



Representation of concepts in brain networks

Włodzisław Duch

Neurocognitive Laboratory,
Center for Modern Interdisciplinary Technologies,
Dept. of Informatics, Faculty of Physics, Astronomy & Informatics,
Nicolaus Copernicus University

Google: Wlodzislaw Duch

Knowledge representation in many-agent systems. Toruń 2018

On the threshold of a dream ...

From analysis of brain processes to neural networks and AI/NLP applications.

Brain – Mind relations.

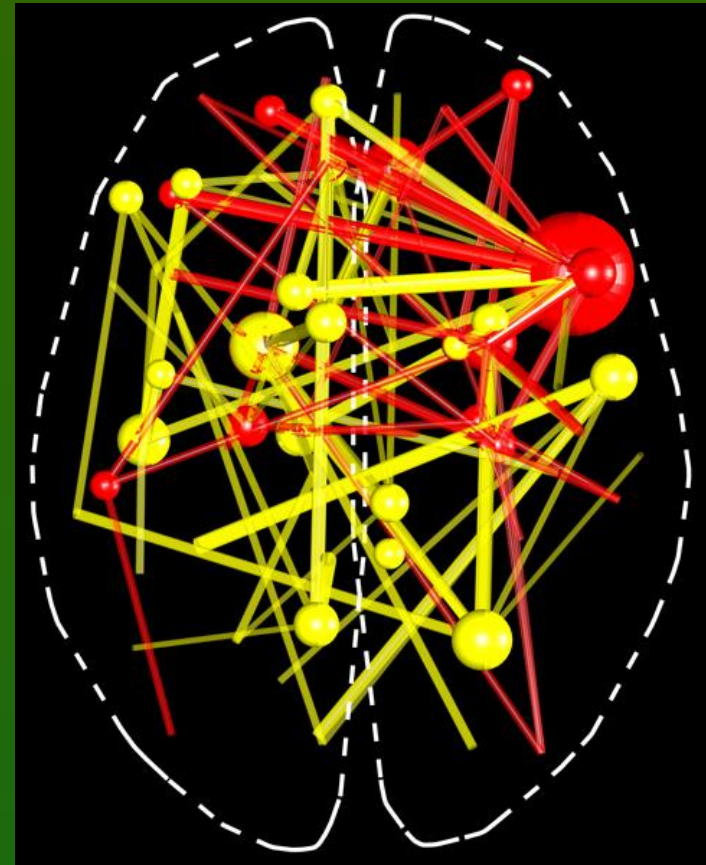
Phenomics.

Development.

Brain simulations.

Fingerprints of real mental activity.

Neurodynamics on real brain networks.



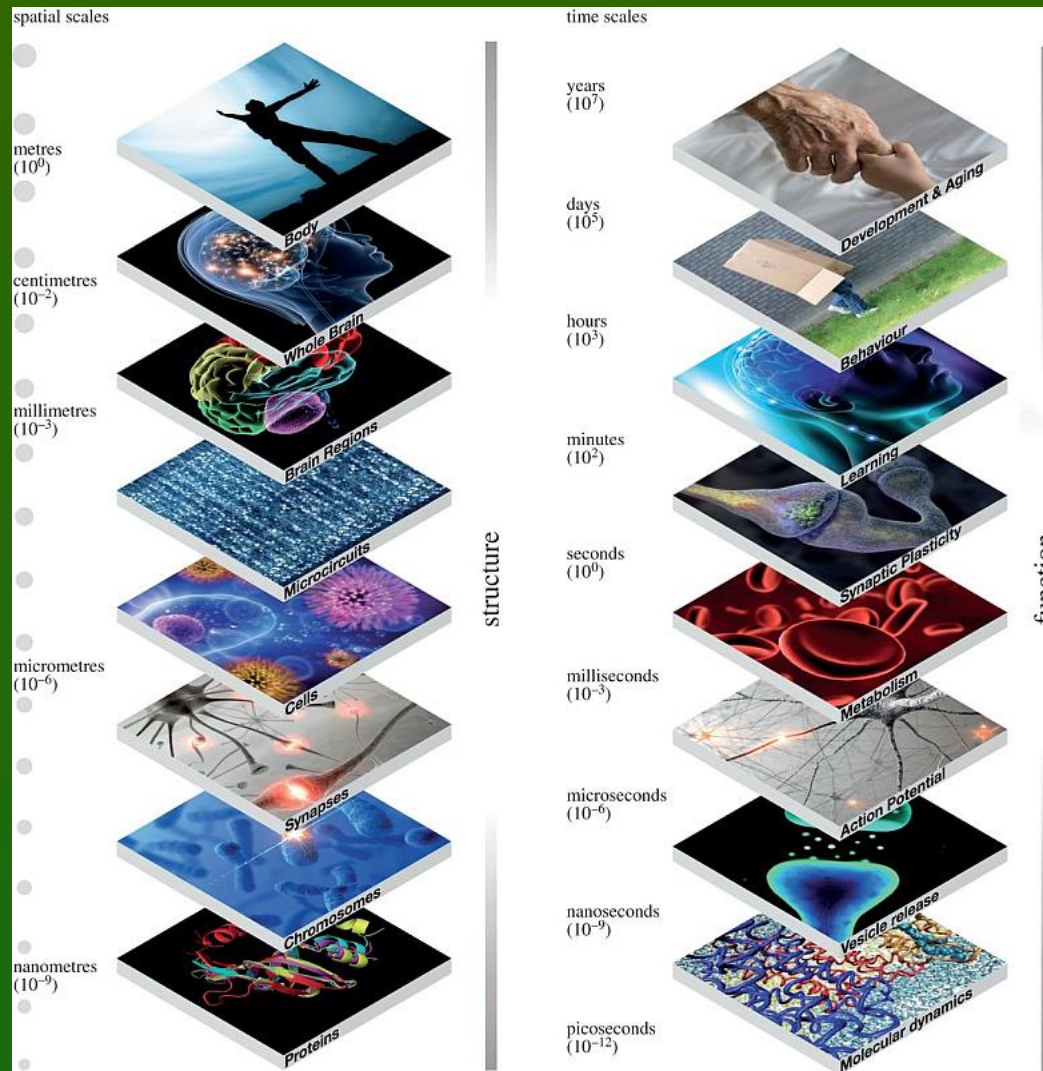
The problem

How do brains, using massively parallel computations, represent knowledge that supports thinking?

- **L. Boltzmann** (1899): “All our ideas and concepts are only internal pictures ... The task of theory consists in constructing an image of the external world that exists purely internally ...”.
- **L. Wittgenstein** (Tractatus 1922): thoughts are pictures of how things are in the world, propositions point to pictures.
- **K. Craik** (1943): the mind constructs "small-scale models" of reality to anticipate events, to reason, and help in explanations.
- **P. Johnson-Laird** (1983): mental models are psychological representations of real, hypothetical or imaginary situations.
- **J. Piaget** (1958): humans develop a context-free deductive reasoning scheme at the level of elementary first-order logic.
- **M. Minsky** (1986), *Society of Mind*: human mind is a vast society of individually simple processes known as agents.
Hierarchical: from simple neurons to whole societies.



Phenomics: levels in space and time



RDoC, neuropsychiatric phenomics, detailed description of major regulatory, affective and cognitive systems at all levels.

A picture is worth a thousand words.

Is verbal description sufficient for recognition?

Experiment. 329 breeds in 10 categories:

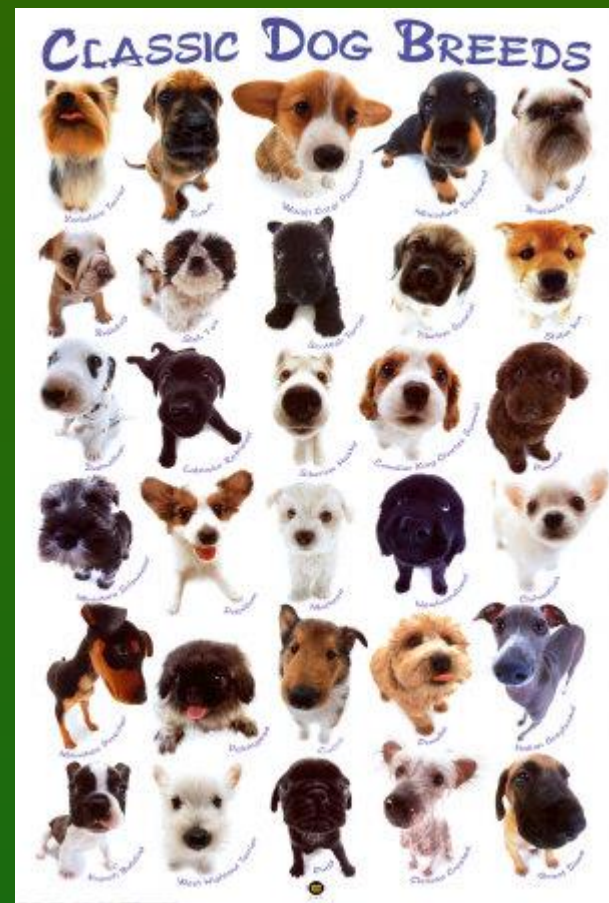
Sheepdogs and Cattle Dogs; Pinscher and Schnauzer;
Spitz and Primitive; Scenthounds; Pointing Dogs;
Retrievers, Flushing Dogs and Water Dogs;
Companion and Toy Dogs; Sighthounds

Write down properties and try to use them in the
20-question game to recognize the breed ... fails!

Visually each category is quite different.

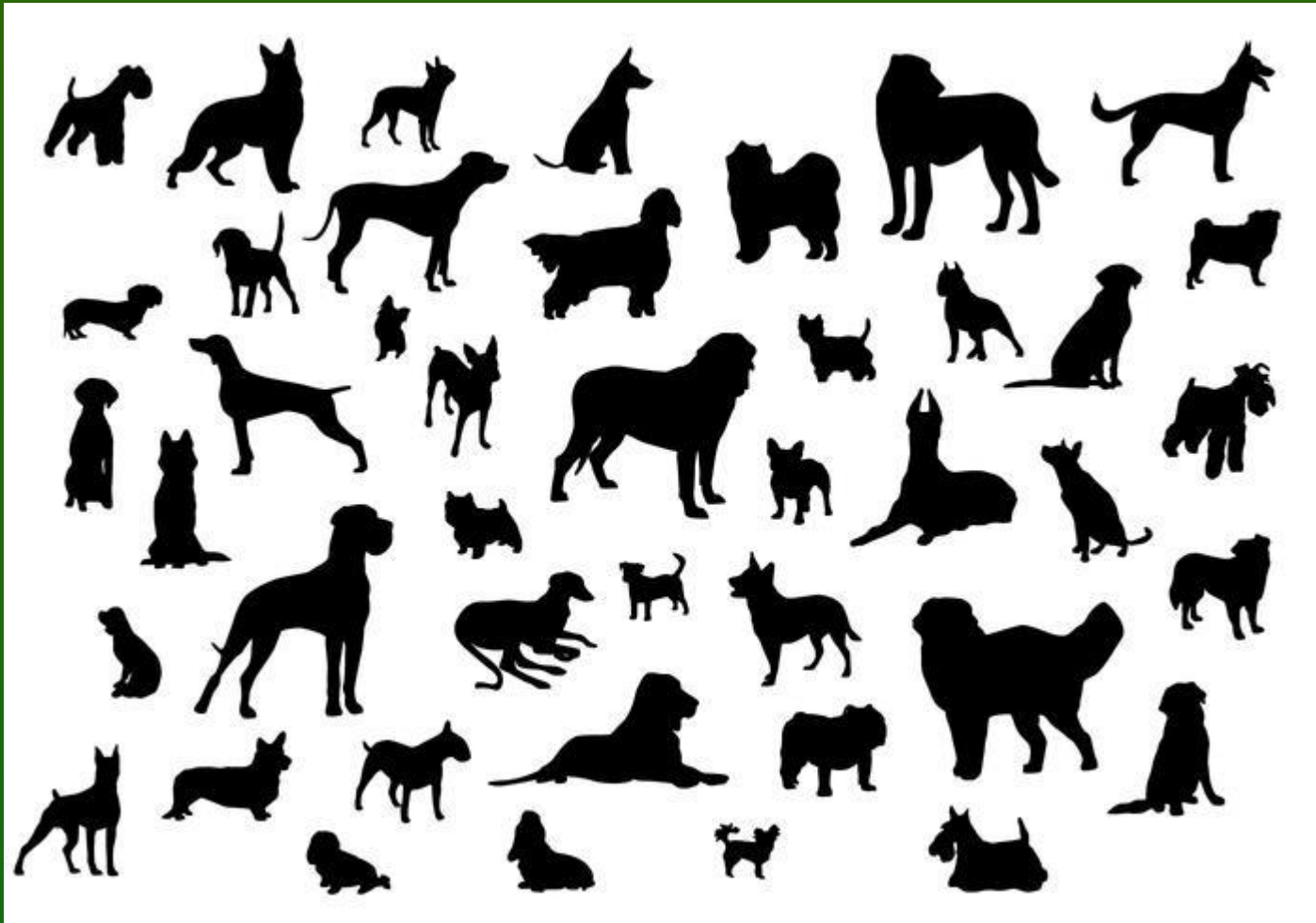
Traditional categorizations are based on behaviors
and features that are not easy to observe.

- Ontologies do not agree with visual similarity.
- Images are important, words are not sufficient even for simple recognition – how are images encoded in the brain?



Dog breeds

Words point to what we already know, silhouettes of dogs images are sufficient for recognition. Brain states have linguistic labels if they are frequently shared.



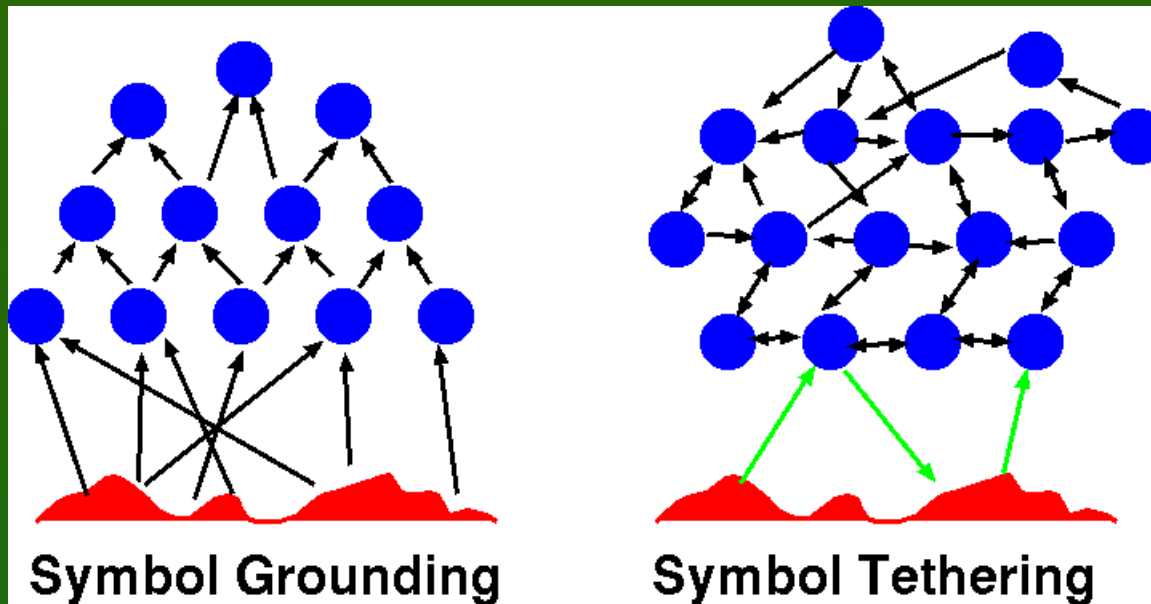
Imitation will get you quite far ...



Where is the meaning?

Symbol grounding problem (Harnad 1990): how can the meaning of concepts be represented in artificial symbolic systems?

- No representations, only sensorimotor embodiment (robotics, Cog). Some concepts have shared meaning through embodiment.

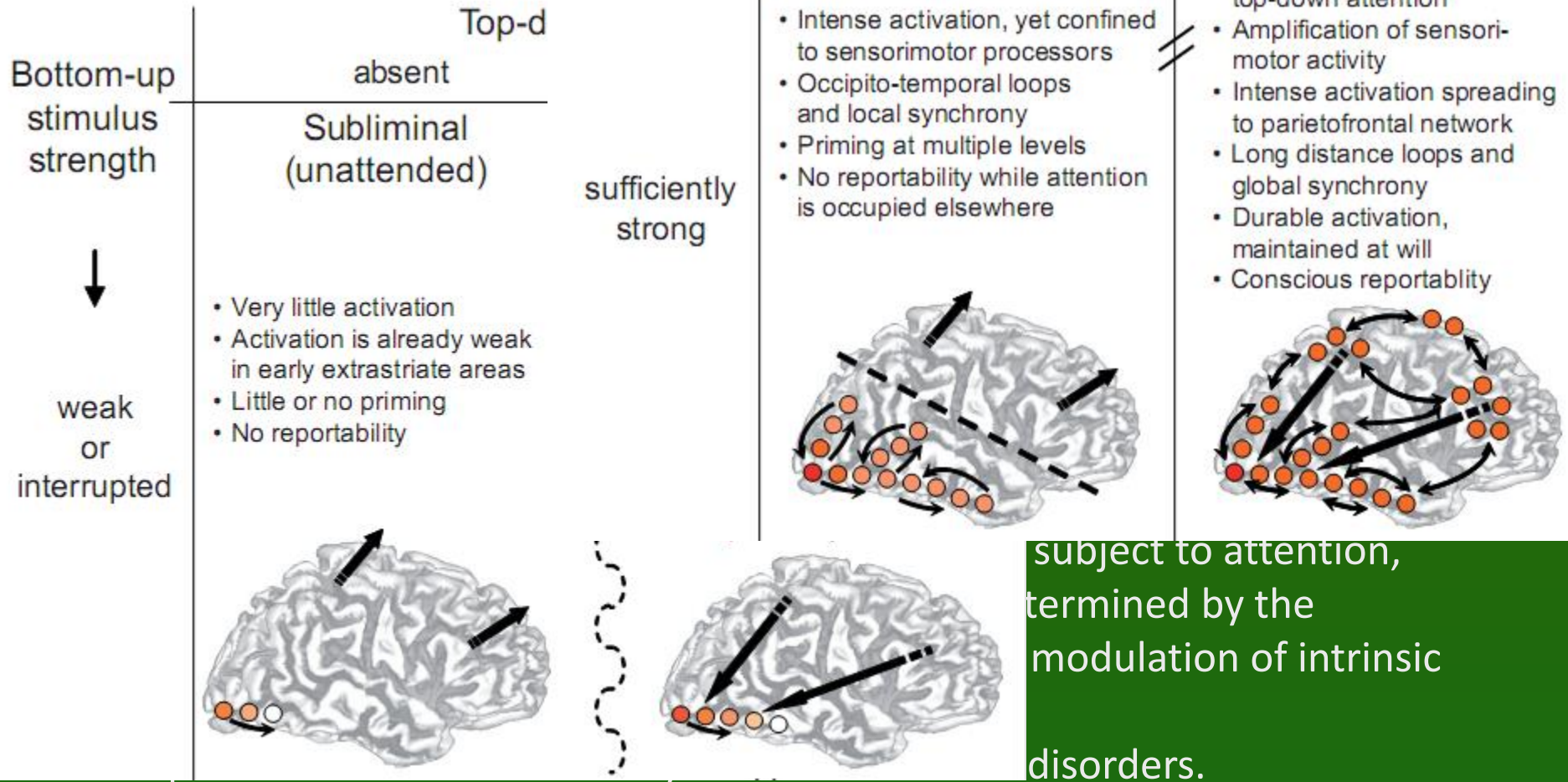


Aaron Sloman (2007): only simple concepts come from our “being in the world” experience, others are compounds, abstract, relational.

David Hume gave a good example: “golden mountain”.

Not symbol grounding but symbol tethering, meaning from mutual interactions.

What is



Dehaene et al, Conscious, preconscious, and subliminal processing, TCS 2006
 Bottom-up strength & top-down attention combined leads to 4 brain states with both stimulus and attention required for conscious reportability. No imagery?

Brains ↔ Minds

Define mapping $S(M) \leftrightarrow S(B)$, as in BCI.

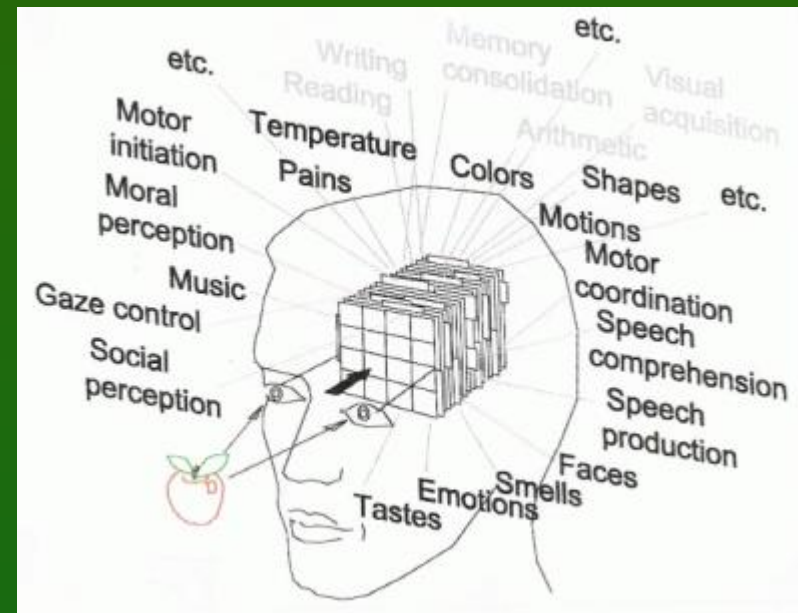
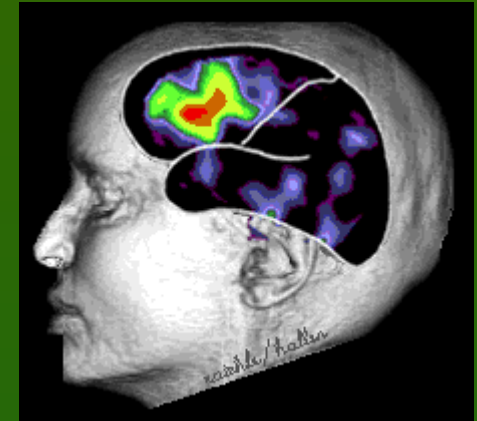
How do we describe the state of mind?

Verbal description is not sufficient unless words are represented in a space with dimensions that measure different aspects of experience.

Stream of mental states, movement of thoughts
↔ trajectories in psychological spaces.

Two problems: discretization of continuous processes for symbolic models, and lack of good phenomenology – we are not able to describe our mental states.

Neurodynamics: bioelectrical activity of the brain, neural activity measured using EEG, MEG, NIRS-OT, PET, fMRI ...



