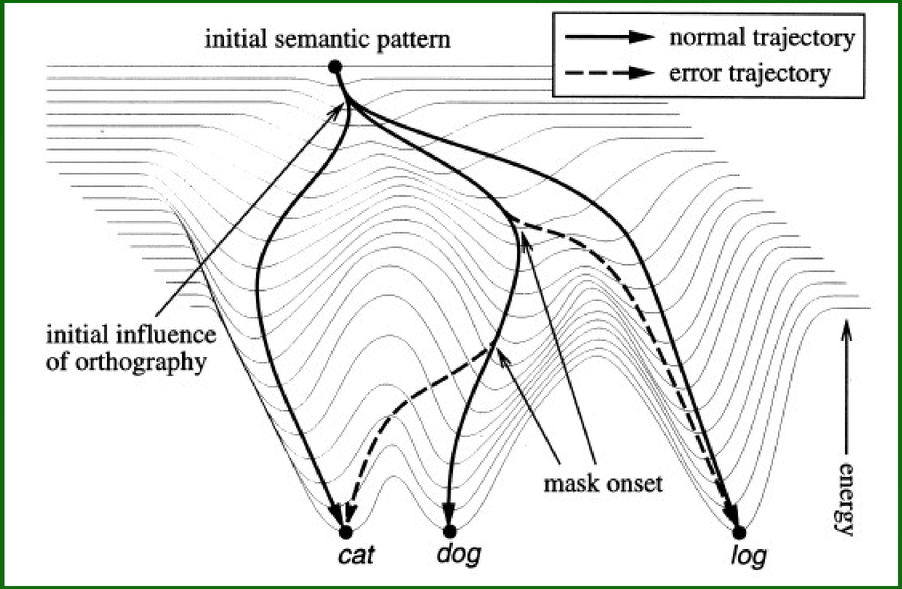
Applications: object recognition, psychophysics, category formation in low-D psychological spaces, case-based reasoning.

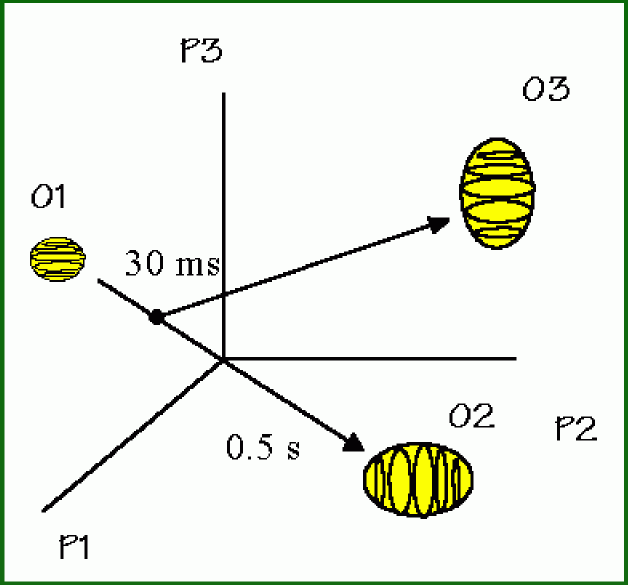
Energies of trajectories

P.McLeod, T. Shallice, D.C. Plaut, Attractor dynamics in word recognition: converging evidence from errors by normal subjects, dyslexic patients and a connectionist model. Cognition 74 (2000) 91-113.

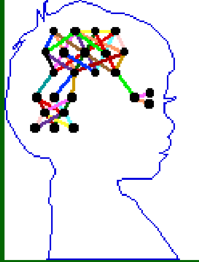
M Spivey, Continuity of Mind (Oxford Uni Press, 2007)

New area in psycholinguistics: investigation of dynamical cognition, influence of masking on semantic and phonological errors.





Neuroimaging words



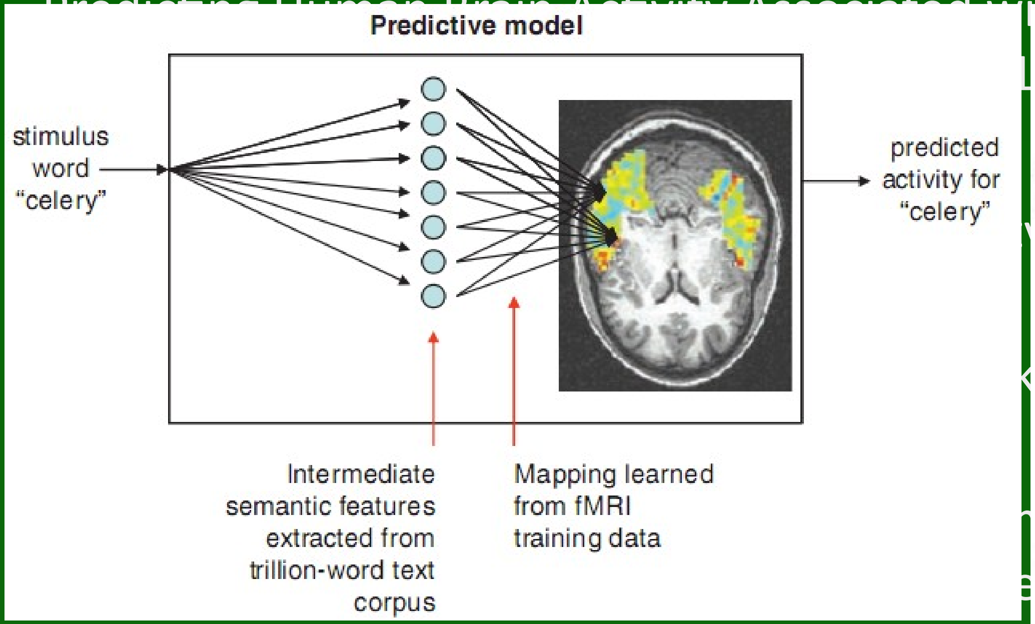
Predicting Human Brain Activity Associated with the Meanings
of Nouns," T. M. Mitchell et al, Science, 320, 1191, May 30, 2008

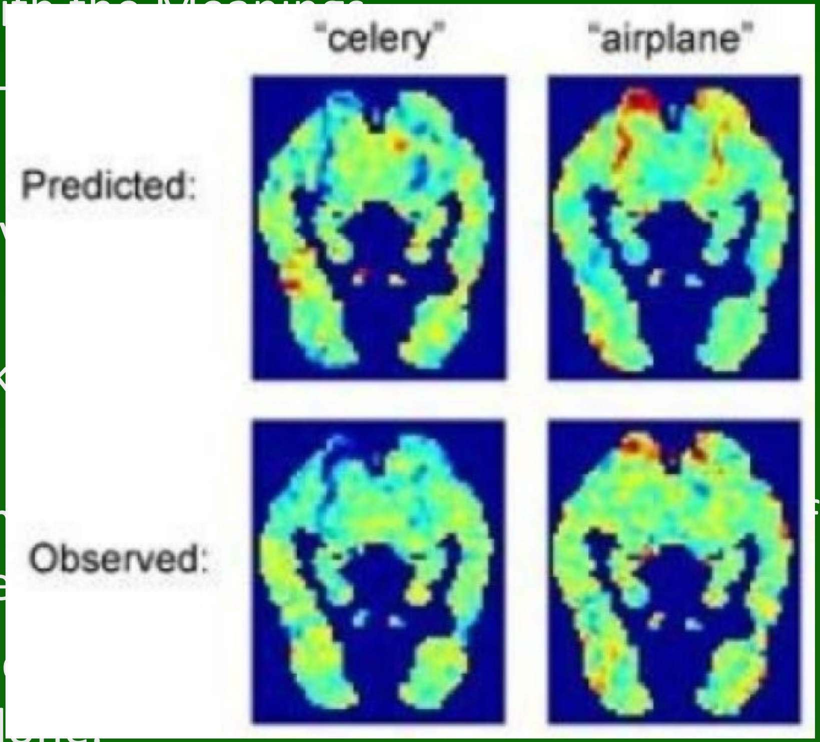
- Clear differences between fMRI brain activity when people read and think about different nouns.
- Reading words and seeing the drawing invokes similar brain activations, presumably reflecting semantics of concepts.
- Although individual variance is significant similar activations are found in brains of different people, a classifier may still be trained on pooled data.
- Model trained on ~10 fMRI scans + very large corpus (10^{12}) predicts brain activity for over 100 nouns for which fMRI has been done.

Transform words => vector of 25 semantic features, perception and action.
Sensory: see, hear, touch, smell, taste, fear.
Motor: eat, manipulate, move, pick, push, stroke, talk, run, walk.
Actions: break, ride, clean, enter, fill, open, carry.

Neuroimaging words

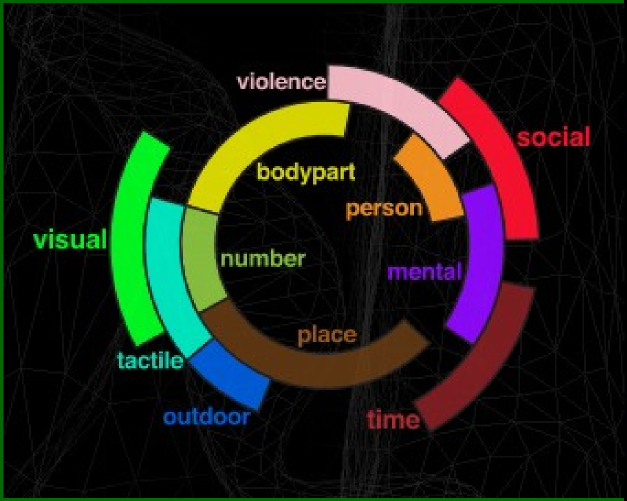
Predicting Human Brain Activity Associated with the Meanings

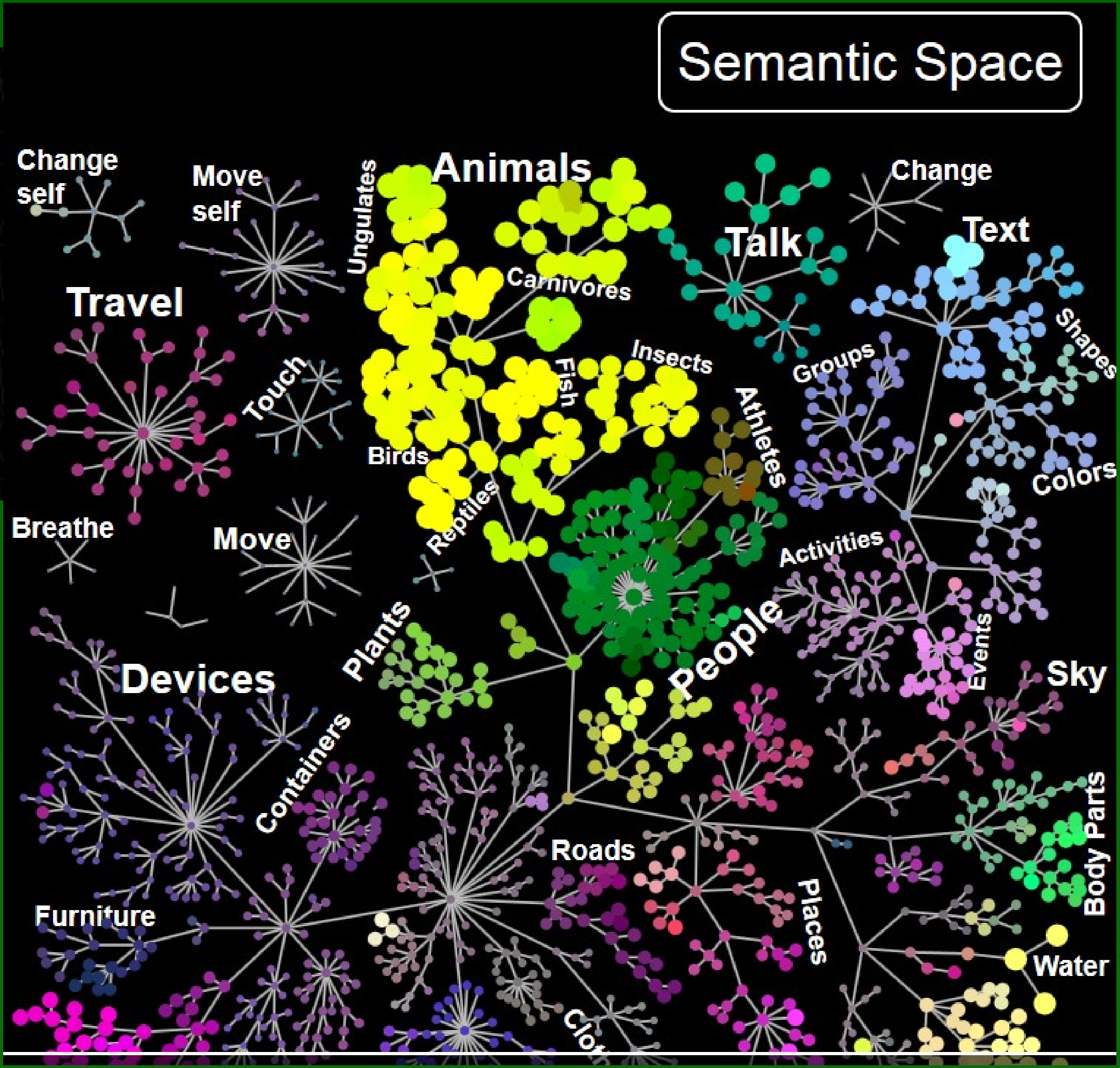

Predictive model
stimulus word "celery"
predicted activity for "celery"
Intermediate semantic features extracted from trillion-word text corpus
Mapping learned from fMRI training data



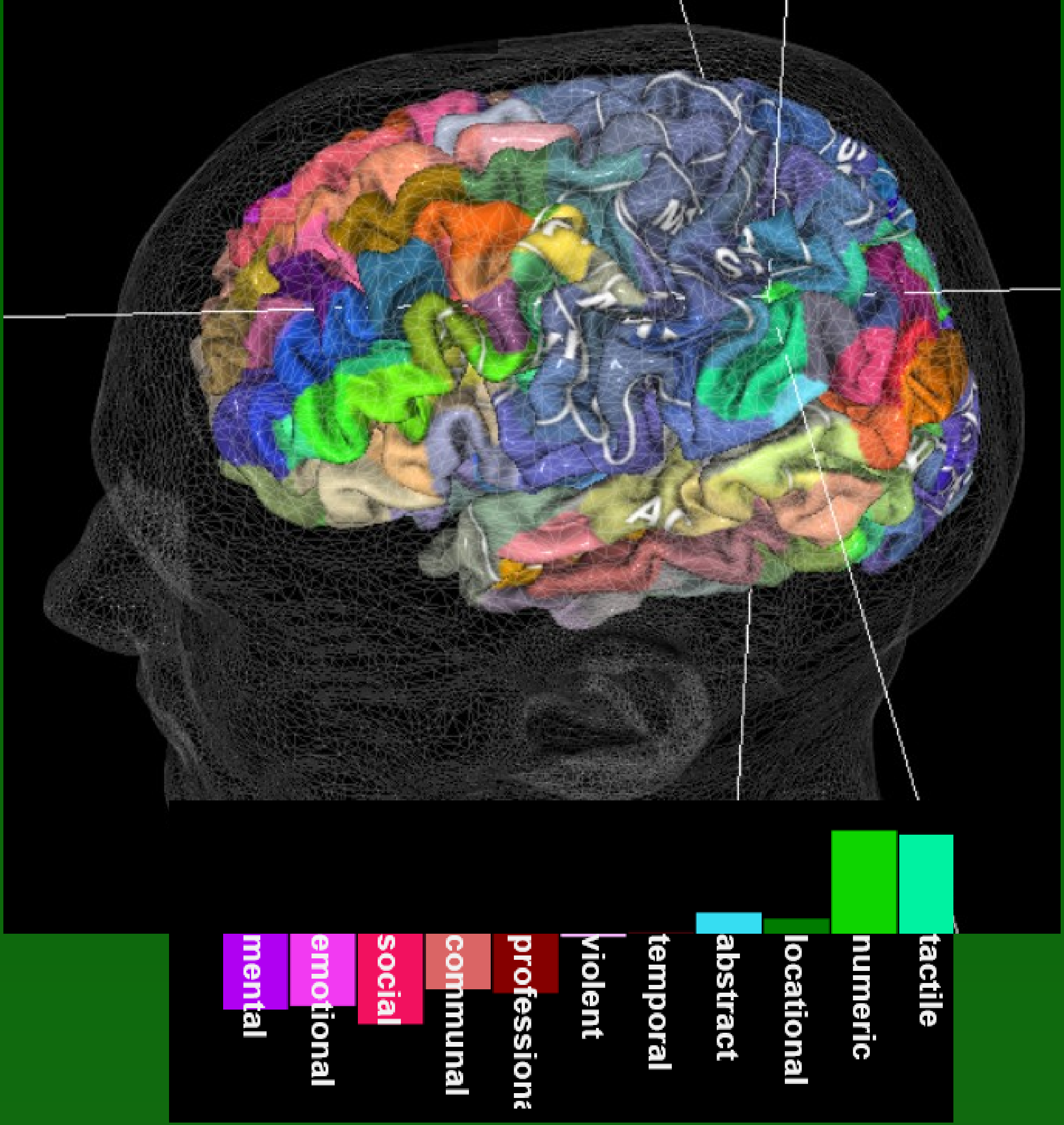
- Model trained on ~10 fMRI scans + very larg for over 100 nouns for which fMRI has been d

Transform words => vector of 25 semantic features, perception and action.
Sensory: see, hear, touch, smell, taste, fear.
Motor: eat, manipulate, move, pick, push, stroke, talk, run, walk.
Actions: break, ride, clean, enter, fill, open, carry.

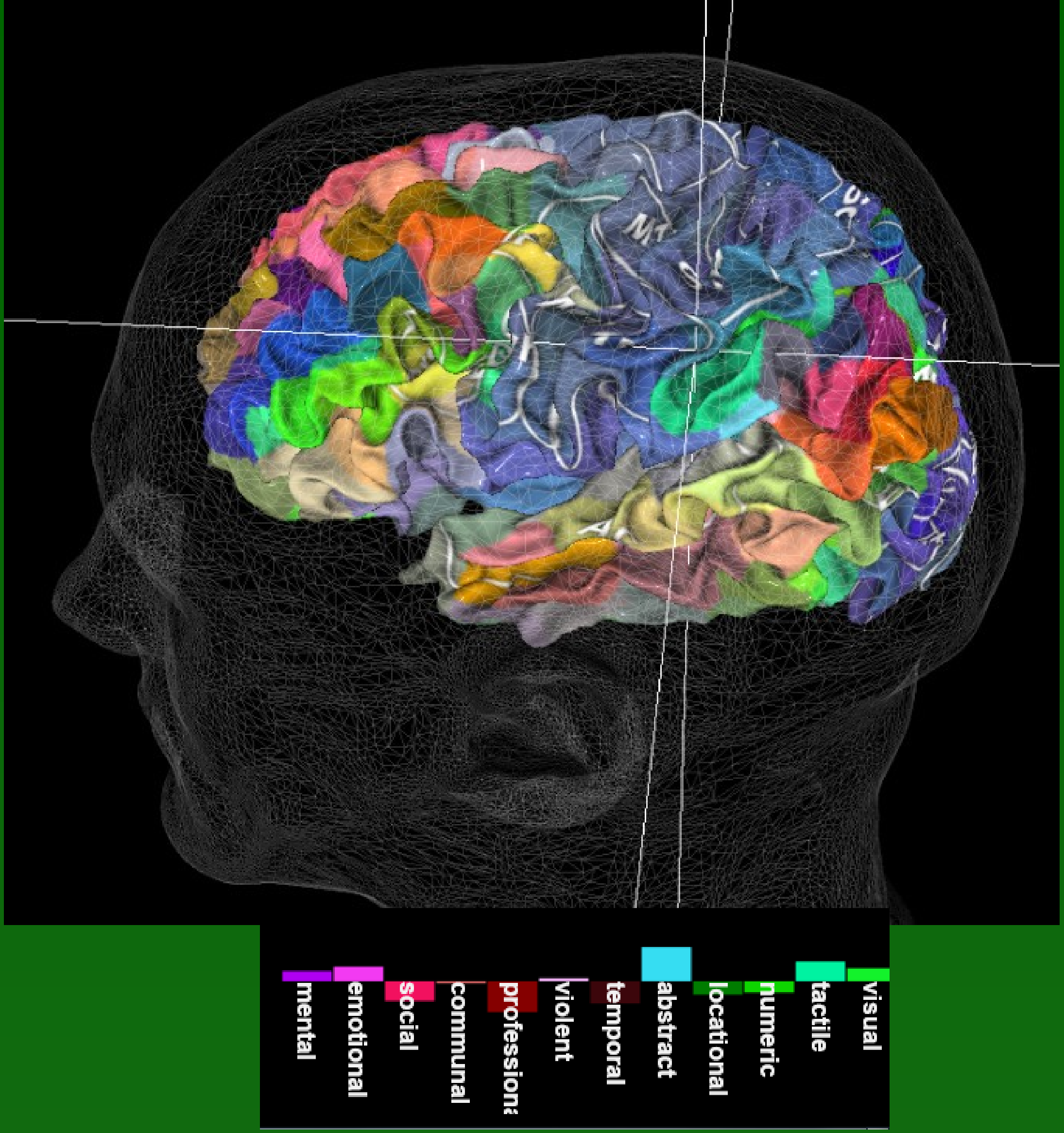




Gallant lab created maps of fMRI brain activity (60K voxels) for a number of words clustered around different concepts. http://gallantlab.org/

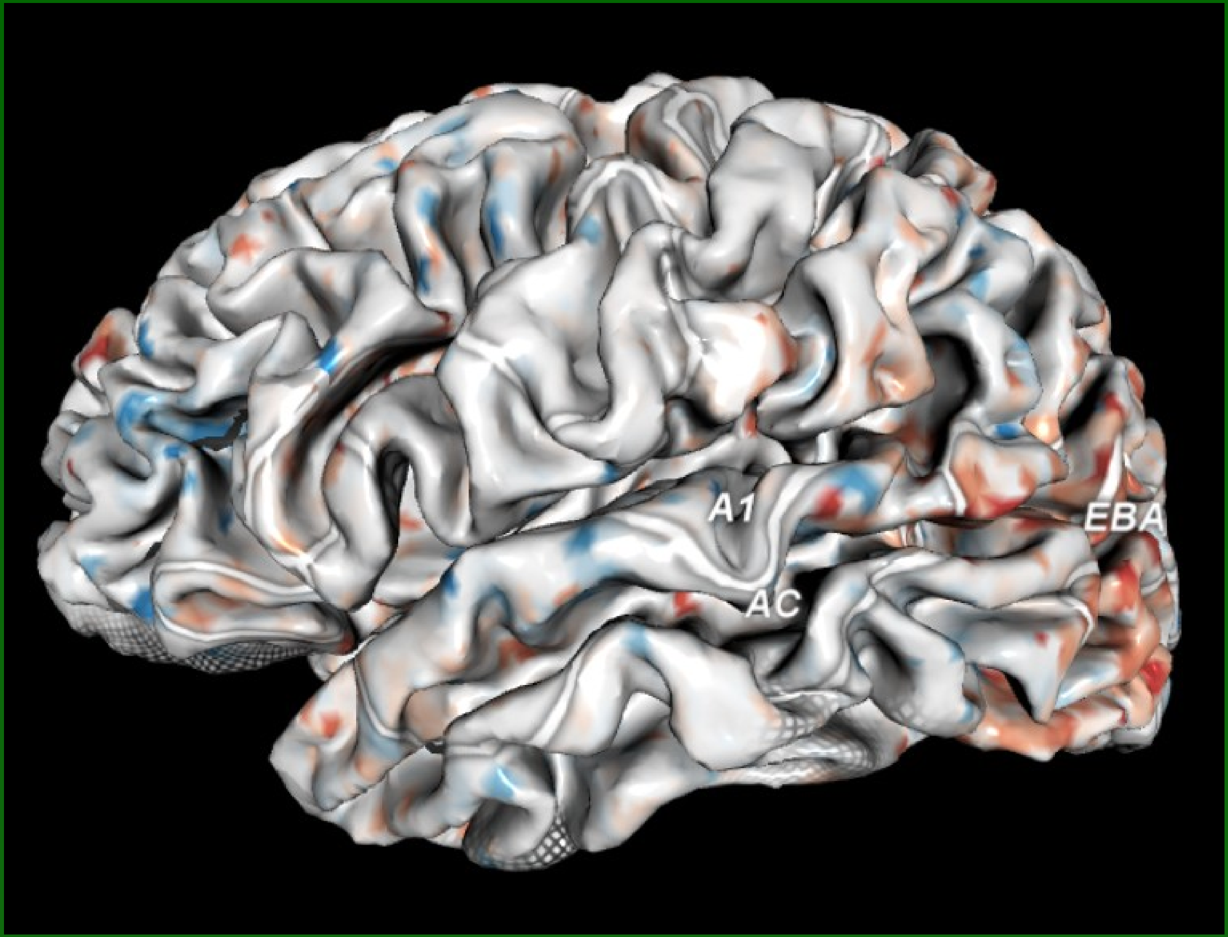


HIPS area voxels shows strong activity with numeric concepts.
http://gallantlab.org/

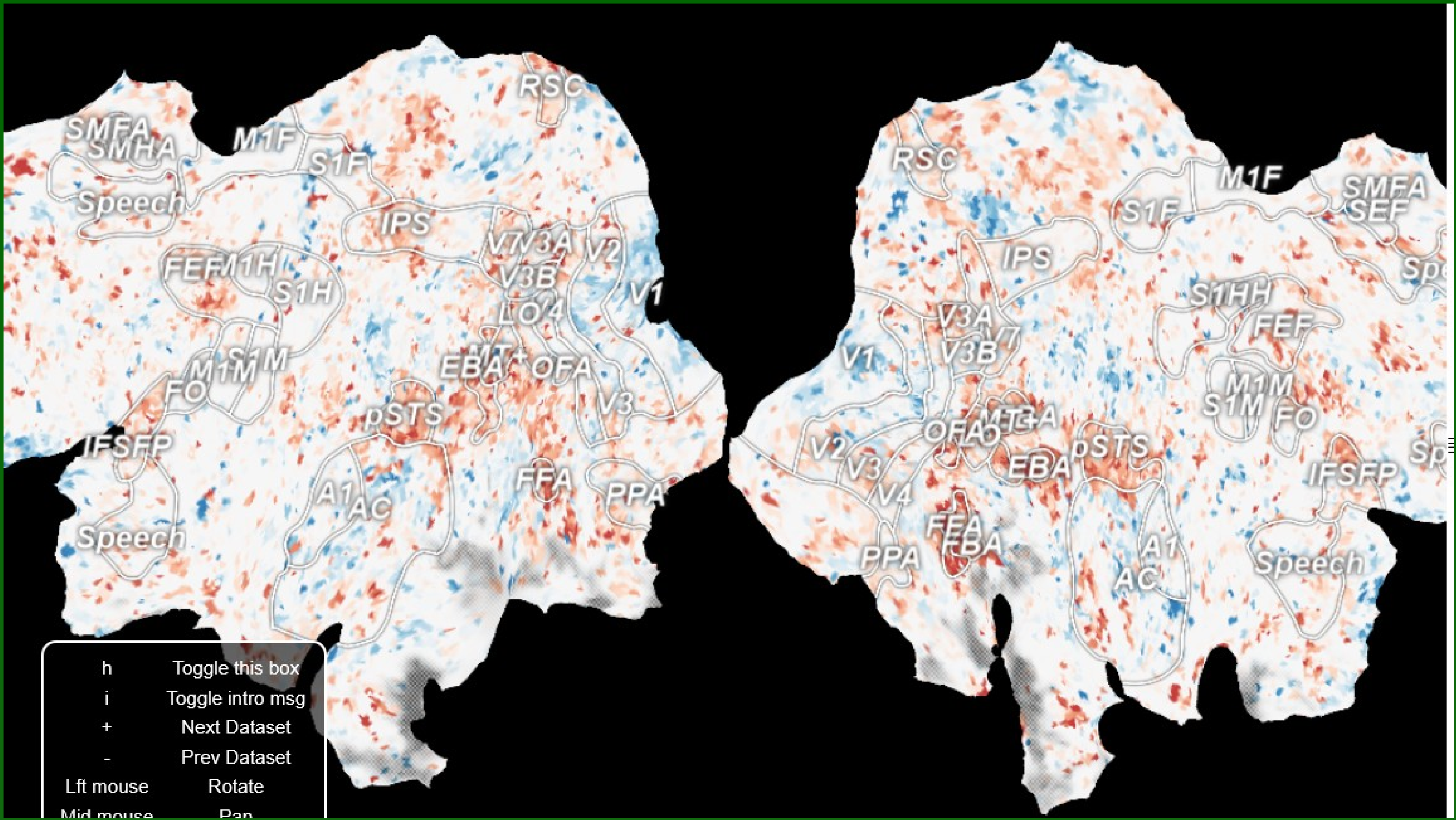


HIPS area voxels shows strong activity also with abstract concepts.