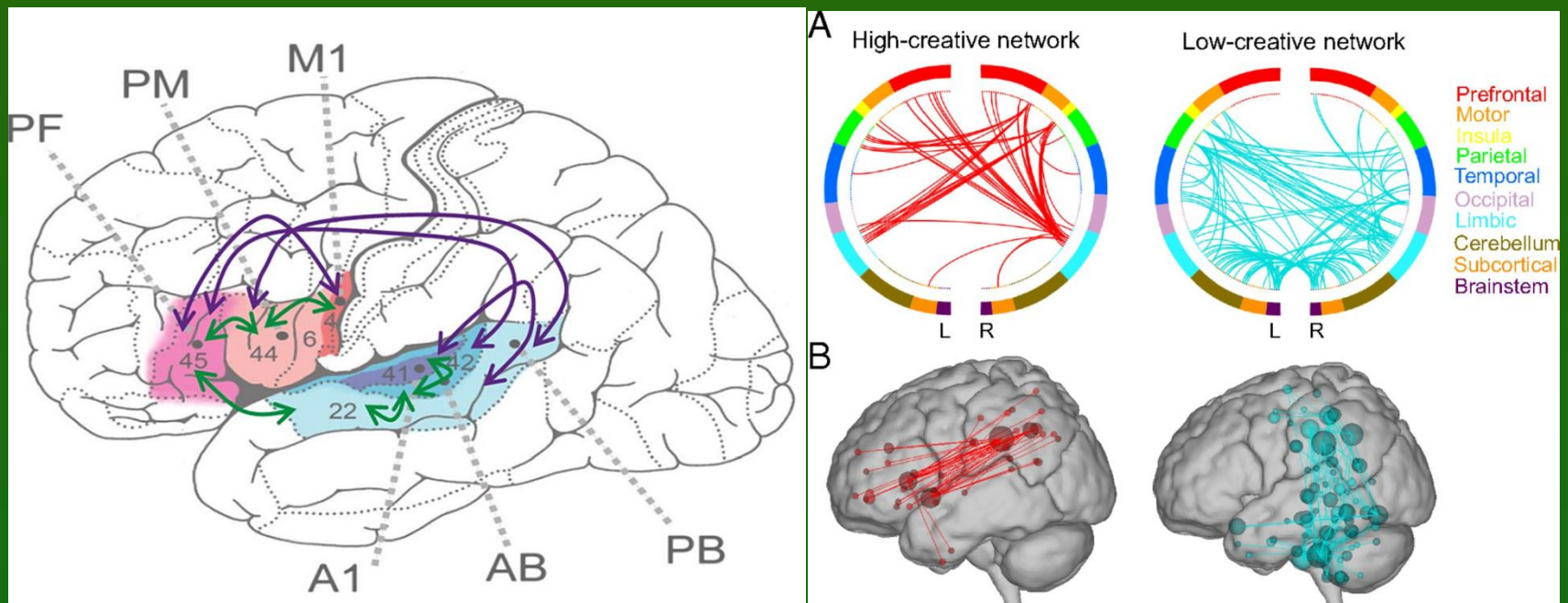
E. Schwitzgabel, Perplexities of Consciousness. MIT Press 2011.

Fluid nature

Development of brain in infancy: first learning how to move, sensorimotor activity organizes brain network processes.

[The Developing Human Connectome Project](#): create a dynamic map of human brain connectivity from 20 to 44 weeks post-conceptual age, which will link together imaging, clinical, behavioral, and genetic information.

Pointing, gestures, pre-linguistic – Monika Boruta-Żywiczyńska (our BabyLab).



Logic and language

Logic arguments: if both X and Z then not Y, or If Y then either not X or not Z, sentential connectives

Linguistic arguments:

It was X that Y saw Z take, or Z was seen by Y taking X, phrasal verbs.

The ability to use logic and understand language may dissociate.

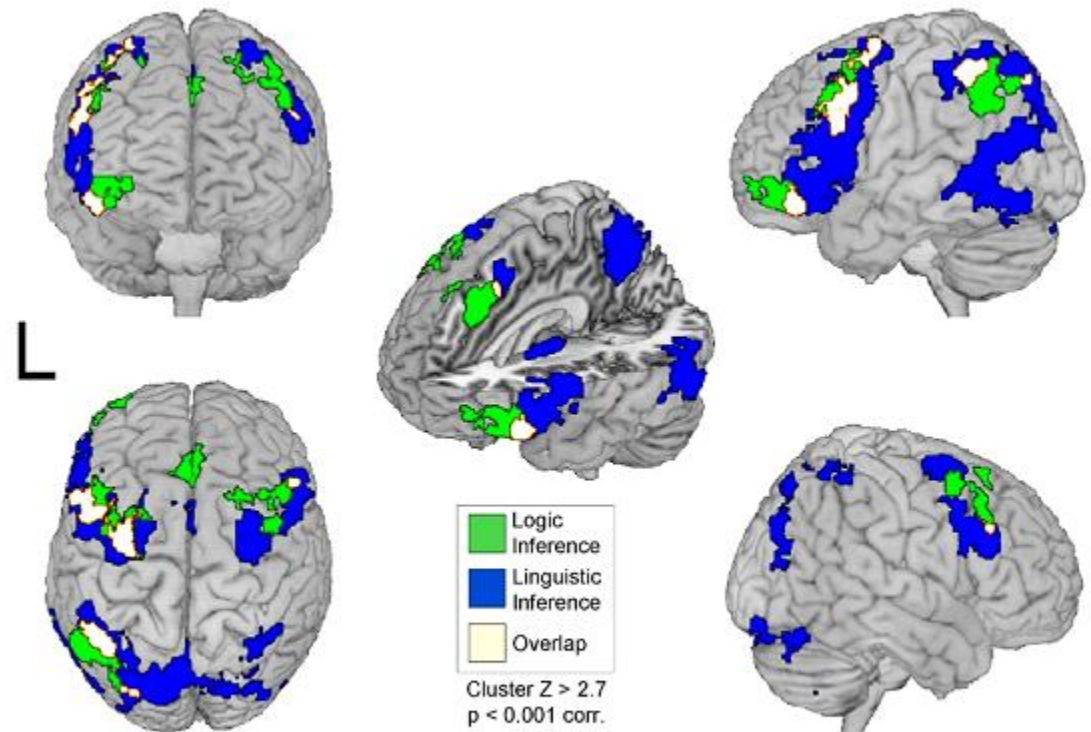
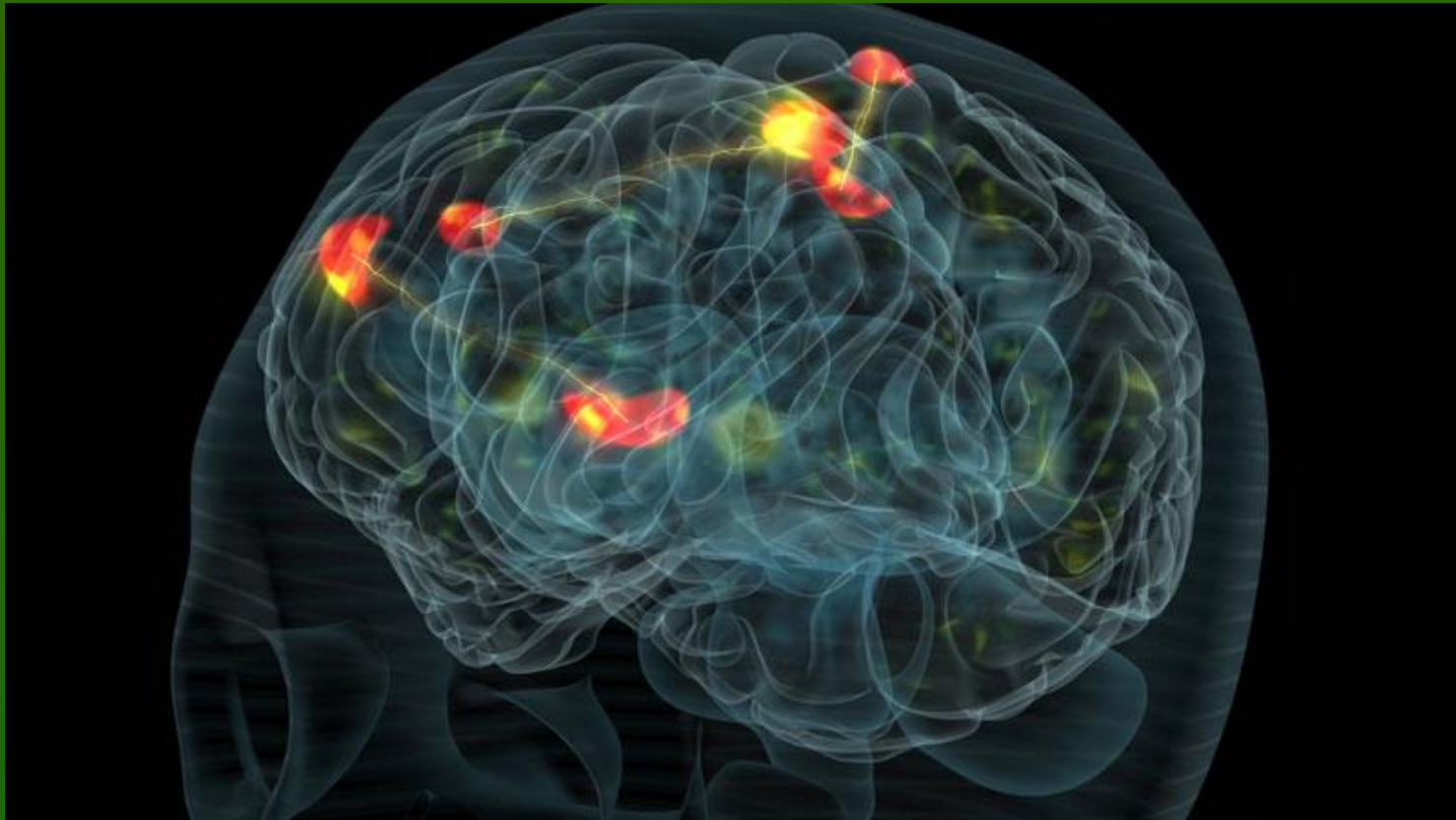


Fig. 1. Inference minus grammar contrast. Mean group activity for logic arguments (green/yellow) and linguistic arguments (blue/yellow).

M.M. Monti, L.M. Parsons, D.N. Osherson, The boundaries of language and thought: neural basis of inference making. PNAS 2009

Mental state: strong coherent activation

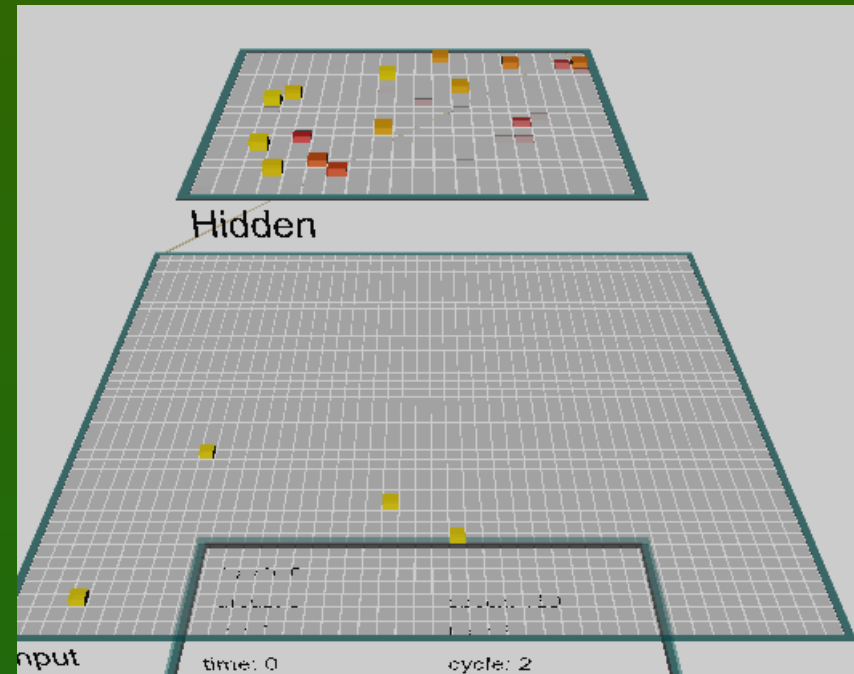


Many processes go on in parallel, controlling homeostasis and behavior. Most are automatic, hidden from our Self. What is noise and what thought? Signal Detection Theory: time is needed to build statistics, many active subnetworks compete for access to consciousness, the winner-takes-most mechanism leaves only the strongest at each moment: percept, though ...

Cognitive Computational Neurodynamics

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 represented by A and B vectors, calculating $\cos(A,B) = \frac{A \cdot B}{|A| |B|}$.

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

For each questions there are 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, enough to pass examination. This should be a base rate for understanding.

Model of reading



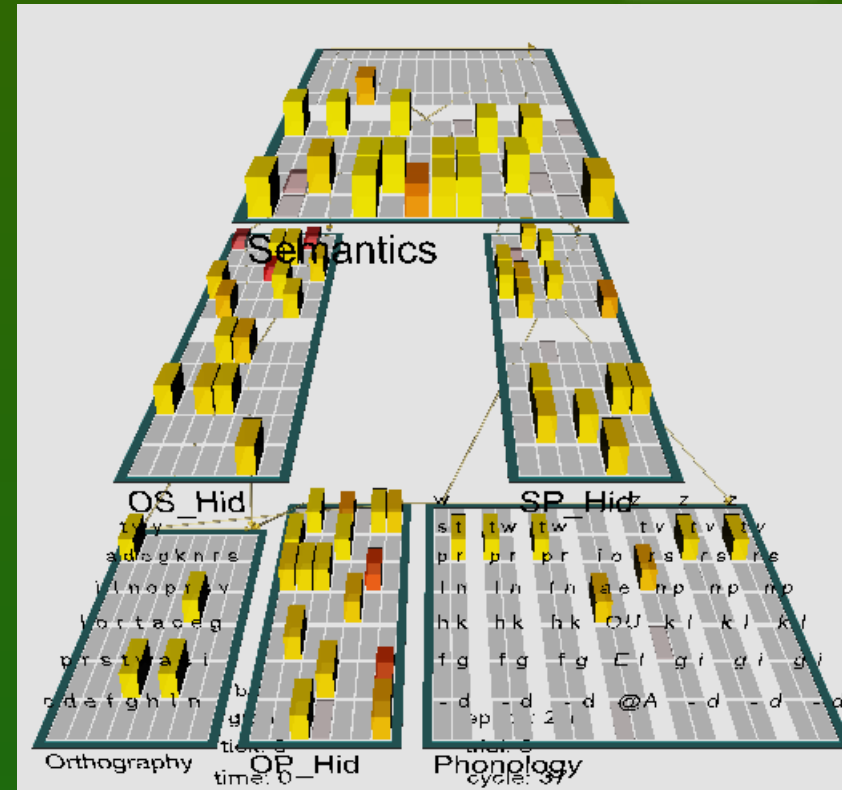
Emergent neural simulator:

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.



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

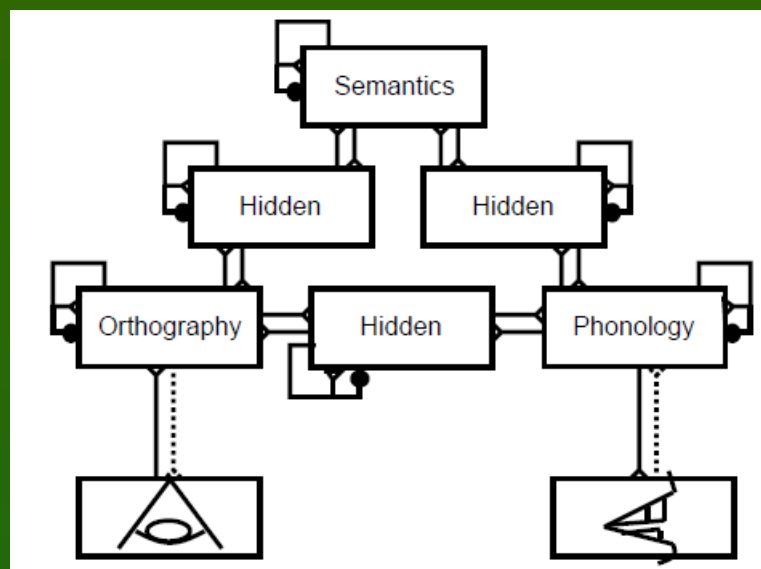
Fluctuations around final configuration = attractors representing concepts.

How to see properties of their basins, their relations?

Reading and dyslexia

Phonological dyslexia: deficit in reading pronounceable nonwords (e.g., “nust” (Wernicke).

Deep dyslexia like phonological dyslexia + significant levels of semantic errors, reading for ex. “dog” as “cat”.

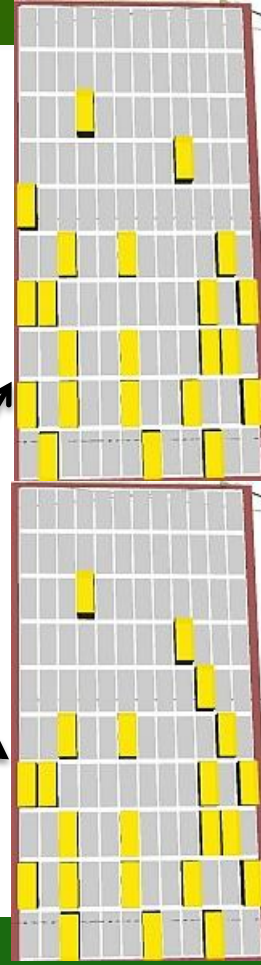
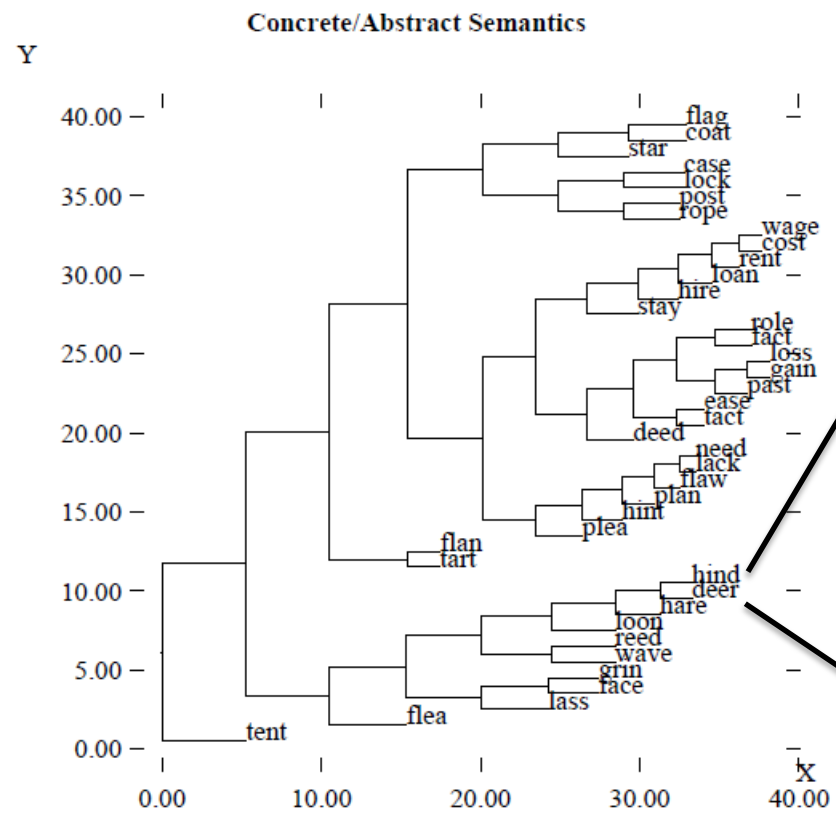


Surface dyslexia: preserved ability to read nonwords, impairments in retrieving semantic information from written words, difficulty in reading exception, low-frequency words, ex. “yacht.”
Surface dyslexia - visual errors, but not semantic errors. .

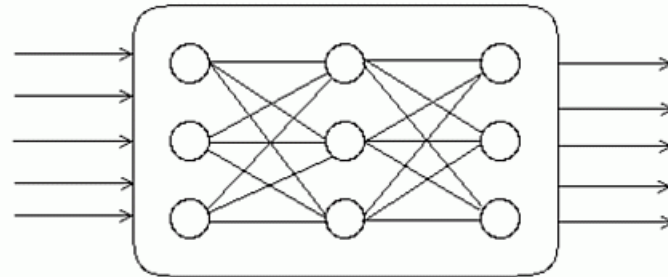
Double route model of dyslexia includes orthography, phonology, and semantic layers, direct ortho=Phono route and indirect ortho => semantics => phono, allowing to pronounce rare words.

Words to read

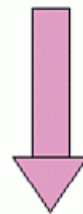
Conc	Phon	Abst	Phon
tart	tttartt	tact	ttt@ktt
tent	tttentt	rent	rrrentt
face	fffAsss	fact	fff@ktt
deer	dddErrr	deed	dddEddd
coat	kkkOttt	cost	kkkostt
grin	grrinnn	gain	gggAnnn
lock	lllakkk	lack	lll@kkk
rope	rrrOppp	role	rrrOlll
hare	hhhArrr	hire	hhhIrrr
lass	lll@sss	loss	lllosss
flan	fllonnn	plan	pll@nnn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



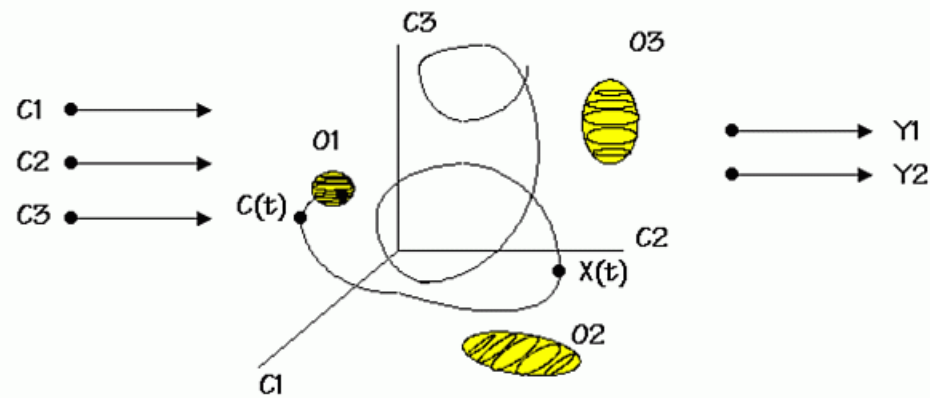
40 words, 20 abstract & 20 concrete; dendrogram shows similarity in phonological and semantic layers after training.