http://gallantlab.org/



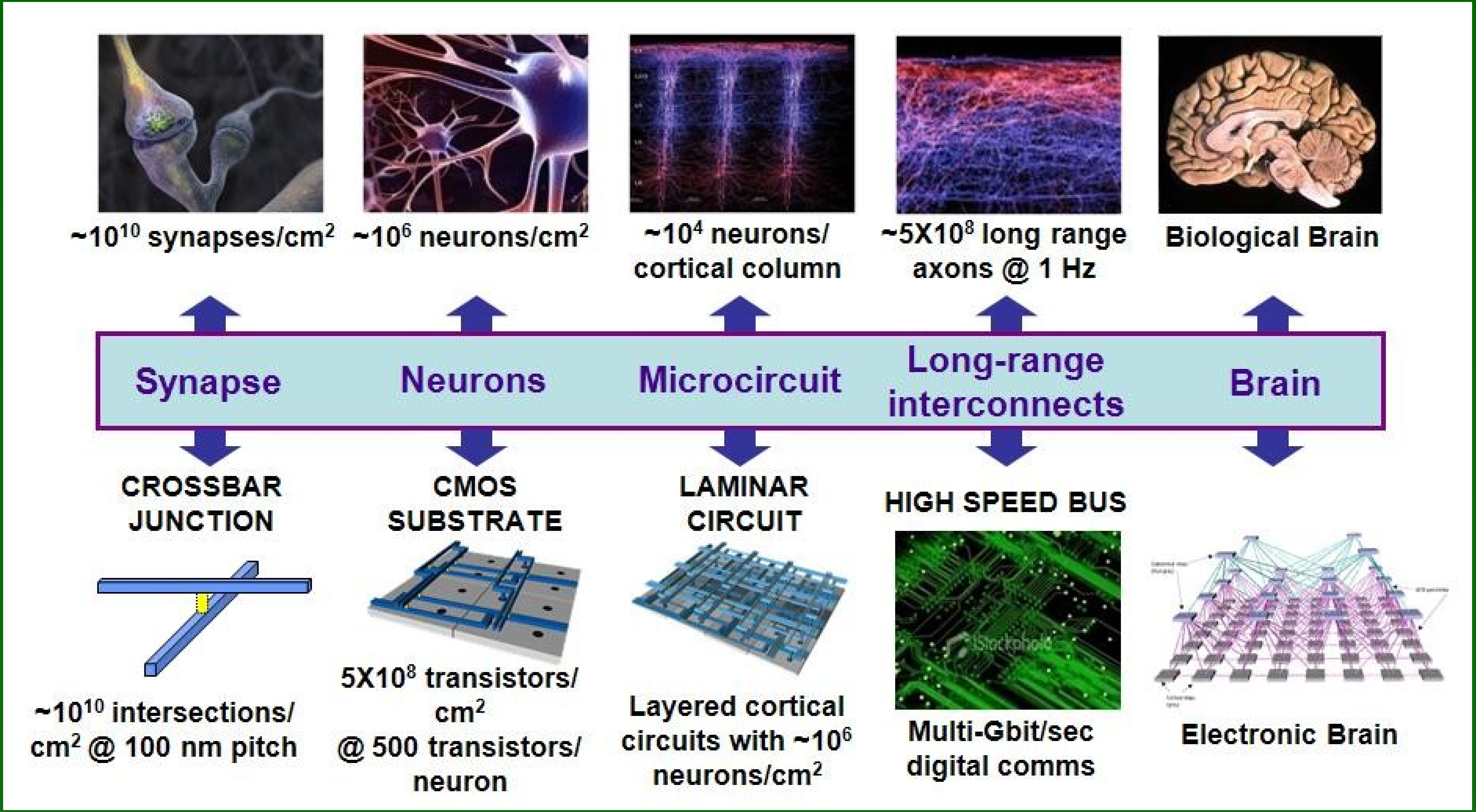
Activation of each word (sound, image) leads to specific brain activity distributions, each brain area may contribute to understanding of a number of words, representing specific qualities and facilitating associations.



h	Toggle this box
i	Toggle intro msg
+	Next Dataset
-	Prev Dataset
Lft mouse	Rotate
Mid mouse	Pan

Aktywacja pojęć prowadzi do aktywacji określonych struktur mózgu.
Każda ze struktur uczestniczy w semantycznej interpretacji wielu pojęć.
http://gallantlab.org/brainviewer/huthetal2012/

Nanotech for model brain



~10^{10} synapses/cm^2

~10^6 neurons/cm^2

~10^4 neurons/ cortical column

~$5X10^8$ long range axons @ 1 Hz

Biological Brain

Synapse

Neurons

Microcircuit

Long-range interconnects

Brain

CROSSBAR JUNCTION

CMOS SUBSTRATE

LAMINAR CIRCUIT

HIGH SPEED BUS

~10^{10} intersections/ cm^2 @ 100 nm pitch

$5X10^8$ transistors/ cm^2 @ 500 transistors/ neuron

Layered cortical circuits with ~10^6 neurons/cm^2

Multi-Gbit/sec digital comms

Electronic Brain

Source: DARPA Synapse, projekt koordynowany przez IBM (2008)

Neuromorficzne komputery

Synapse 2015: IBM TrueNorth chip
~1M neurons and ¼G synapses, ok 5.4G tranzystorów.
NS16e module=16 chips=16M neurons, >4G synapses, requires only 1.1 W!
Scaling: 256 modules, ~4G neurons, ~1T= 10^{12} synapses < 300 W power!
IBM Neuromorphic System can reach complexity of the human brain.

Recurrence plots

Trajectory of dynamical system (neural activities) may be visualized using recurrence plots (RP).

$$\mathbf{x}(t) = \{x_i(t)\}_{i=1..n}^{t=1..N}$$

Poincaré (1890) proved recurrence theorem:

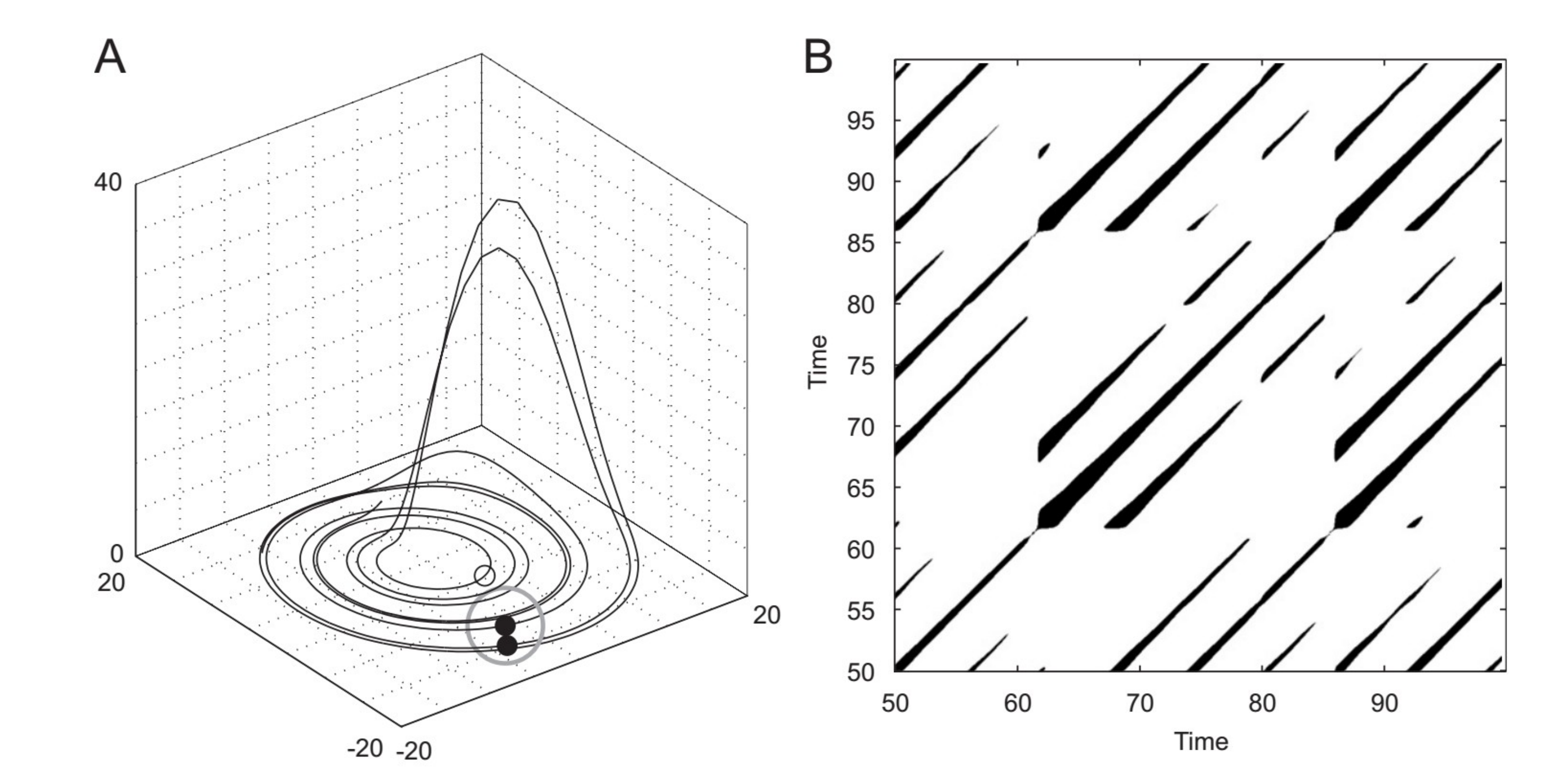
If we have a measure preserving transformation, the trajectory will eventually come back to the neighbourhood of any former point with probability one.

$$R_{ij} = \begin{cases} 1 : x_i \approx x_j \\ 0 : x_i \not\approx x_j \end{cases} \qquad i, j = 1 \cdots N$$

R is recurrence matrix based on approximate equality of N trajectory points. For discretized time steps binary matrix ***R***$_{\mathrm{ij}}$ is obtained.

Many measures of complexity and dynamical invariants are derived from RP matrices: generalized entropies, correlation dimensions, mutual information, redundancies, etc. Great intro: N. Marwan et al, Recurrence plots for the analysis of complex system. Physics Reports 438 (2007) 237–329

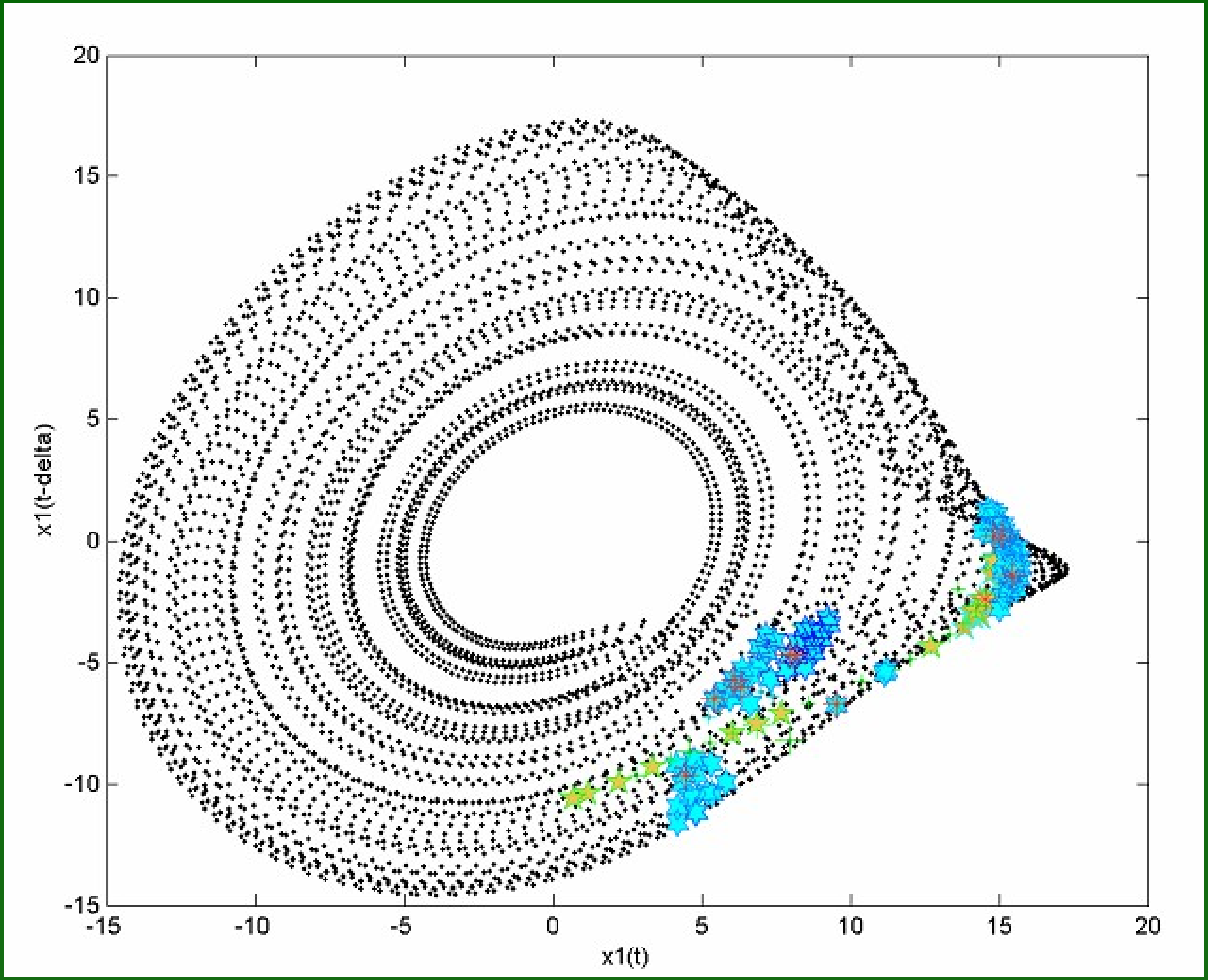
Rössler system



$$\dot{x} = -y - z; \dot{y} = x + ay; \dot{z} = b + z(x - c)$$

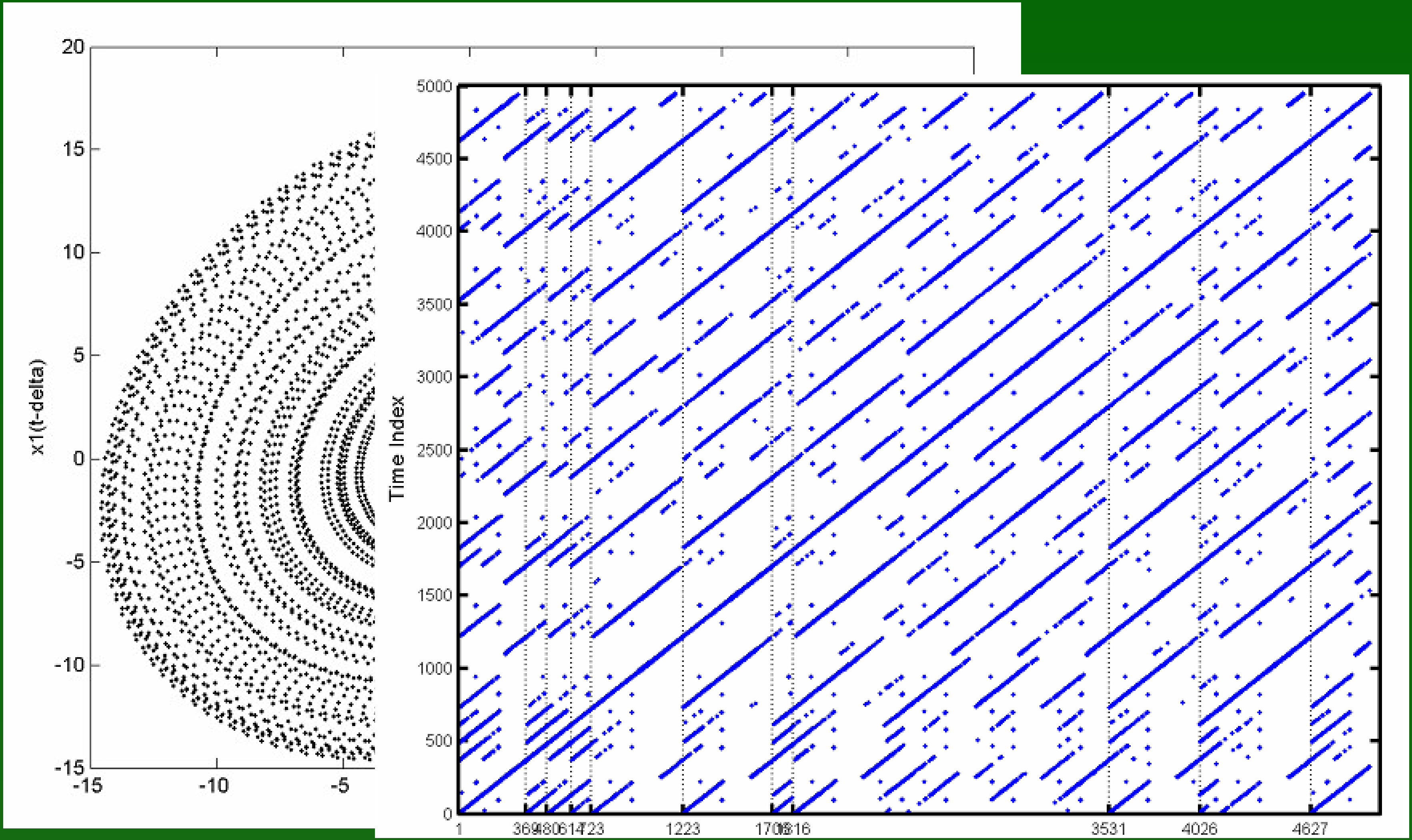
Recurrence plots

Embedding of time series via Taken's theorem: $x(t)$ is replaced by a vector $Y(t)=(x(t);x(t-\delta); \ldots x(t-k\delta))$. This recreates original dynamics for $k<2d+1$.



Recurrence plots

Embedding of time series via Taken's theorem: $x(t)$ is replaced by a vector $Y(t)=(x(t);x(t-\delta);$ … $x(t-k\delta))$. This recreates original dynamics for $k<2d+1$.



Symbolic Dynamics (SD)

SD: dynamical system is modeled by a discrete space of sequences of abstract symbols (states of the system). Dynamics is given by the shift operator, generating a discrete-time Markov process. In practice:

- Phase space is partitioned into regions labeled with different symbols A_i
- Every time the system trajectory is found in one of these regions appropriate symbol is emitted.
- Sequence of symbols gives a coarse-grained description of dynamics that can be analyzed using statistical tools.
- Although discretization of continuous dynamical states looses the fluid nature of cognition, symbolic dynamics gives an appropriate framework for cognitive representations (Spivey, Continuity of mind, 2007)
- SD is used for low-d systems. In high-d partitioning phase spaces will contain a huge number of regions with sharply defined boundaries, and sequences are not easy to comprehend.
- We are mostly interested in high-d dynamical systems, d>100.

Fuzzy Symbolic Dynamics (FSD)

$$R(t,t';\varepsilon)=\Theta\left(\varepsilon-\left\|x(t)-x(t')\right\|\right)$$

R matrix with real distances, or distances from reference points:

$$S(\mathbf{x}(t),\mathbf{x}_0)=\Theta\left(\varepsilon-\left\|\mathbf{x}(t)-\mathbf{x}_0\right\|\right) => \exp\left(-\left\|\mathbf{x}(t)-\mathbf{x}_0\right\|\right)$$

1. Standardize original data in high dimensional space.

2. Find cluster centers (e.g. by k-means algorithm): $\mu_1, \mu_2 \ldots \mu_d$

3. Use non-linear mapping to reduce dimensionality to d, for example:

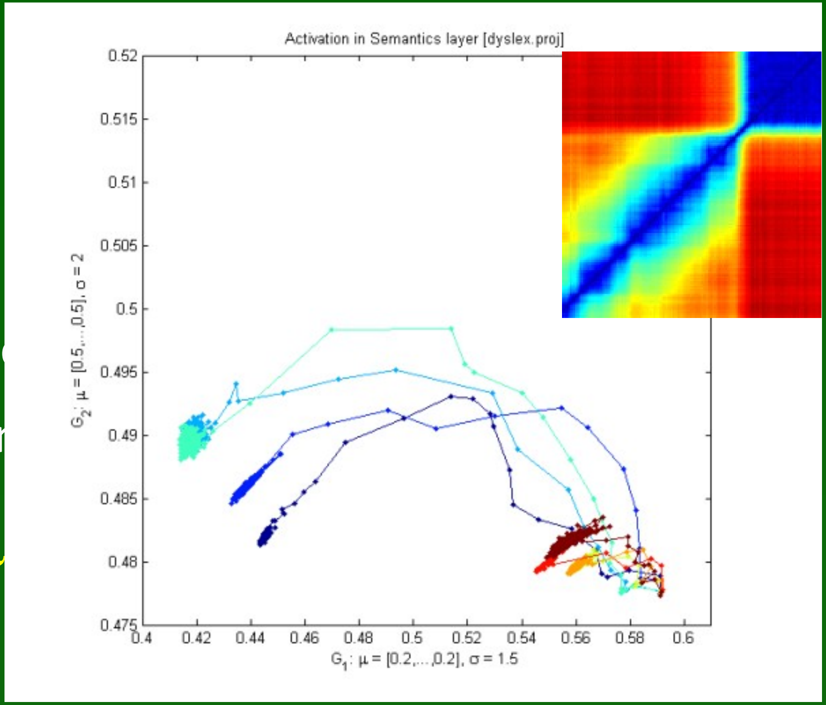
$$y_k(t;\mu_k,\Sigma_k)=\exp\left(-\left(x-\mu_k\right)^{\mathrm{T}}\Sigma_k^{-1}\left(x-\mu_k\right)\right)$$

Localized membership functions $y_k(t;W)$:

sharp indicator functions => symbolic dynamics; $x(t)$ => strings of symbols;

soft functions => fuzzy symbolic dynamics, dimensionality reduction
$Y(t)=(y_1(t;W), y_2(t;W))$ => visualization of high-dim data.

Fuzzy Symbolic Dynamics (FSD)



Complementing information in RPs:

RP plots $S(t,t_0)$ values as a matrix; FSD

1. Standardize data.
2. Find cluster centers (e.g. by k-means alg
3. Use non-linear mapping to reduce dimer

$$y_k(t;\boldsymbol{\mu}_k,\Sigma_k) = \exp\Big(-\big(\mathbf{x}(t)-\boldsymbol{\mu}$$

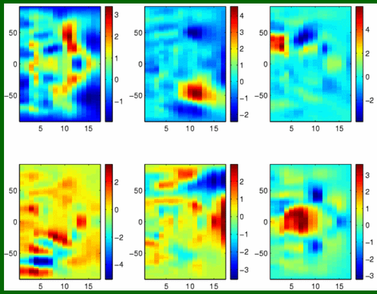
Localized membership functions $y_k(t;W)$:

sharp indicator functions => symbolic dynamics; $x(t)$ => strings of symbols;

soft membership functions => fuzzy symbolic dynamics, dimensionality reduction $Y(t)=(y_1(t;W), y_2(t;W))$ => visualization of high-dim data.

We may then see visualization of trajectory in some basin of attraction. Such visualizations are simply referred to as “attractors”.

FSD is good for you!



- Fuzzy symbolic dynamics is a natural way to generalize the notion of symbolic dynamics and recurrence plots.
- FSD provides dimensionality reduction, non-linear mapping for visualization of trajectories, shows various aspects of dynamics that are difficult to discover looking at individual components, local trajectory clusters and their relations.
- FSD can be applied to raw signals, transformed signals (ex. ICA/PCA components), or to signals in the time-frequency domain.
- Key: good reference points for membership functions.

Dobosz K, Duch W. (2010) Understanding Neurodynamical Systems via Fuzzy Symbolic Dynamics. Neural Networks Vol. 23 (2010) 487-496

Duch W, Dobosz K, *Visualization for Understanding of Neurodynamical Systems.* Cognitive Neurodynamics 5(2), 145-160, 2011.

FSD development

- Optimization of parameters of membership functions to see more structure from the point of view of relevant task.
- Learning: supervised clustering, projection pursuit based on quality of clusters => projection on interesting directions.
- Measures to characterize dynamics: position and size of basins of attractors, transition probabilities, types of oscillations around each attractor (follow theory of recurrent plots for more).
- Visualization in 3D and higher (lattice projections etc).
- Tests on model data and on the real data.

Model of reading

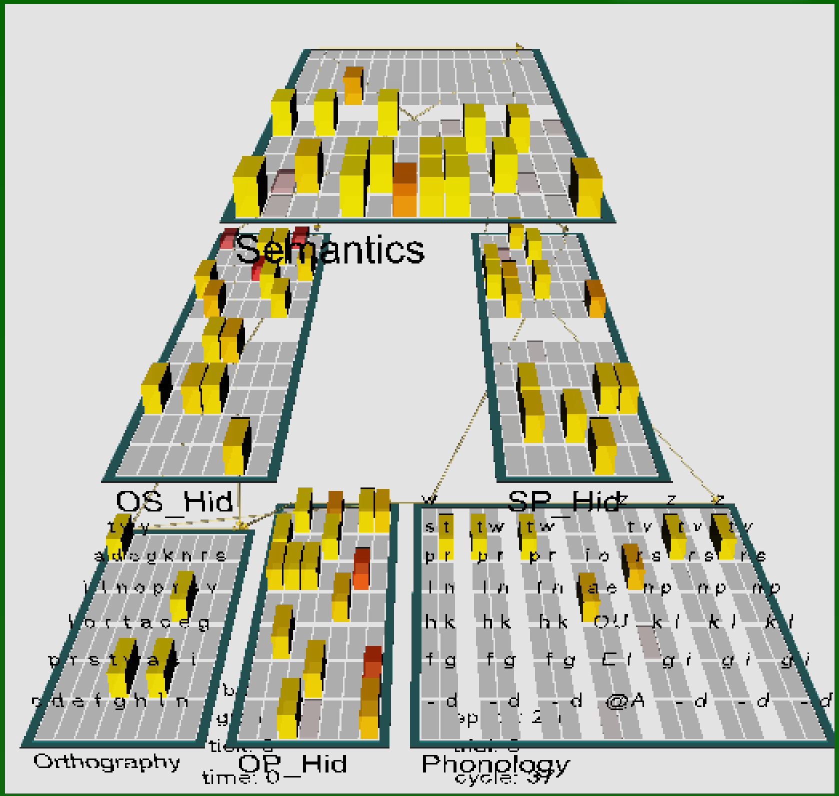
Emergent neural simulator:

Aisa, B., Mingus, B., and O'Reilly, R. The emergent neural modeling system. Neural Networks, 21, 1045-1212, 2008.

3-layer model of reading:

orthography, phonology, semantics, or distribution of activity over 140 microfeatures of concepts.

Hidden layers in between.



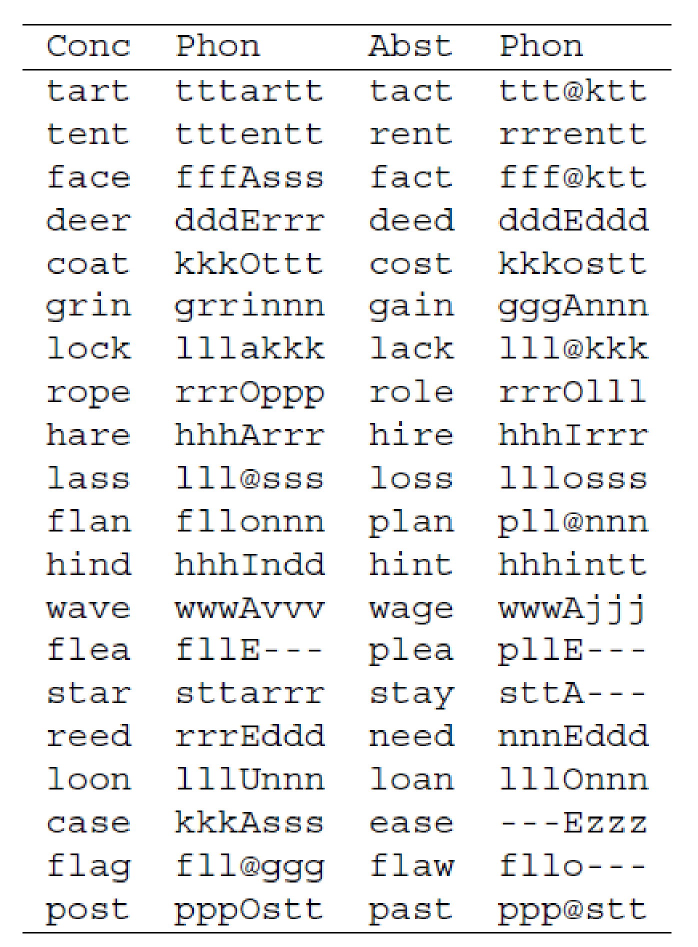
Neurons ~20 parameters, excitation, inhibition, leak currents.