Neurodynamics



Psychological space



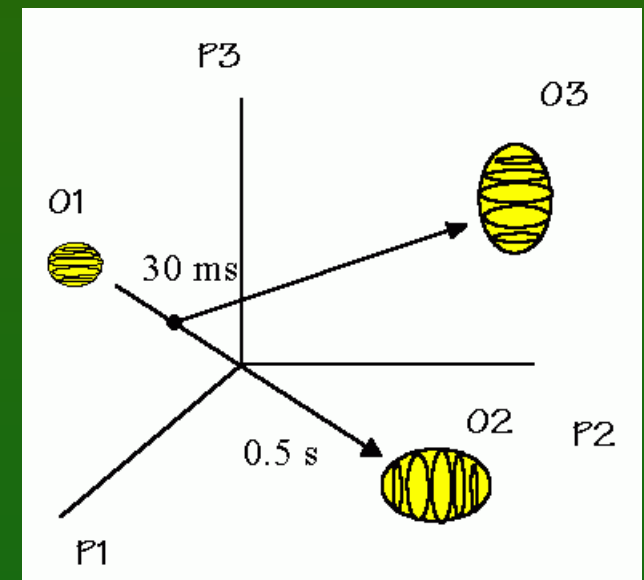
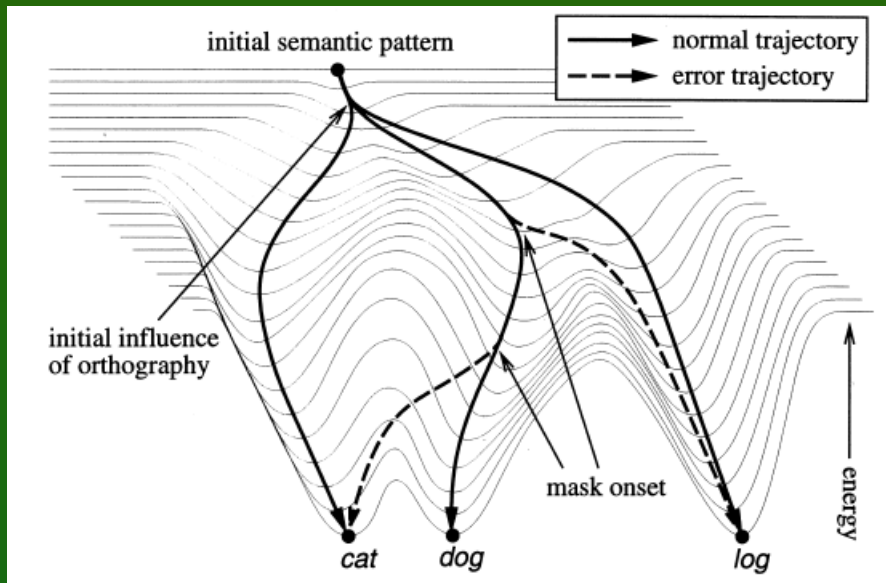
Energies of trajectories

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.

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



Fuzzy Symbolic Dynamics (FSD)

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

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

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

1. Standardize original data in high dimensional space.
2. Find cluster centers (e.g. by k-means algorithm): $\mu_1, \mu_2 \dots \mu_d$
3. Use non-linear mapping to reduce dimensionality to d, for example:

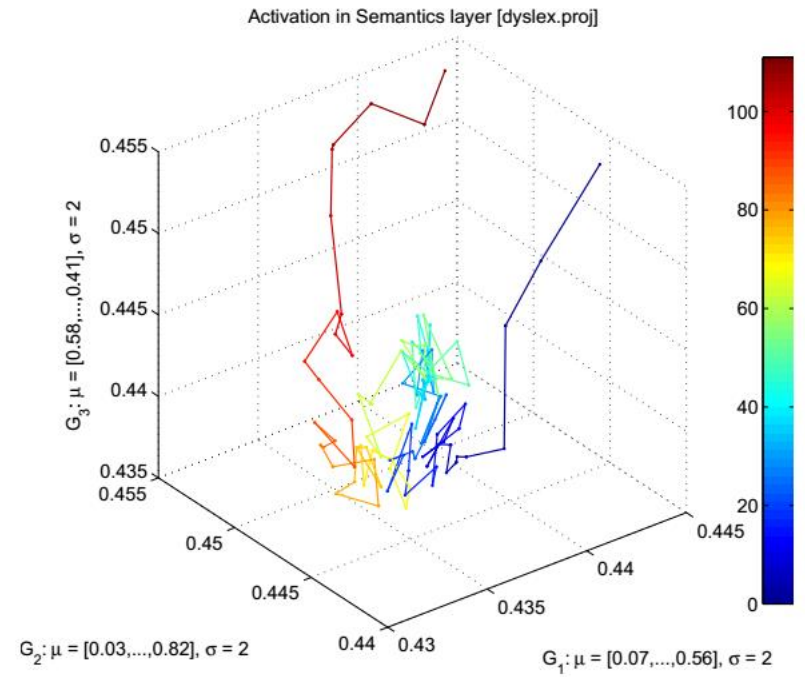
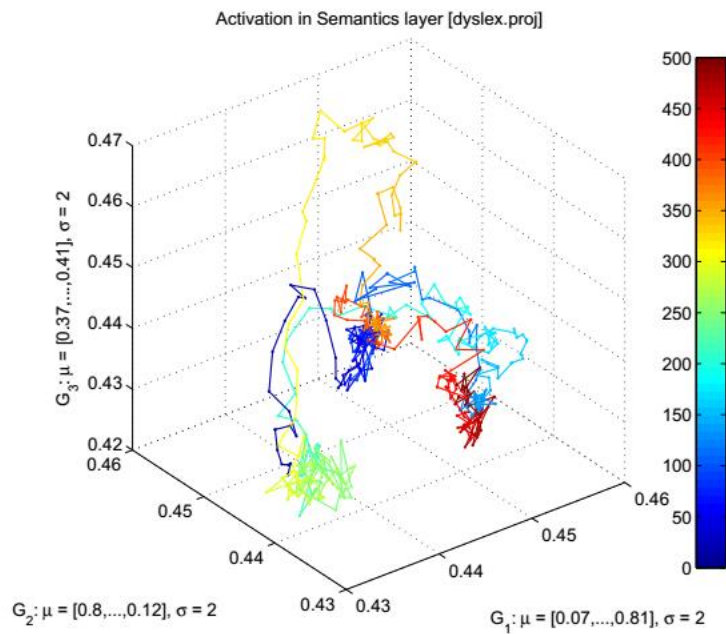
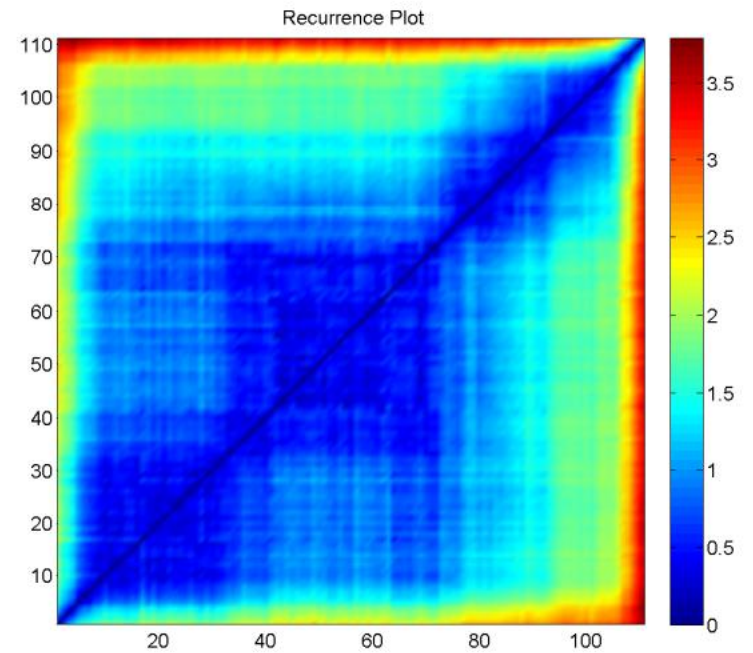
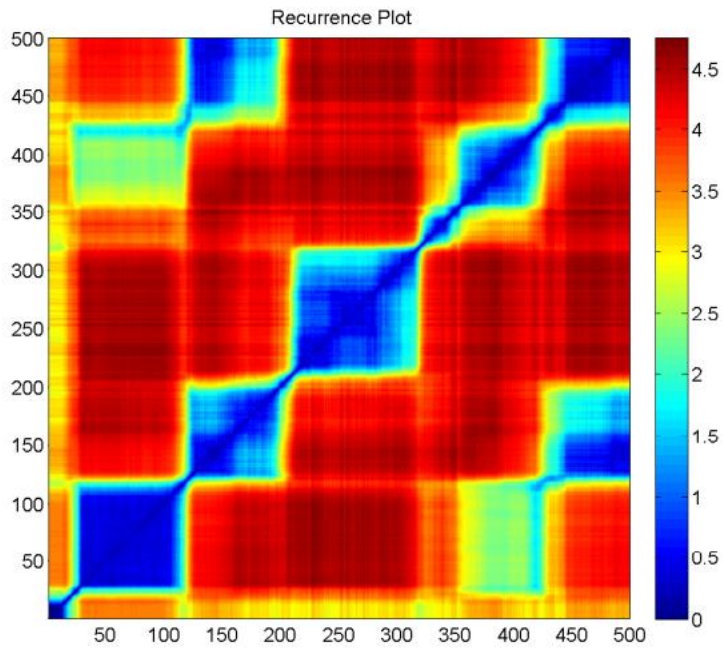
$$y_k(t; \mu_k, \Sigma_k) = \exp\left(-\left(x - \mu_k\right)^T \Sigma_k^{-1} \left(x - \mu_k\right)\right)$$

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

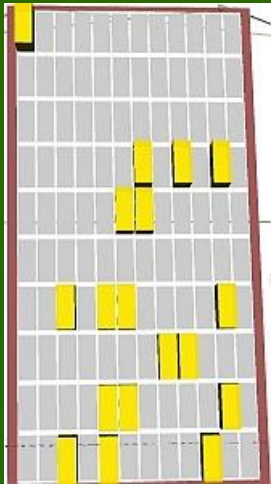
sharp indicator functions \Rightarrow symbolic dynamics; $\mathbf{x}(t) \Rightarrow$ strings of symbols;

soft functions \Rightarrow fuzzy symbolic dynamics, dimensionality reduction

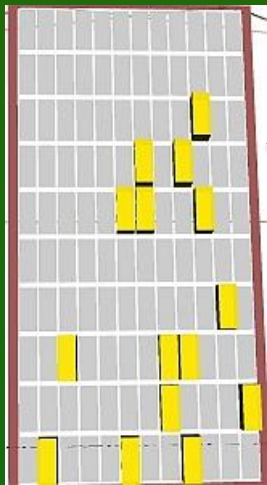
$Y(t) = (y_1(t; W), y_2(t; W)) \Rightarrow$ visualization of high-dim data.



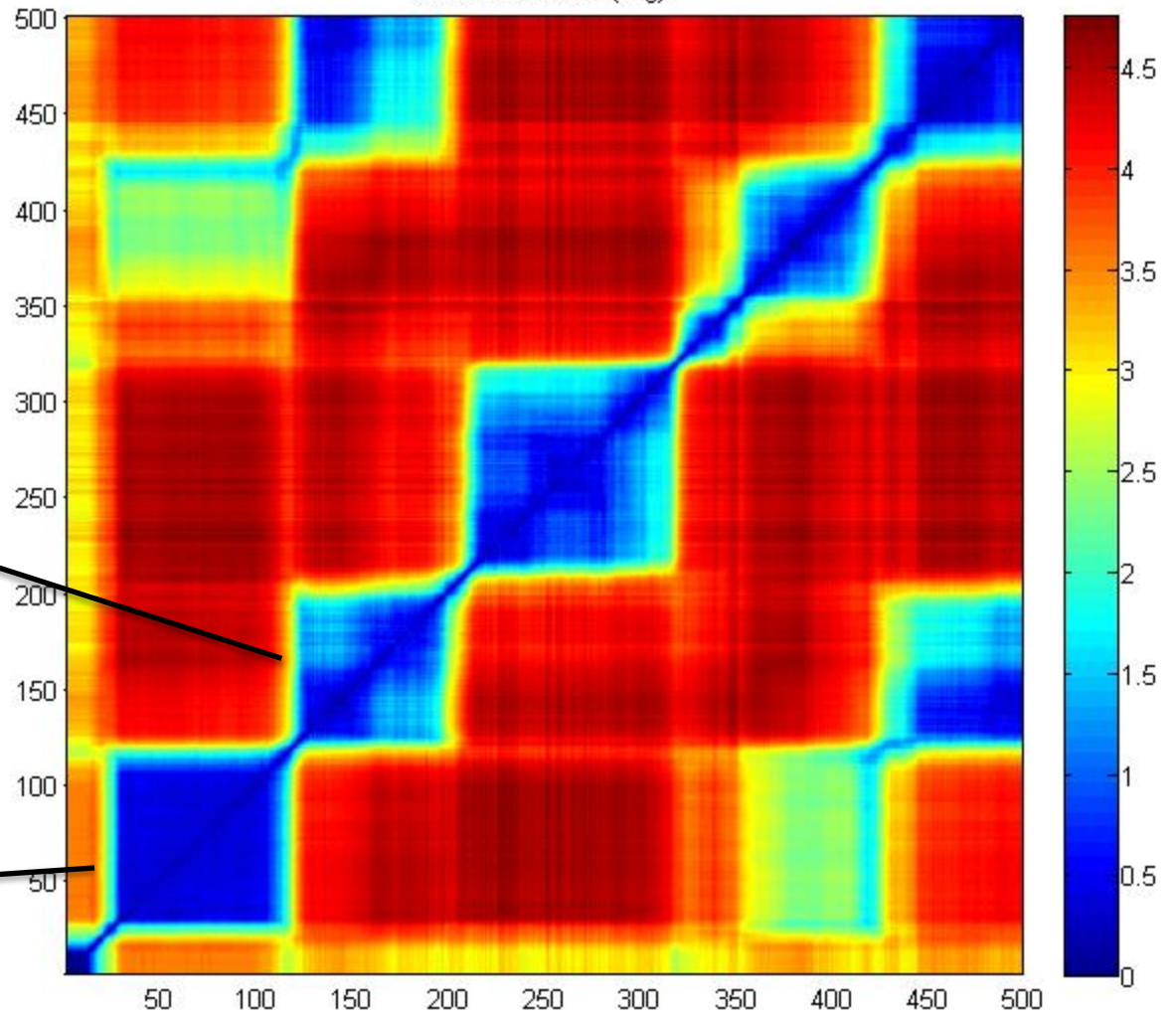
rope



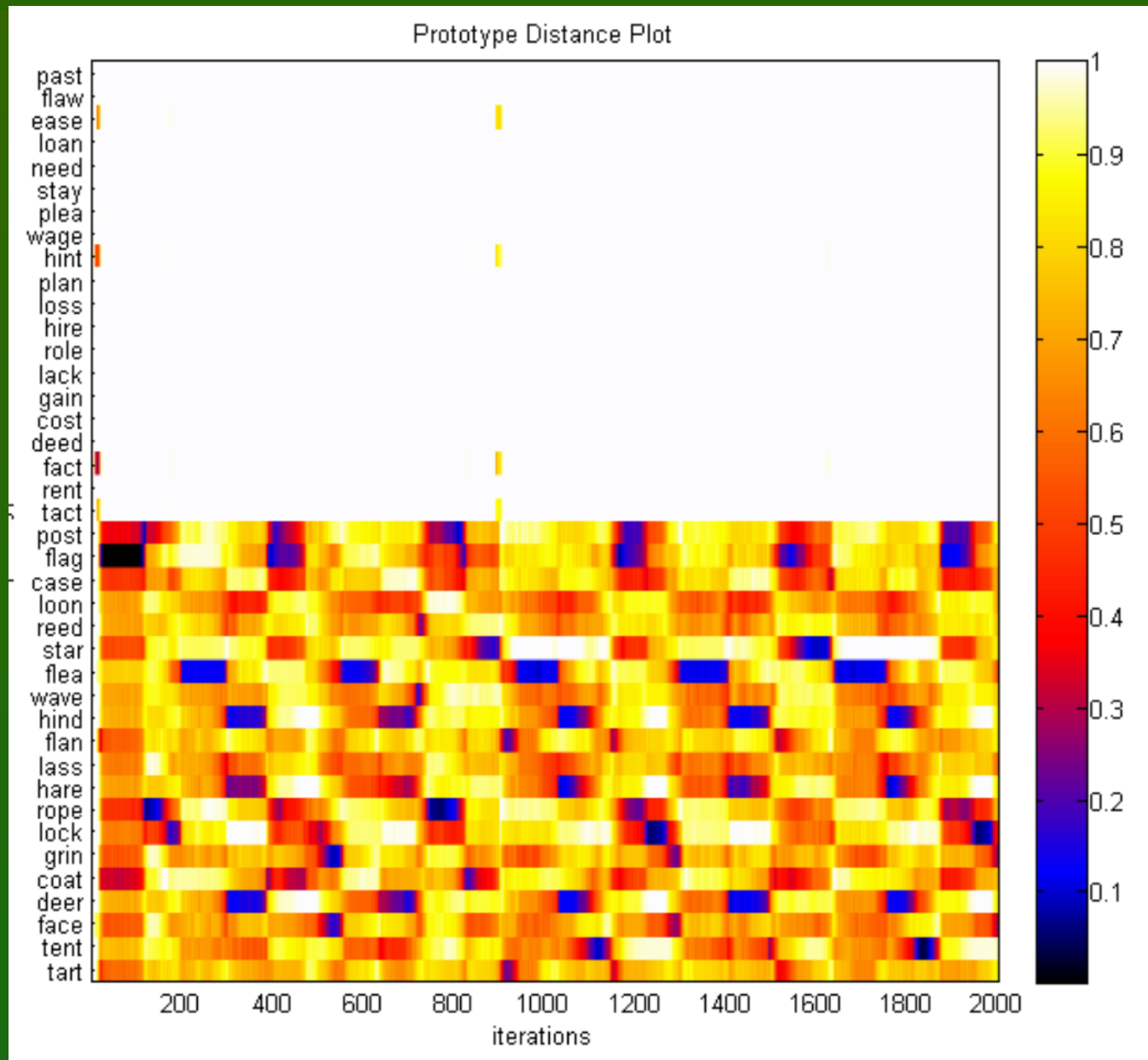
flag



Recurrence Plot (flag)

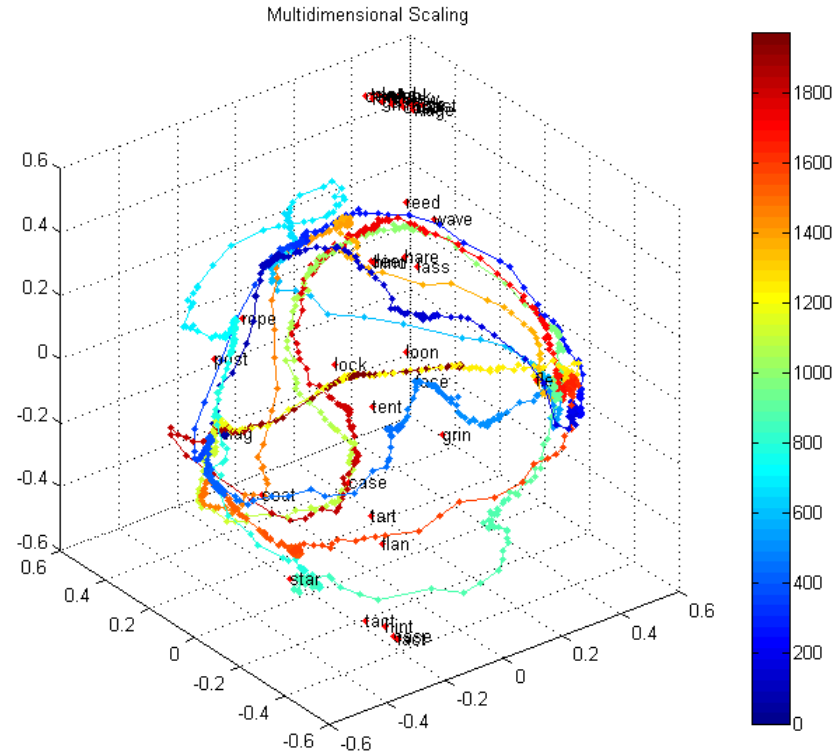
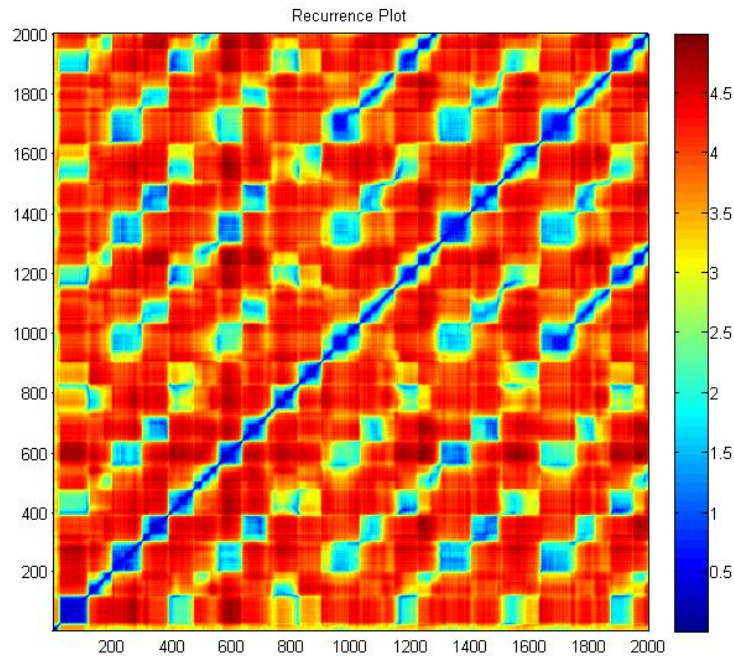


Transitions to new patterns that share some active units (microfeatures).



PDP shows how far is the current state from basins of attractors.
 Abstract concepts have different set of microfeatures, not activated here.

Trajectory visualization



Recurrence plots and MDS/FSD/SNE visualization of trajectories of the brain activity. Here data from 140-dim semantic layer activity during spontaneous associations in the 40-words microdomain, starting with the word “flag”.

Our toolbox: <http://fizyka.umk.pl/~kdobosz/visertoolbox/>

Attract

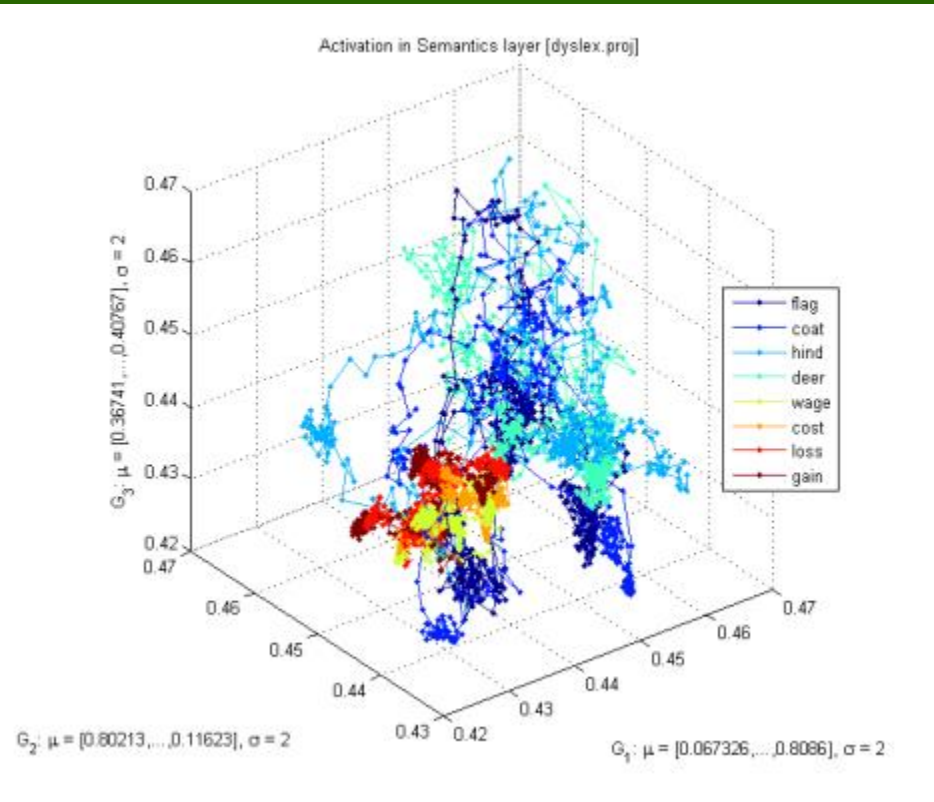
Attention results from:

- inhibitory competition,
- bidirectional interactive processing,
- multiple constraint satisfaction.