Learning: mapping one of the 3 layers to the other two.

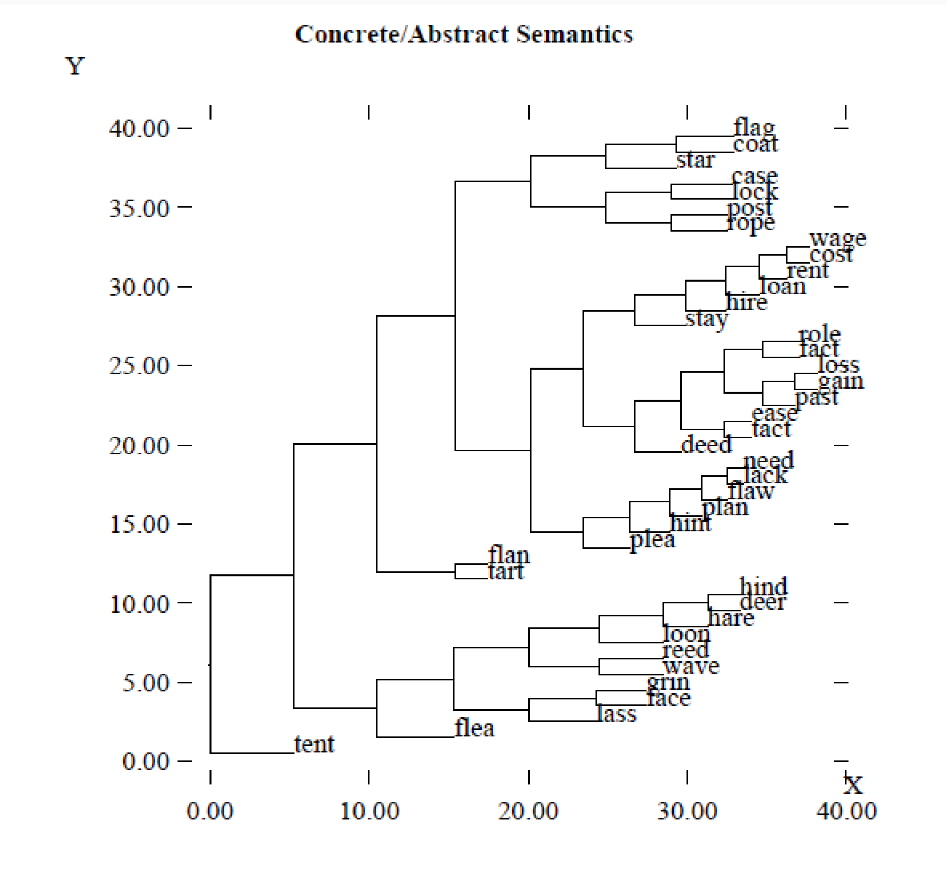
Fluctuations around final configuration = attractors representing concepts.

How to see properties of their basins, their relations?

Words to read



Conc	Phon	Abst	Phon
tart	tttartt	tact	ttt@ktt
tent	tttentt	rent	rrrentt
face	fffAsss	fact	fff@ktt
deer	dddErrr	deed	dddEddd
coat	kkkOttt	cost	kkkostt
grin	grrinnn	gain	gggAnnn
lock	lllakkk	lack	lll@kkk
rope	rrrOppp	role	rrrOlll
hare	hhhArrr	hire	hhhIrrr
lass	lll@sss	loss	lllosss
flan	fllonnn	plan	pll@nnn
hind	hhhIndd	hint	hhhintt
wave	wwwAvvv	wage	wwwAjjj
flea	fllE---	plea	pllE---
star	sttarrr	stay	sttA---
reed	rrrEddd	need	nnnEddd
loon	lllUnnn	loan	lllOnnn
case	kkkAsss	ease	---Ezzz
flag	fll@ggg	flaw	fllo---
post	pppOstt	past	ppp@stt

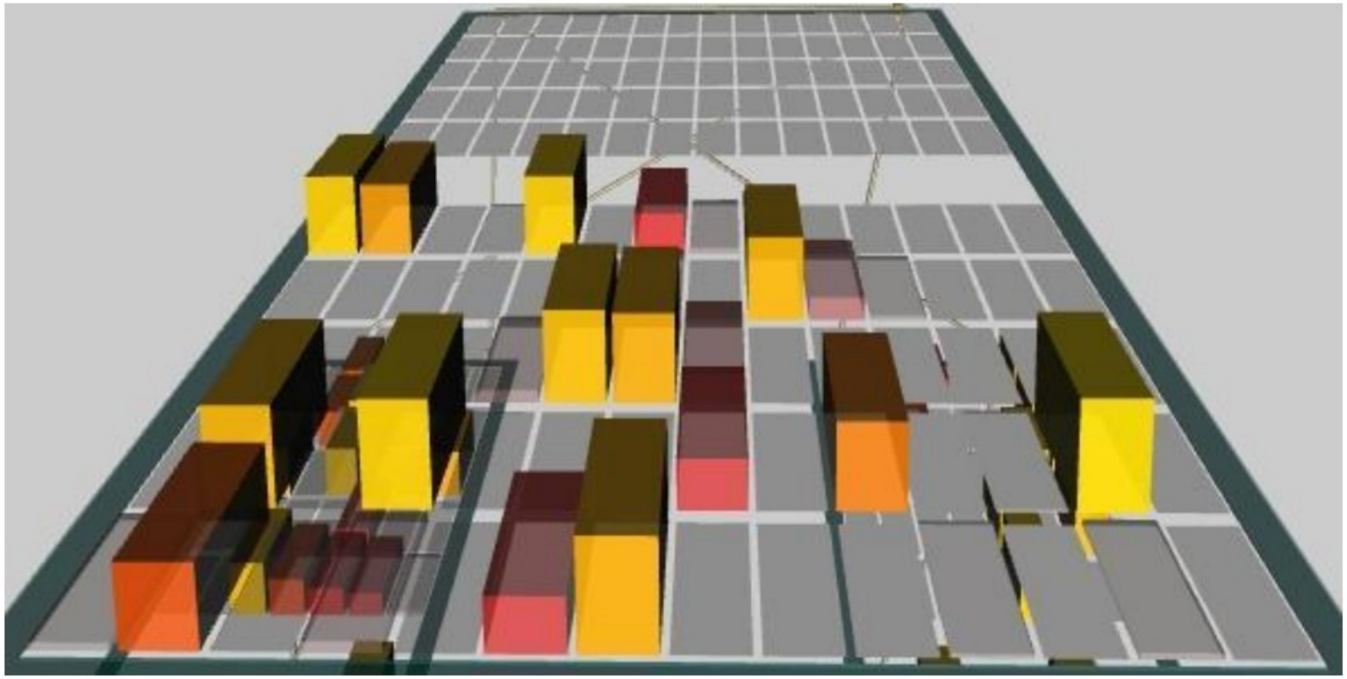


40 words, 20 abstract & 20 concrete;

dendrogram shows similarity in semantic layers after training.

Semantic layer

Semantic layer has 140 units; here activity for the “case” word is shown, upper 70 units code abstract microfeatures, lower physical.



Rys 22 Warstwa semantyczna sieci w trakcie sto dziesiątej iteracji (jeszcze „case” – odległość około 0,9326)

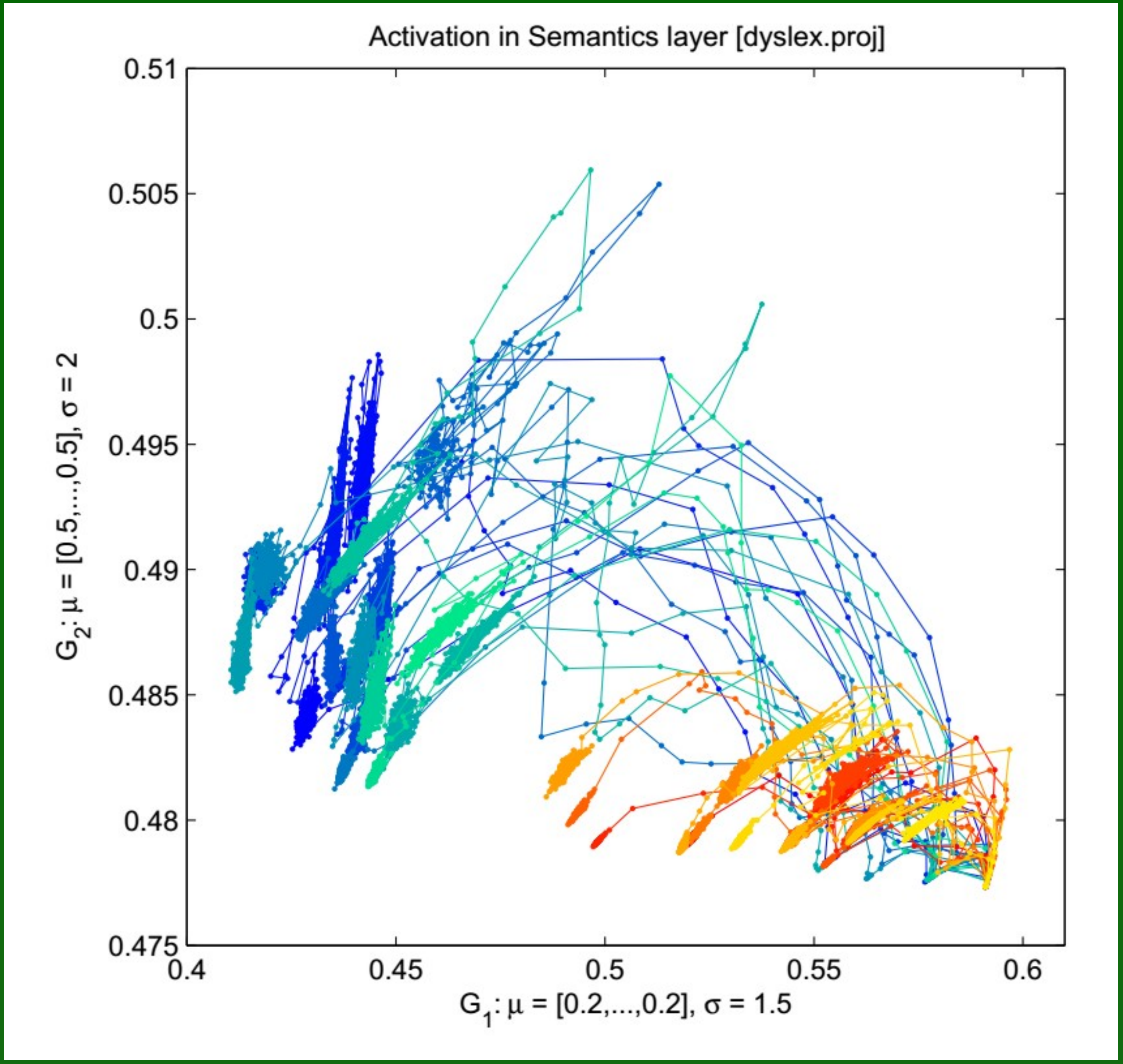
Attractors for words

FSD representation of 140-dim. trajectories in 2 or 3 dimensions for 40 words.

Attractor landscape changes in time due to neuron accommodation.

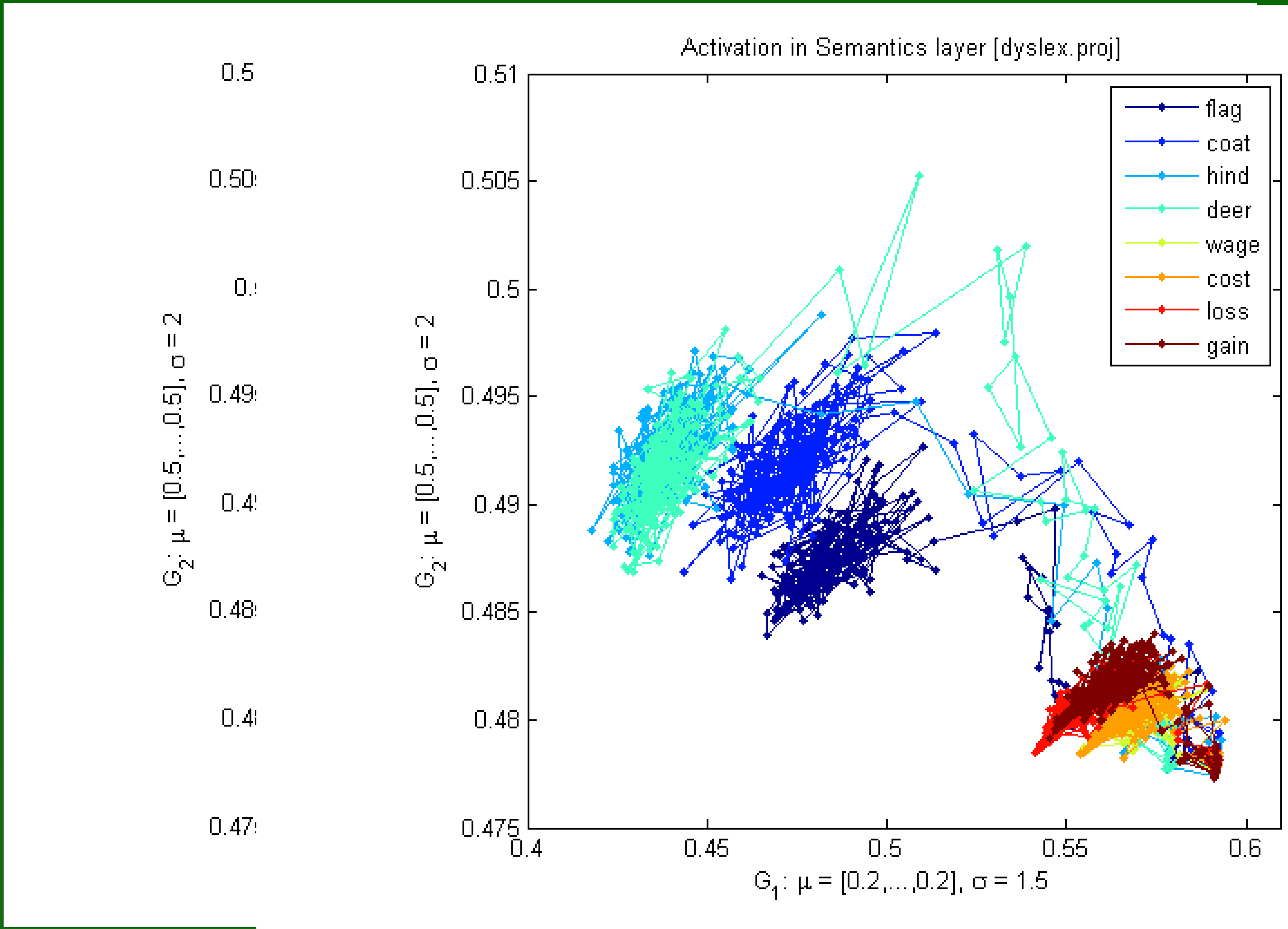
Cost and *rent* have semantic associations, attractors are close to each other, but without noise transitions between their basins of attractions do not occur.

Associations require some noise.



2D attractors for words

8 words, more synaptic noise, orthographic input, initially there is no activity in semantic layer. Hind and deer are almost identical, adding noise creates a single attractor basin.



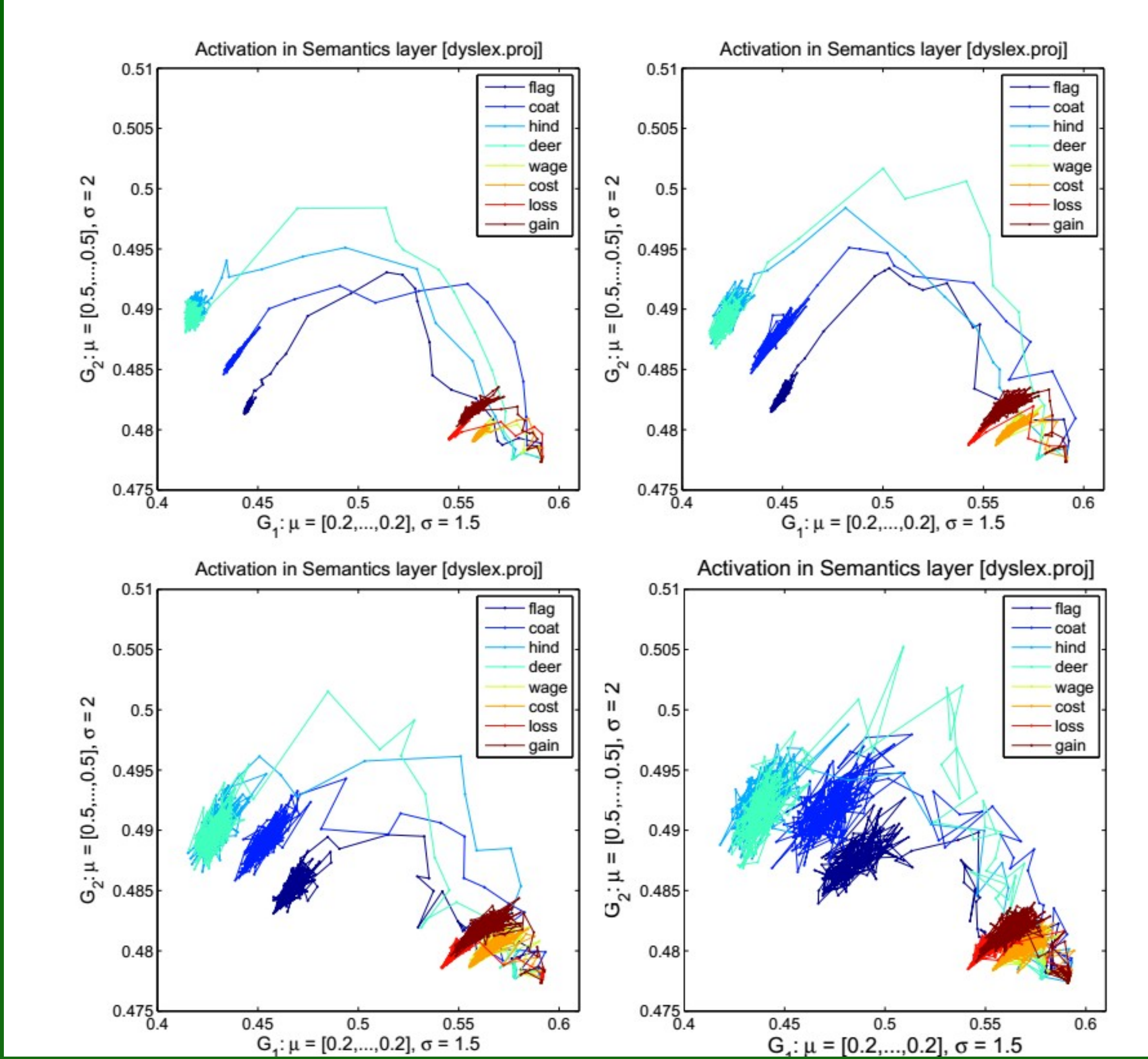
Influence of noise

Trajectories and basins of attraction for two correlated words

(*flag–coat* i *hind–deer*) and two abstract words (*wage–cost* i *lost–gain*)

Gaussian synaptic noise was increased from

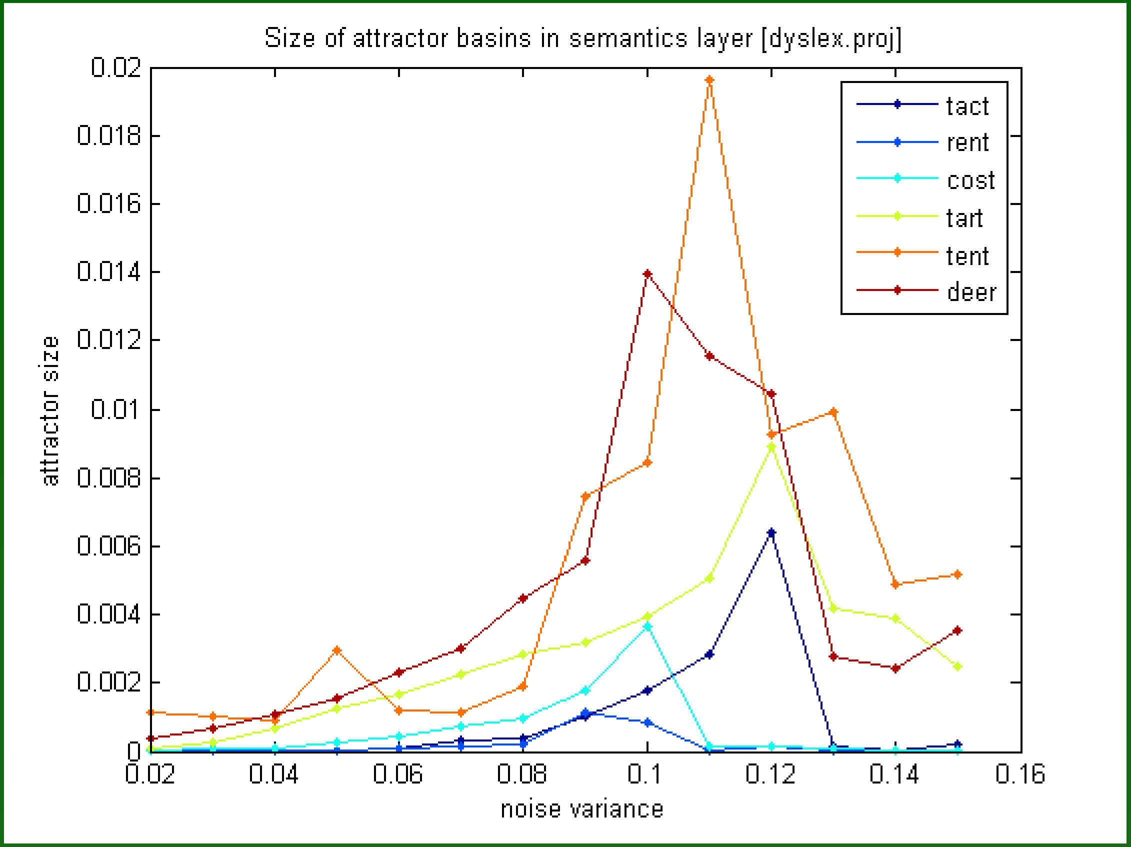
0.02, 0.04, 0.06 to 0.09.



Depth of attractor basins

Variance around the center of a cluster grows with synaptic noise; for narrow and deep attractors it will grow slowly, but for wide basins it will grow fast.

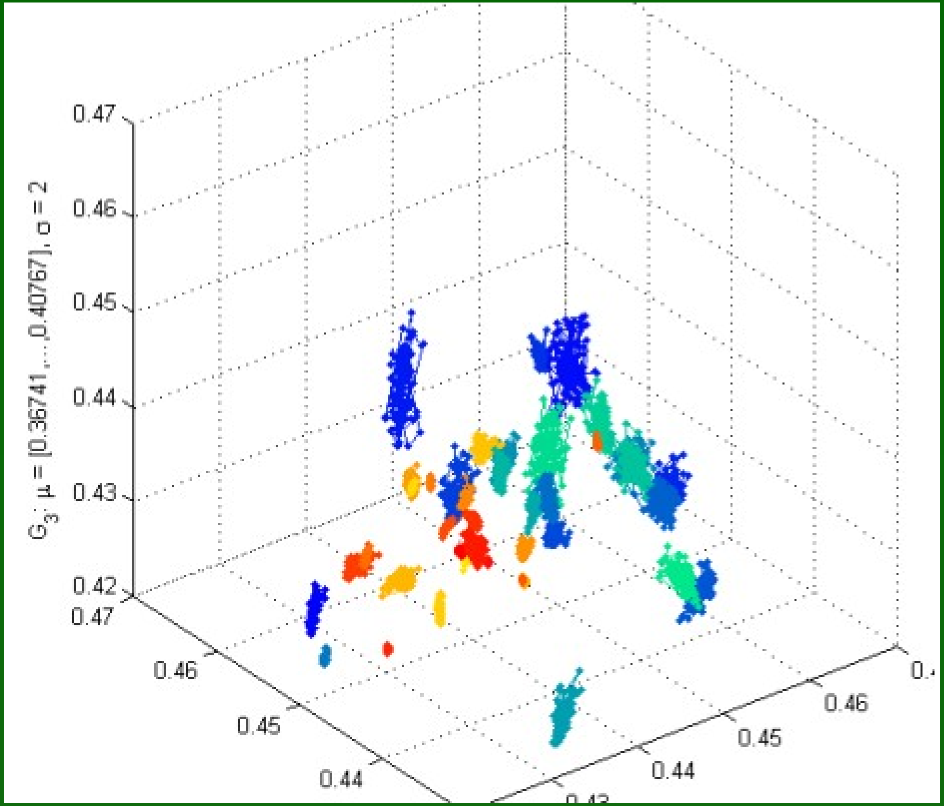
Jumping out of the attractor basin reduces the variance due to mutual inhibition of all desynchronized neurons.



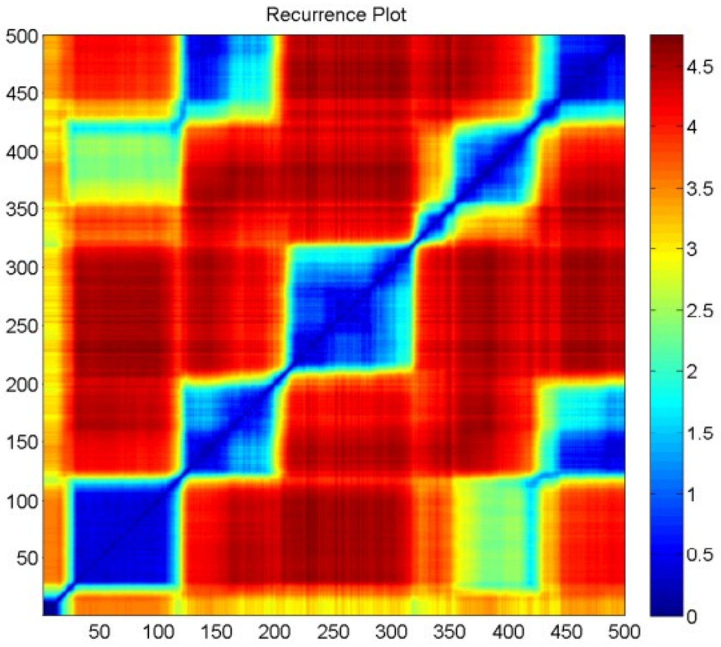
3D attractors for words

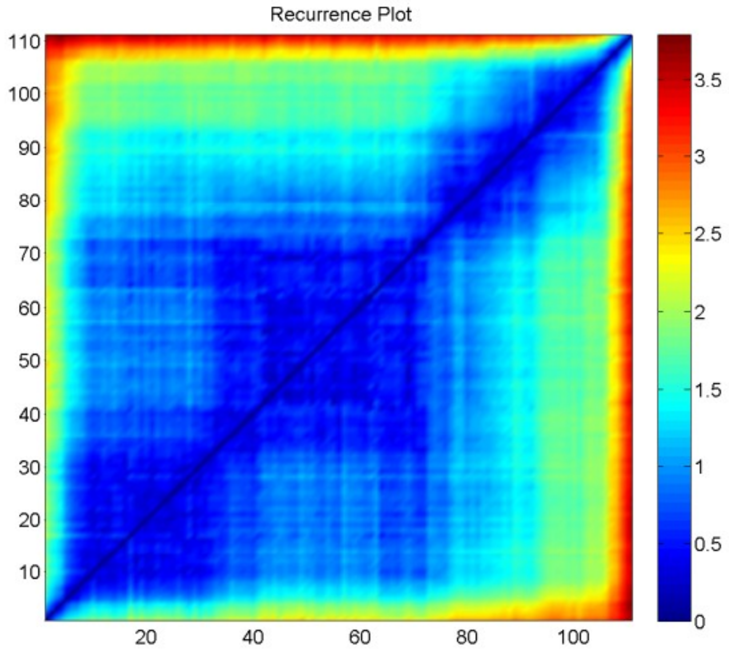
Non-linear visualization of activity of the semantic layer with 140 units for the model of reading that includes phonological, orthographic and semantic layers + hidden layers.

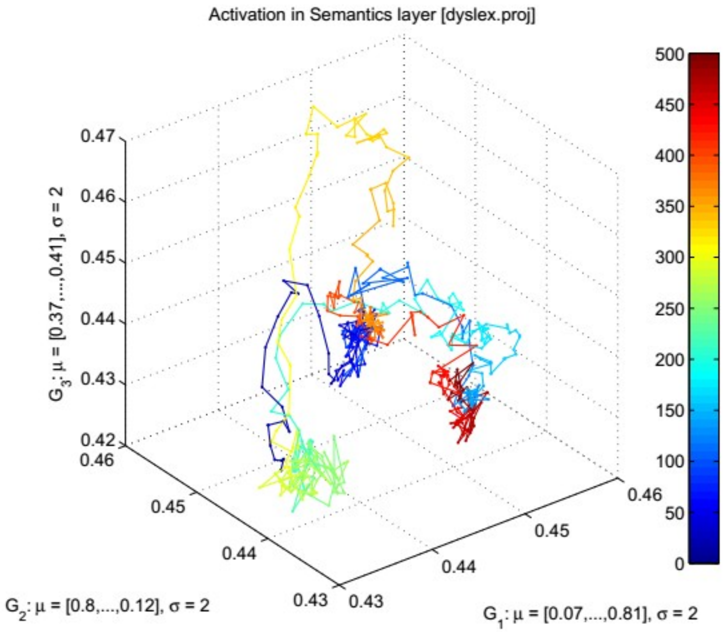
Cost /wage, hind/deer have semantic associations, attractors are close to each other, but without neuron accommodation attractor basins are tight and narrow, poor generalization expected.

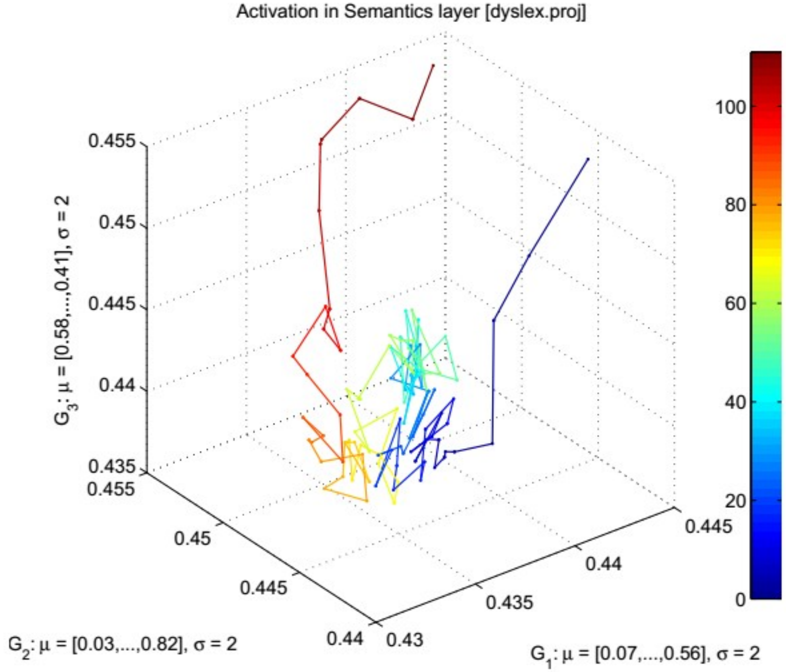


Training with more variance in phonological and written form of words may help to increase attractor basins and improve generalization.

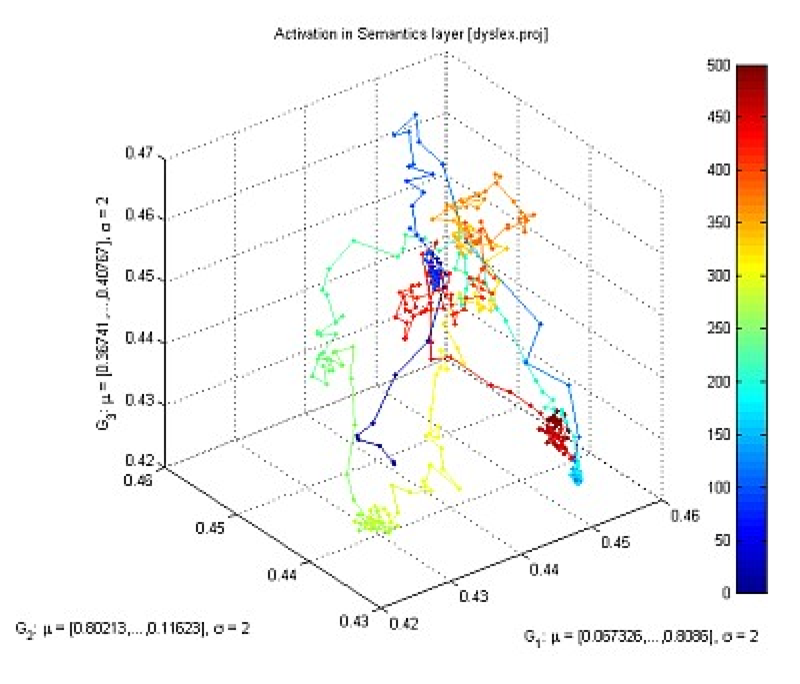

Recurrence Plot
500
450
400
350
300
250
200
150
100
50
50 100 150 200 250 300 350 400 450 500
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0

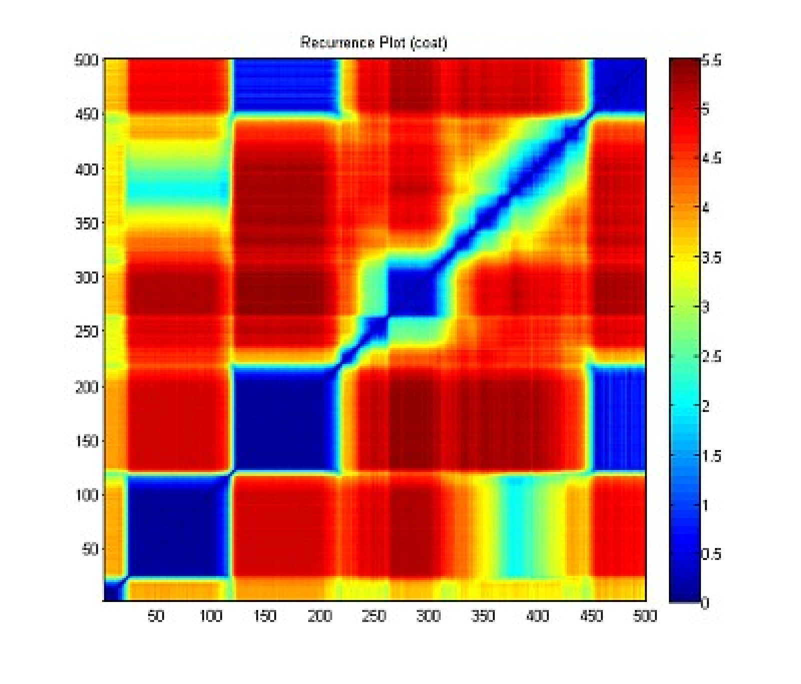

Recurrence Plot
110
100
90
80
70
60
50
40
30
20
10
20 40 60 80 100
3.5
3
2.5
2
1.5
1
0.5
0


Activation in Semantics layer [dyslex.proj]
G$_3$: μ = [0.37,...,0.41], σ = 2
0.47
0.46
0.45
0.44
0.43
0.42
G$_2$: μ = [0.8,...,0.12], σ = 2
0.46
0.45
0.44
0.43
G$_1$: μ = [0.07,...,0.81], σ = 2
0.43
0.44
0.45
0.46
500
450
400
350
300
250
200
150
100
50
0


Activation in Semantics layer [dyslex.proj]
G$_3$: μ = [0.58,...,0.41], σ = 2
0.455
0.45
0.445
0.44
0.435
G$_2$: μ = [0.03,...,0.82], σ = 2
0.455
0.45
0.445
0.44
G$_1$: μ = [0.07,...,0.56], σ = 2
0.43
0.435
0.44
0.445
100
80
60
40
20
0

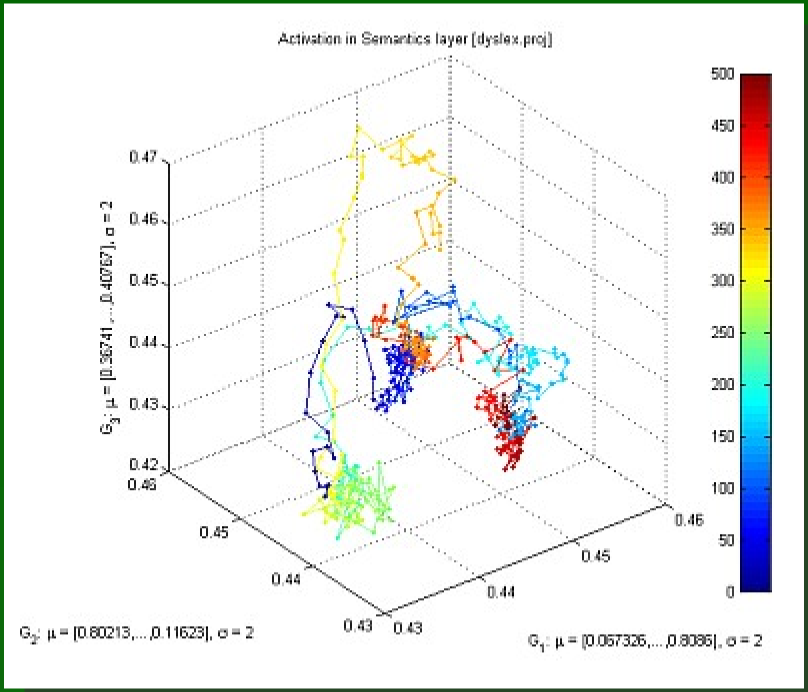
Fast transitions

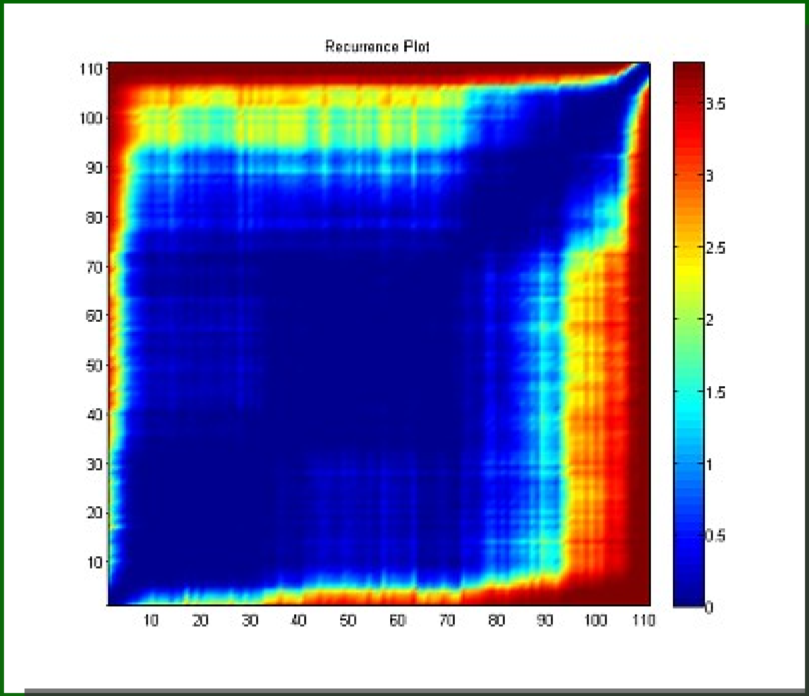




Attention is focused only for a brief time and than moved to the next attractor basin, some basins are visited for such a short time that no action may follow, no chance for other neuronal groups to synchronize. This corresponds to the feeling of confusion, not being conscious of fleeting thoughts.

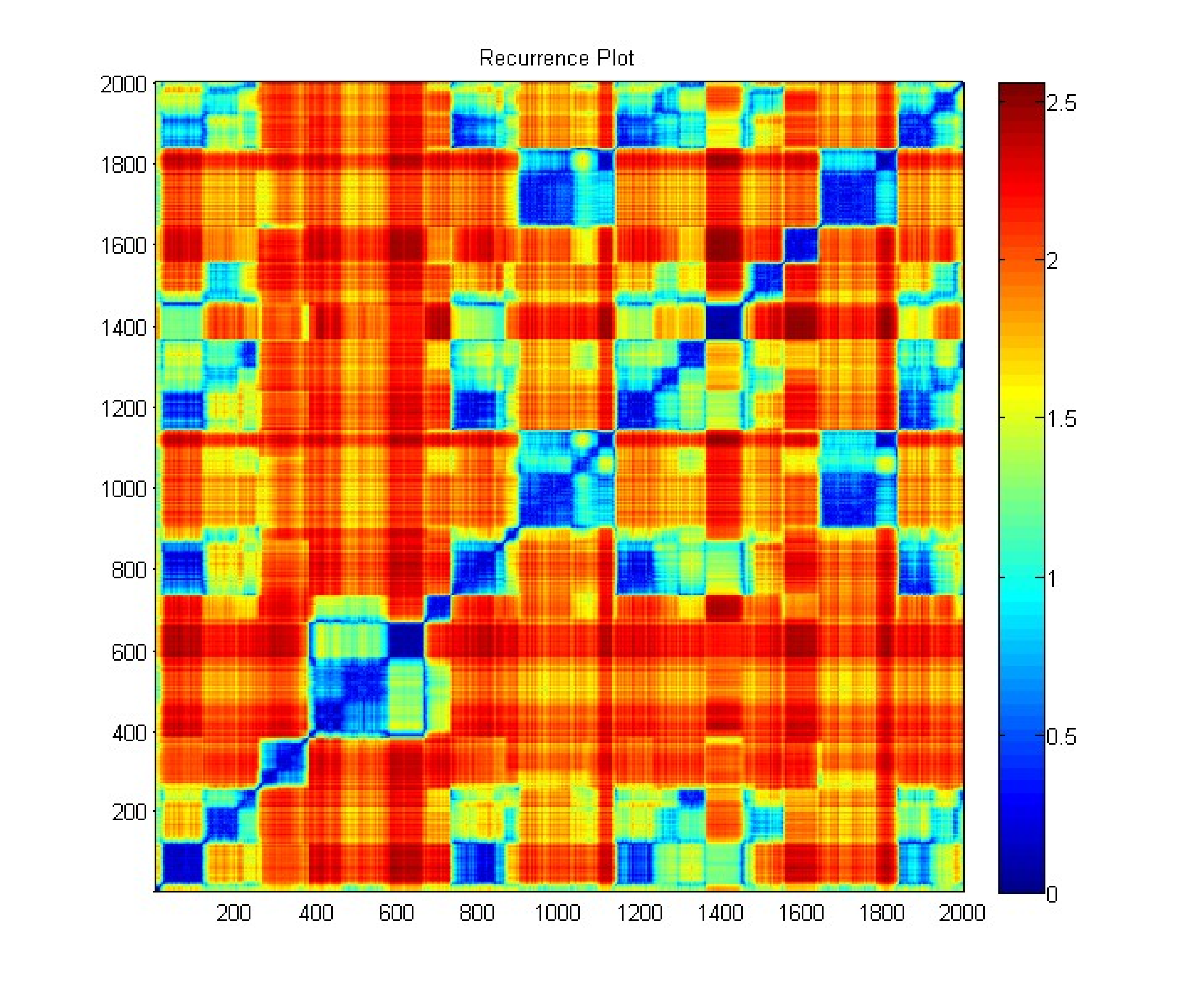
Recurrence plots





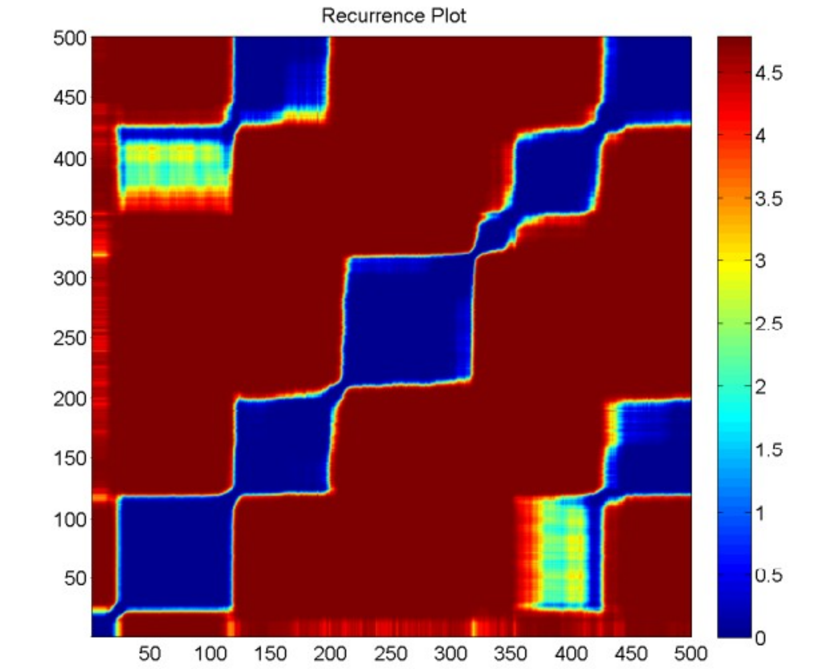
Starting from the word "flag", with small synaptic noise (var=0.02), the network starts from reaching an attractor and moves to another one (frequently quite distant), creating a "chain of thoughts".

Same trajectories displayed with recurrence plots, showing roughly 5 larger basins of attractors and some transient points.

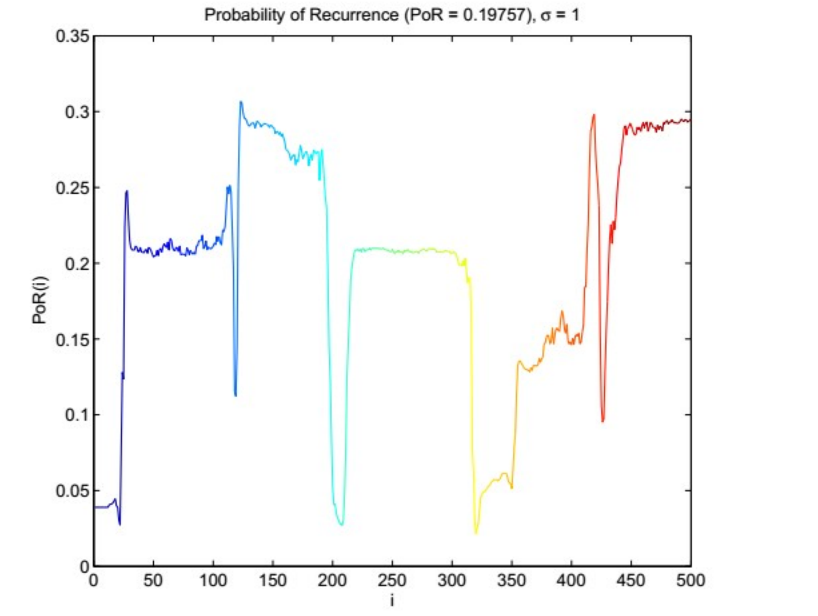


„Gain”: trajectory of semantic activations quickly changes to new prototype synchronized activity, periodically returns at 800, 1200, 1900.

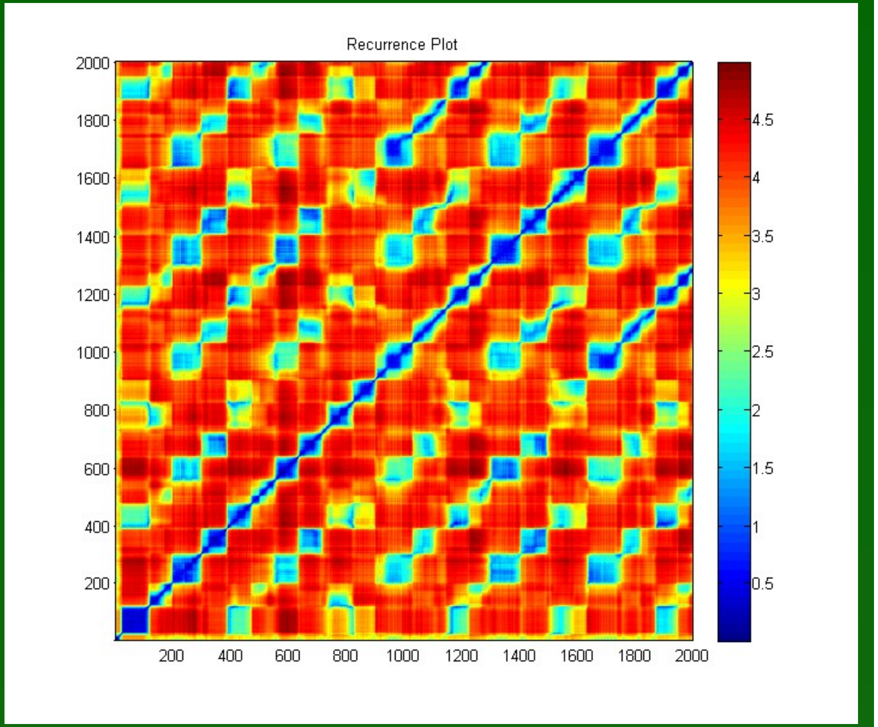
Aggressive smoothing.