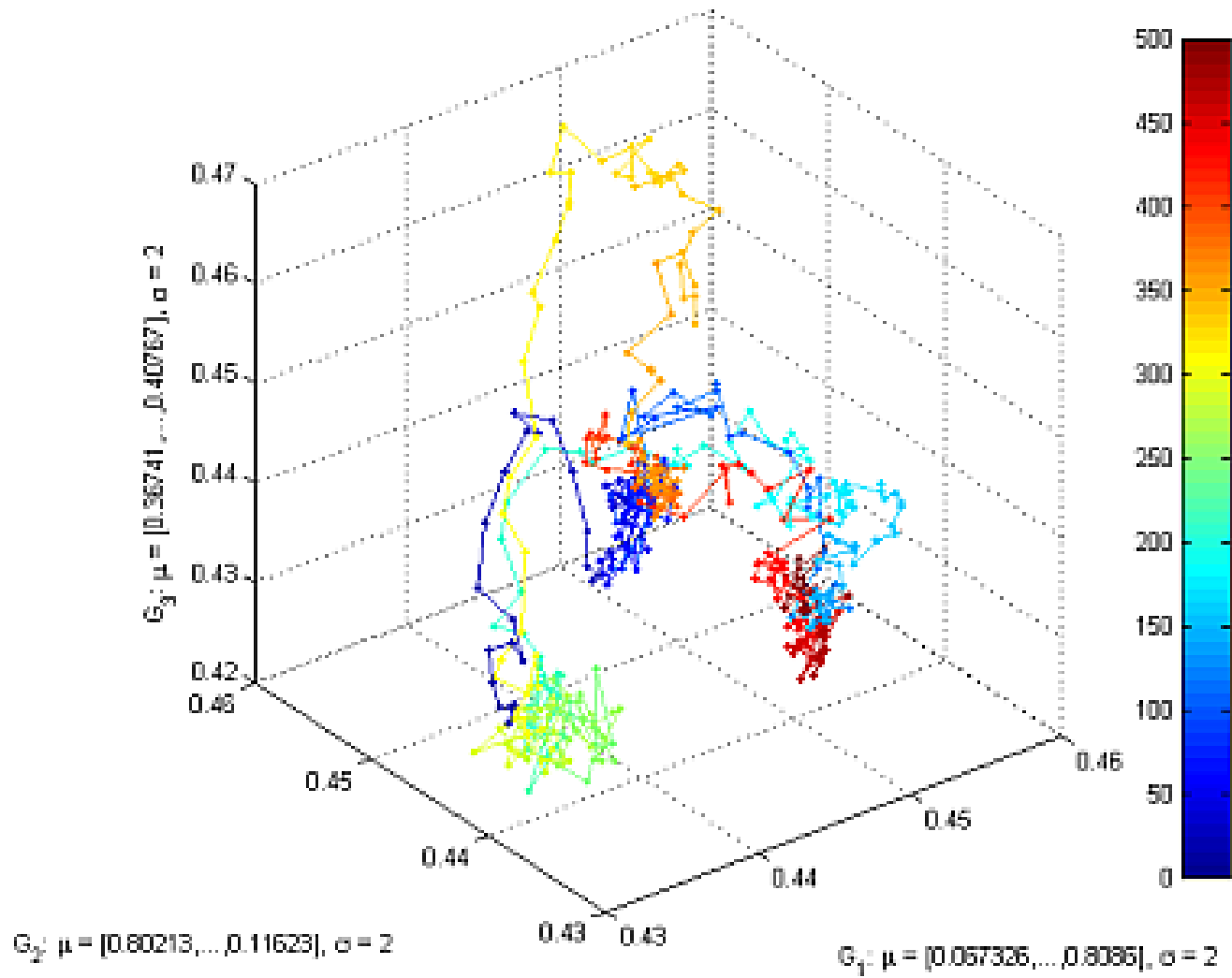
Basins of attractors: input activations {L

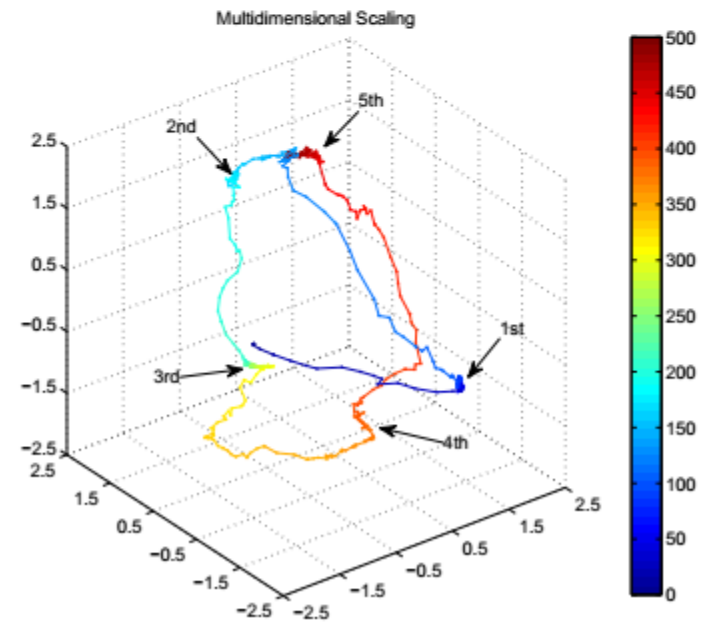
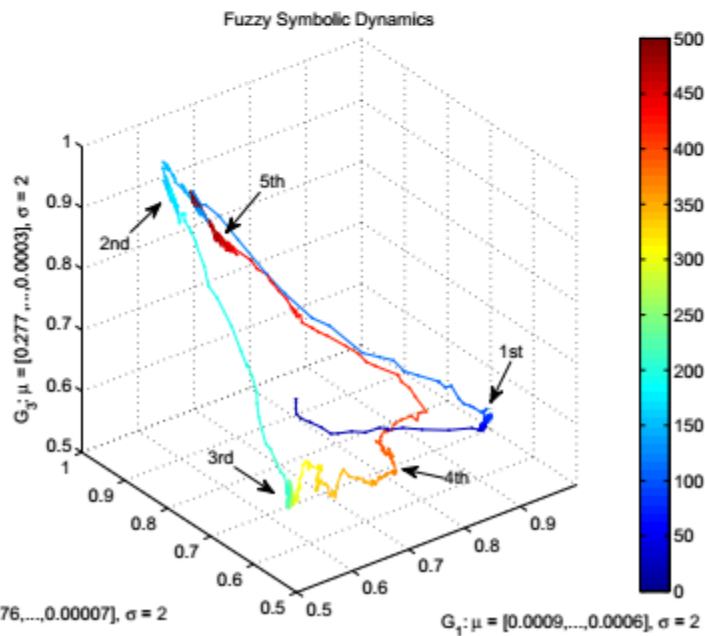
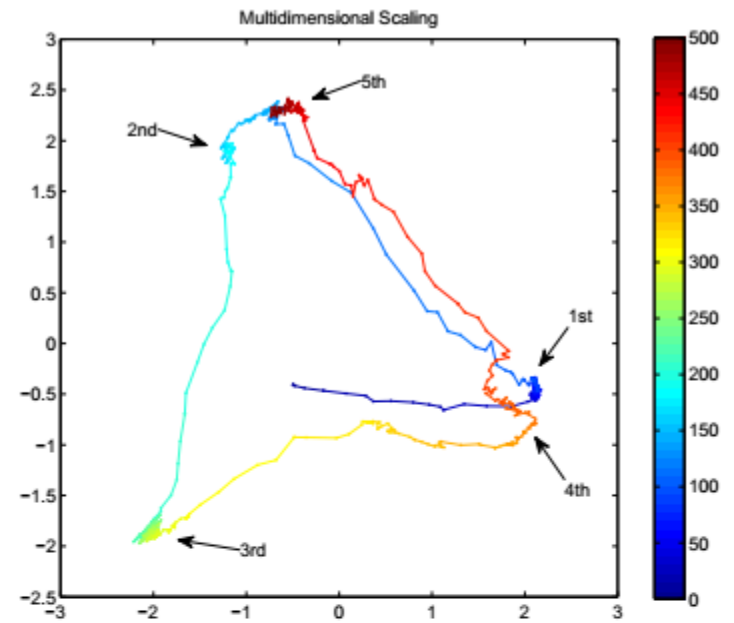
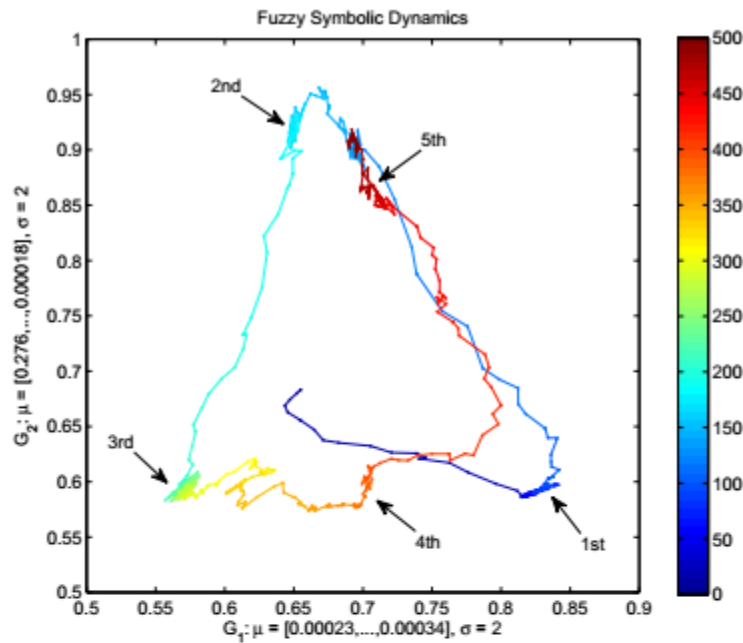
- Normal case: relatively large, easy associations, moving from one basin of attraction to another, exploring the activation space.
- Without accommodation (voltage-dependent K^+ channels): deep, narrow basins, hard to move out of the basin, associations are weak.

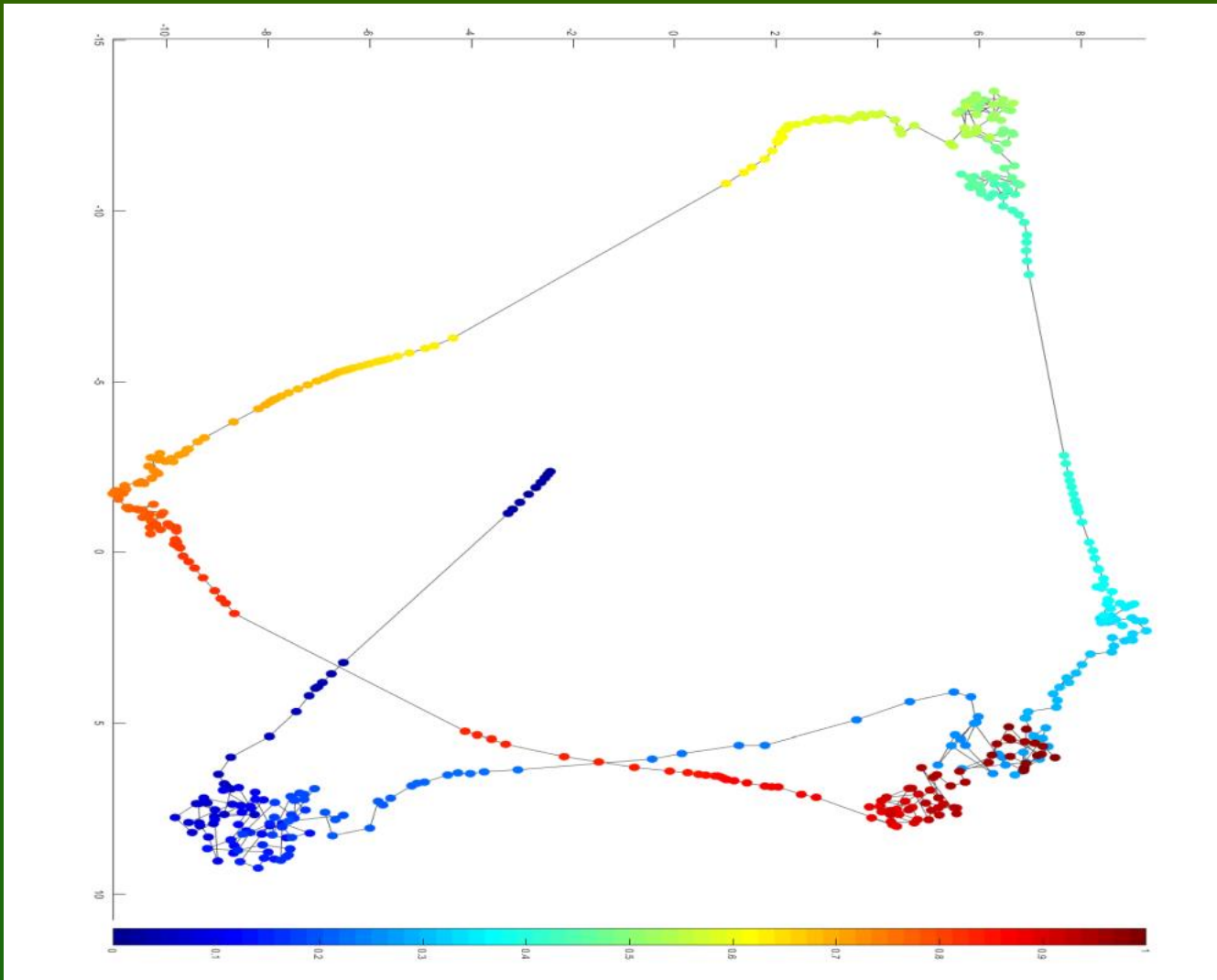
Accommodation: basins of attractors shrink and vanish because neurons desynchronize due to the fatigue; this allows other neurons to synchronize, leading to quite unrelated concepts (thoughts).



Activation in Semantics layer [dyslex.proj]





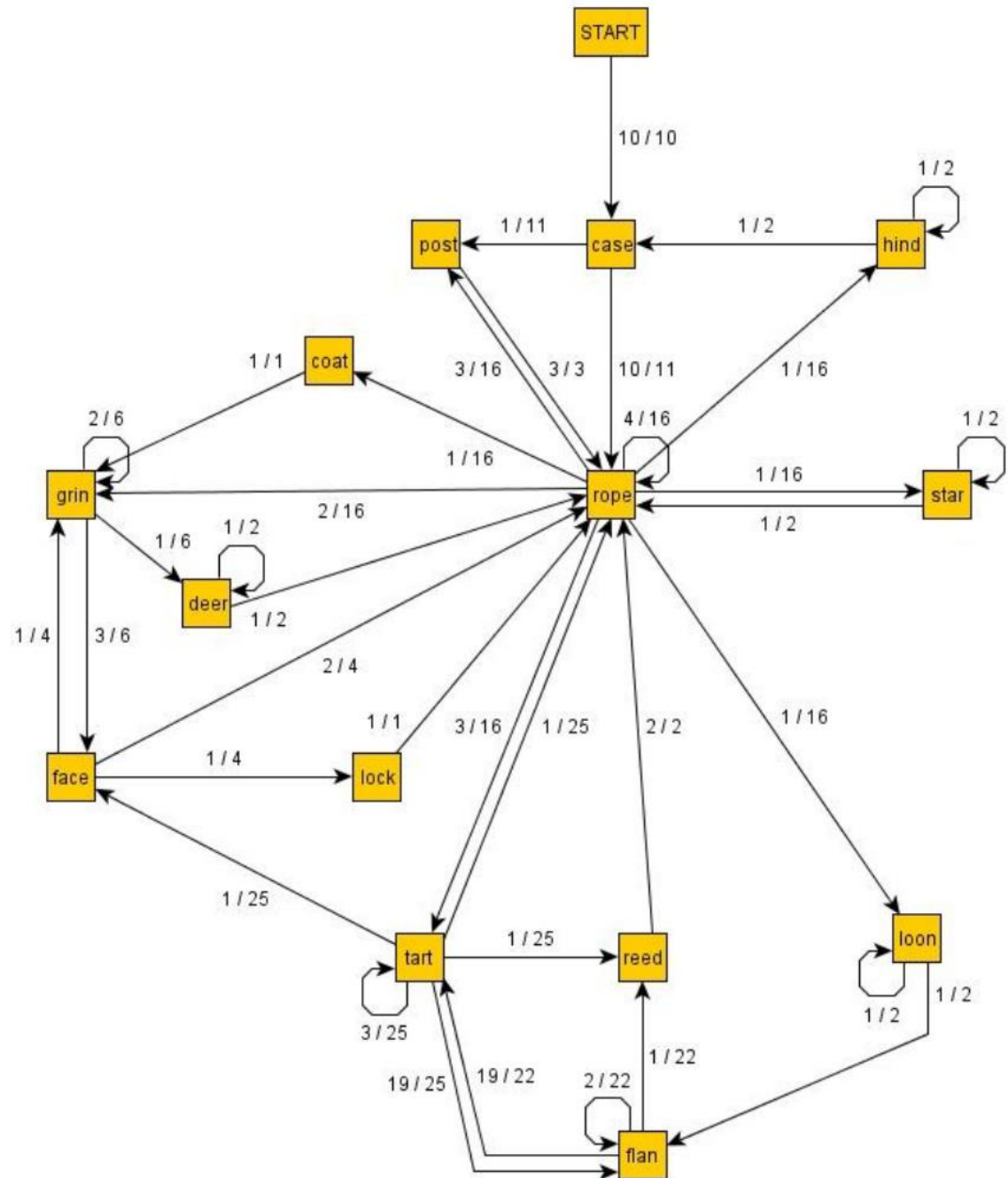


Stochastic Neighbor Embedding plots.

Discretization showing transitions between attractors, 10 runs.

Why these particular transitions?

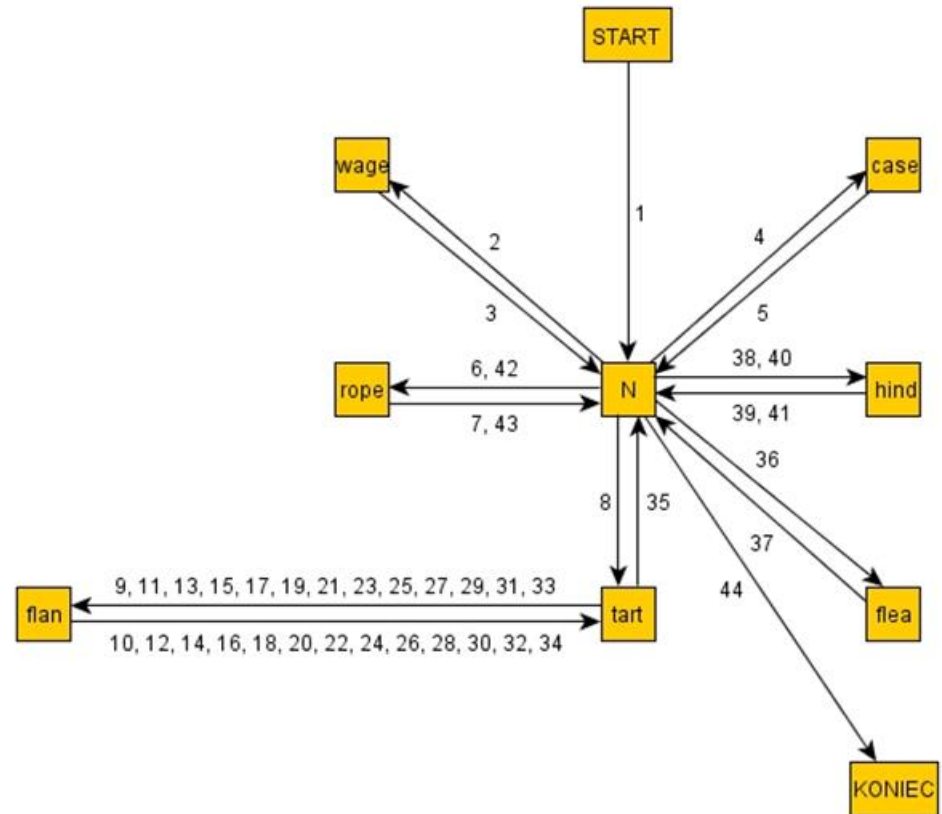
Connected attractors share some microfeatures, some are deactivated, but visualization using RP or FSD does not show such details. In the phase space dimensions are rescaled during dynamics.



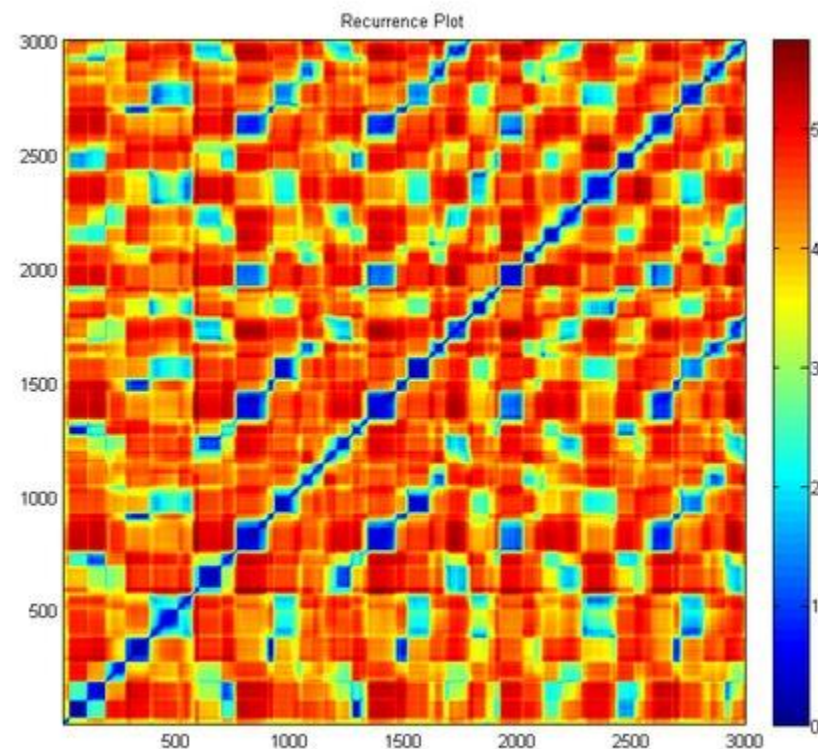
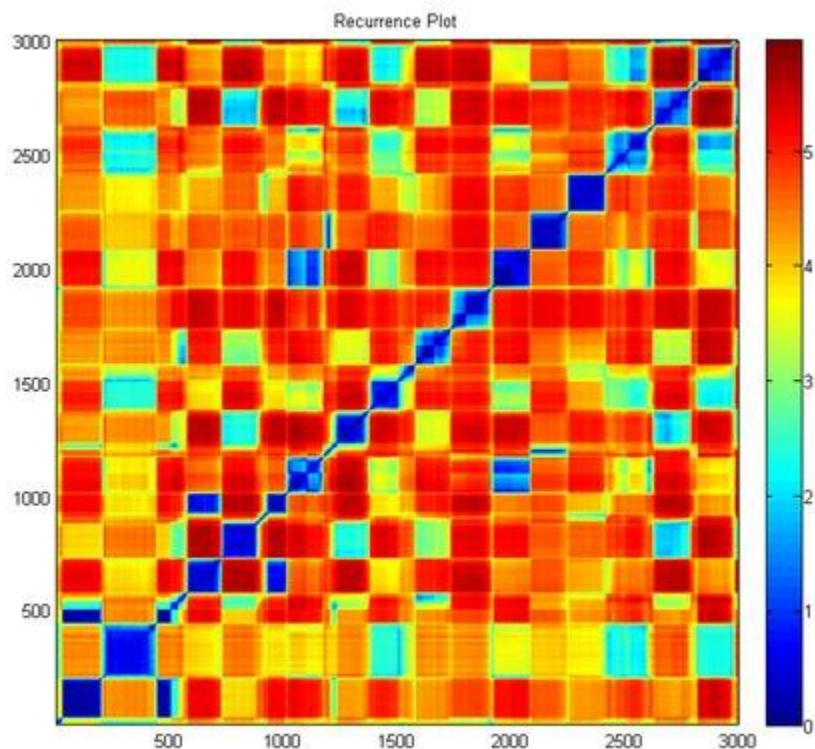
Transition graphs

Like in molecular dynamics, long time is needed to explore various potential transitions between attractor basins – depending on priming (previous dynamics or context) and noise in the system.