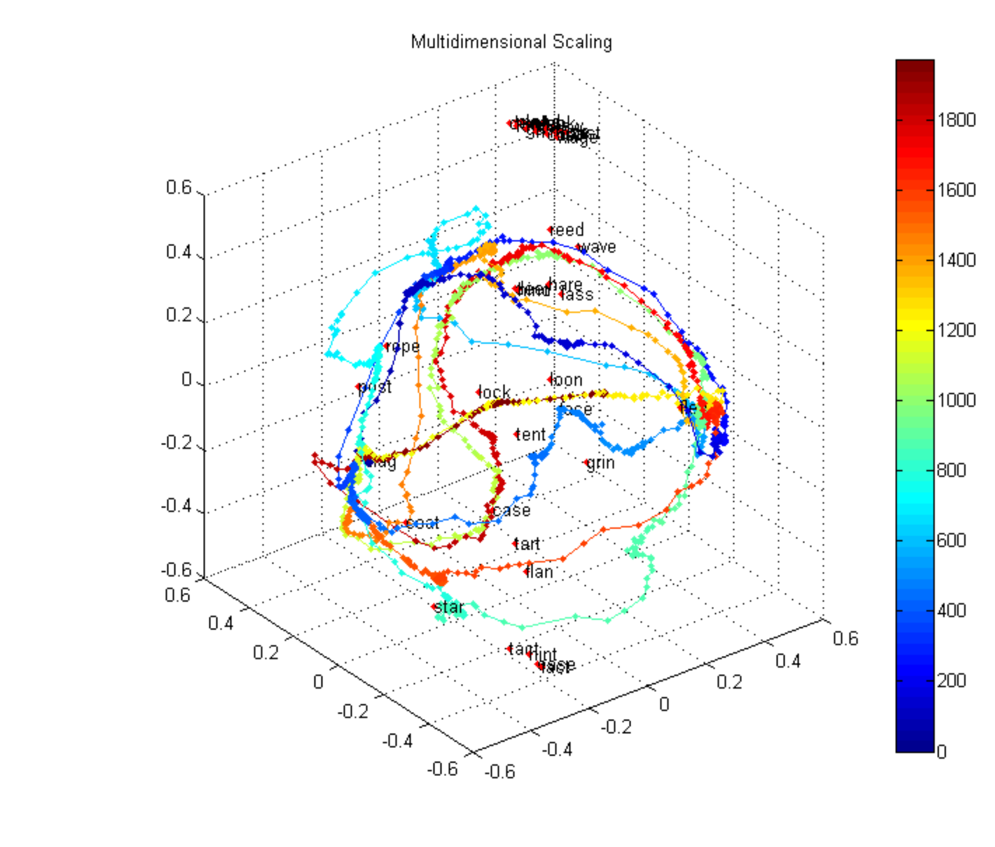


Probability of recurrence.
Total probability of recurrence is ~0.2

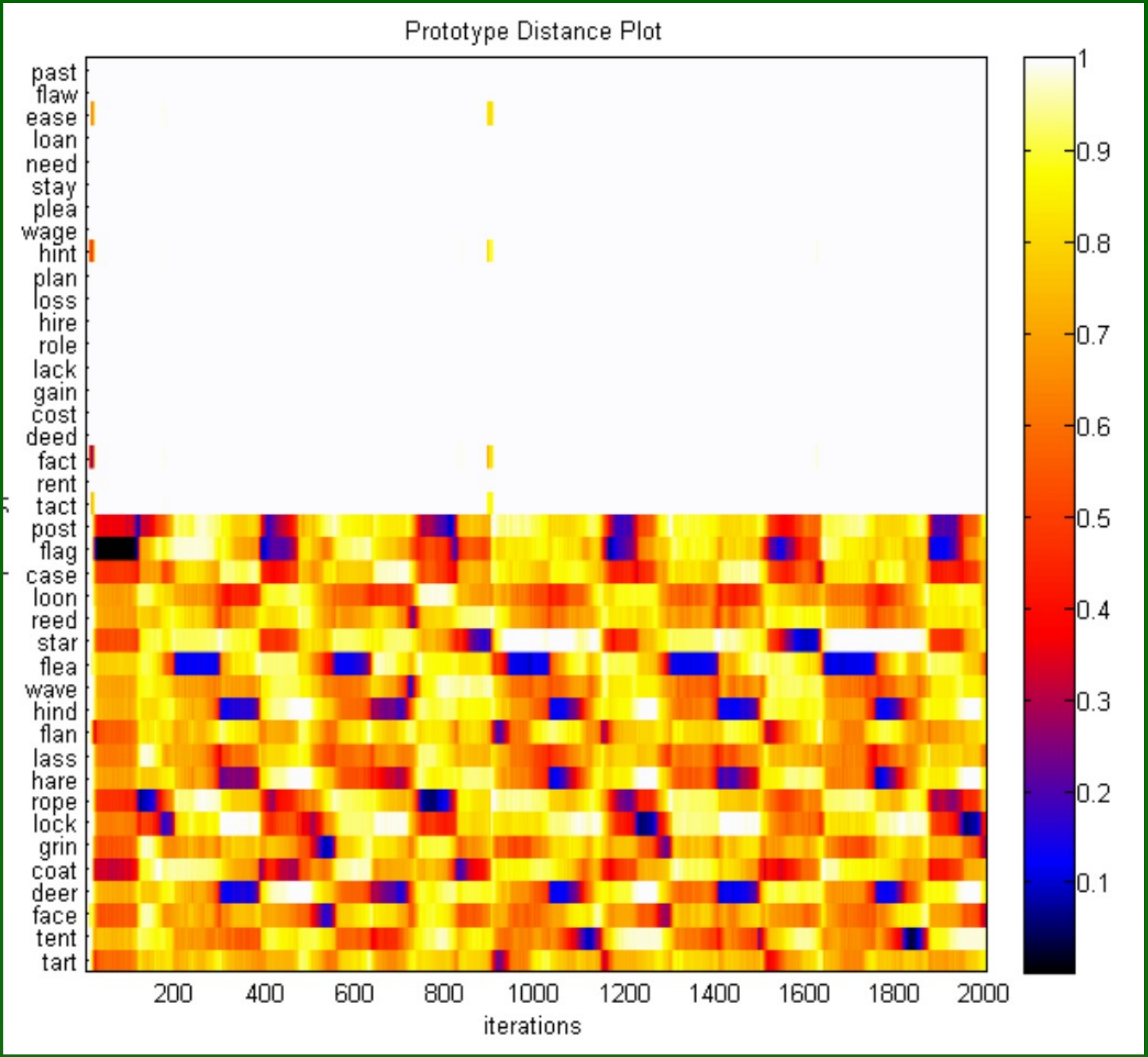


Long trajectories

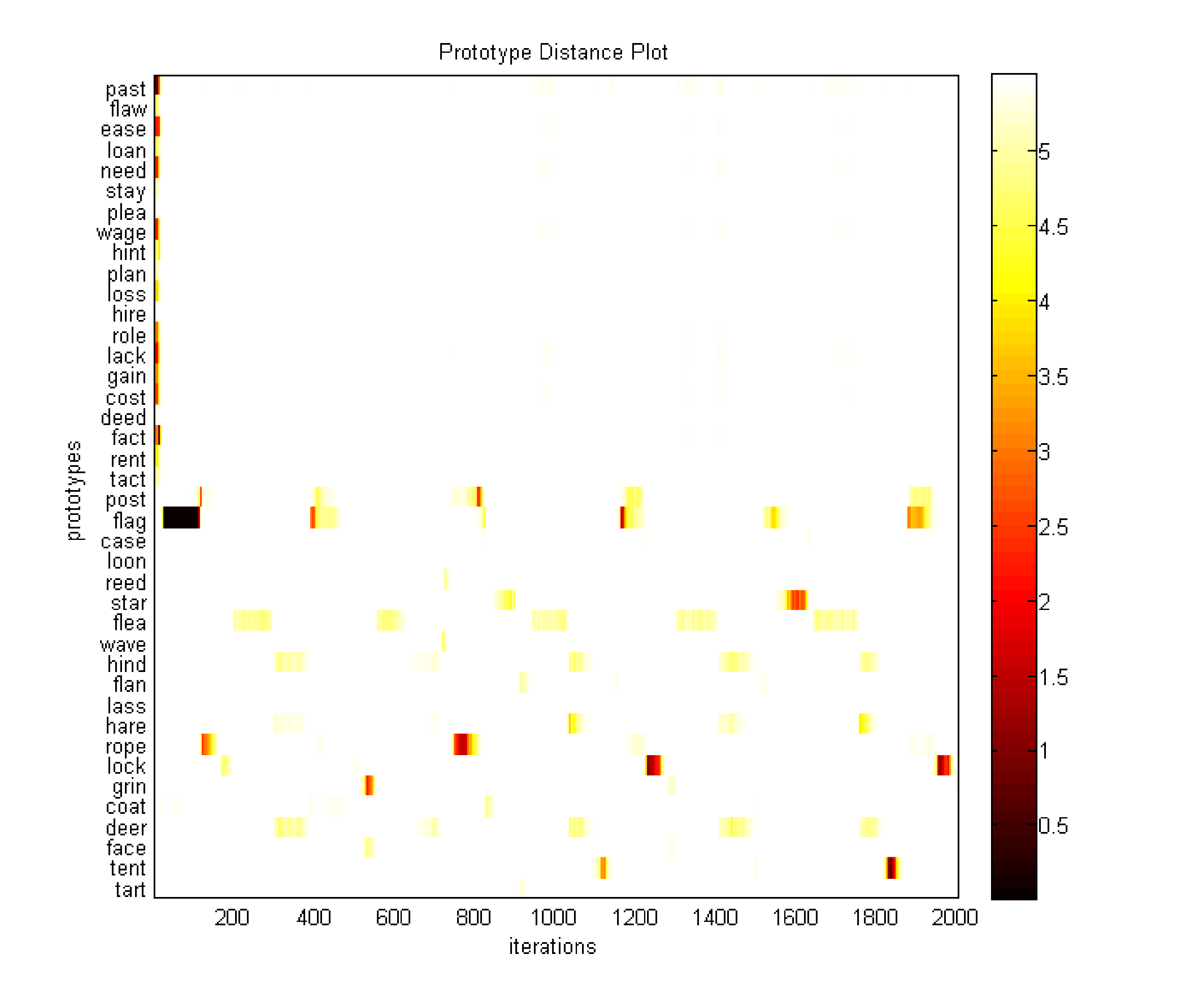




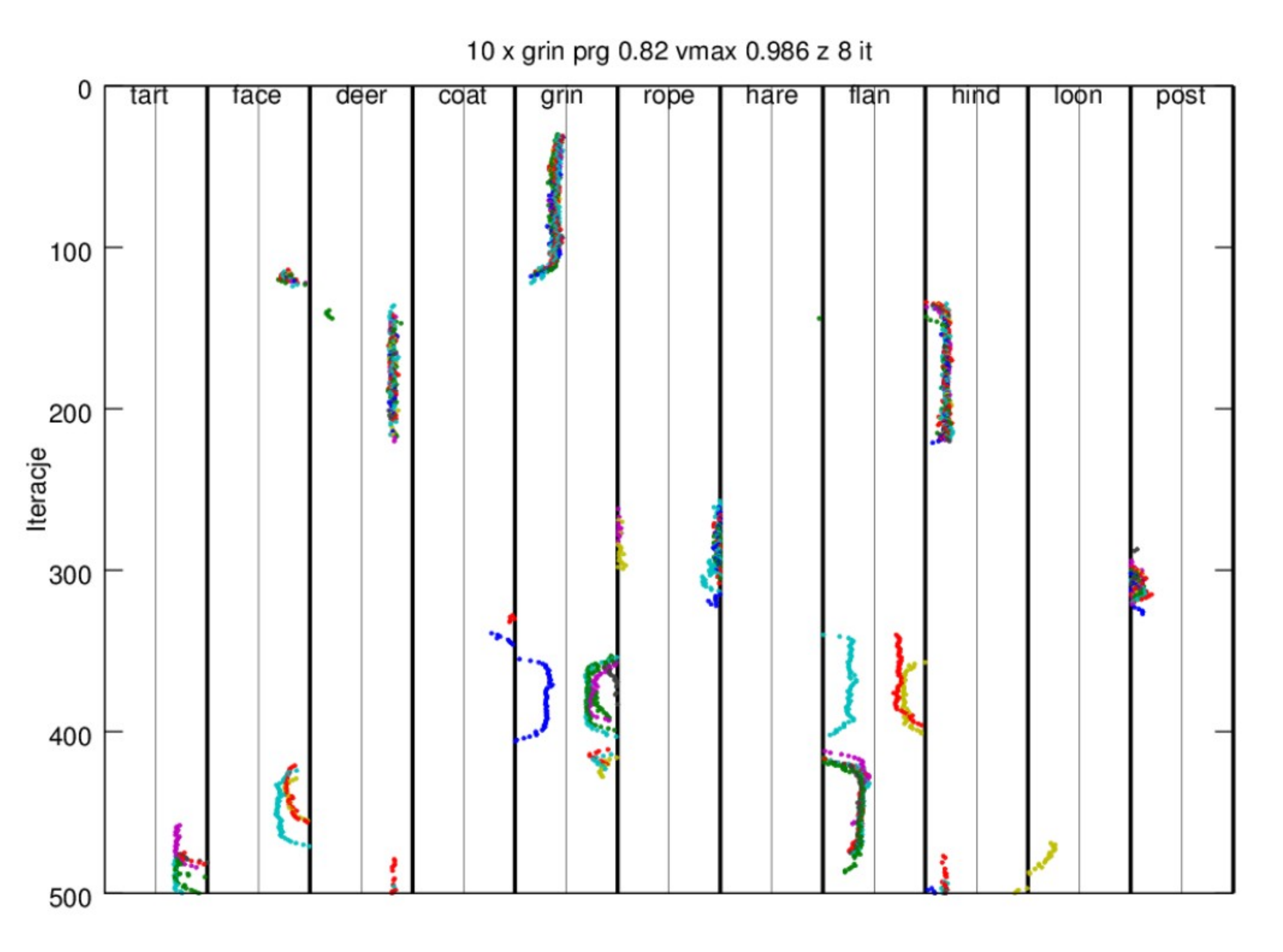
Recurrence plots vs MDS, starting with the word “flag” in 40-words microdomain.



PDP for transitions starting from „flag”



Long trajectory and stronger smoothing.

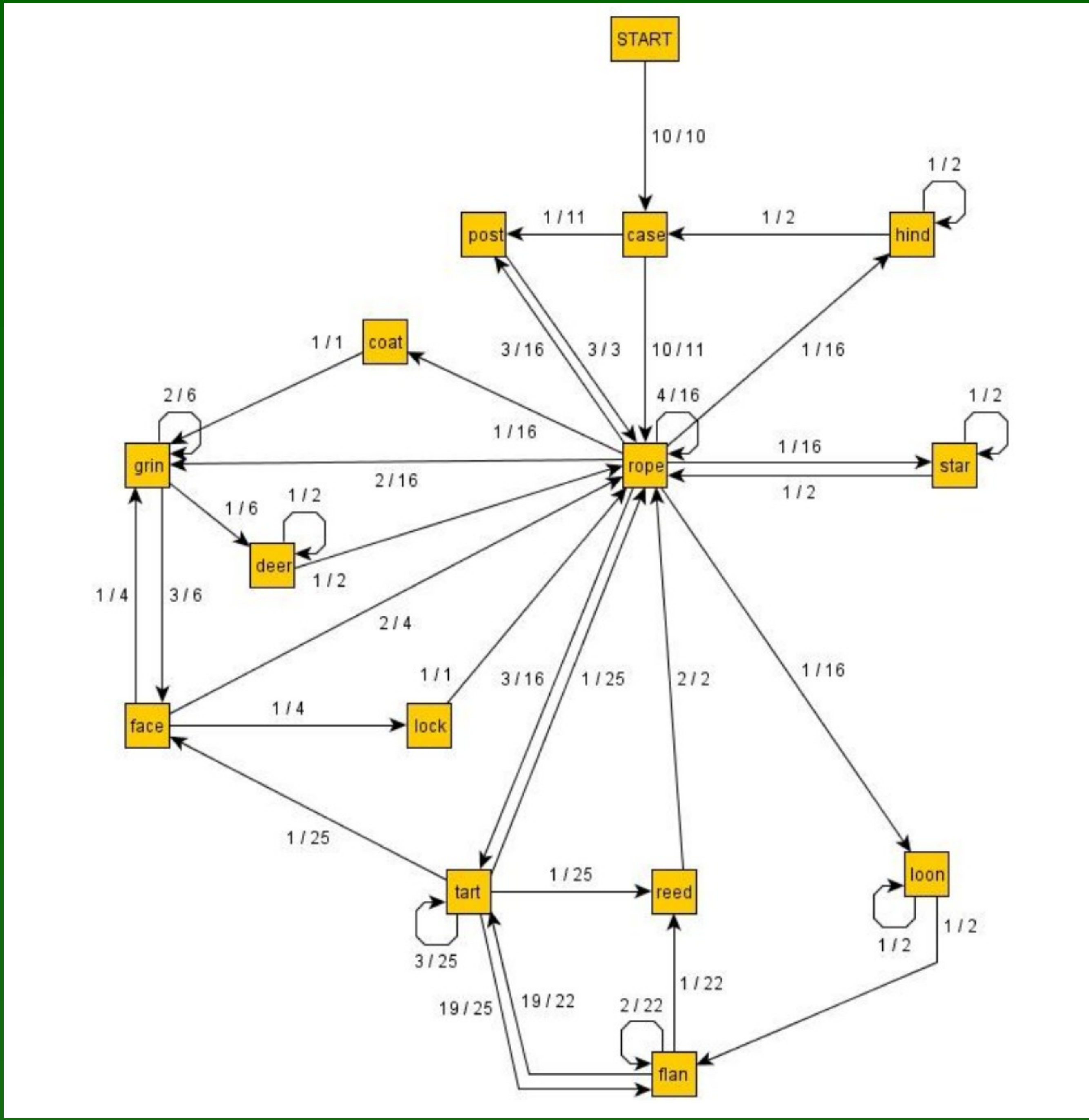


Plots showing only closest attractors for several runs, each plotted with different colors, side shows from which side trajectory came.
Starting with grin => face => deer or hind => …

Transitions between attractors after 10 runs.

Why these particular transitions?

Connected attractors share some microfeatures, some are deactivated, but visualization using RP or FSD does not show such details.
In the phase space dimensions are rescaled during dynamics.



MDS word mapping

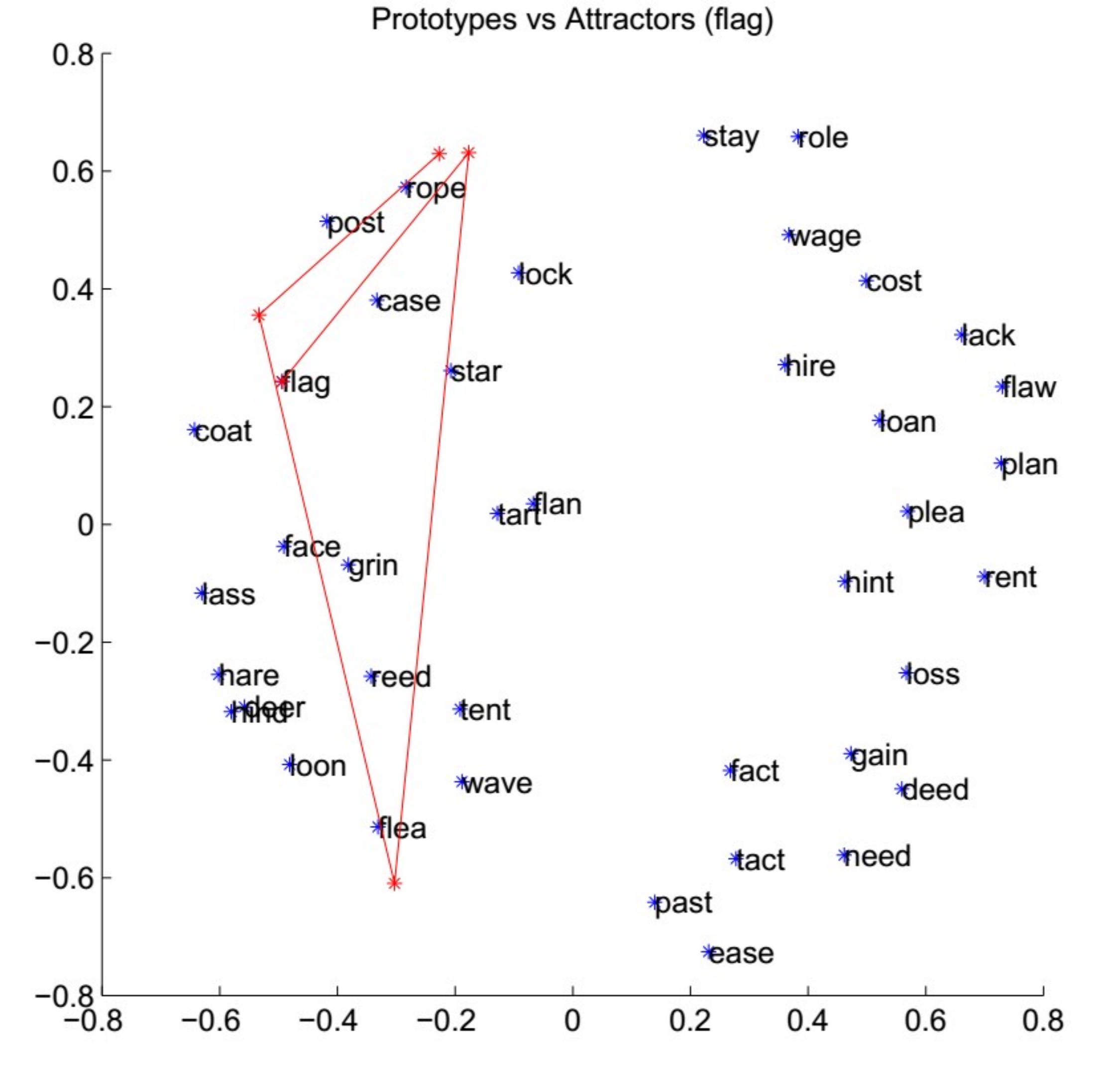
MDS representation of all 40 words, showing similarities of their 140 dimensional vectors.

Attractors are in some cases far from words.

Transition

Flag => rope => flea,

not clear why such big jump.



Optimization of FSD parameters

- Find centers of attractors (clusters) **P**$_i$ and calculated distances:

$$\mathbf{D}_{ij} = \|\mathbf{P}_i - \mathbf{P}_j\| \, .$$

- Map **P**$_i$ to low (2 or 3) $\bar{d}$ space using Gaussian functions:

$$\mathbf{G}(\mathbf{P}_i; \mathbf{Q}_k, \sigma_k) = \left[G_k(\mathbf{P}_i; \mathbf{Q}_k, \sigma_k) \right]_{k=1,\dots,\bar{d}} \, ,$$

- To preserve distances D$_{ij}$ in low-d space calculate g_{ij}

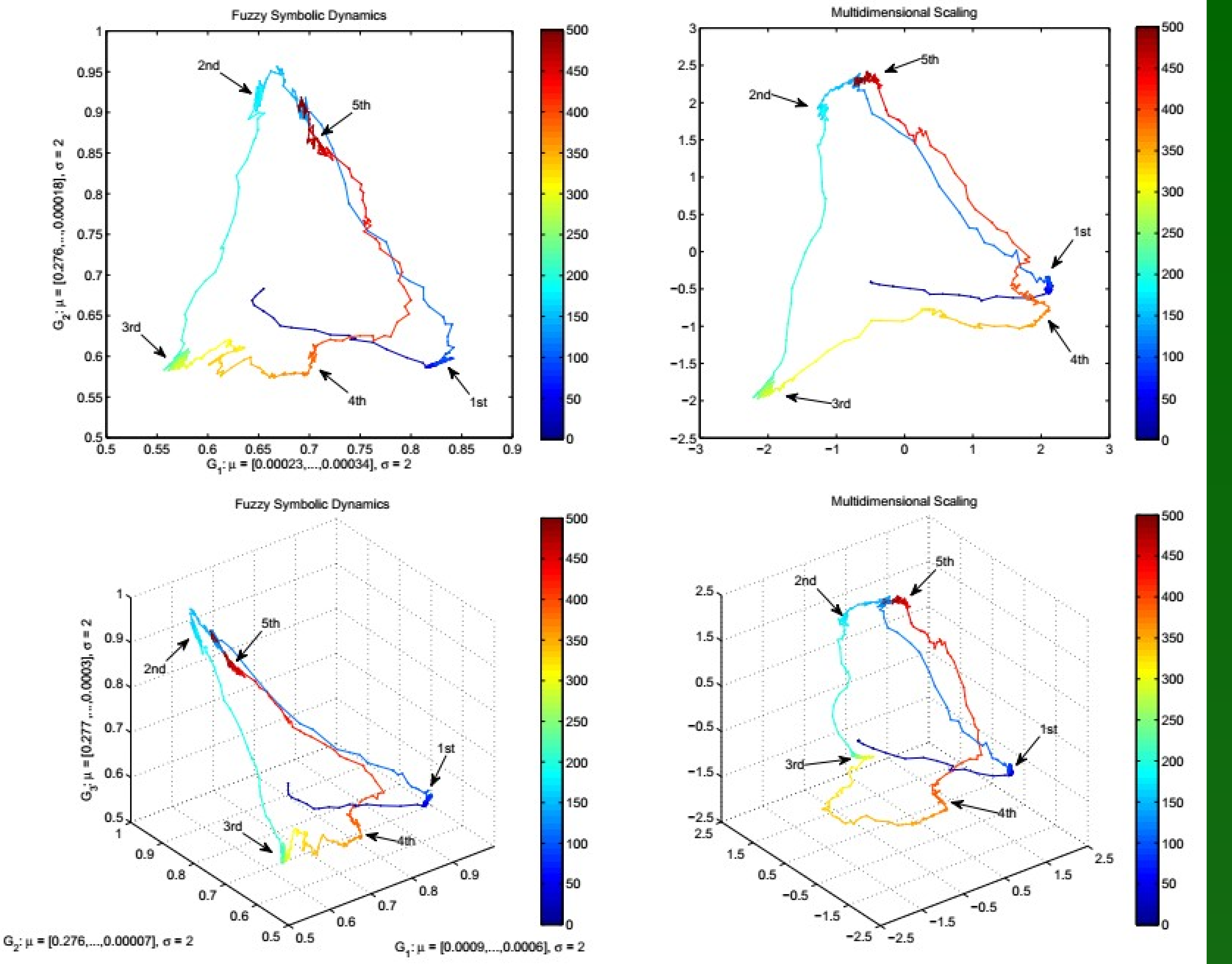
$$g_{ij}(\mathbf{Q}_k, \sigma_k) = \|\mathbf{G}(\mathbf{P}_i; \mathbf{Q}_k, \sigma_k) - \mathbf{G}(\mathbf{P}_j; \mathbf{Q}_k, \sigma_k)\| \, .$$

- and minimize stress function:

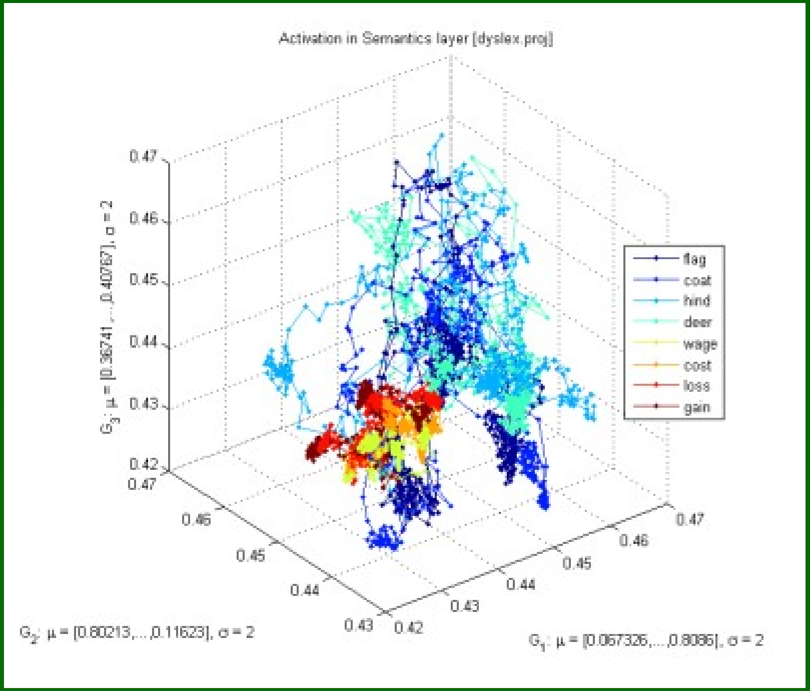
$$\mathcal{I}(\mathbf{Q}_k, \sigma_k) = \sum_{i>j} \|g_{ij}(\mathbf{Q}_k, \sigma_k) - \mathbf{D}_{ij}\| \, ,$$

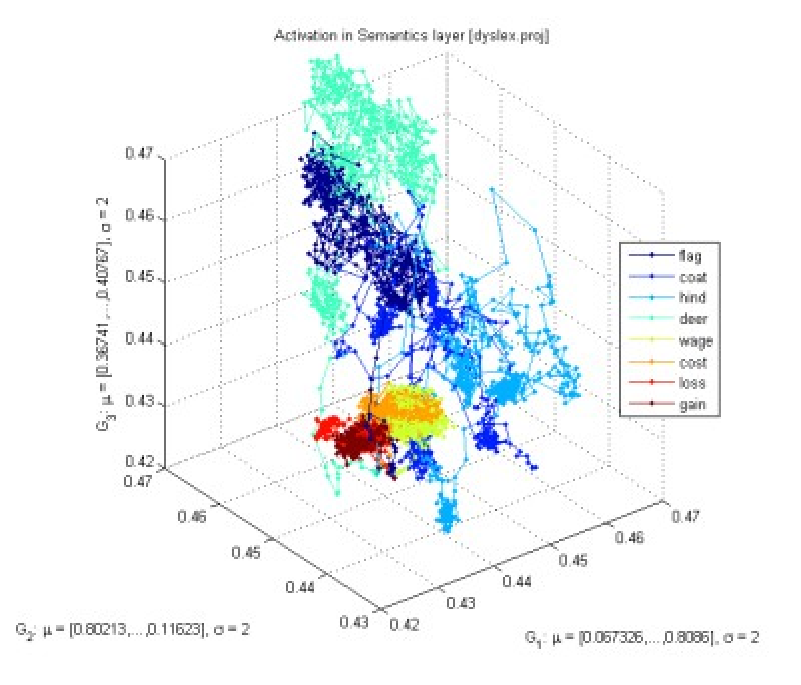
to find best positions **Q**$_k$ (+ and dispersions σ_k), k = 1, ..., $\bar{d}$

- Simpler minimization: select only a few smallest distances.


Fuzzy Symbolic Dynamics
G_2: μ = [0.276,...,0.00018], σ = 2
G_1: μ = [0.00023,...,0.00034], σ = 2
1st
2nd
3rd
4th
5th
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0
50
100
150
200
250
300
350
400
450
500
Multidimensional Scaling
1st
2nd
3rd
4th
5th
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5
3
−3
−2
−1
0
1
2
3
Fuzzy Symbolic Dynamics
G_3: μ = [0.277 ,...,0.0003], σ = 2
G_2: μ = [0.276,...,0.00007], σ = 2
G_1: μ = [0.0009,...,0.0006], σ = 2
1st
2nd
3rd
4th
5th
0.5
0.6
0.7
0.8
0.9
1
Multidimensional Scaling
1st
2nd
3rd
4th
5th
−2.5
−1.5
−0.5
0.5
1.5
2.5

Connectivity effects

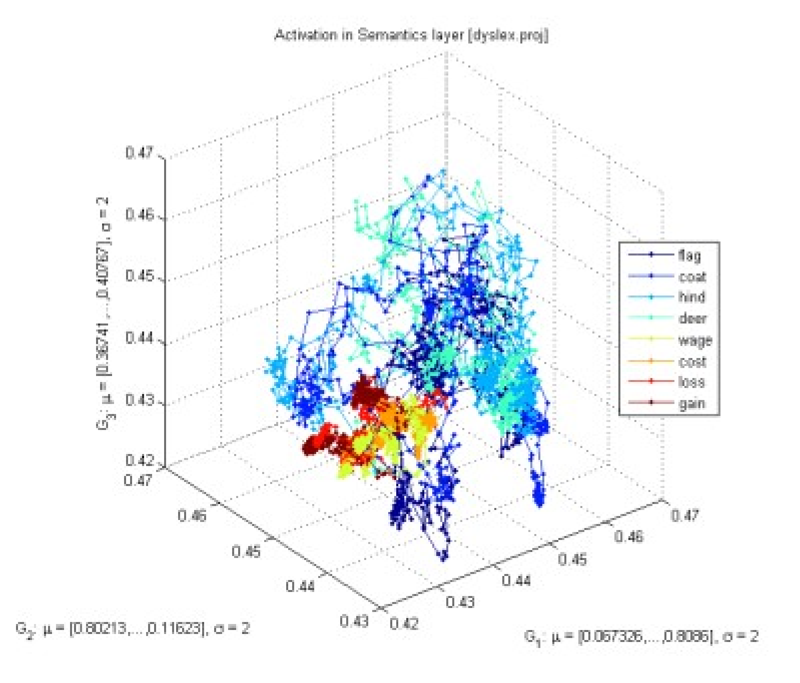


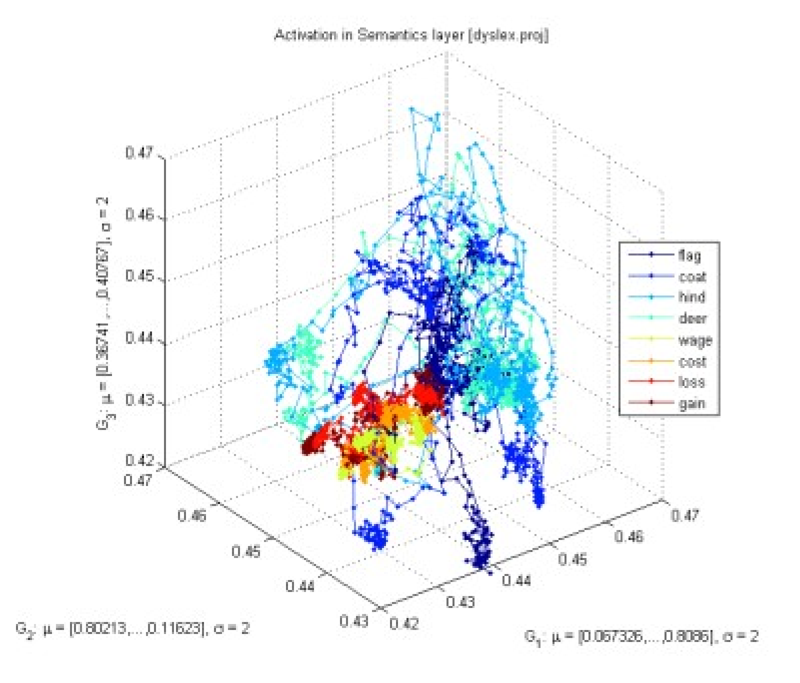


With small synaptic noise (var=0.02) the network starts from reaching an attractor and moves to another one (frequently quite distant), creating a “chain of thoughts”.

Same situation but recurrent connections within layers are stronger, fewer but larger attractors are reached, more time is spent in each attractor.

Exploration

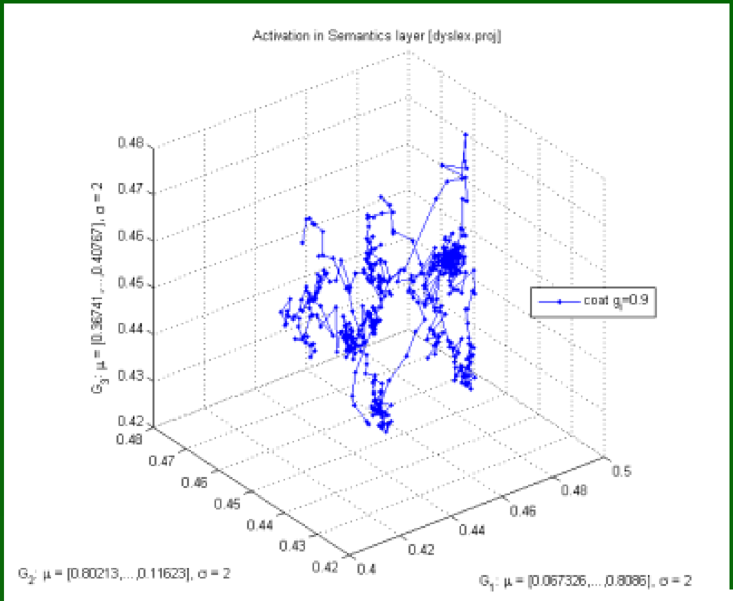


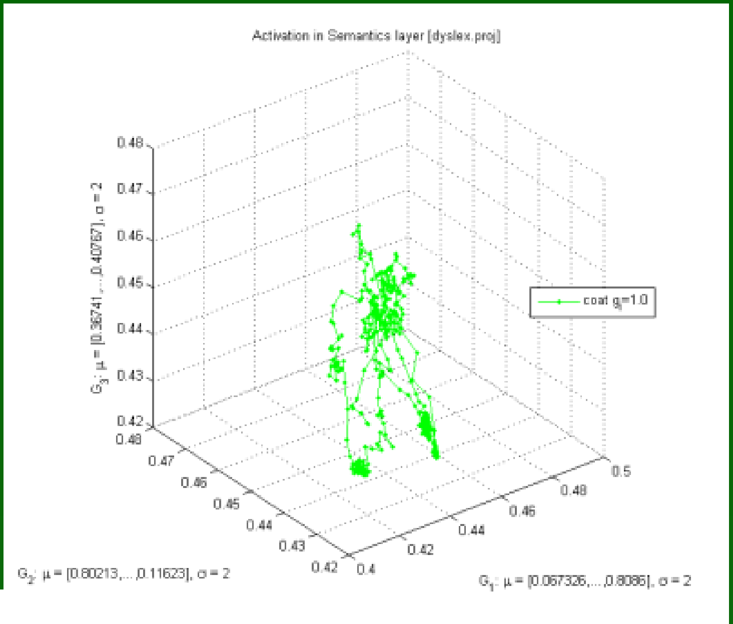


Same parameters but different runs: each time a single word is presented and dynamics run exploring different attractors.

Like in molecular dynamics, long time is needed to explore various potential attractors – depending on priming (previous dynamics or context) and chance.

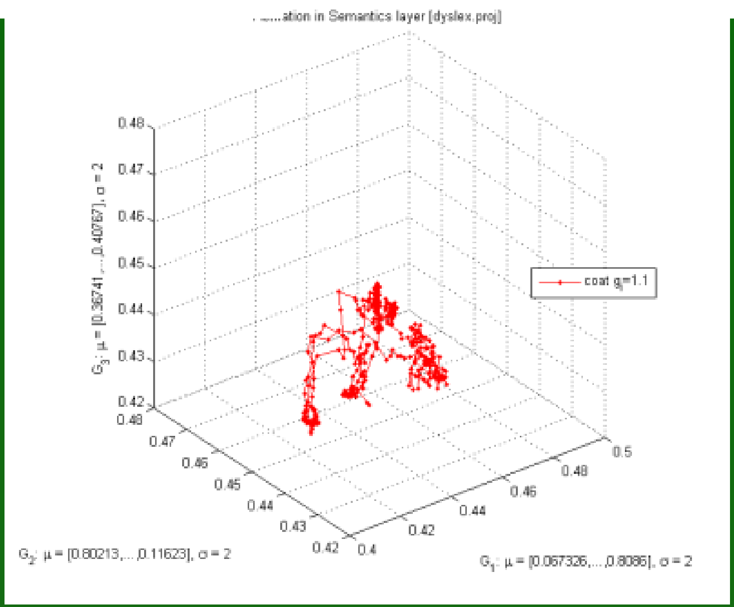
Inhibition effects





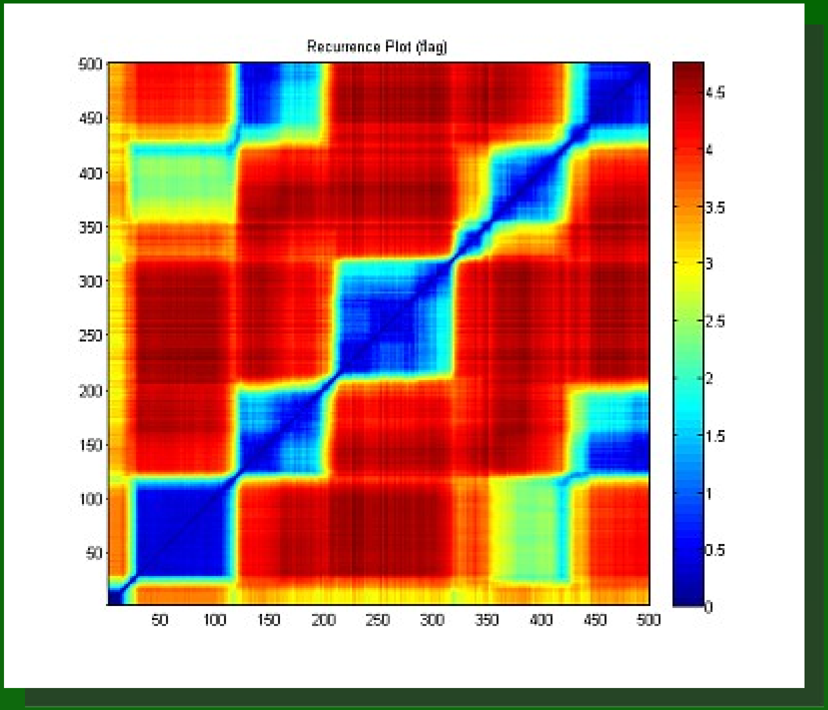
Increasing g_i from 0.9 to 1.1 reduces the attractor basin sizes and simplifies trajectories.

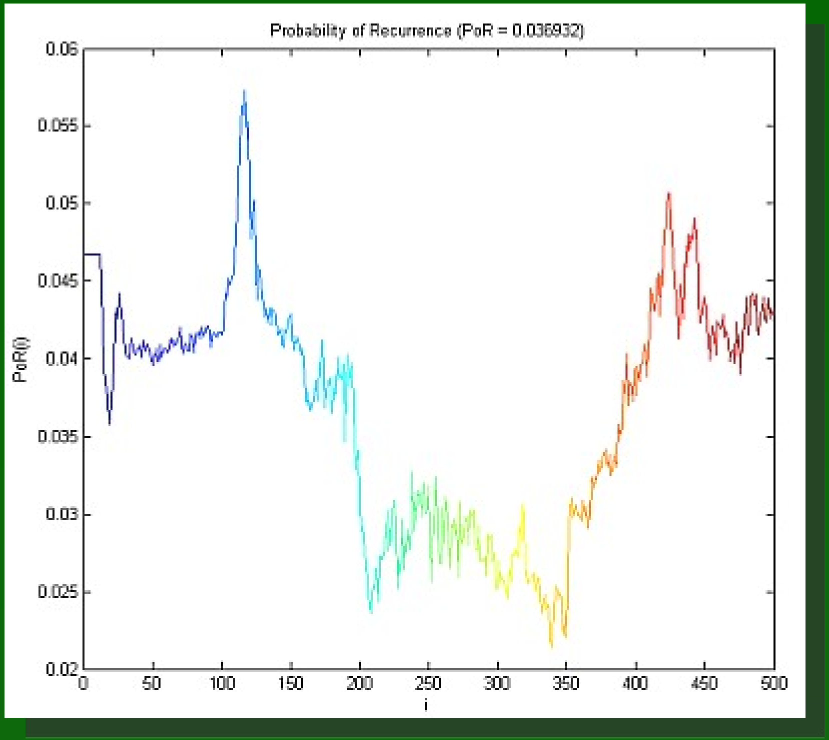
Not all attractors are real words.



Strong inhibition, empty head …

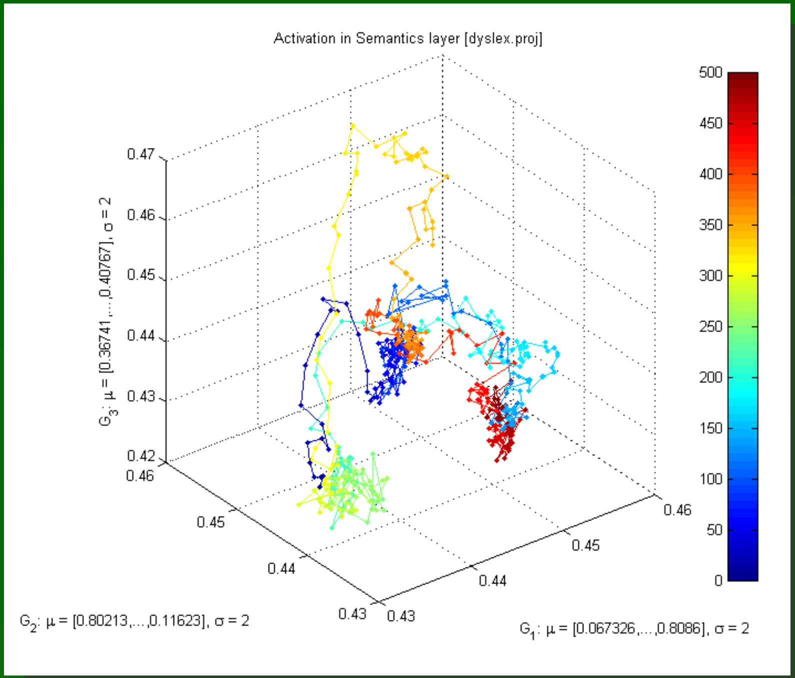
Probability of recurrence

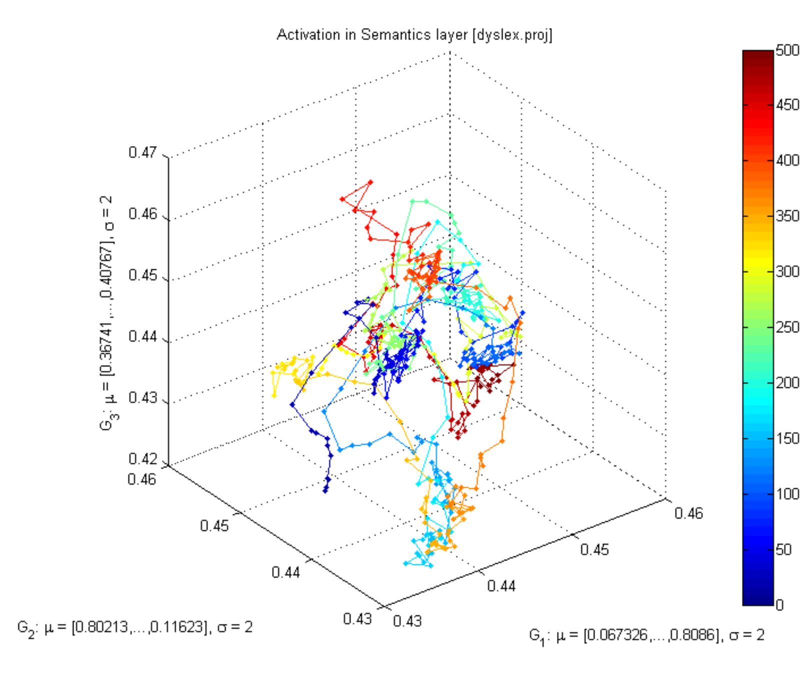




Probability of recurrence may be computed from recurrence plots, allowing for evaluation how strongly some basins of attractors capture neurodynamics.

Normal-ADHD



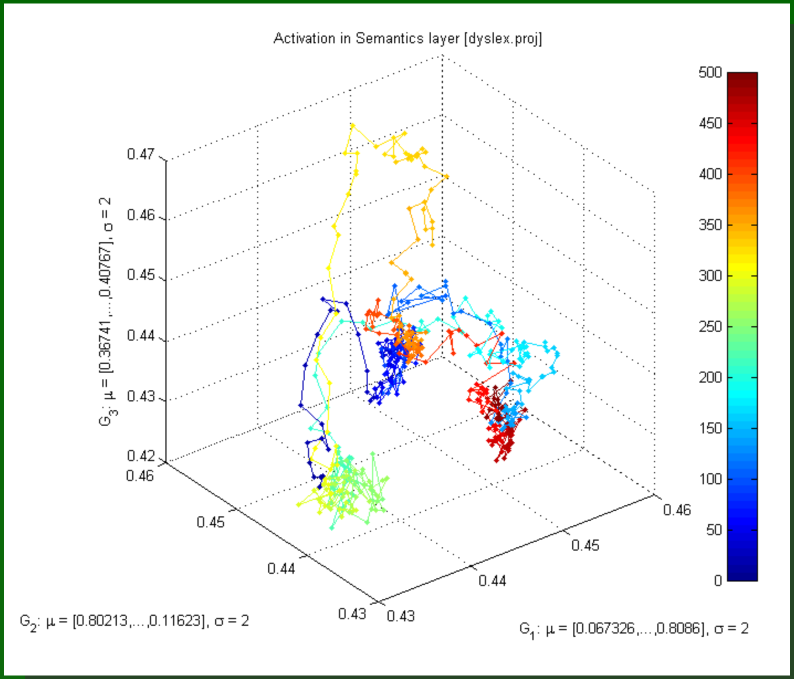


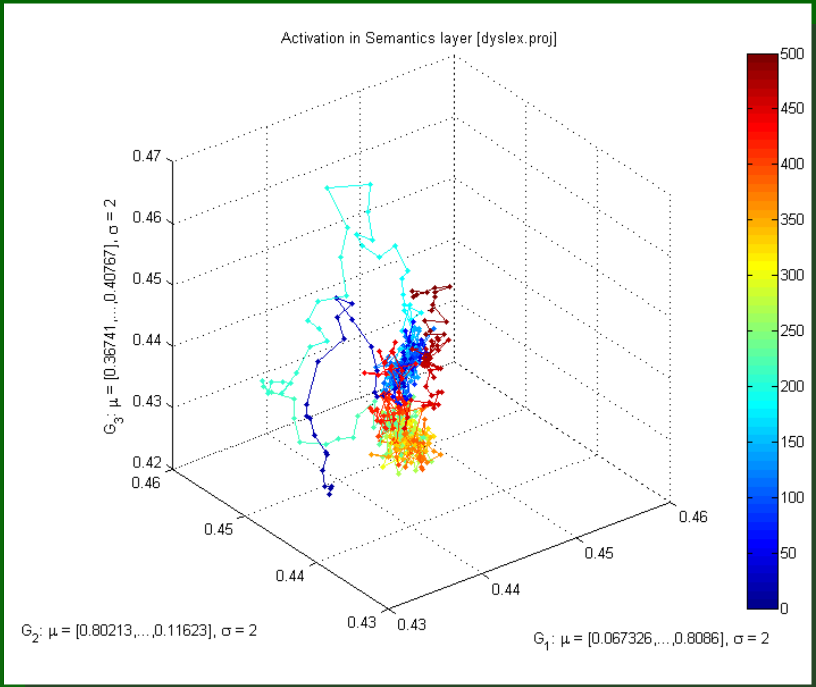
All plots for the flag word, different values of b_inc_dt parameter in the accommodation mechanism, b_inc_dt = 0.01 & b_inc_dt = 0.02

b_inc_dt = time constant for increases in intracellular calcium which builds up slowly as a function of activation.

http://kdobosz.wikidot.com/dyslexia-accommodation-parameters

Normal-Autism





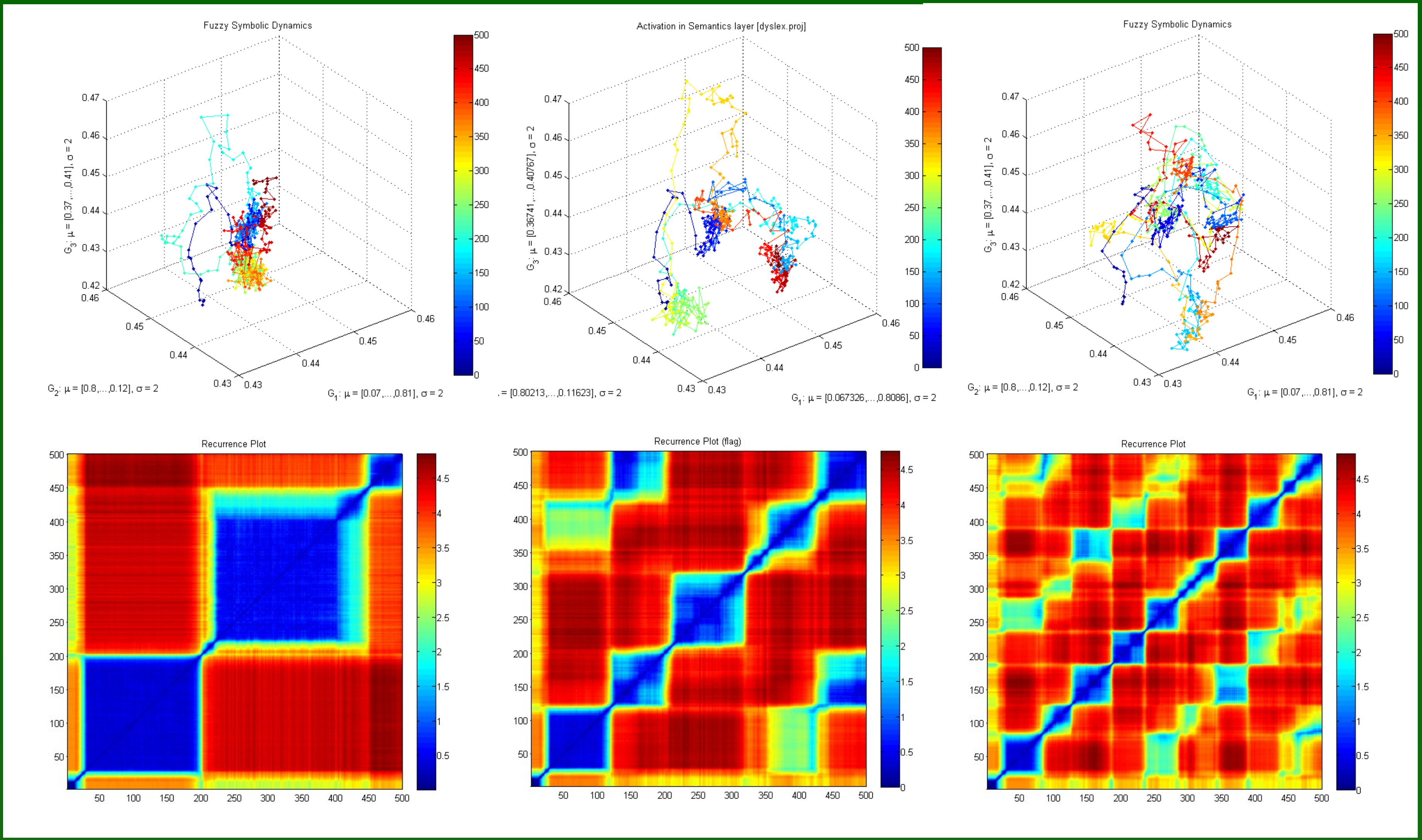
All plots for the flag word, different values of b_inc_dt parameter in the accommodation mechanism. b_inc_dt = 0.01 & b_inc_dt = 0.005

b_inc_dt = time constant for increases in intracellular calcium which builds up slowly as a function of activation.

http://kdobosz.wikidot.com/dyslexia-accommodation-parameters

Autism-Normal-ADHD

b_inc_dt = 0.005 *b_inc_dt* = 0.01 *b_inc_dt* = 0.02



Some questions

Stream of mental states = attractor states + transitions between them.

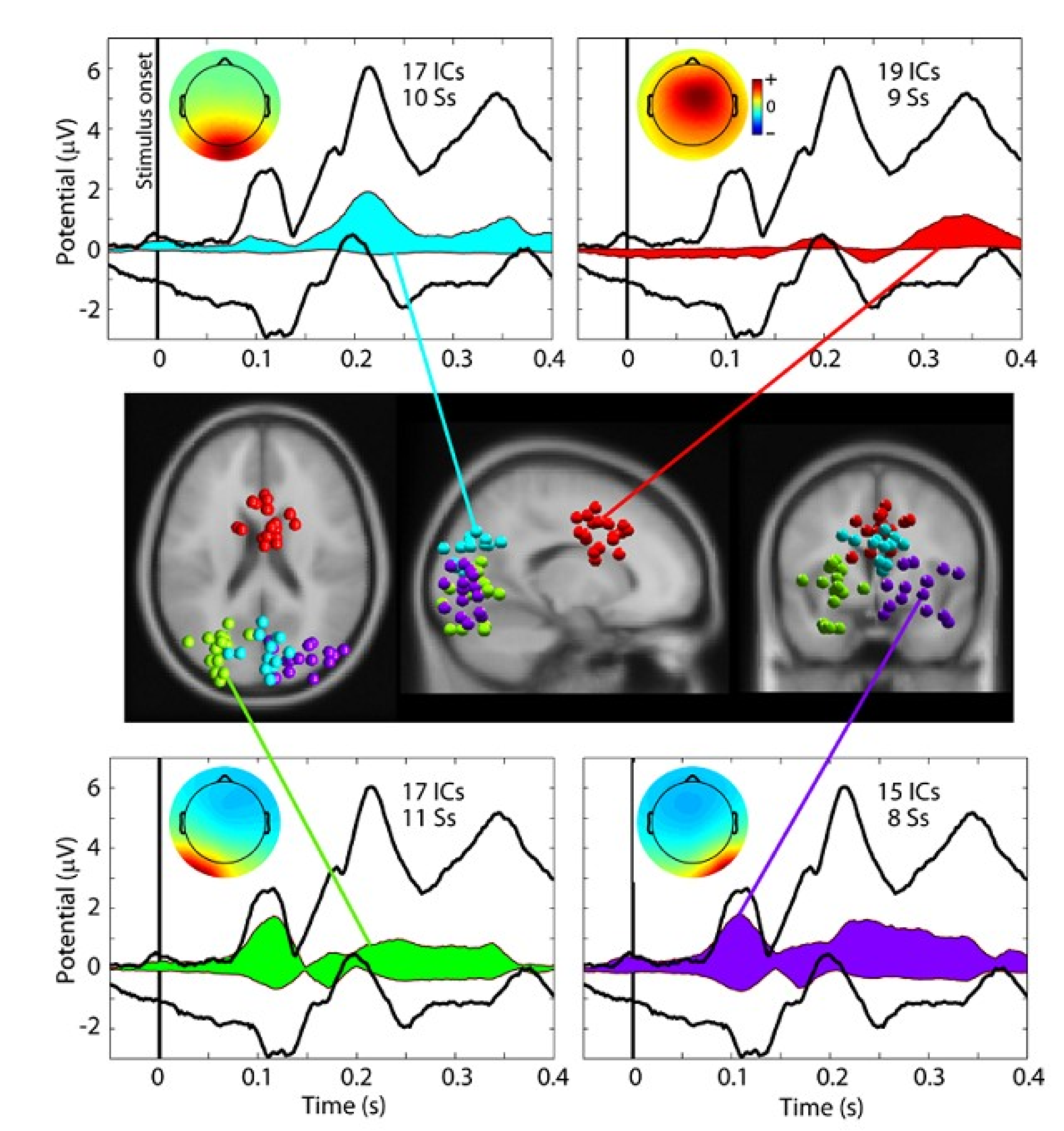
Problems:

1. Jumping between subspaces of different subsets of dimensions; rescaling dimensions? Manifold learning?
2. How to imagine multidimensional attractors? Trajectory has many "escape channels" requiring different energy.
3. Real EEG dynamics is oscillatory, how to transform it to attractor dynamics? First use source localization?
4. Distances (transition probabilities) in neural space are not symmetric – use Finsler spaces?
5. Natural Language Processing based on spreading activation in networks?
6. How to use attractor dynamics to construct mental models?

Source localization maps brain activity to attractor dynamics.

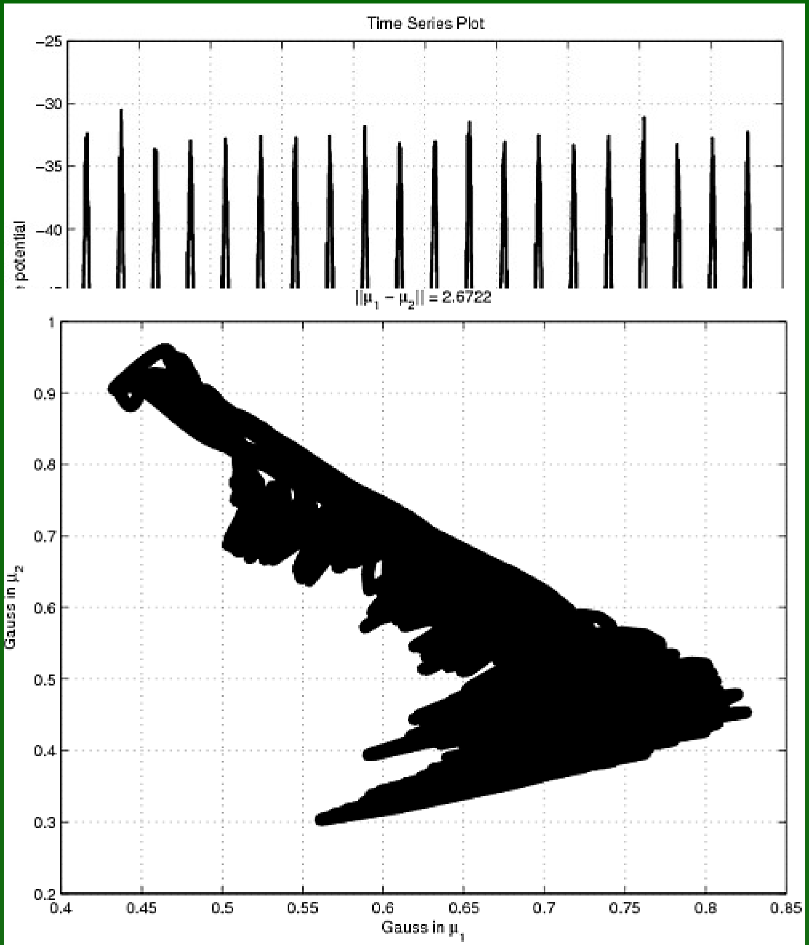
Problem: these sources pop up and vanish in different places.