In some cases this model may get into obsessive kind of loop, like here, alternating between “tart” and “flan”.



RSVP simulations: HFA



Normal presentation: 500 it/word

Fast presentation: 100 it/word

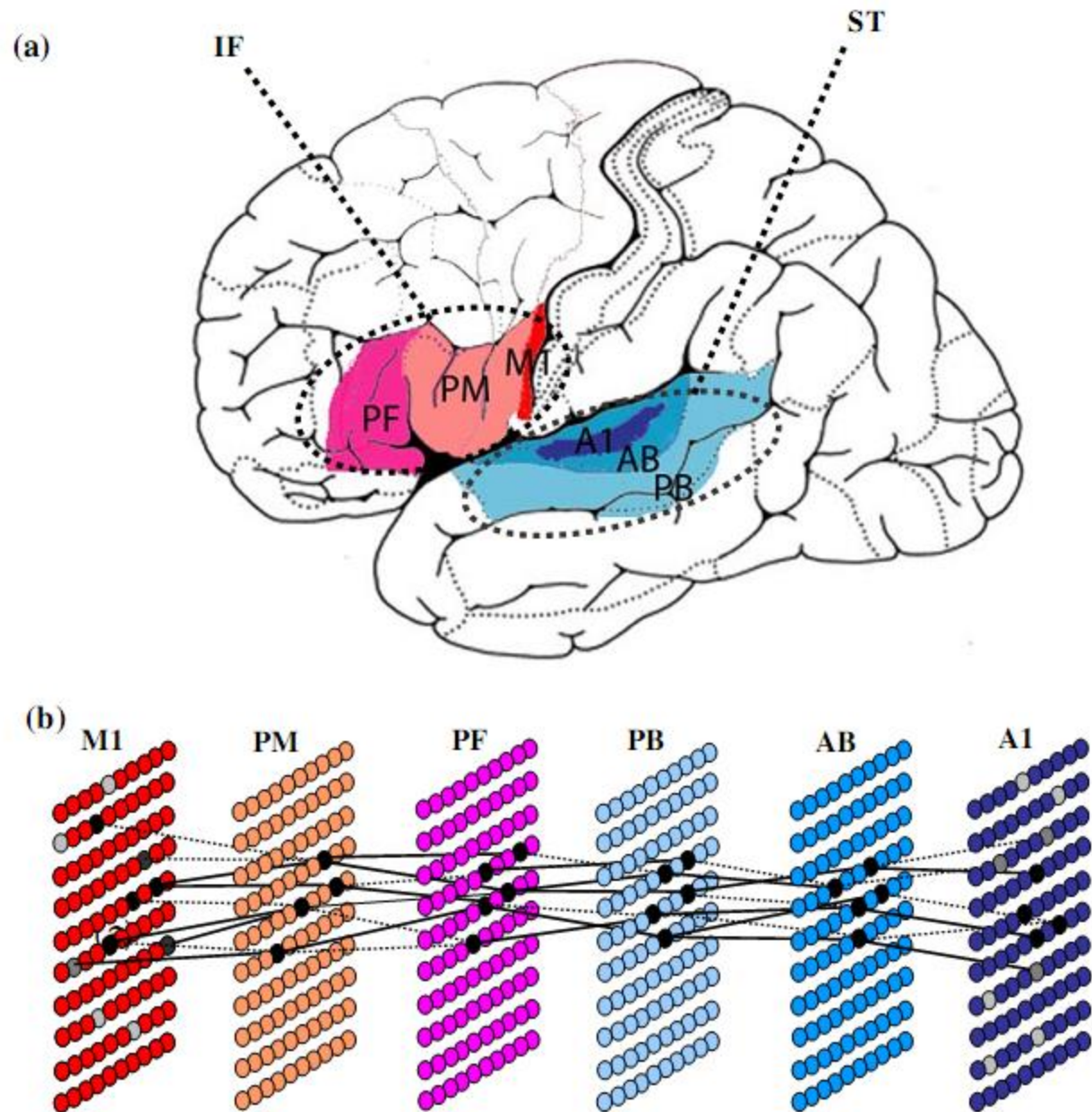
Difference between fast and slow resynchronization of brain networks.

High functioning ASD case (HFA) – brain activity returns to previous states, skips some stimuli during rapid serial visual presentation.

A better model

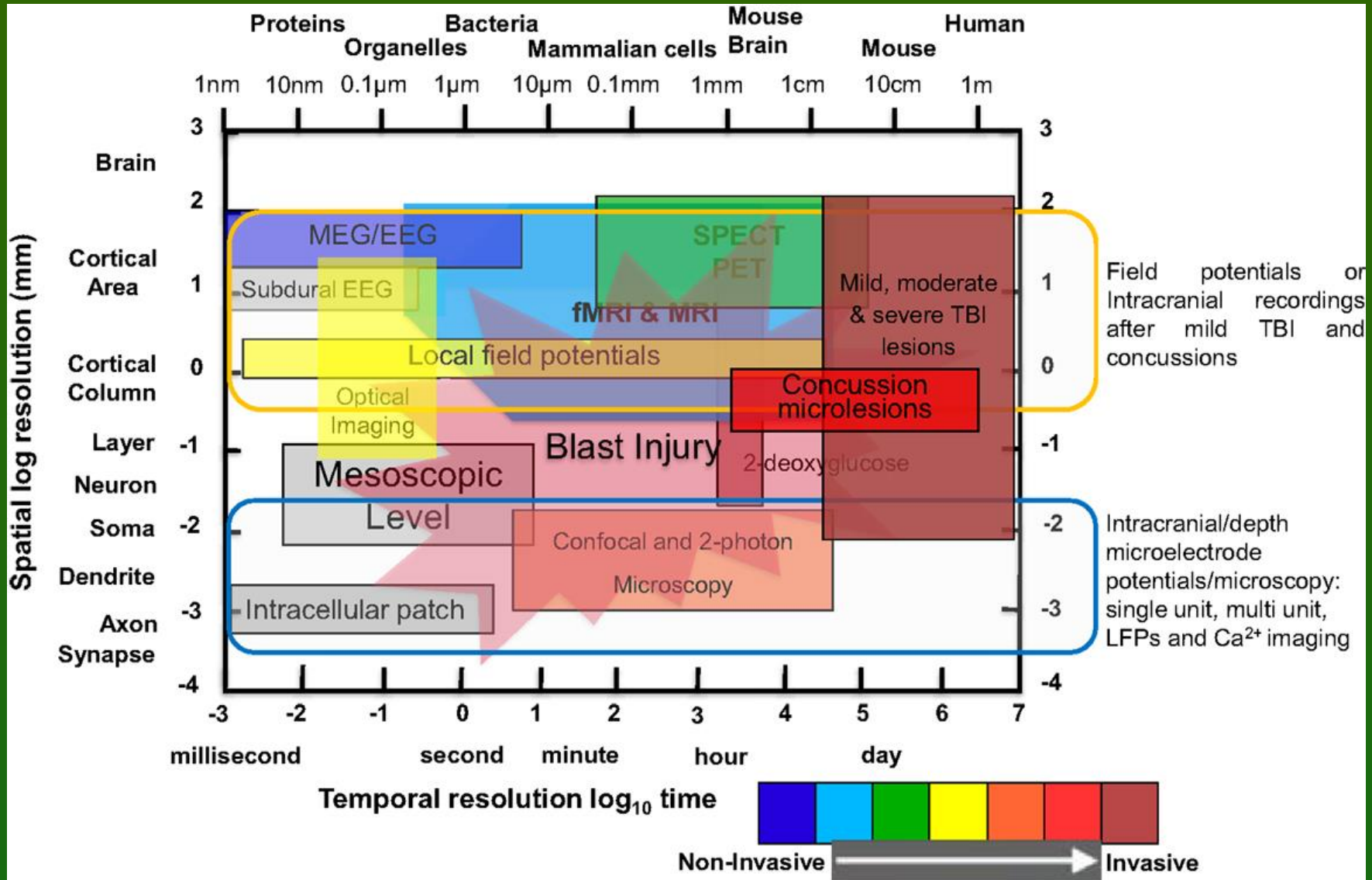
Garagnani et al.
Recruitment and consolidation of cell assemblies for words by way of Hebbian learning and competition in a multi-layer neural network, *Cognitive Comp.* 1(2), 160-176, 2009.

Primary auditory cortex (A1), auditory belt (AB), parabelt (PB, Wernicke's area), inferior pre-frontal (PF) and premotor (PM, Broca), primary motor cortex (M1).

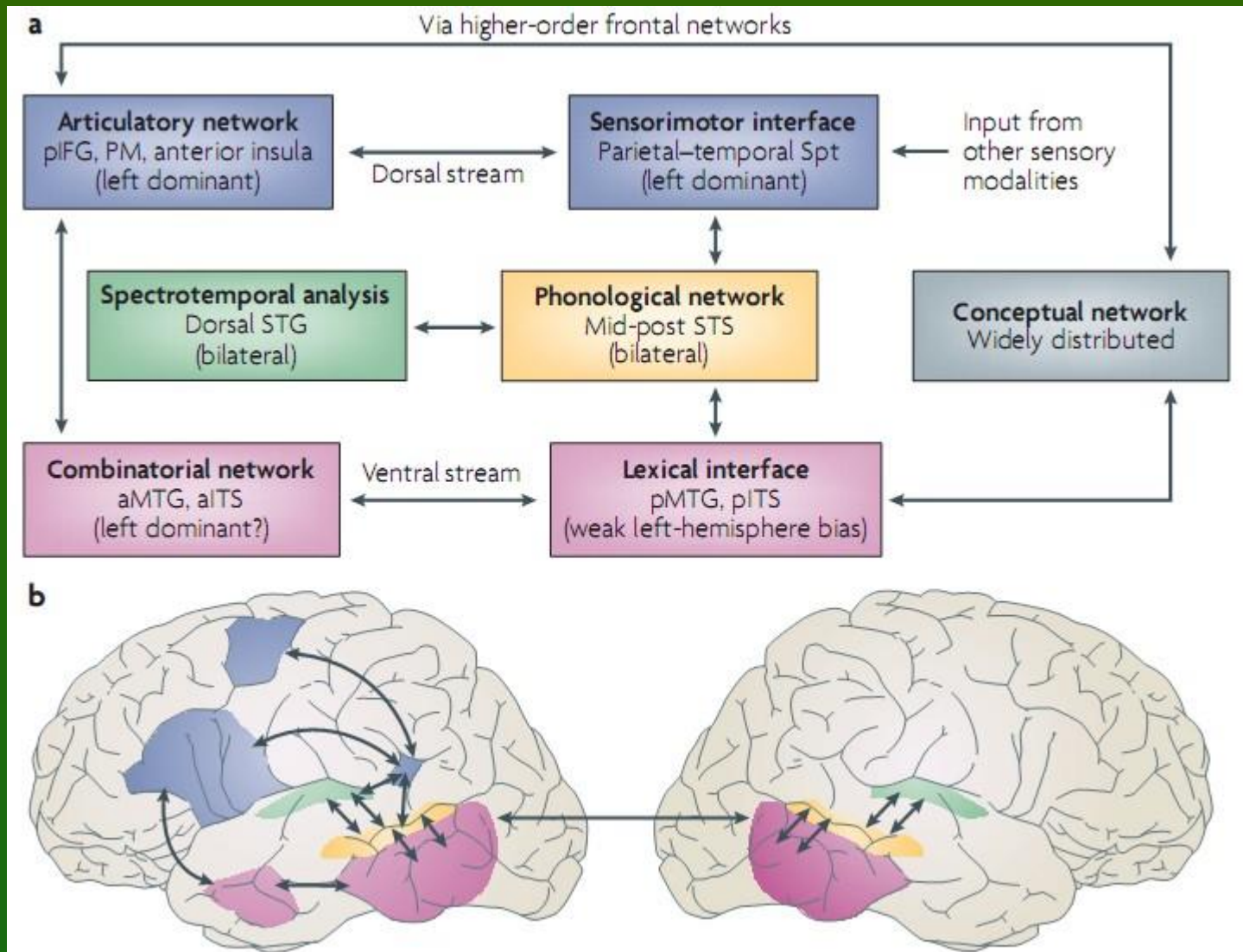


Brain networks:
neurolinguistics.

Neuroimaging techniques

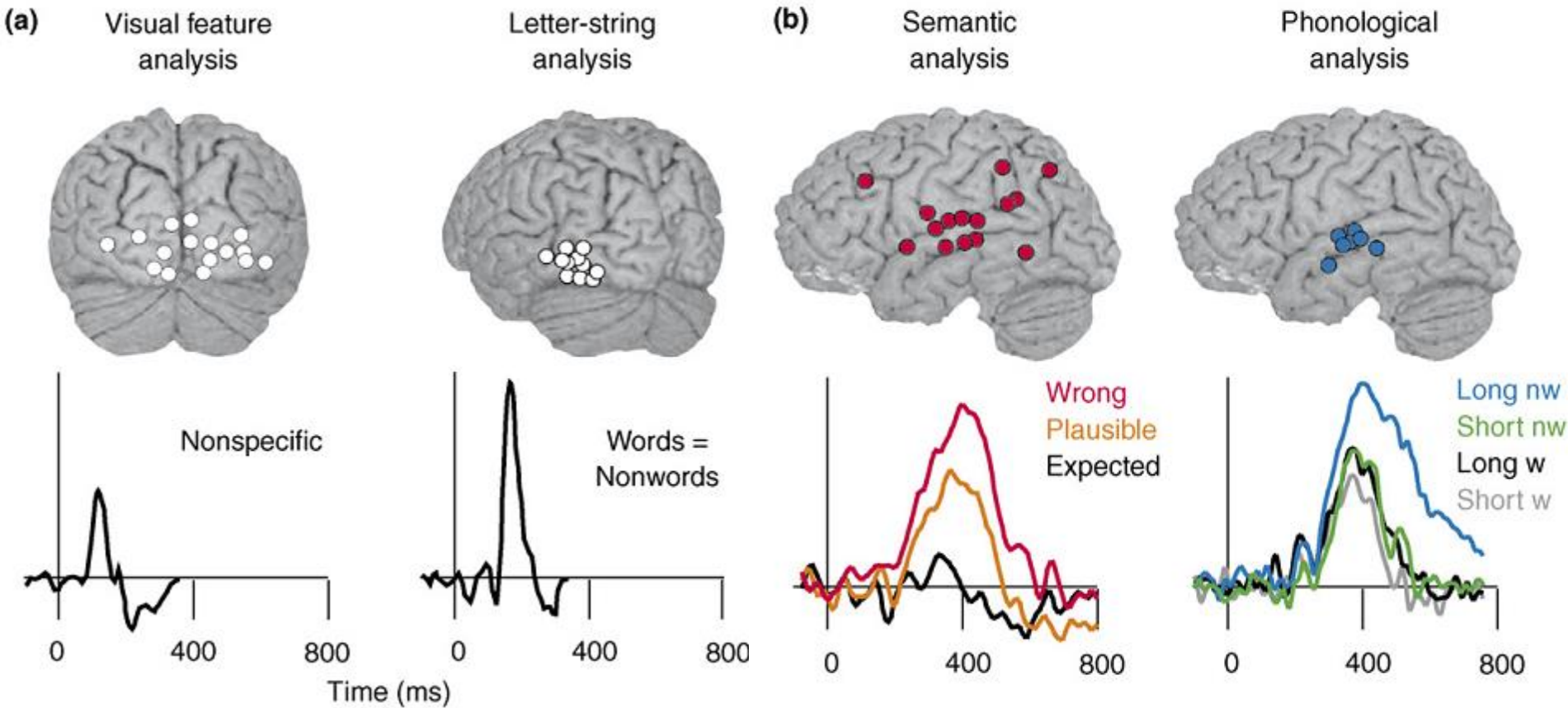


Speech in the brain



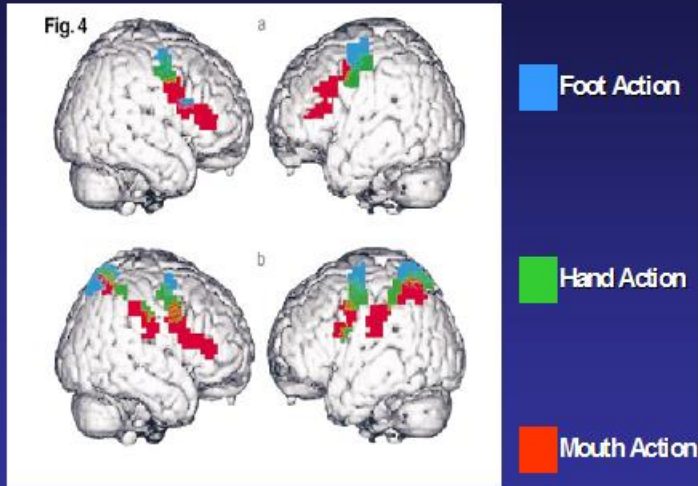
How should a concept meaning be represented?

Reading Brain



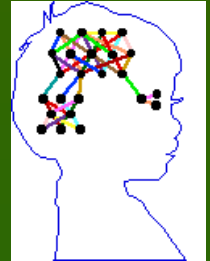
MEG activity patches for single word reading, time course of activations.
R. Salmelin, J. Kujala, Neural representation of language: activation versus long-range connectivity. TICS 10(11), 519-525, 2006

Somatotopy of Action Observation



Buccino et al. Eur J Neurosci 2001

s in the brain



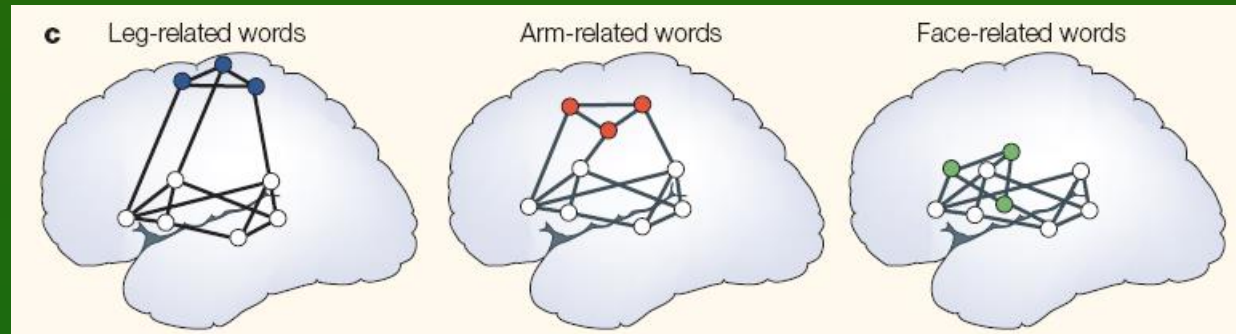
show that most likely categorical, are used, not the acoustic input.

> words => semantic concepts.

des semantic by 90 ms (from N200 ERPs).

uroscience of Language. On Brain Circuits of
bridge University Press.

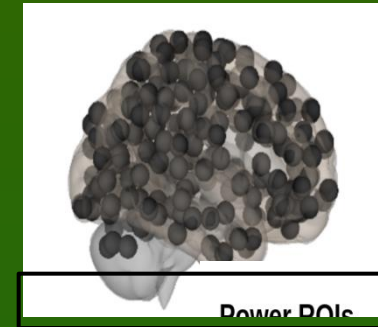
Action-perception
networks inferred
from ERP and fMRI



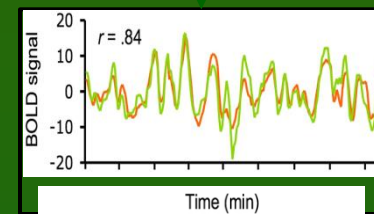
Left hemisphere: precise representations of symbols, including phonological components; right hemisphere? Sees clusters of concepts.

Human connectome and MRI/fMRI

Node definition (parcelation)



Signal extraction

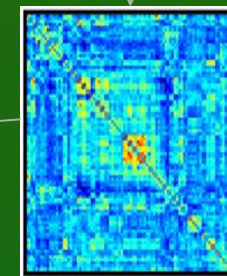


Correlation calculation

Binary matrix



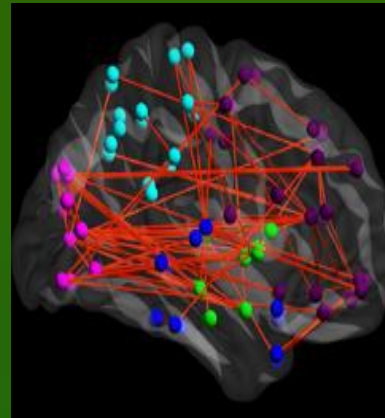
Correlation matrix



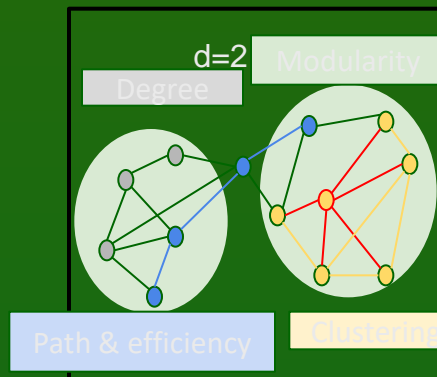
Structural connectivity



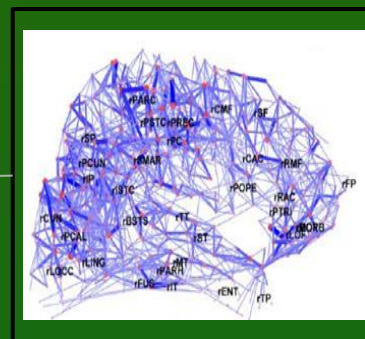
Functional connectivity



Graph theory



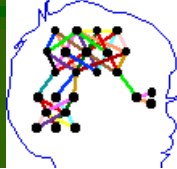
Whole-brain graph



Many toolboxes are available for such analysis.