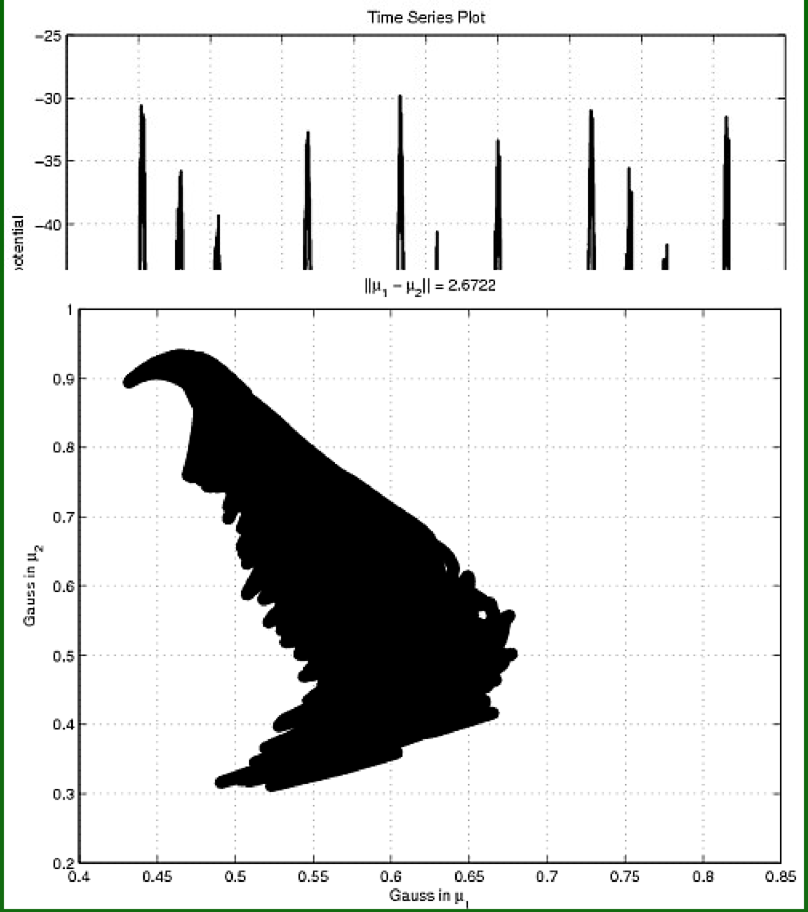
Fig. from:
Makeig, Onton, ERP Features and EEG Dynamics: An ICA Perspective, 2009

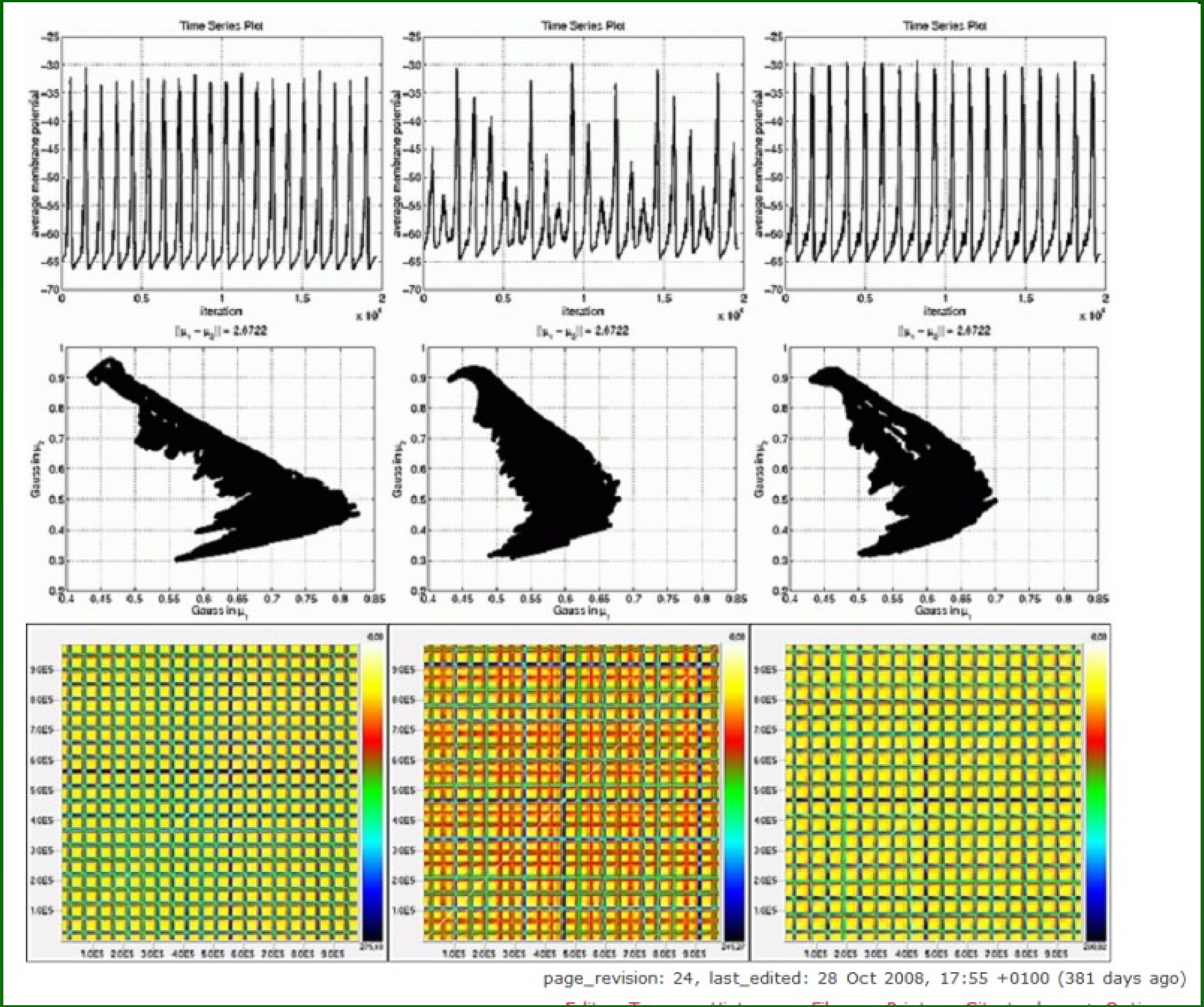


Respiratory Rhythm Generator

3 layers, spiking neurons, output layer with 50 neurons



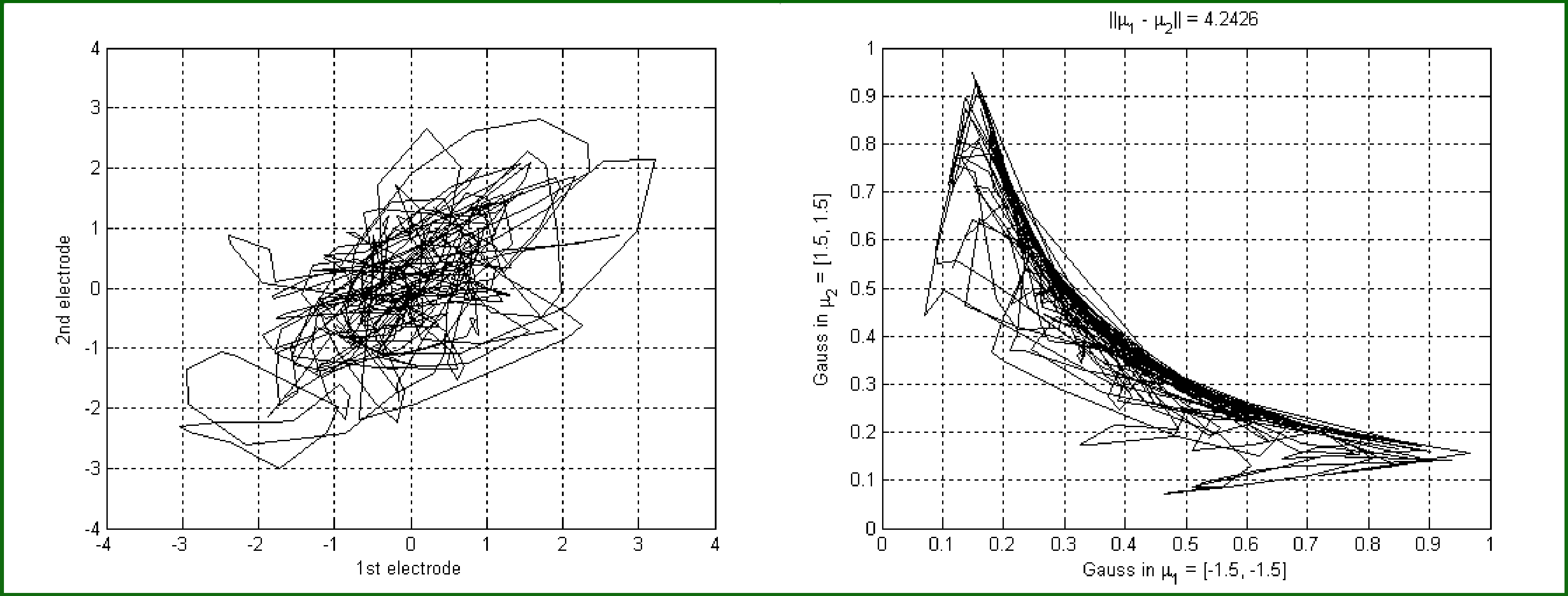




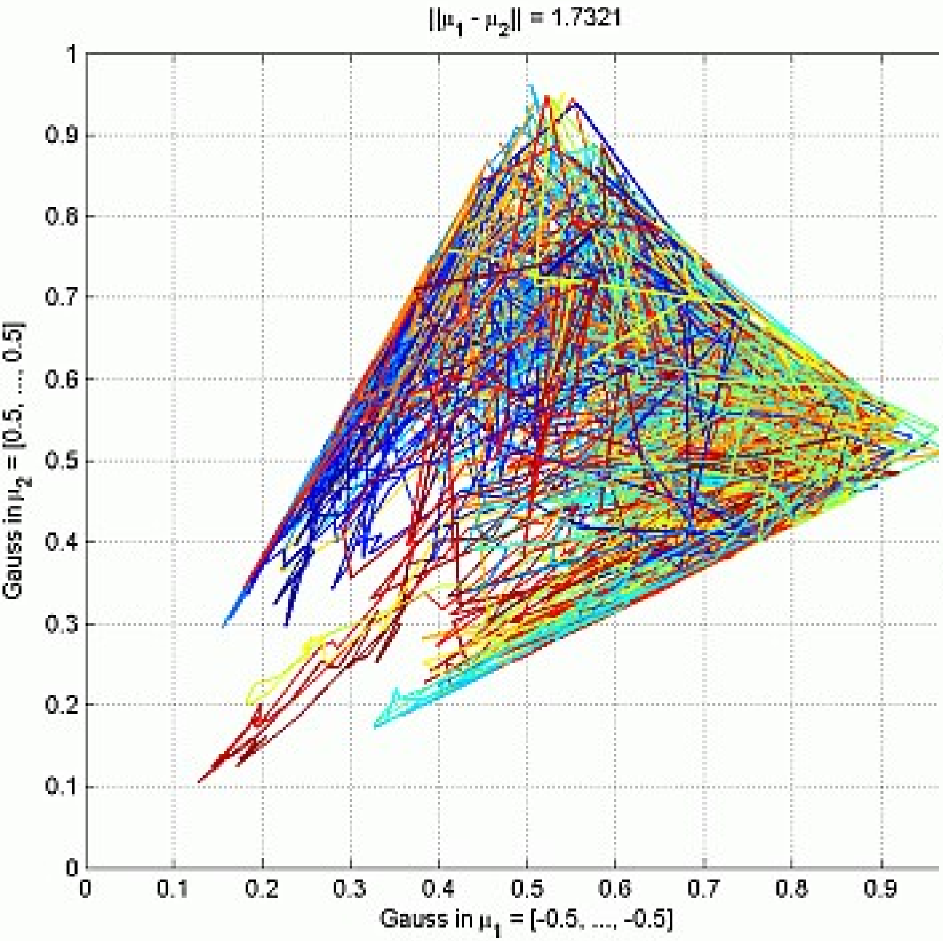
page_revision: 24, last_edited: 28 Oct 2008, 17:55 +0100 (381 days ago)

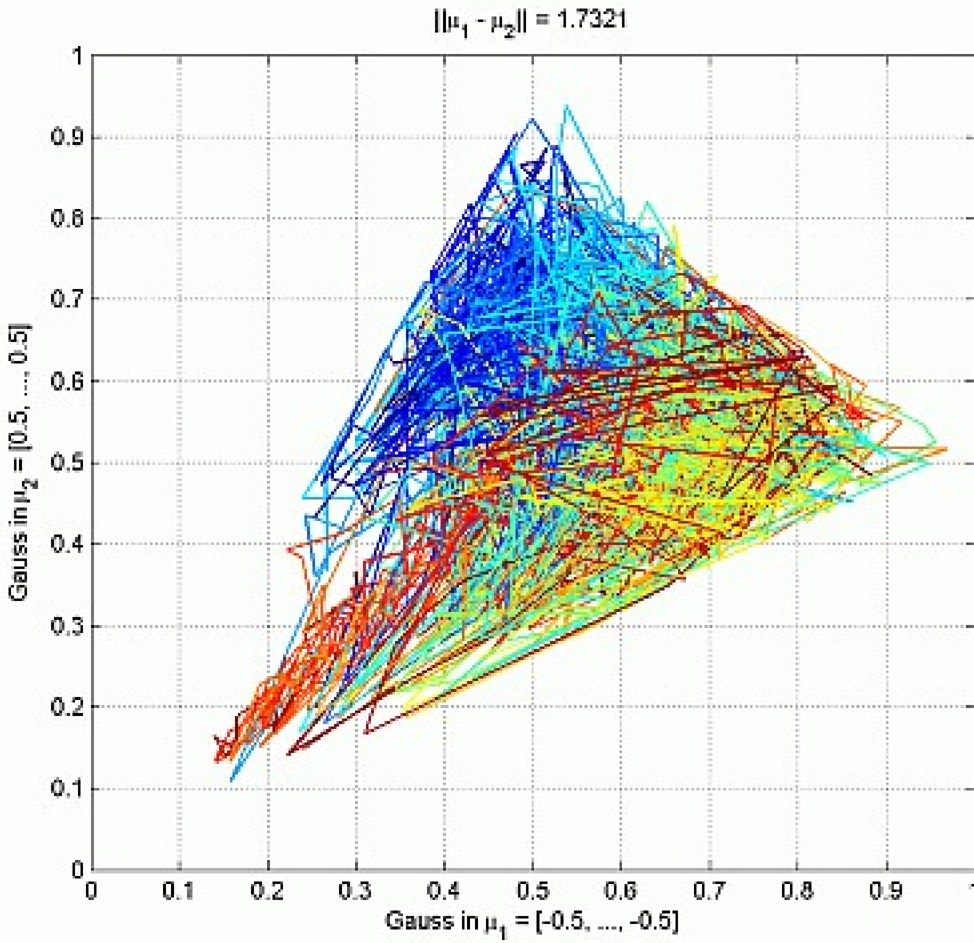
BCI EEG example

- Data from two electrodes, BCI IIIa



Alcoholics vs. controls

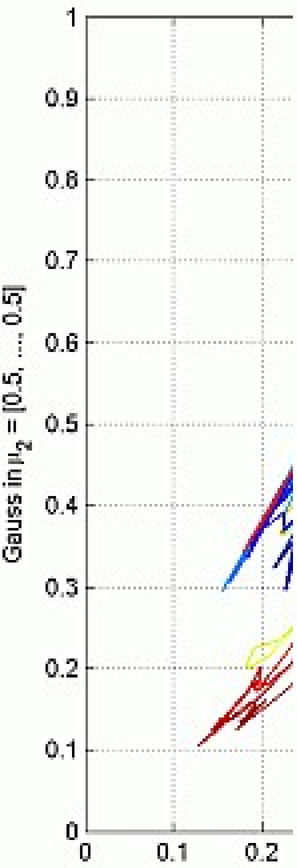


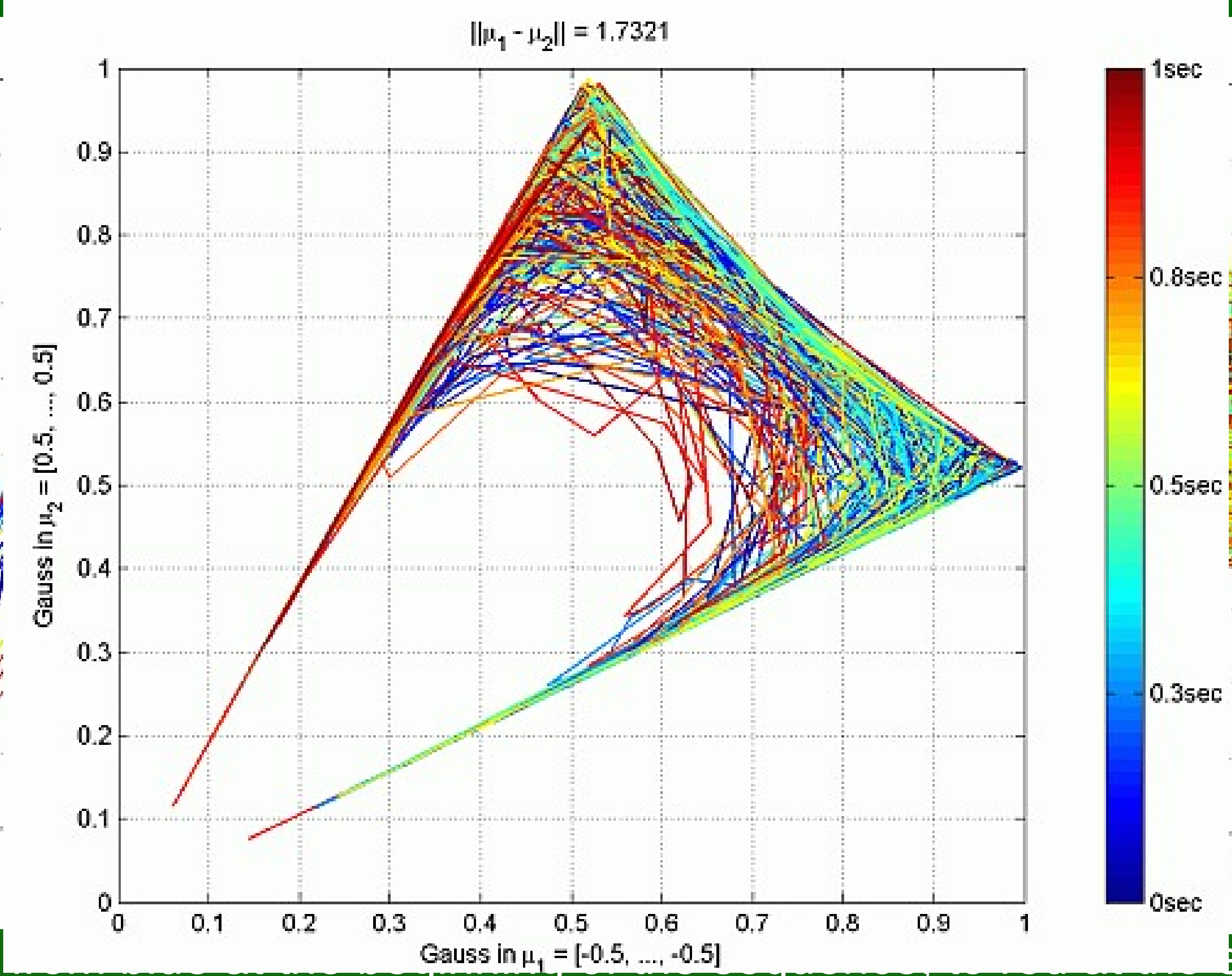


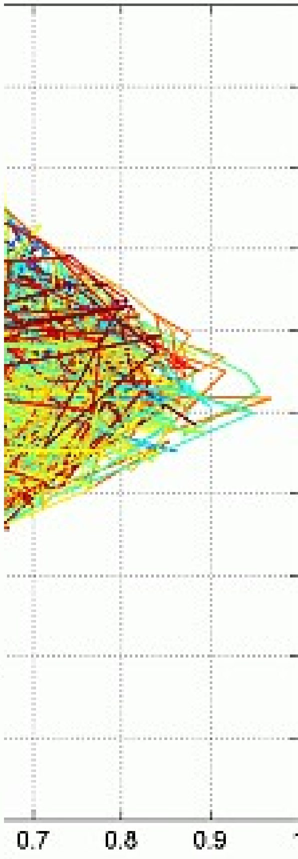
Colors: from blue at the beginning of the sequence, to red at the end.

Left: normal subject; right: alcoholic; task: two matched stimuli, 64 channels (3 after PP), 256 Hz sampling, 1 sec, 10 trials; single st alc.

Alcoholics vs. controls







Colors: end.

Left: normal subject; right: alcoholic; task: two matched stimuli, 64 channels (3 after PP), 256 Hz sampling, 1 sec, 10 trials; single st alc.

Mental models

Kenneth Craik, 1943 book “The Nature of Explanation”, G-H Luquet attributed mental models to children in 1927.

P. Johnson-Laird, 1983 book and papers.

Imagination: mental rotation, time ~ angle, about 60^{o}/sec.

Internal models of relations between objects, hypothesized to play a major role in cognition and decision-making.

AI: direct representations are very useful, direct in some aspects only!

Reasoning: imaging relations, “seeing” mental picture, semantic?

Systematic fallacies: a sort of cognitive illusions.

- If the test is to continue then the turbine must be rotating fast enough to generate emergency electricity.
- The turbine is not rotating fast enough to generate this electricity.
- What, if anything, follows? Chernobyl disaster …

If A=>B; then ~B => ~A, but only about 2/3 students answer correctly..

Mental models summary

The mental model theory is an alternative to the view that deduction depends on formal rules of inference.

1. MM represent explicitly what is true, but not what is false; this may lead naive reasoner into systematic error.
2. Large number of complex models => poor perf
3. Tendency to focus on a few possible models = decisions.

Cognitive illusions are just like visual illusions.

M. Piattelli-Palmarini, Inevitable Illusions: How Mist (1996)

R. Pohl, Cognitive Illusions: A Handbook on Fallaci Judgement and Memory (2005)

Amazing, but mental models theory ignores everyth learning in any form! How and why do we reason th I'm innocent! My brain made me do it!

Mental models

Easy reasoning A=>B, B=>C, so A=>C

- All mammals suck milk.
- Humans are mammals.
- => Humans suck milk. Simple associative process, easy to simulate.

... but almost no-one can draw conclusion from:

- All academics are scientist.
- No wise men is an academic.
- What can we say about wise men and scientists?

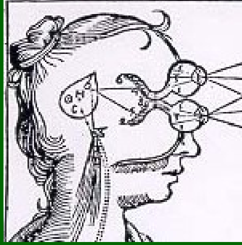
Surprisingly only ~10% of students get it right after days of thinking.

No simulations explaining why some mental models are so difficult.

Why is it so hard? What really happens in the brain?

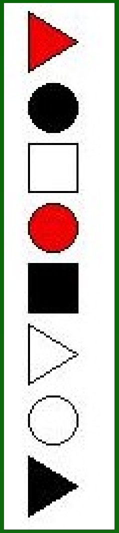
Try to find a new point of view to illustrate it.

Learning complex categories



Categorization is quite basic, many psychological models/experiments.
Multiple brain areas involved in different categorization tasks.
Classical experiments on rule-based category learning:
Shepard, Hovland and Jenkins (1961), replicated by Nosofsky *et al.* (1994).

Problems of increasing complexity; results determined by logical rules.
3 binary-valued dimensions:
shape (square/triangle), color (black/white), size (large/small).
4 objects in each of the two categories presented during learning.

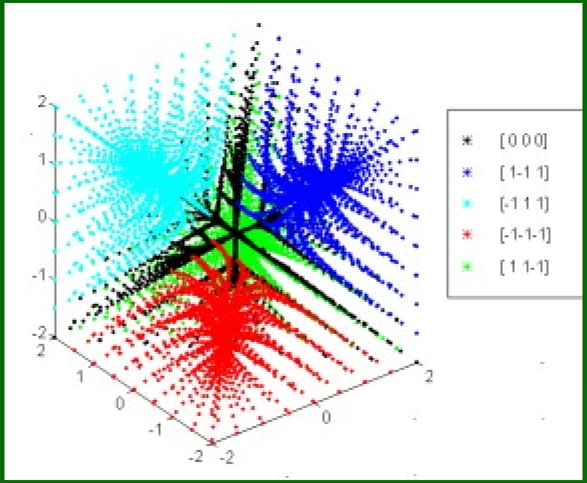


Type I - categorization using one dimension only.
Type II - two dim. are relevant, including exclusive or (XOR) problem.
Types III, IV, and V - intermediate complexity between Type II - VI.
All 3 dimensions relevant, "single dimension plus exception" type.
Type VI - most complex, 3 dimensions relevant, enumerate, no simple rule.

Difficulty (number of errors made): Type I < II < III ~ IV ~ V < VI
For n bits there are 2^n binary strings 0011…01; how complex are the rules (logical categories) that human/animal brains still can learn?

Canonical neurodynamics.

What happens in the brain during category learning?
Complex neurodynamics <=> simplest, canonical dynamics.
For all logical functions one may write corresponding equations.



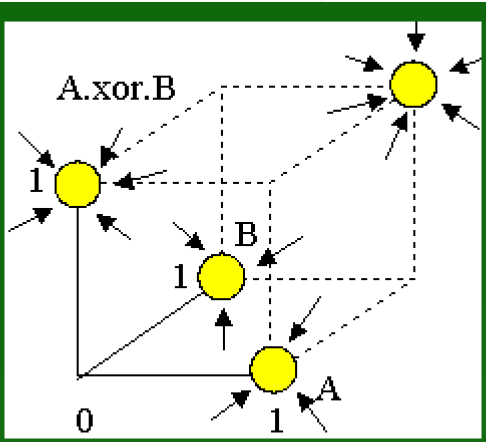
For XOR (type II problems) equations are:

$$V(x,y,z) = 3xyz + \frac{1}{4}\left(x^2+y^2+z^2\right)^2$$

$$\dot{x} = -\frac{\partial V}{\partial x} = -3yz - \left(x^2+y^2+z^2\right)x$$

$$\dot{y} = -\frac{\partial V}{\partial y} = -3xz - \left(x^2+y^2+z^2\right)y$$

$$\dot{z} = -\frac{\partial V}{\partial z} = -3xy - \left(x^2+y^2+z^2\right)z$$

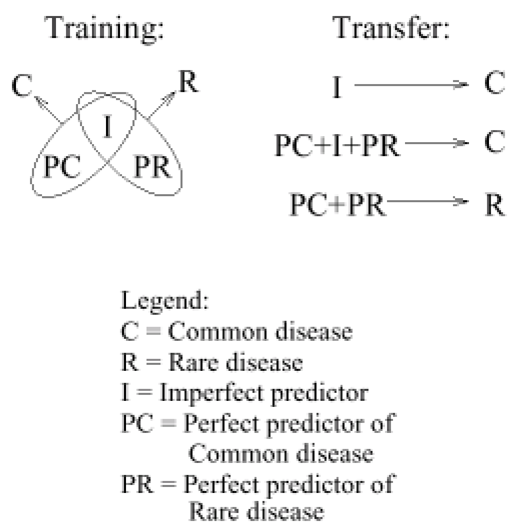


Corresponding feature space for relevant dimensions A, B

Inverse based rates

Relative frequencies (base rates) of categories are used for classification:

if on a list of disease and symptoms disease C associated with (PC, I) symptoms is 3 times more common as R,
then symptoms PC => C, I => C (base rate effect).



Legend:
C = Common disease
R = Rare disease
I = Imperfect predictor
PC = Perfect predictor of Common disease
PR = Perfect predictor of Rare disease

Predictions contrary to the base:
inverse base rate effects (Medin, Edelson 1988).

Although PC + I + PR => C (60% answers)
PC + PR => R (60% answers)

Why such answers?
Psychological explanations are not convincing.

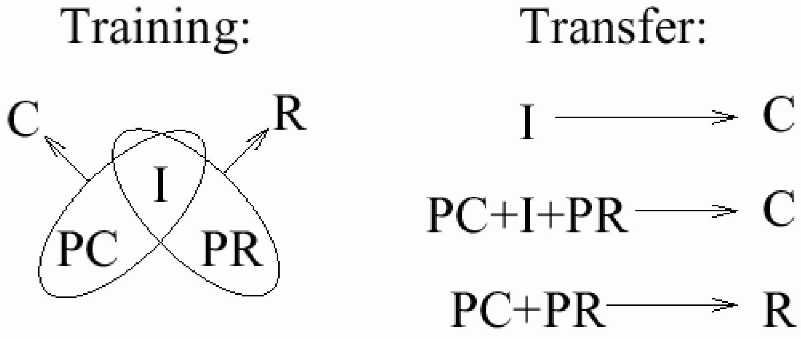
Effects due to the neurodynamics of learning?

I am not aware of any dynamical models of such effects.

Inverse based

Relative frequencies (base rates) of categor

if on a list of disease and symptoms disease
symptoms is 3 times more common as R,
then symptoms PC => C, I => C (base rate e



Legend:
C = Common disease
R = Rare disease
I = Imperfect predictor
PC = Perfect predictor of
Common disease
PR = Perfect predictor of
Rare disease

Predictions contrary to the base:
inverse base rate effects (Medin, Edelson 19

Although PC + I + PR => C (60% answers)
PC + PR => R (60% answers)

Legend:
C = Common disease
R = Rare disease
I = Imperfect predictor
PC = Perfect predictor of
Common disease
PR = Perfect predictor of
Rare disease

Why such answers?
Psychological explanations are not convincing.

Effects due to the neurodynamics of learning?

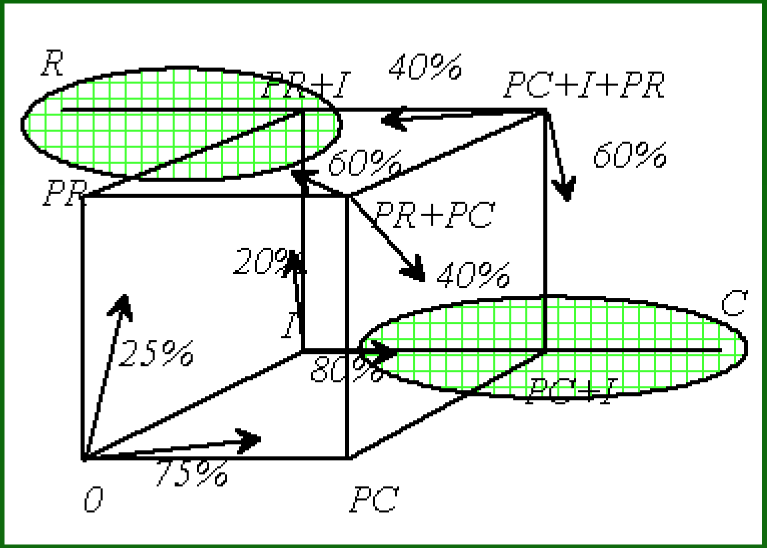
I am not aware of any dynamical models of such effects.

IBR neurocognitive explanation

Psychological explanation:
J. Kruschke, Base Rates in Category Learning (1996).

PR is attended to because it is a distinct symptom, although PC is more common.

Basins of attractors - neurodynamics;
PDFs in P-space {C, R, I, PC, PR}.



PR + PC activation leads more frequently to R because the basin of attractor for R is deeper.

Construct neurodynamics, get PDFs.
Unfortunately these processes are in 5D.

Prediction: weak effects due to order and timing of presentation (PC, PR) and (PR, PC), due to trapping of the mind state by different attractors.

Learning

Point of view

Neurocognitive	Psychology
I+PC more frequent => stronger synaptic connections, larger and deeper basins of attractors.	Symptoms I, PC are typical for C because they appear more often.
To avoid attractor around I+PC leading to C, deeper, more localized attractor around I+PR is created.	Rare disease R - symptom I is misleading, attention shifted to PR associated with R.

Probing

Point of view

Neurocognitive	Psychology
Activation by I leads to C because longer training on I+PC creates larger common basin than I+PR.	I => C, in agreement with base rates, more frequent stimuli I+PC are recalled more often.
Activation by I+PC+PR leads frequently to C, because I+PC puts the system in the middle of the large C basin and even for PR geadients still lead to C.	I+PC+PR => C because all symptoms are present and C is more frequent (base rates again).
Activation by PR+PC leads more frequently to R because the basin of attractor for R is deeper, and the gradient at (PR,PC) leads to R.	PC+PR => R because R is distinct symptom, although PC is more common.

Mental model dynamics

Why is it so hard to draw conclusions from:

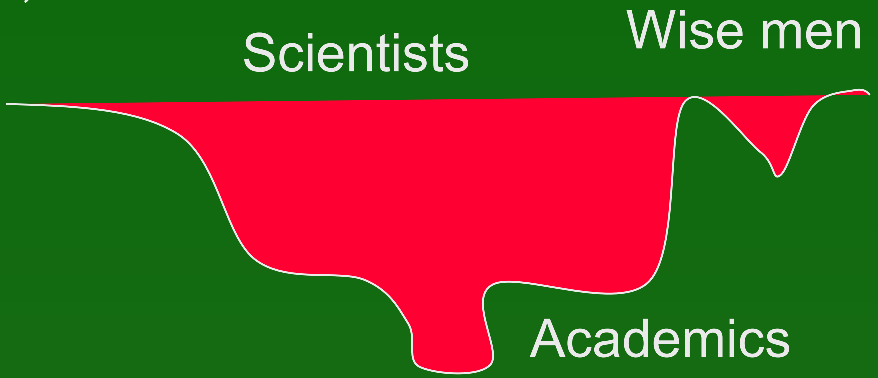
- All academics are scientist.
- No wise men is an academic.
- What can we say about wise men and scientists?

All A’s are S, ~ W is A;   relation S <=> W ?

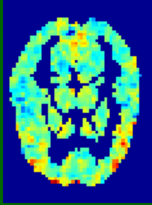
What happens with neural dynamics?

Basins of A is larger than B, as B is a subtype of A, and thus has to inherit most properties that are associated with A.
Attractor for B has to be within A.
Thinking of B makes it hard to think of A, as the

Basins of attractors for the
3 concepts involved;
basin for “Wise men” has unknown
relation to the other basins.



Neurocognitive reps.



How to approach modeling of word (concept) w representations in the brain? Word $w = (w_f, w_s)$ has

- phonological (+visual) component w_f, word form;
- extended semantic representation w_s, word meaning;
- is always defined in some context *Cont* (enactive approach).

$\Psi(w,Cont,t)$ evolving prob. distribution (pdf) of brain activations.

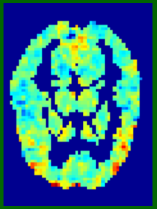
Hearing or thinking a word w , or seeing an object labeled as w adds to the overall brain activation in a non-linear way.

How? Maximizing overall self-consistency, mutual activations, meanings that don't fit to current context are automatically inhibited.

Result: almost continuous variation of word meaning.

This process is rather difficult to approximate using typical knowledge representation techniques, such as vector NLP models, connectionist models, semantic networks, frames or probabilistic networks.

Approximate reps.



States $\Psi(w,Cont)$ ⇔ lexicographical meanings:

- clusterize $\Psi(w,Cont)$ for all contexts;
- define prototypes $\Psi(w_k,Cont)$ for different meanings w_k.

A1: use spreading activation in semantic networks to define Ψ.
A2: take a snapshot of activation Ψ in discrete space (vector approach).

Meaning of the word is a result of priming, spreading activation to speech, motor and associative brain areas, creating affordances.

$\Psi(w,Cont)$ ~ quasi-stationary wave, with phonological/visual core activations w_f and variable extended representation w_s selected by *Cont*.

$\Psi(w,Cont)$ state into components, because the semantic representation

E. Schrödinger (1935): best possible knowledge of a whole does not include the best possible knowledge of its parts! Not only in quantum case. Left semantic network *LH* contains w_f coupled with the *RH*.

Semantic => vector reps

Some associations are subjective, some are universal.

How to find the activation pathways in the brain? Try this algorithm:

- Perform text pre-processing steps: stemming, stop-list, spell-checking ...
- Map text to some ontology to discover concepts (ex. UMLS ontology).
- Use relations (Wordnet, ULMS), selecting those types only that help to distinguish between concepts.
- Create first-order cosets (terms + all new terms from included relations), expanding the space – acts like a set of filters that evaluate various aspects of concepts.
- Use feature ranking to reduce dimensionality of the first-order coset space, leave all original features.
- Repeat last two steps iteratively to create second- and higher-order enhanced spaces, first expanding, then shrinking the space.

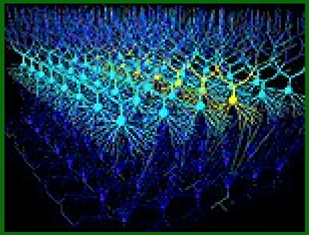
Result: a set of **X** vectors representing concepts in enhanced spaces, partially including effects of spreading activation.

Some connections

Geometric/dynamical ideas related to mind may be found in many fields:

Neuroscience:
D. Marr (1970) “probabilistic landscape”.
C.H. Anderson, D.C. van Essen (1994): Superior Colliculus PDF maps
S. Edelman: “neural spaces”, object recognition, global representation space approximates the Cartesian product of spaces that code object fragments, representation of similarities is sufficient.

Psychology:
K. Levin, psychological forces.
G. Kelly, Personal Construct Psychology.
R. Shepard, universal invariant laws.
P. Johnson-Laird, mind models.



Folk psychology: to put in mind, to have in mind, to keep in mind (mindmap), to make up one's mind, be of one mind ... (space).

More connections

AI: problem spaces - reasoning, problem solving, SOAR, ACT-R, little work on continuous mappings (MacLennan) instead of symbols.

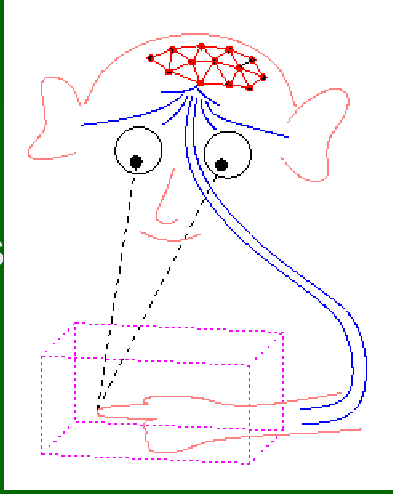
Engineering: system identification, internal models inferred from input/output observations – this may be done without any parametric assumptions if a number of identical neural modules are used!

Philosophy:
P. Gärdenfors, Conceptual spaces
R.F. Port, T. van Gelder, ed. Mind as motion (MIT Press 1995)

Linguistics:
G. Fauconnier, Mental Spaces (Cambridge U.P. 1994).
Mental spaces and non-classical feature spaces.
J. Elman, Language as a dynamical system; J. Feldman neural basis;
Stream of thoughts, sentence as a trajectory in P-space.

Psycholinguistics: T. Landauer, S. Dumais, Latent Semantic Analysis, Psych. Rev. (1997) Semantic for 60 k words corpus requires about 300 dim.

Conclusions



Understanding of reasoning requires a model of brain processes => logic and reasoning.

Simulations of the brain may lead to mind functions, but we still need conceptual understanding.

Psychological interpretations and models are confabulations! They provide wrong conceptualization of real brain processes.

Low-dimensional representation of mental/brain events are needed.

Complex neurodynamics => dynamics in P-spaces, visualization helps.

Is this a good bridge between mind and brain?

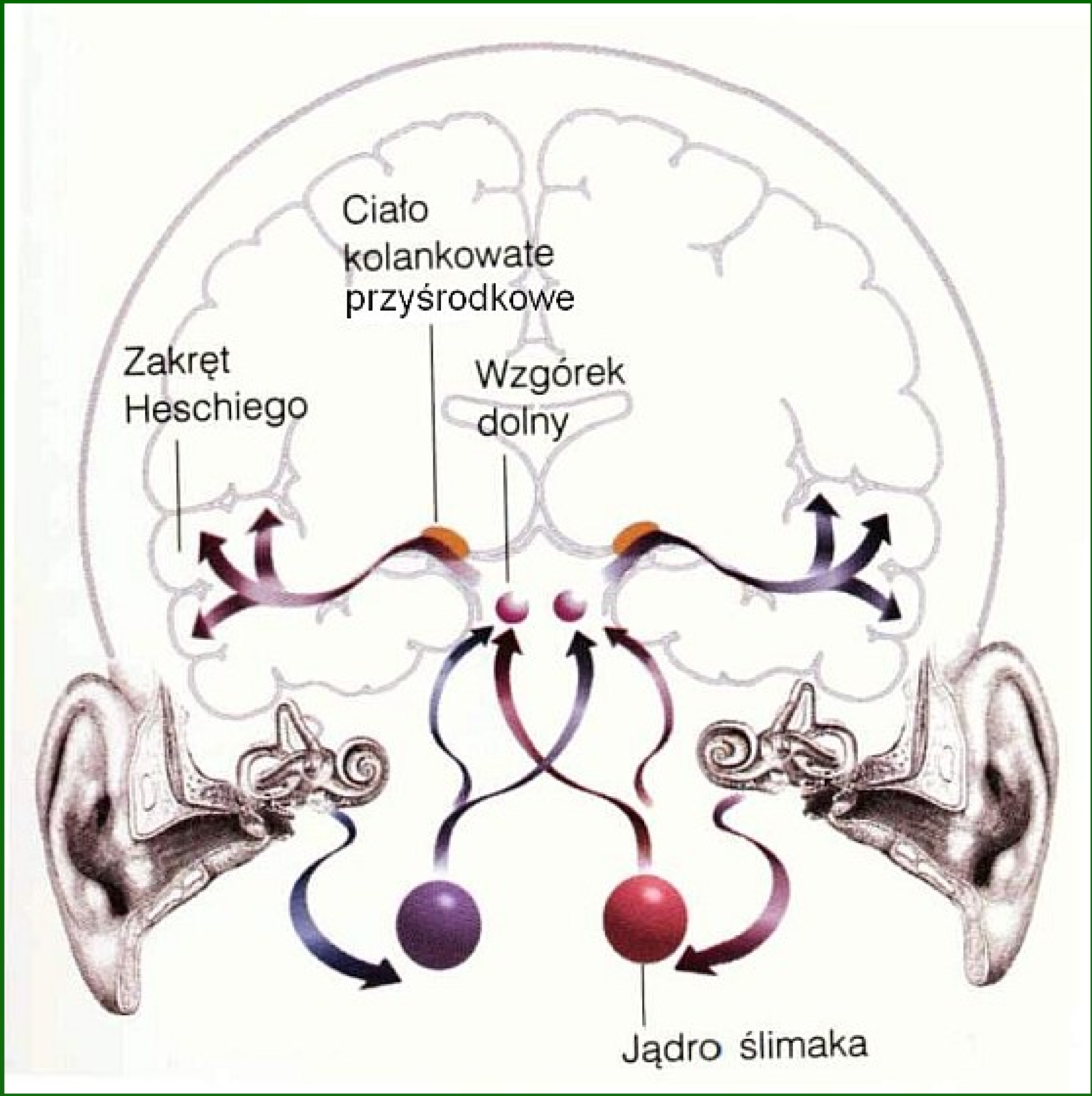
Mind models, psychology, logic … do not even touch the truth.

However, P-spaces may be high-dimensional, so hard to visualize.

How to describe our inner experience (Hurlburt & Schwitzgebel 2007)?

Still I hope that at the end of the road physics-like theory of events in mental spaces will be possible, explaining higher cognitive functions.

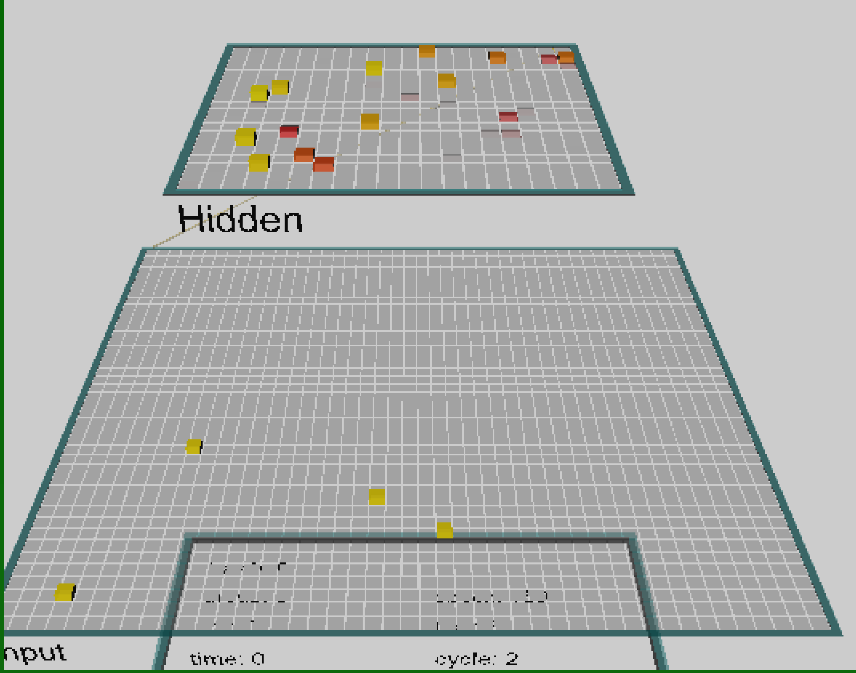
BRAIN AWARENESS WEEK
JUST USE IT



Google: Wlodzislaw Duch => presentations, papers …

Simple mindless network

Inputs = words, 1920 selected from a 500 pages book (O'Reilly, Munakata, Explorations book, this example is in Chap. 10). 20x20=400 hidden elements, with sparse connections to inputs, each hidden unit trained using Hebb principle, learns to react to correlated or similar words. For example, a unit may point to synonyms: act, activation, activations.

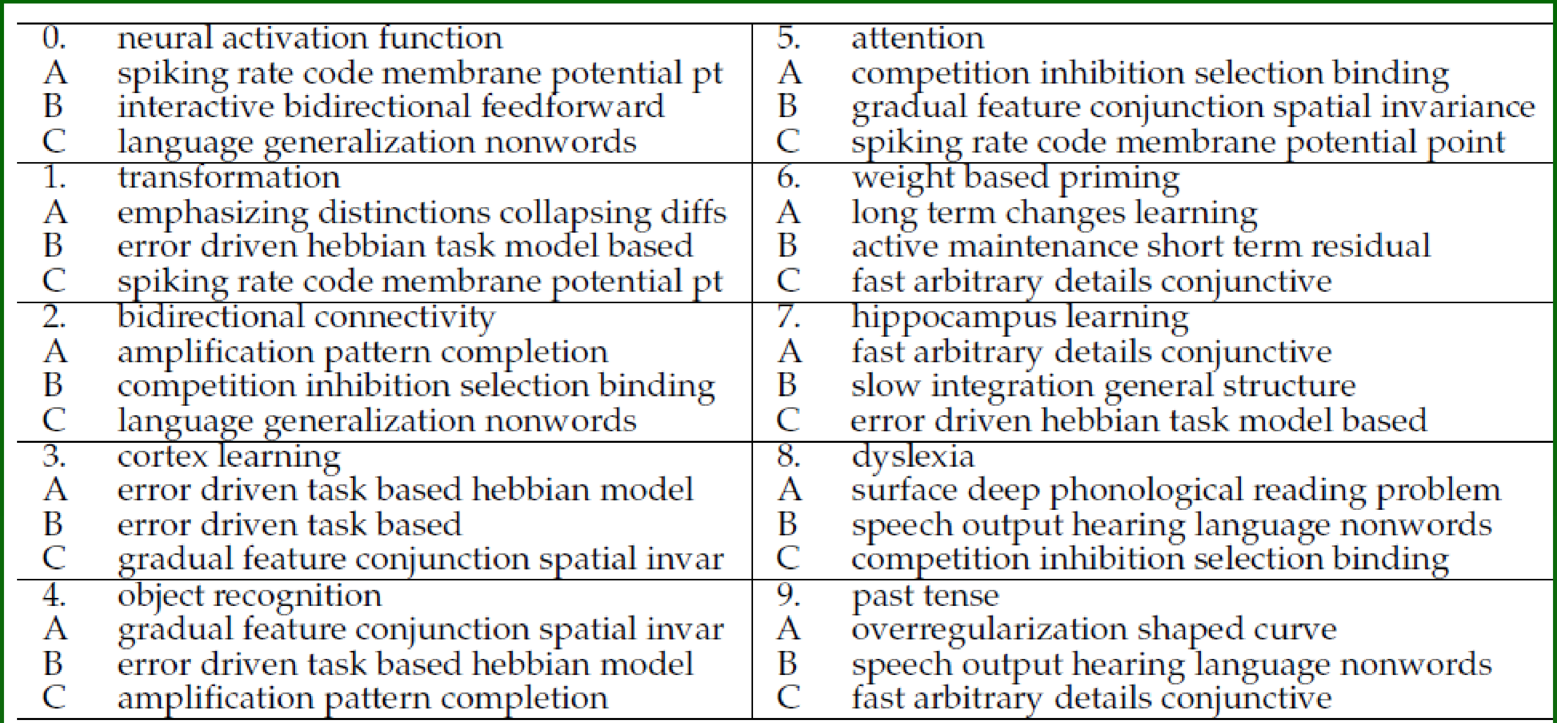


Compare distribution of activities of hidden elements for two words A, B, calculating cos(A,B) = A*B/|A||B|.

Activate units corresponding to several words: A="attention", B="competition", gives cos(A,B)=0.37. Adding "binding" to "attention" gives cos(A+C,B)=0.49.

This network is used on multiple choice test.

Multiple-choice Quiz



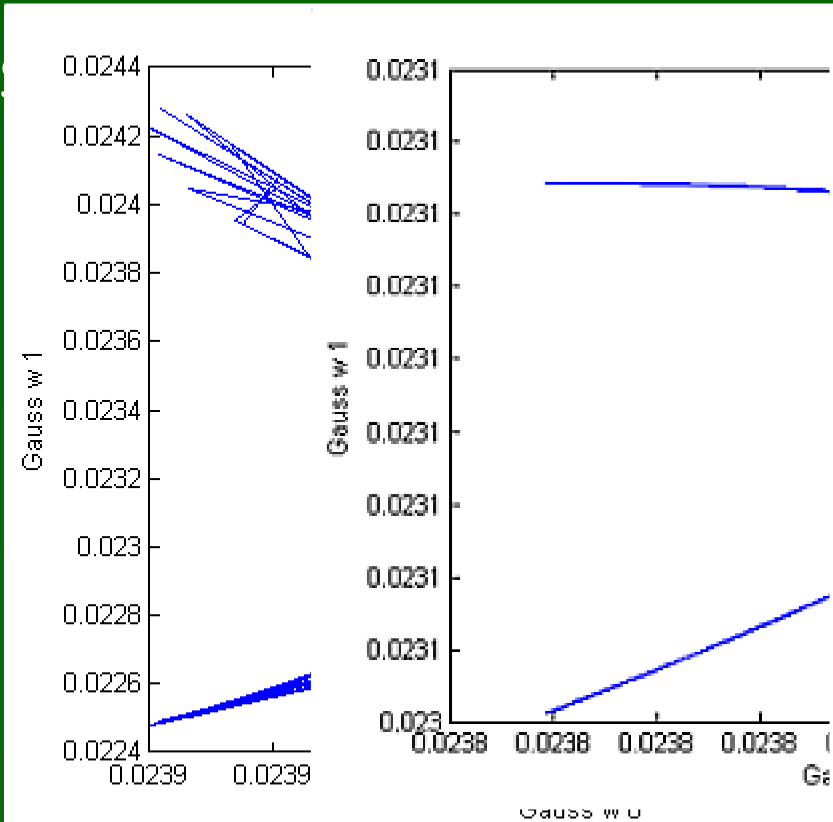
0.	neural activation function	5.	attention
A	spiking rate code membrane potential pt	A	competition inhibition selection binding
B	interactive bidirectional feedforward	B	gradual feature conjunction spatial invariance
C	language generalization nonwords	C	spiking rate code membrane potential point
1.	transformation	6.	weight based priming
A	emphasizing distinctions collapsing diffs	A	long term changes learning
B	error driven hebbian task model based	B	active maintenance short term residual
C	spiking rate code membrane potential pt	C	fast arbitrary details conjunctive
2.	bidirectional connectivity	7.	hippocampus learning
A	amplification pattern completion	A	fast arbitrary details conjunctive
B	competition inhibition selection binding	B	slow integration general structure
C	language generalization nonwords	C	error driven hebbian task model based
3.	cortex learning	8.	dyslexia
A	error driven task based hebbian model	A	surface deep phonological reading problem
B	error driven task based	B	speech output hearing language nonwords
C	gradual feature conjunction spatial invar	C	competition inhibition selection binding
4.	object recognition	9.	past tense
A	gradual feature conjunction spatial invar	A	overregularization shaped curve
B	error driven task based hebbian model	B	speech output hearing language nonwords
C	amplification pattern completion	C	fast arbitrary details conjunctive

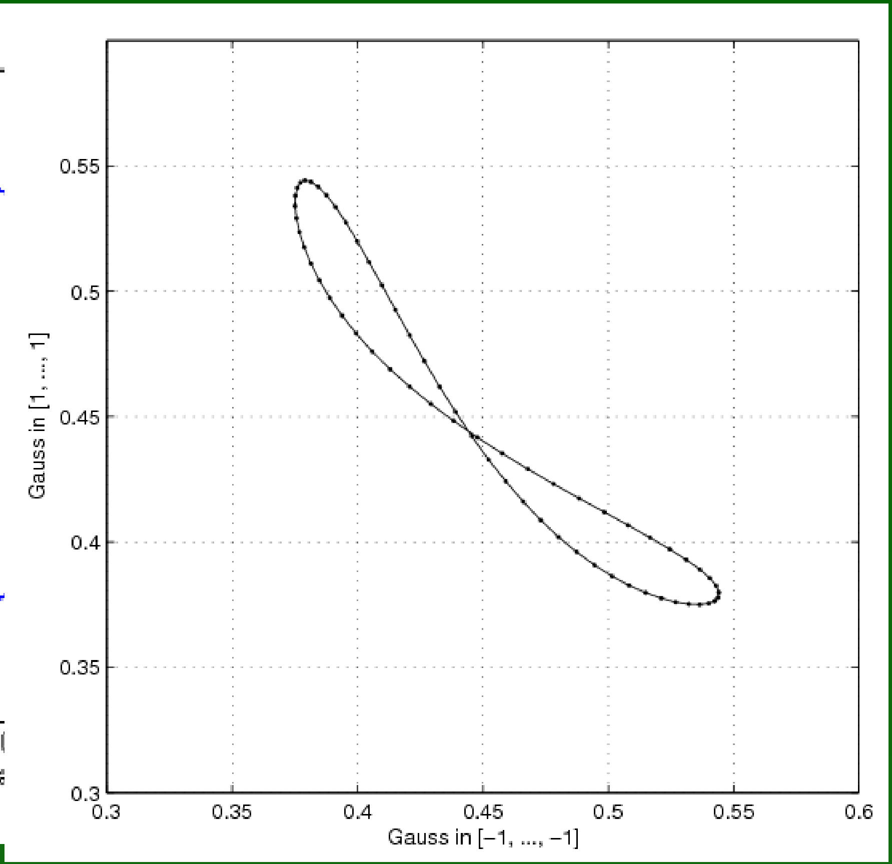
Questions are numbered, each has 3 choices.

Network gives an intuitive answer, based purely on associations, for example what is the purpose of “transformation”: A, B or C.

Network correctly recognizes 60-80% of such questions, more than that requires some understanding …

Model, radial/linear sources



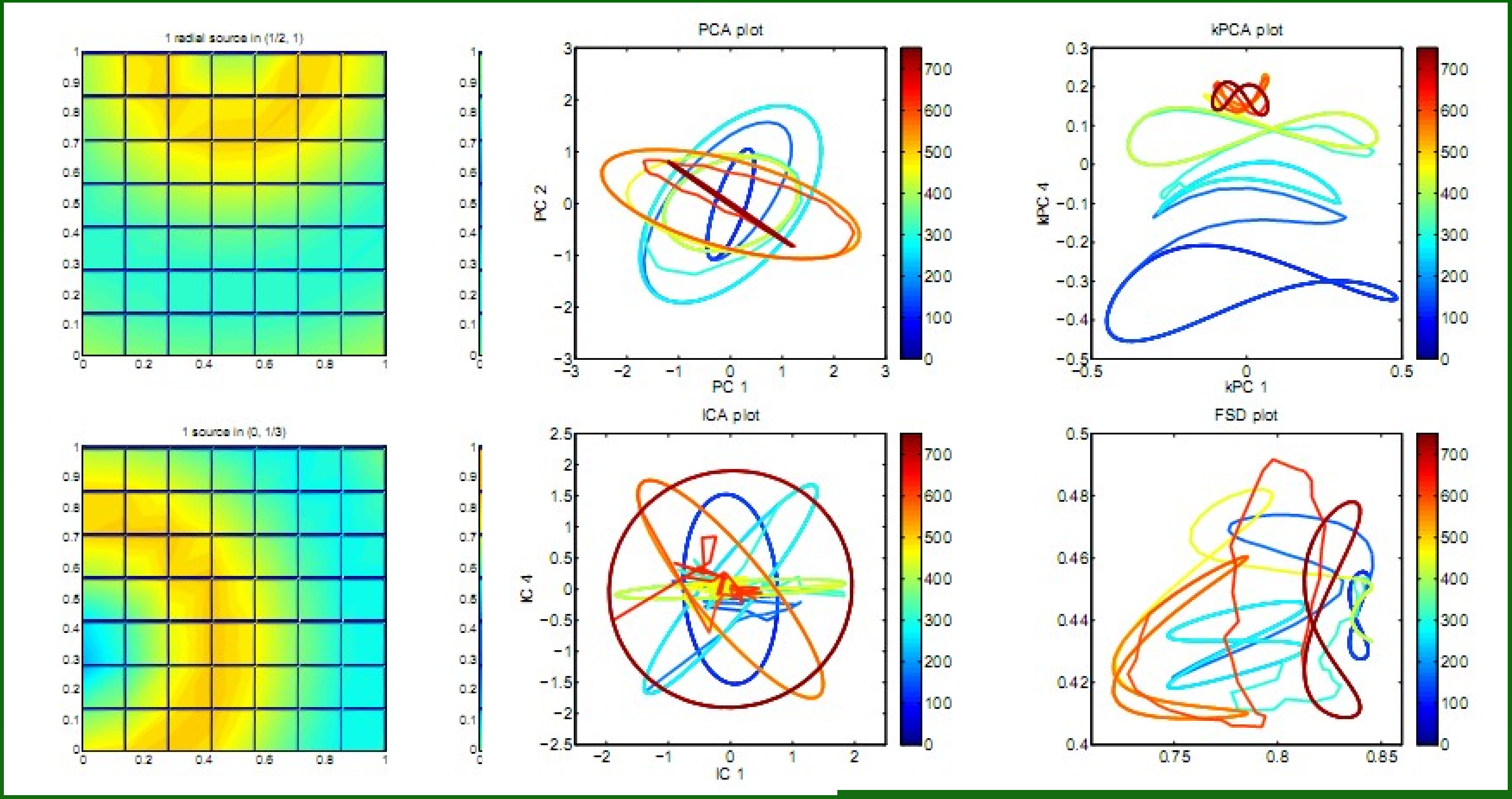


Relatively simple patterns arise, but o

Ex: one and two radial waves.

Radial + plane waves

Radial sources are turned on and off, 5 events+transients.


1 radial source in (1/2, 1)
1 source in (0, 1/3)
PCA plot
PC 1
PC 2
kPCA plot
kPC 1
kPC 4
ICA plot
IC 1
IC 4
FSD plot