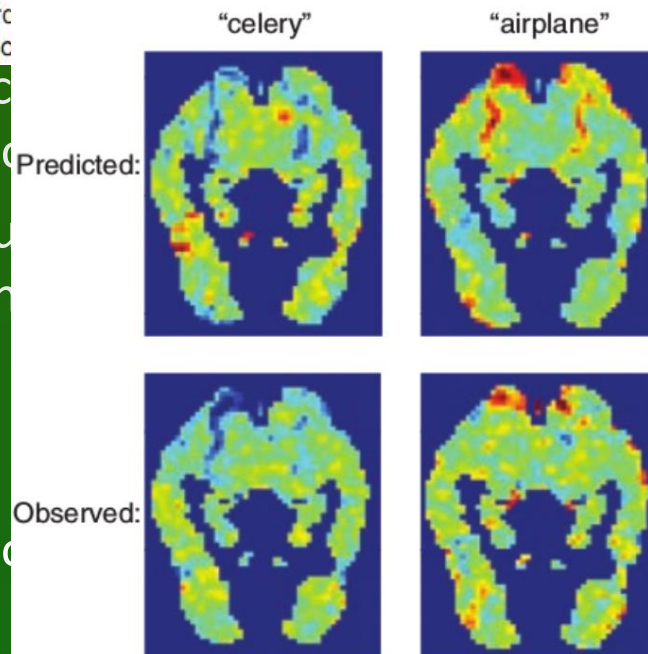
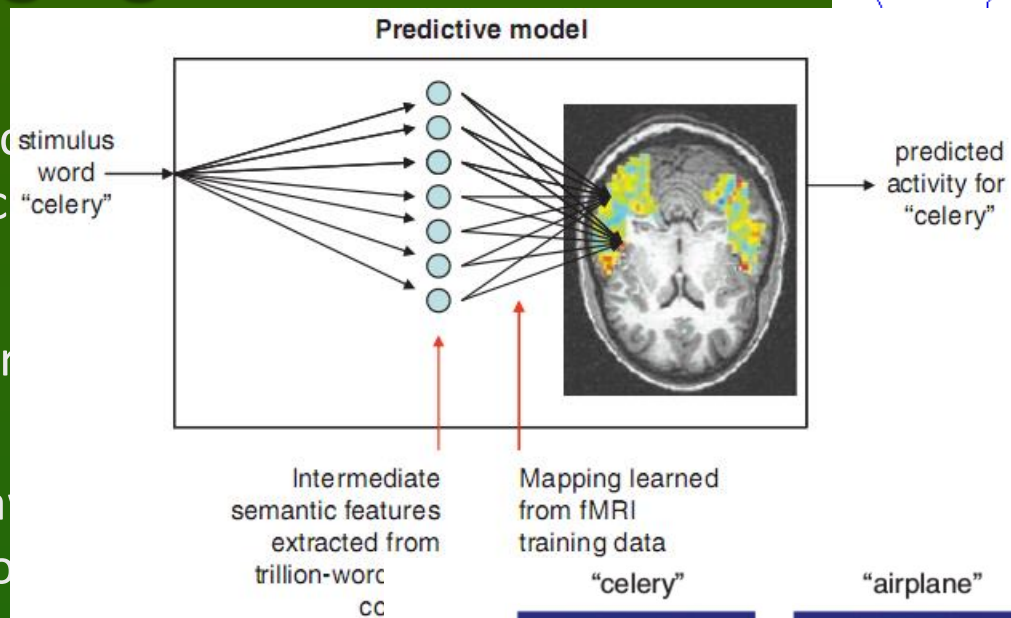
Bullmore & Sporns (2009)

Neuroimaging words



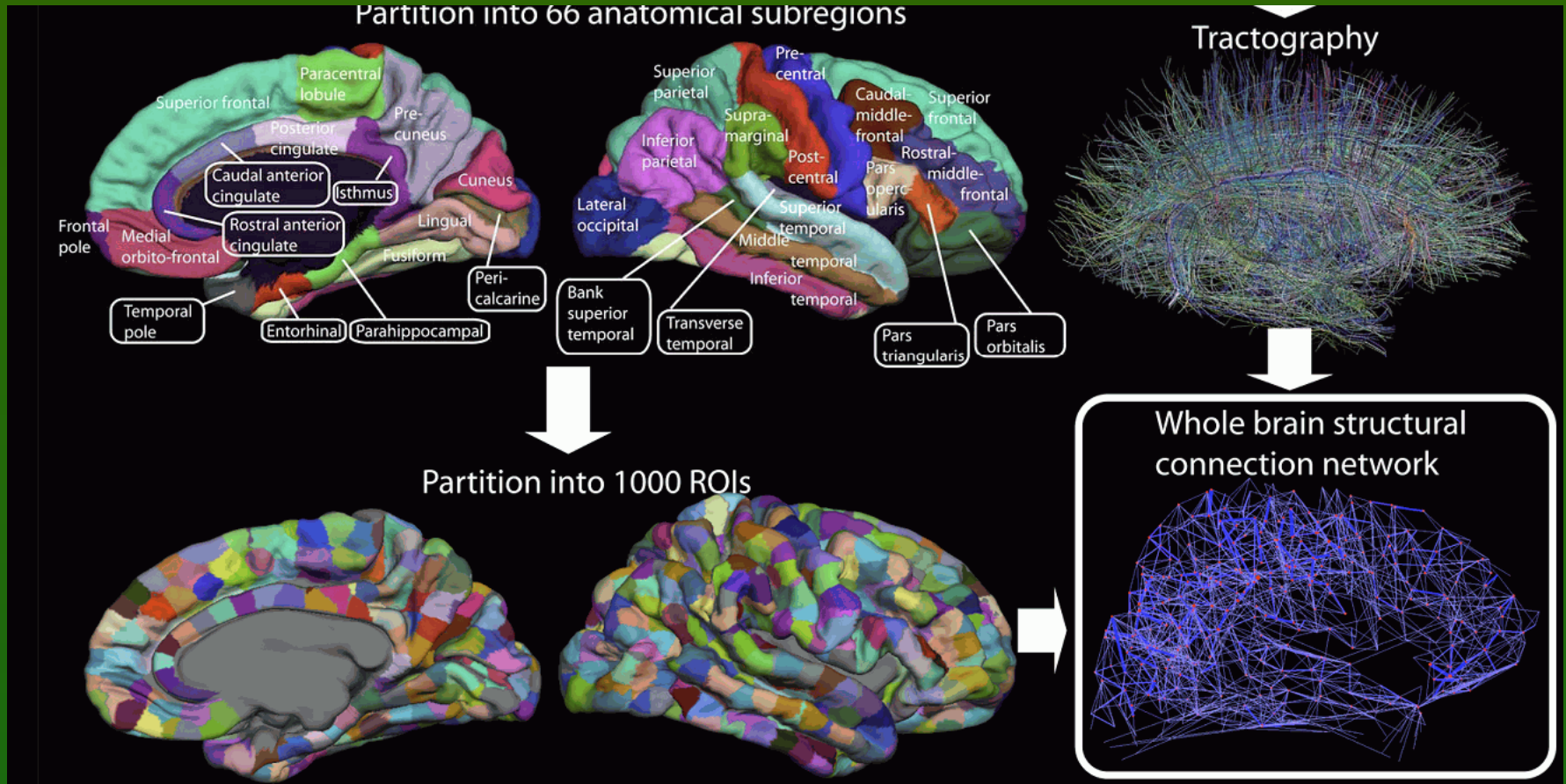
Predicting Human Brain Activity Associated with Words: "Predicting Human Brain Activity Associated with Words: A Case of Nouns," T. M. Mitchell et al, Science

- Clear differences between fMRI brain activity patterns for different nouns.
- Reading words and seeing the drawing of the word, presumably reflecting semantics of the word.
- Although individual variance is significant across subjects, a classifier trained on data of different people, a classifier may still be trained to predict brain activity for a given word.
- Model trained on ~10 fMRI scans + very large corpus of text to predict brain activity for over 100 nouns for which fMRI has been recorded.

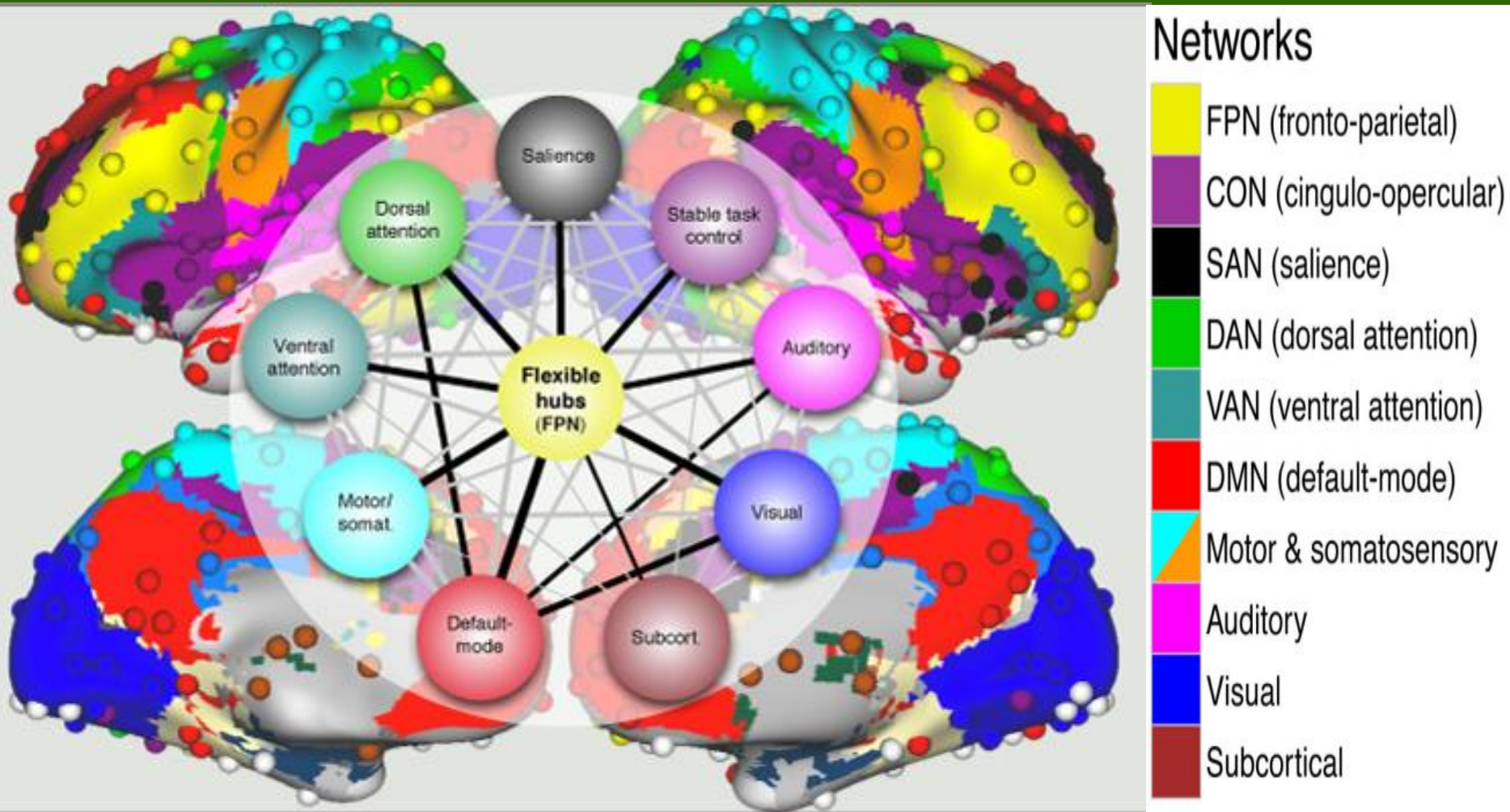


Sensory: fear, hear, listen, see, smell, taste, touch
Motor: eat, lift, manipulate, move, push, rub, run, say
Abstract: approach, break, clean, drive, enter, fill, near, of
Are these 25 features defining brain-based semantics?

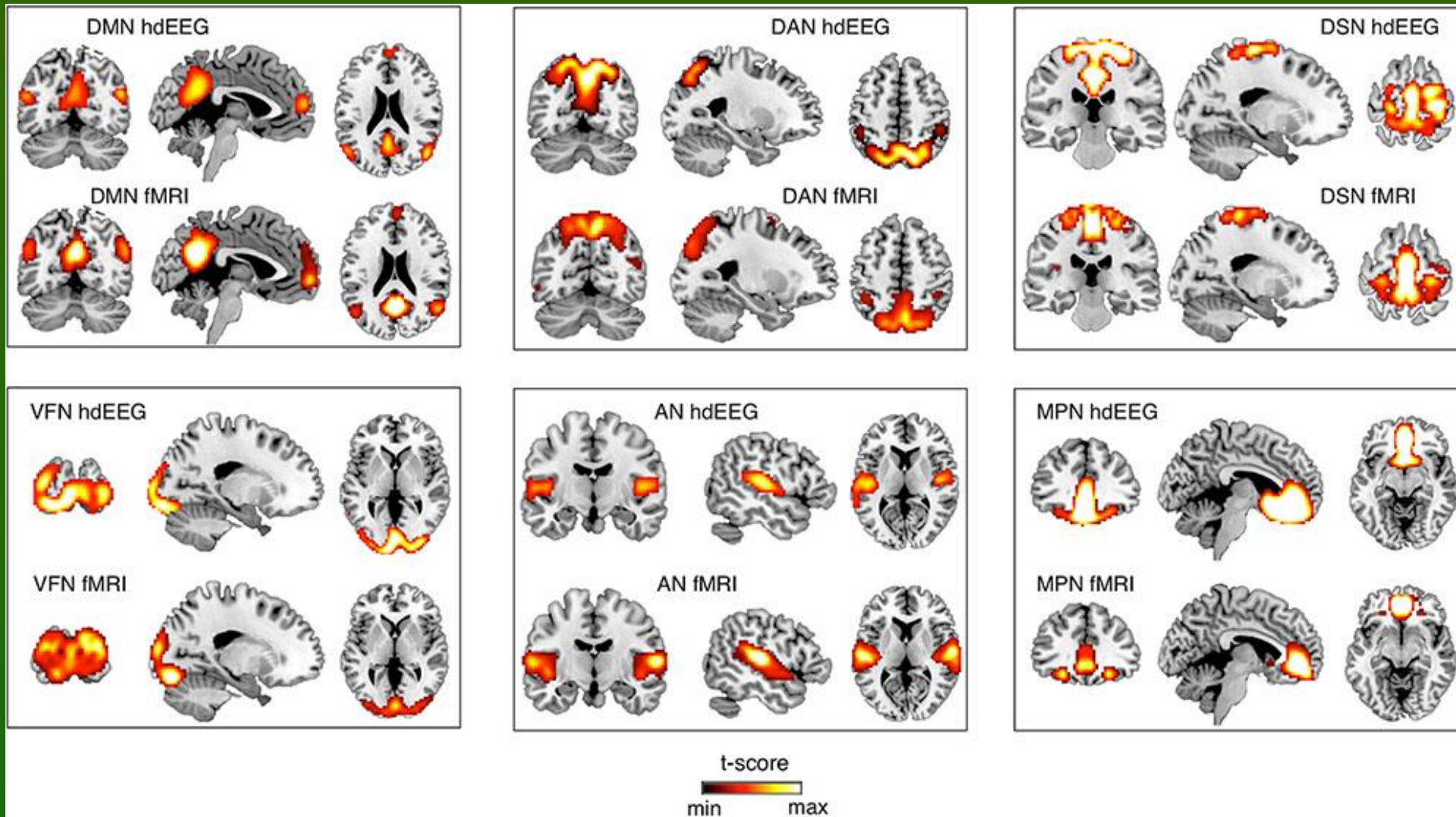
Connectome



Neurocognitive Basis of Cognitive Control



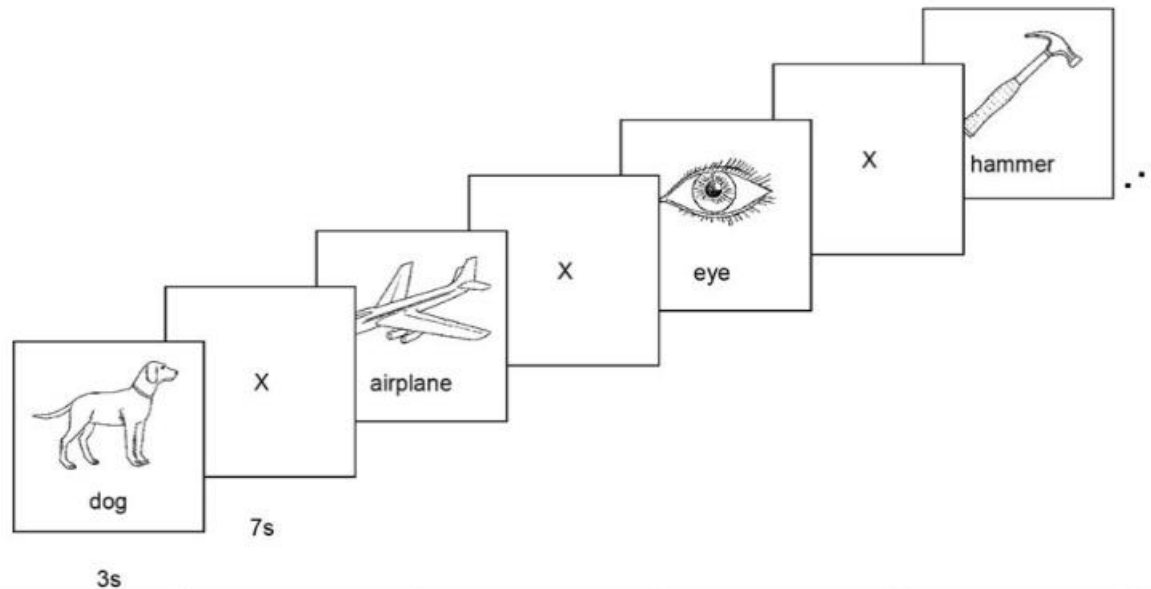
Central role of fronto-parietal (FPN) flexible hubs in cognitive control and adaptive implementation of task demands (black lines=correlations significantly above network average). Cole et al. (2013).



sICA on 10-min fMRI data ($N = 24$, threshold: $p < 0.01$, TFCE corrected). DMN, default mode network; DAN, dorsal attention network; DSN, dorsal somatomotor network; VFN, visual foveal network; AN, auditory network; MPN, medial prefrontal network.

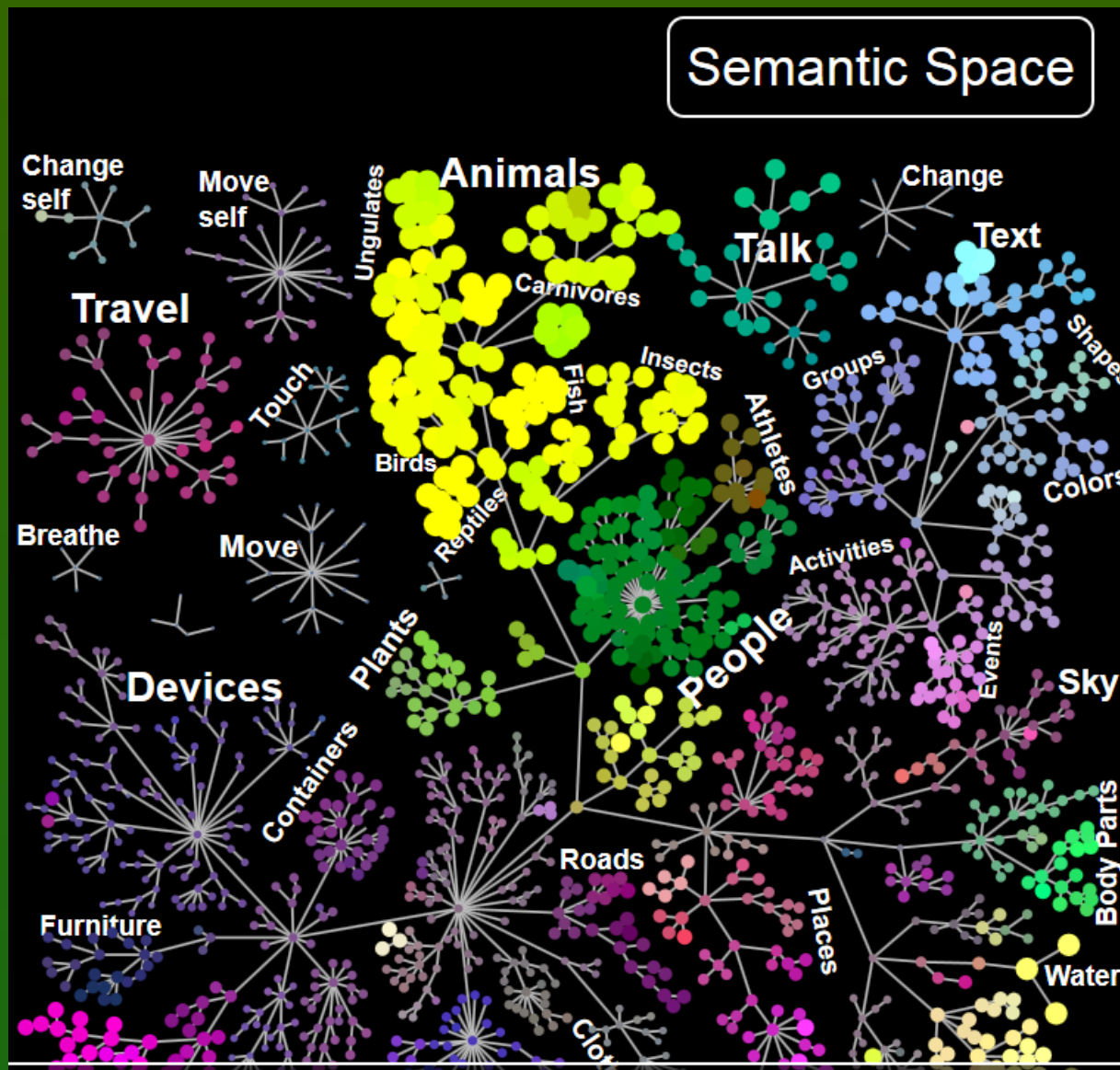
Quasi-stable brain activations?

Maintain brain activation for longer time. Use pictures, video, sounds ...



Category	Exemplar 1	Exemplar 2	Exemplar 3	Exemplar 4	Exemplar 5
animals	bear	cat	cow	dog	horse
body parts	arm	eye	foot	hand	leg
buildings	apartment	barn	church	house	igloo

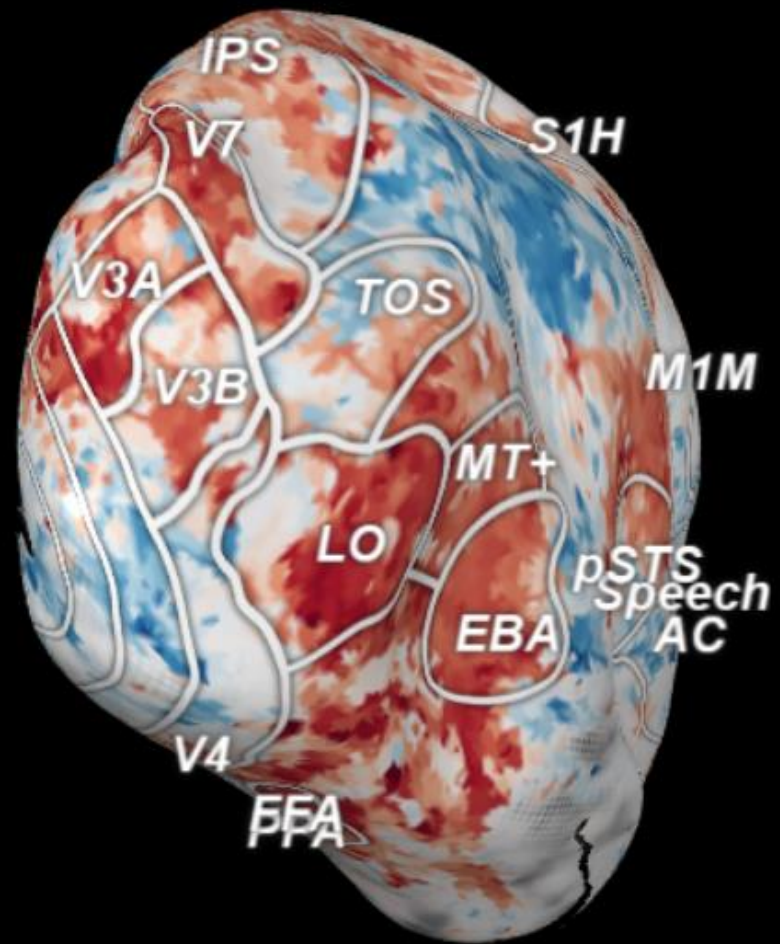
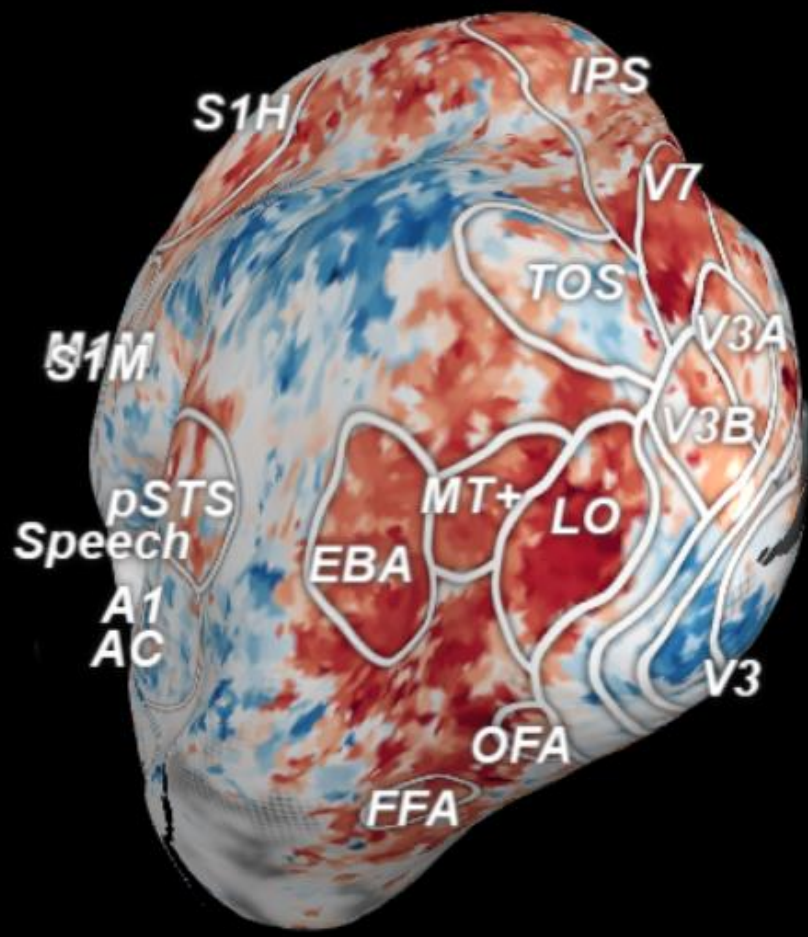
Can we induce stable cortical activation? Locate sources in similar areas as BOLD? Interpret brain activations in terms of brain-based semantics?



Words in the semantic space are grouped by their similarity (Gallant Lab, 2016). Words activate specific ROIs, similar words create similar maps of brain activity. Each voxel may be activated by many words. Video or audio stimuli, fMRI scans.

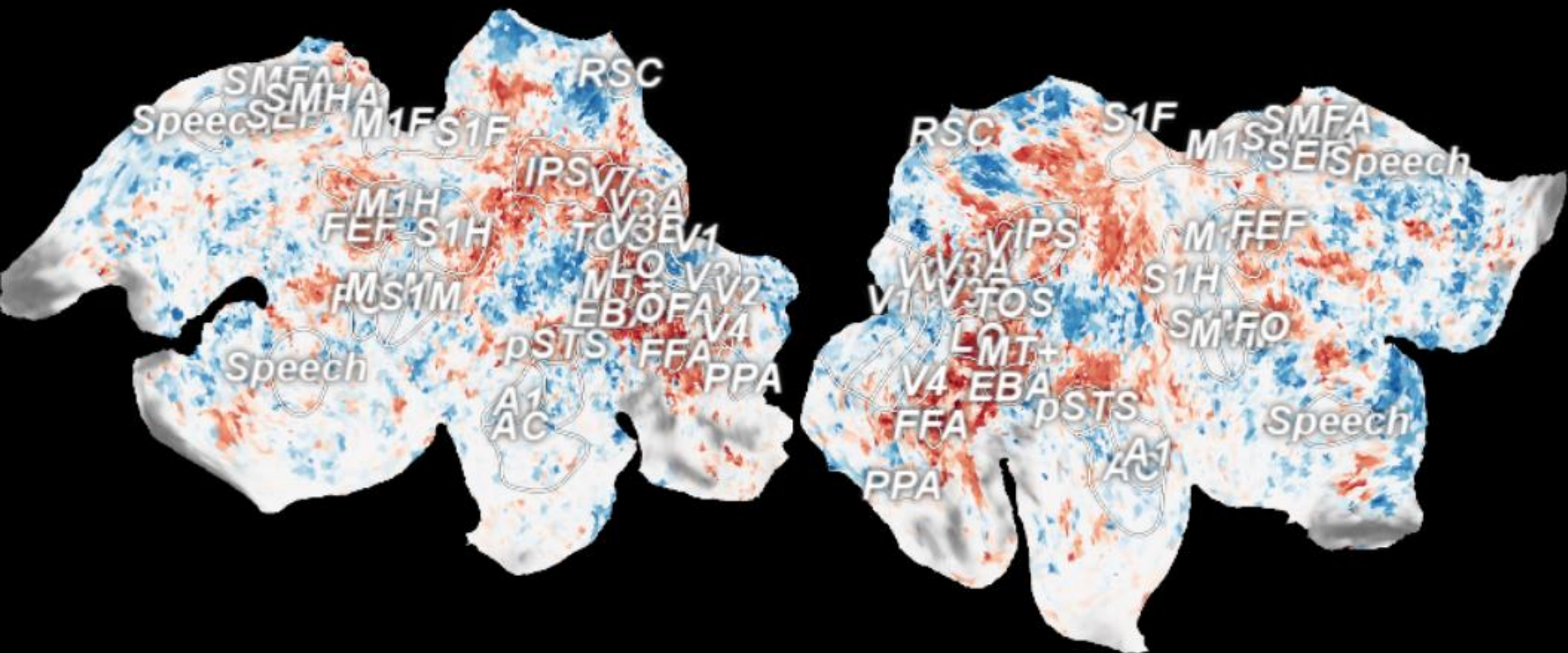


Category zebra: Passive Viewing

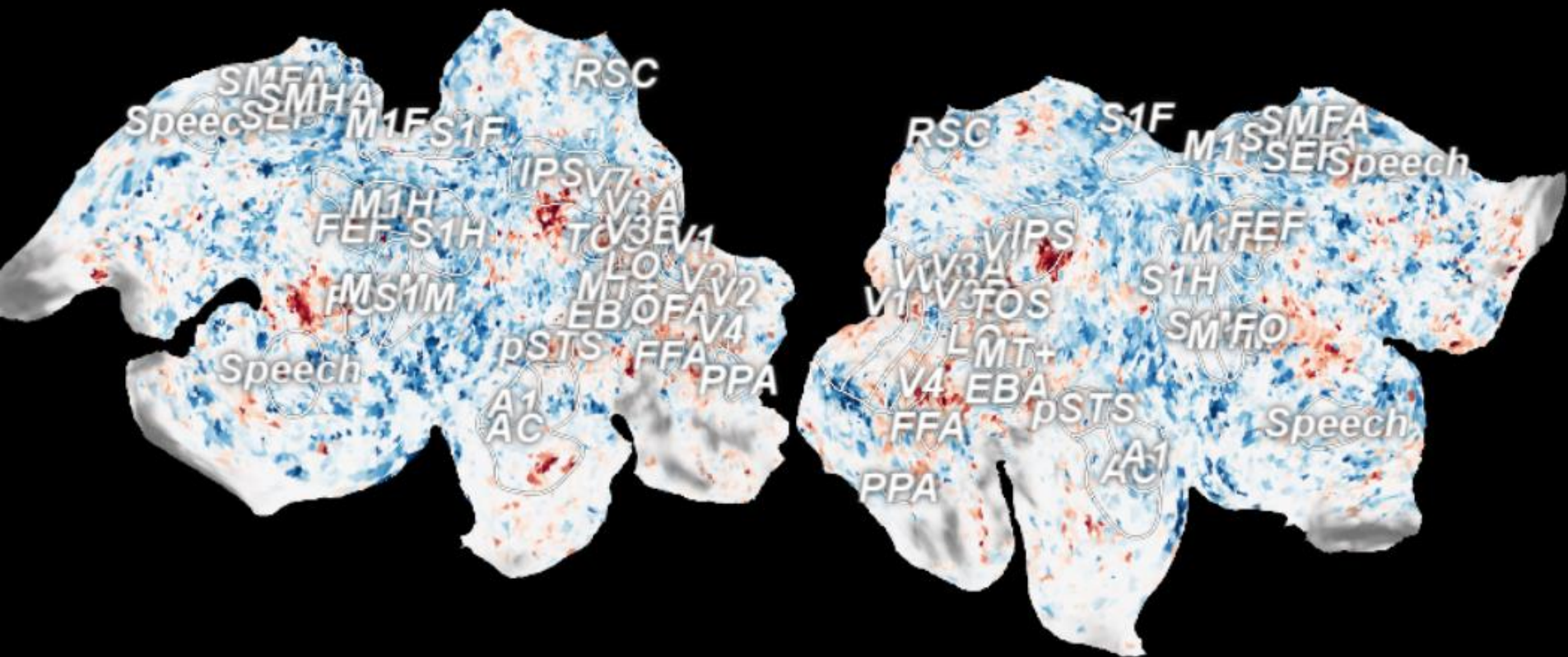


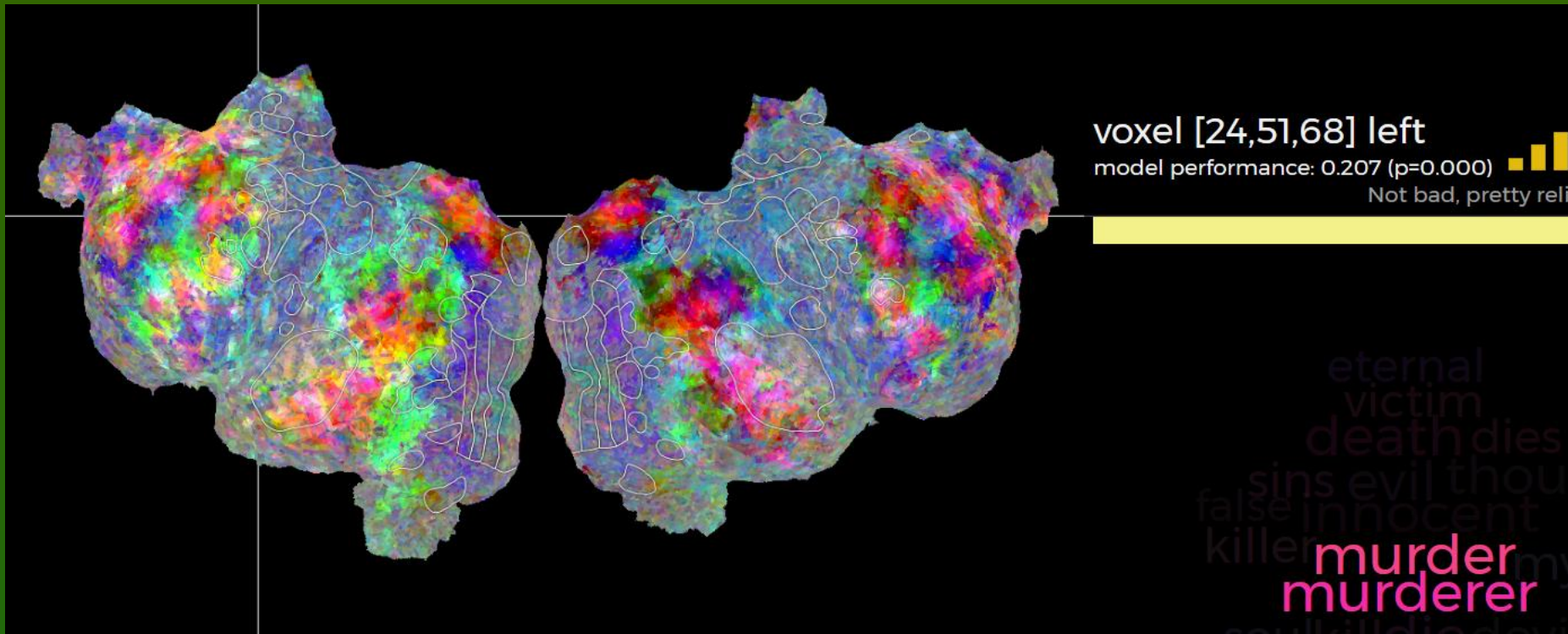


Category zebra: Passive Viewing



Category traffic light: Passive Viewing





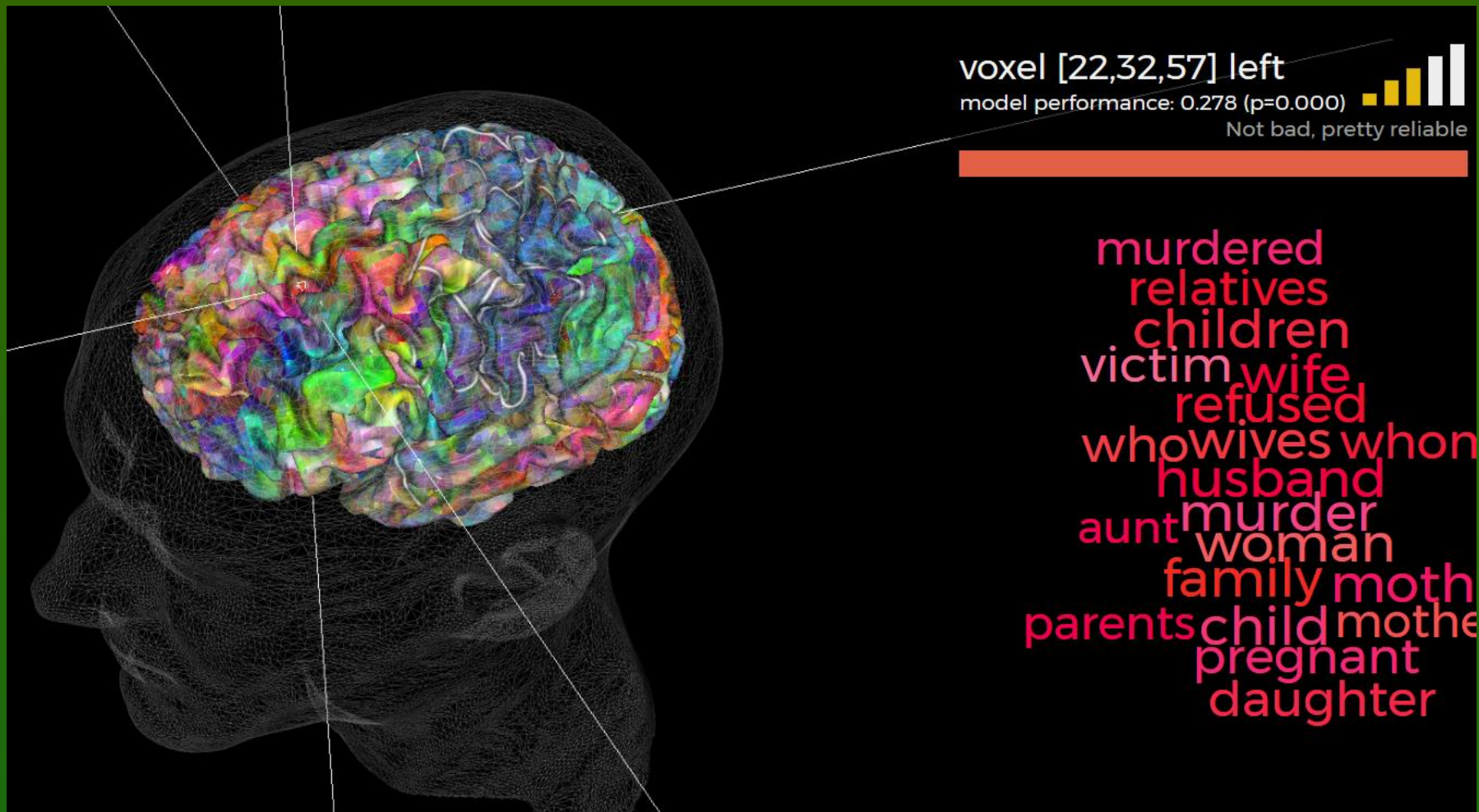
Whole fMRI activity map for the word “murder” shown on the flattened cortex.

Each word activates a whole map of activity in the brain, depending on sensory features, motor actions and affective components associated with this word.

Why such activity patterns arise? Brain subnetworks connect active areas.

<http://gallantlab.org/huth2016/> and [short movie intro](#).

Can one do something like that with EEG or MEG?



Each voxel responds usually to many related words, whole categories.

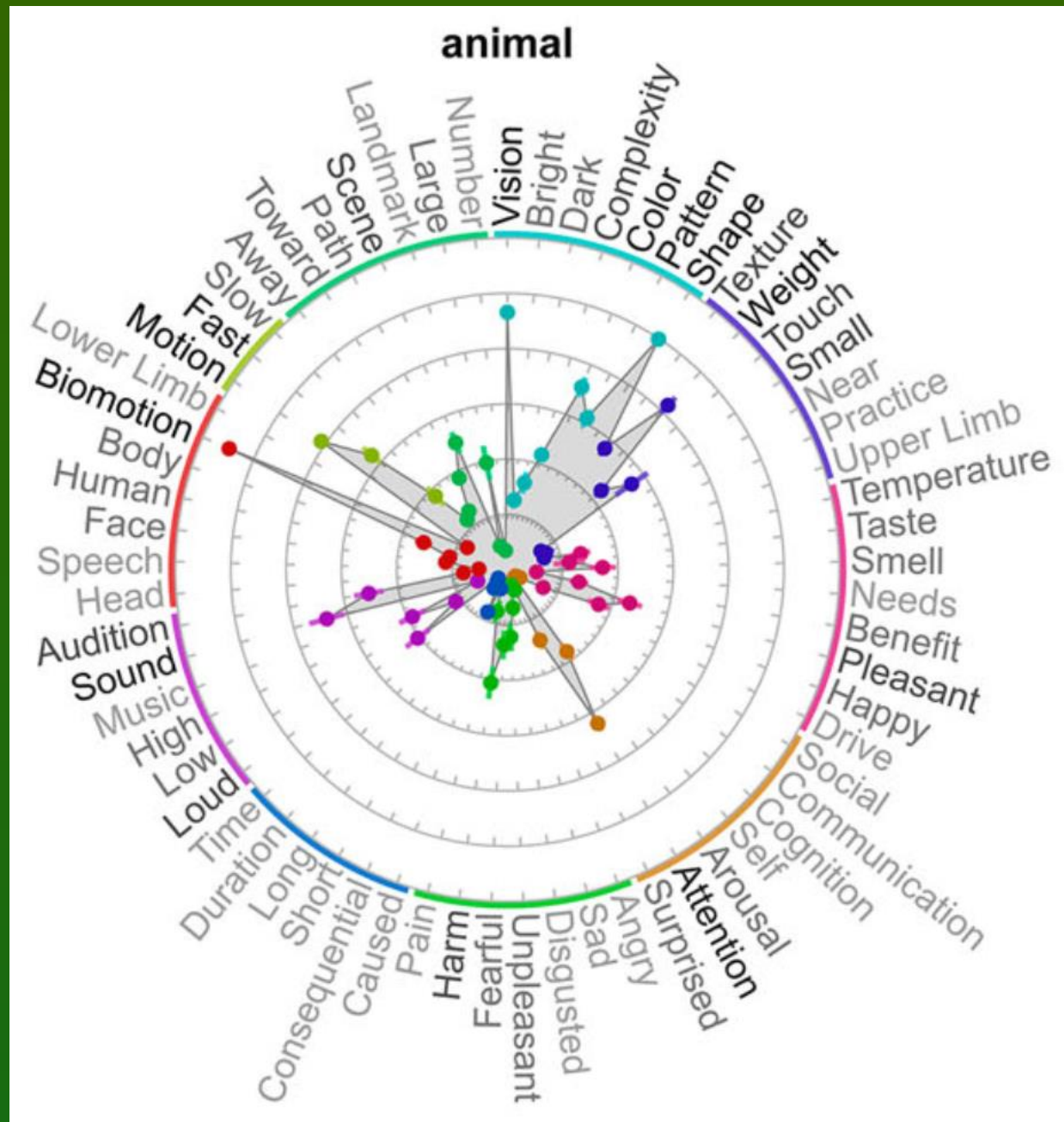
<http://gallantlab.org/huth2016/>

Huth et al. (2016). Decoding the Semantic Content of Natural Movies from Human Brain Activity. *Frontiers in Systems Neuroscience* 10, pp. 81

65 attributes related to neural processes;
Colors on circle: general domains.

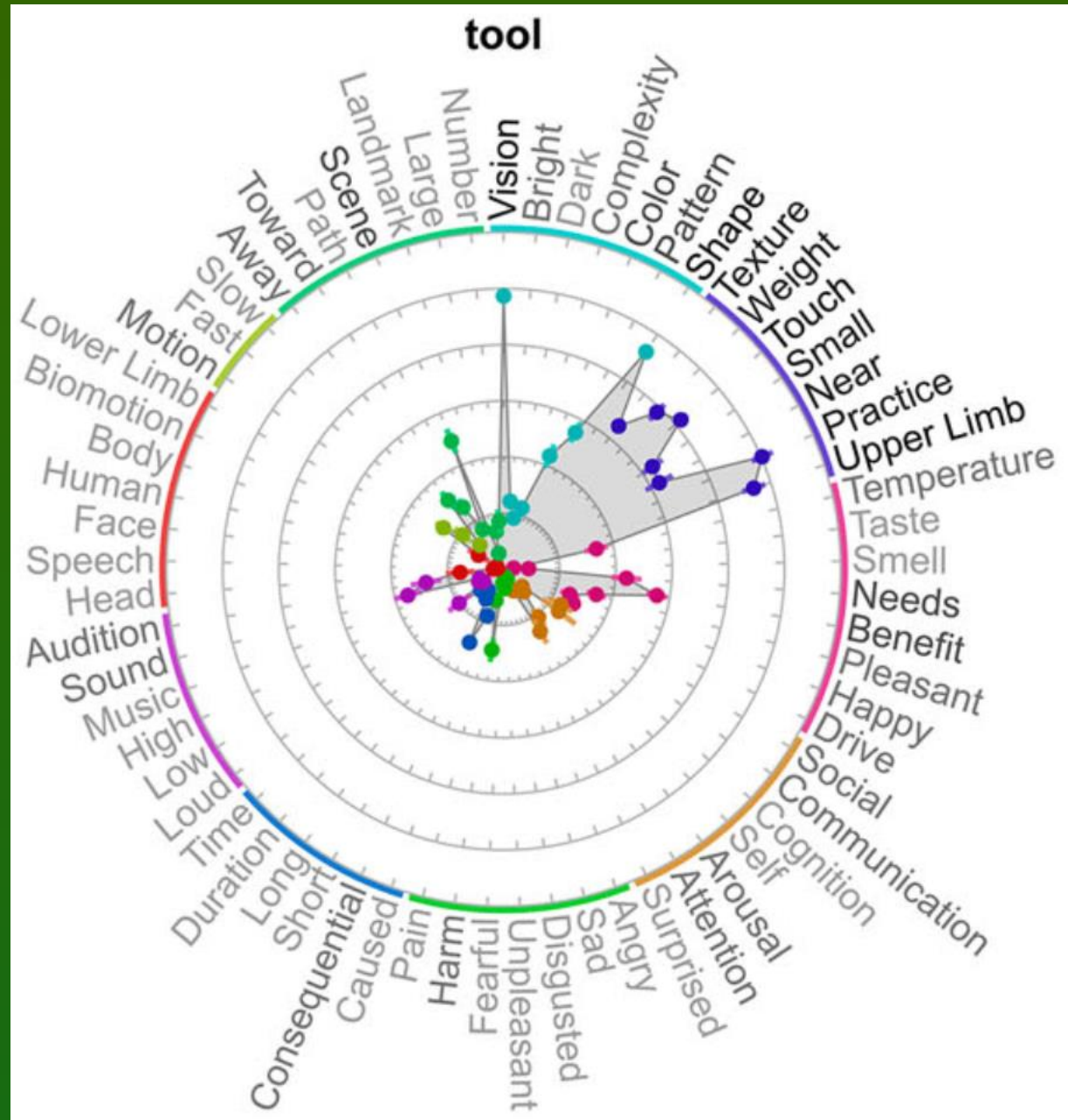
J.R. Binder et al
Toward a Brain-Based
Componential Semantic
Representation, 2016

More than just
visual objects!



65 attributes related to neural processes.
Brain-Based
Representation of tools.

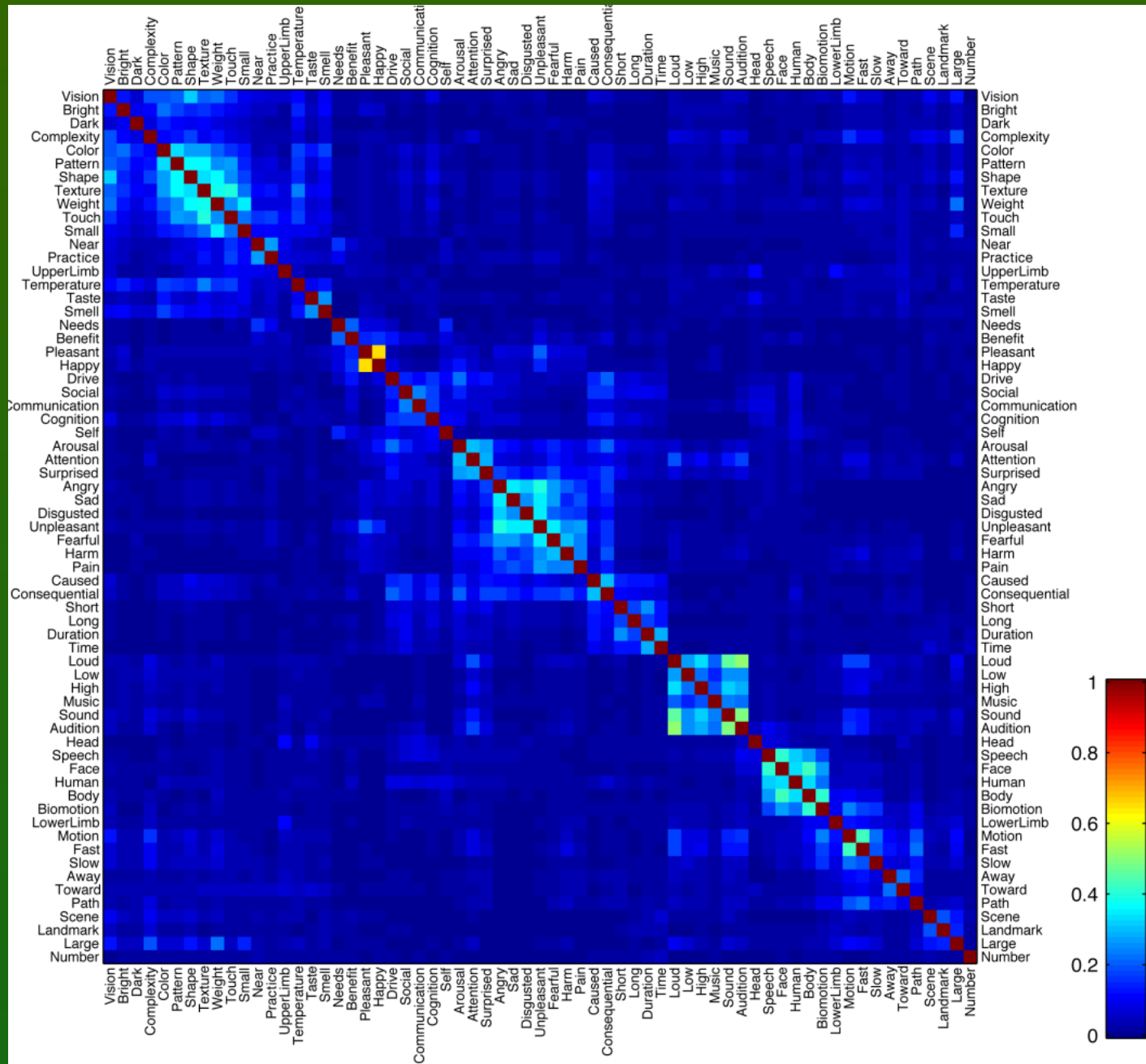
J.R. Binder et al
Toward a Brain-Based
Componential Semantic
Representation
Cognitive
Neuropsychology
2016

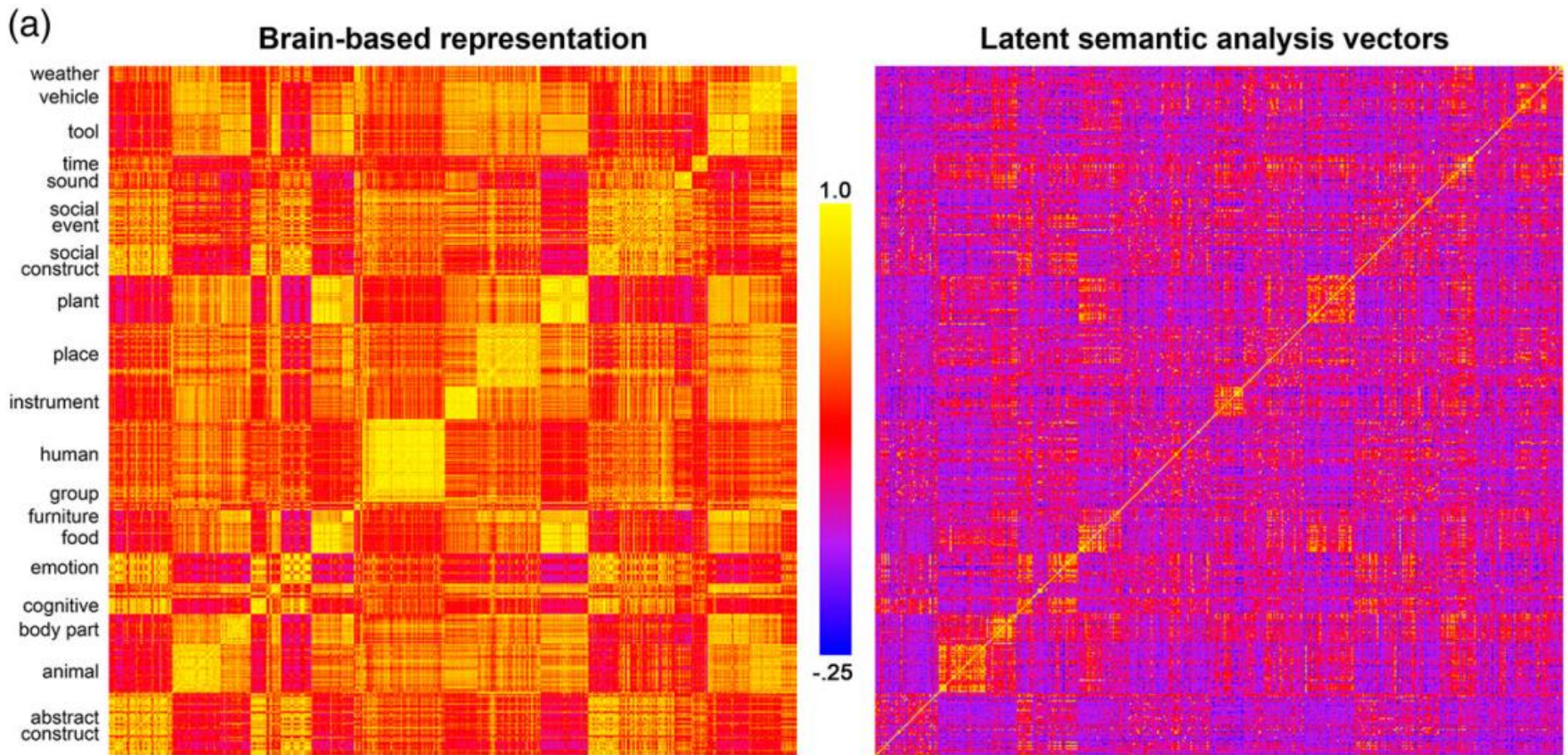


Mutual Information Matrix -
unique BBR, with
low redundancy

65 BBR attributes related to
neural processes.
Spanning the
space in which
concepts may be
represented.

J.R. Binder et al
2016





Cosine similarities, 434 nouns grouped by superordinate category.

Left: brain-based vectors, right latent semantic analysis vectors from large corpus (typical NLP). Yellow = greater similarity.

Similarities within categories are much stronger for BBR.

Wang, S., Zhang, J., Lin, N., & Zong, C. (2017). Investigating Inner Properties of **Multimodal Representation** and Semantic Compositionality with Brain-based Componential Semantics. (no brain signals, just NLP).

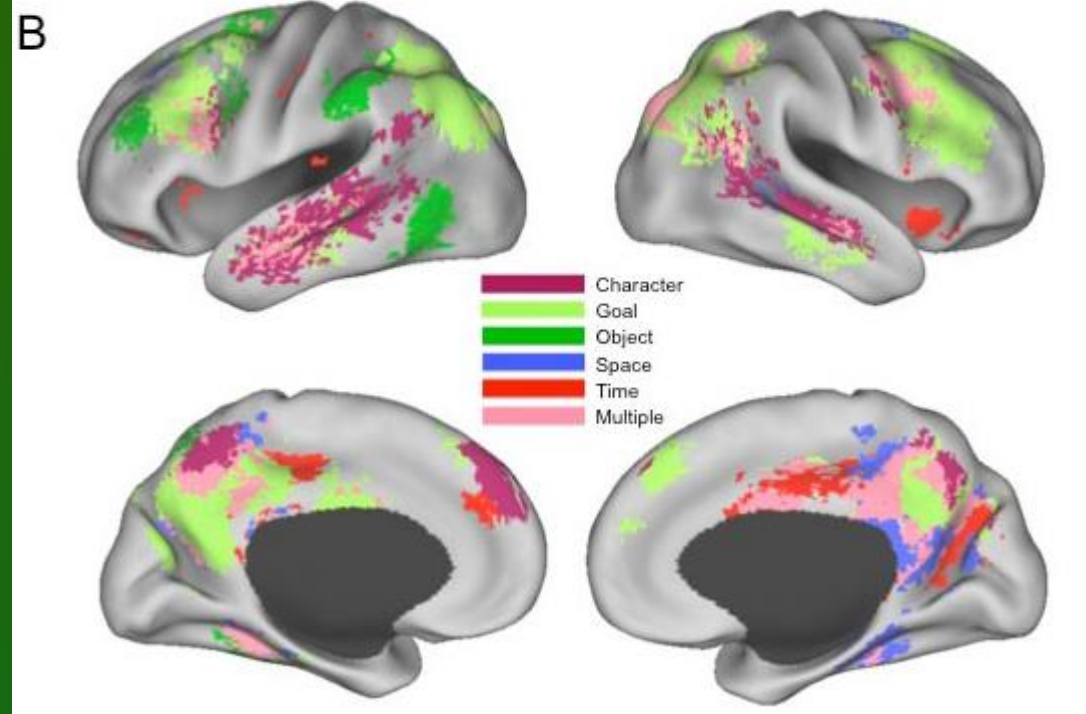
Understanding Brain Activity Near Future

Nicole Speer et al.
 Reading Stories Activates
 Neural Representations of
 Visual and Motor
 Experiences.
Psychological Science
 (2010, in print).

Meaning: always slightly
 different, depending on the
 context, but still may be
 clusterized into relatively
 small number of distinct
 meanings.

A

Clause	Cause	Character	Goal	Object	Space	Time
...[Mrs. Birch] went through the front door into the kitchen.	●				●	
Mr. Birch came in	●	●			●	
and, after a friendly greeting,	●					●
chatted with her for a minute or so.	●					●
Mrs. Birch needed to awaken Raymond.		●				
Mrs. Birch stepped into Raymond's bedroom,			●		●	
pulled a light cord hanging from the center of the room,				●		
and turned to the bed.						
Mrs. Birch said with pleasant casualness, "Raymond, wake up."						
With a little more urgency in her voice she spoke again:						
Son, are you going to school today?						
Raymond didn't respond immediately.		●				●
He screwed up his face			●			
And whimpered a little.						



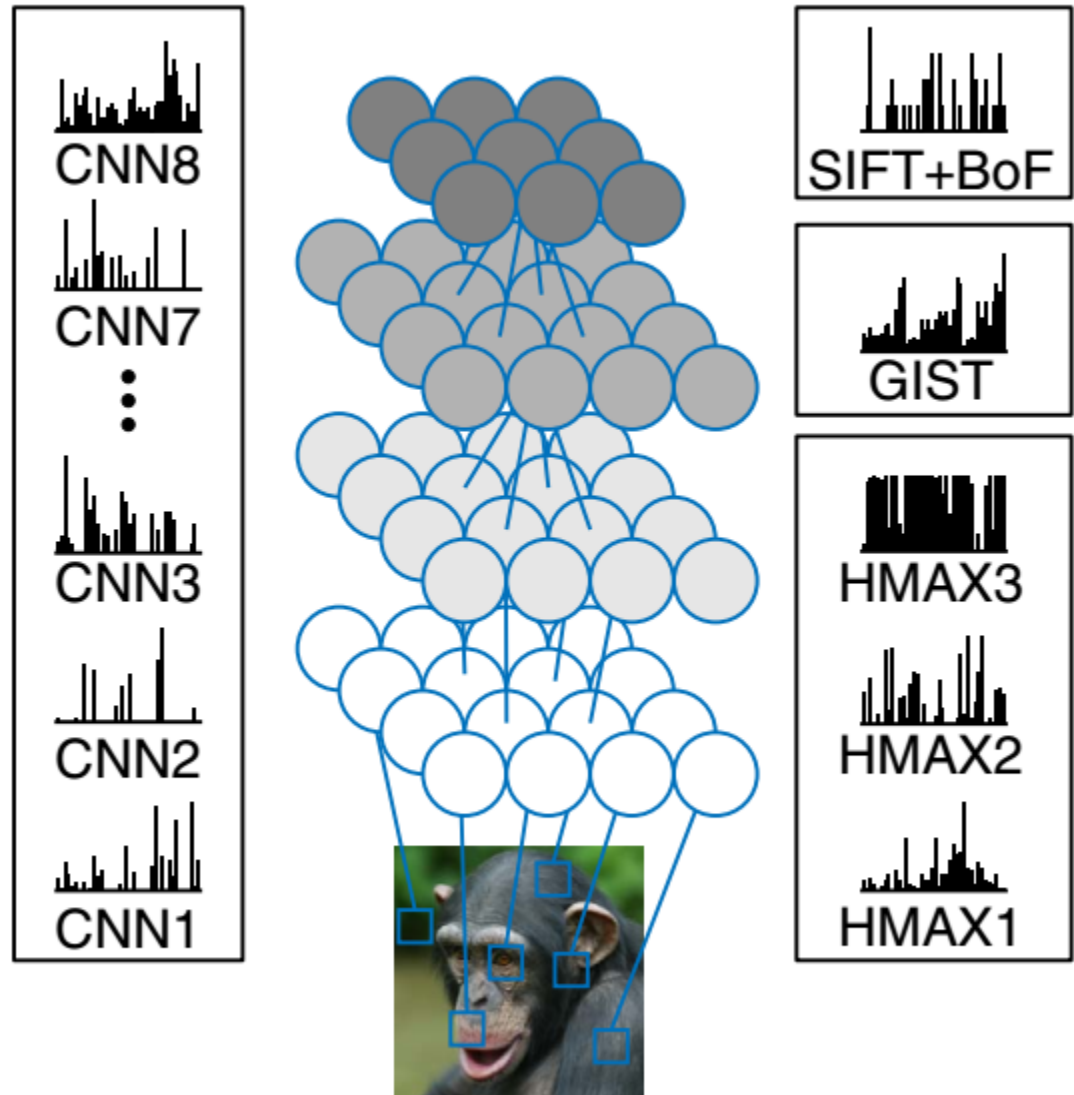
Mental images from brain activity

Can we convert activity of the brain into the mental images that we are conscious of?

Try to estimate features at different layers.

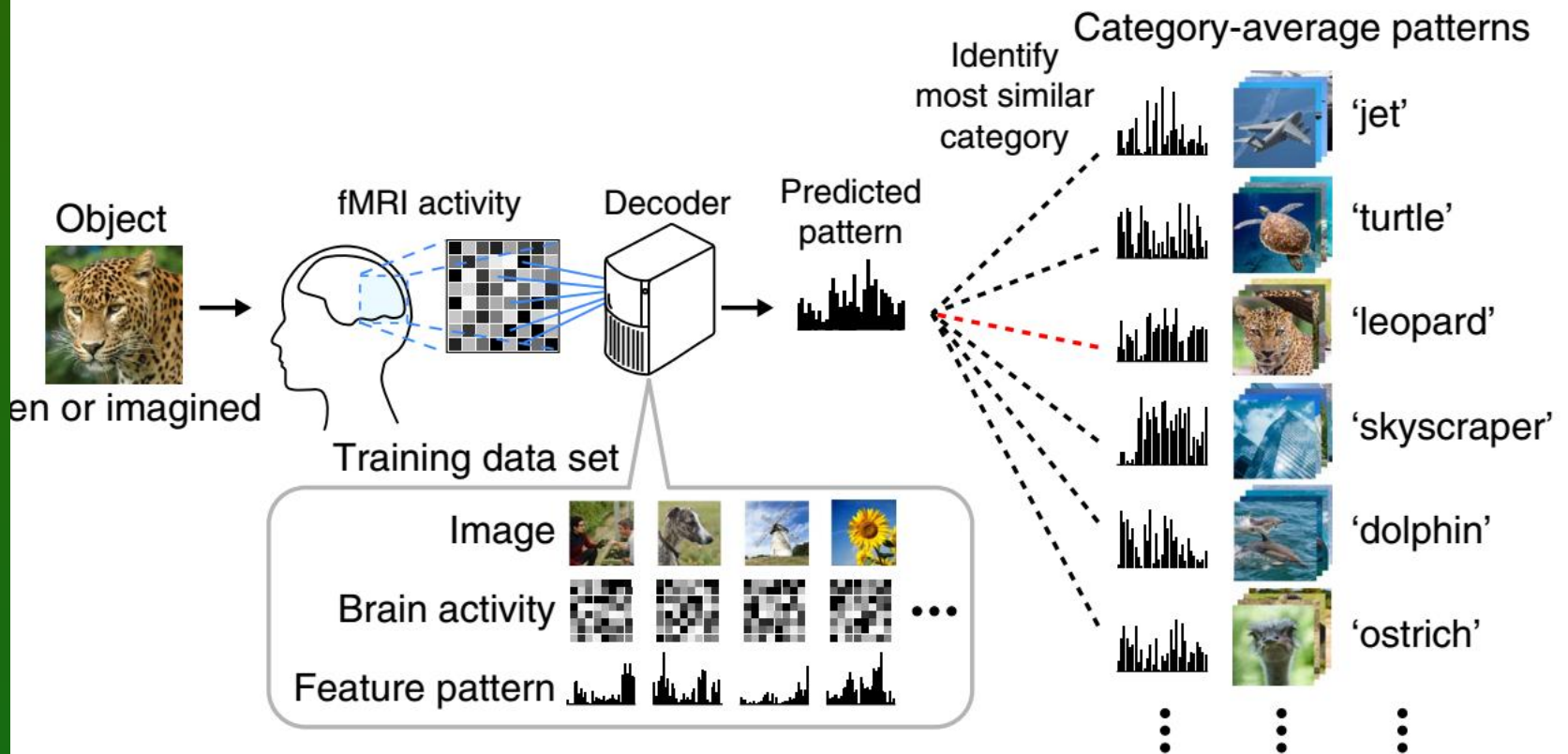
8-layer convolution network, ~60 mln parameters, feature vectors from randomly selected 1000 units in each layer to simplify calculations.

Output: 1000 images.



Brain activity \leftrightarrow Mental image

fMRI activity can be correlated with deep CNN network features; using these features closest image from large database is selected. Horikawa, Kamitani, Generic decoding of seen and imagined objects using hierarchical visual features. Nature Comm. 2017.



Decoding Dreams



Decoding Dreams, ATR Kyoto, Kamitani Lab. fMRI images analysed during REM phase or while falling asleep allows for dream categorisation.

Dreams, thoughts ... can one hide what has been seen and experienced?

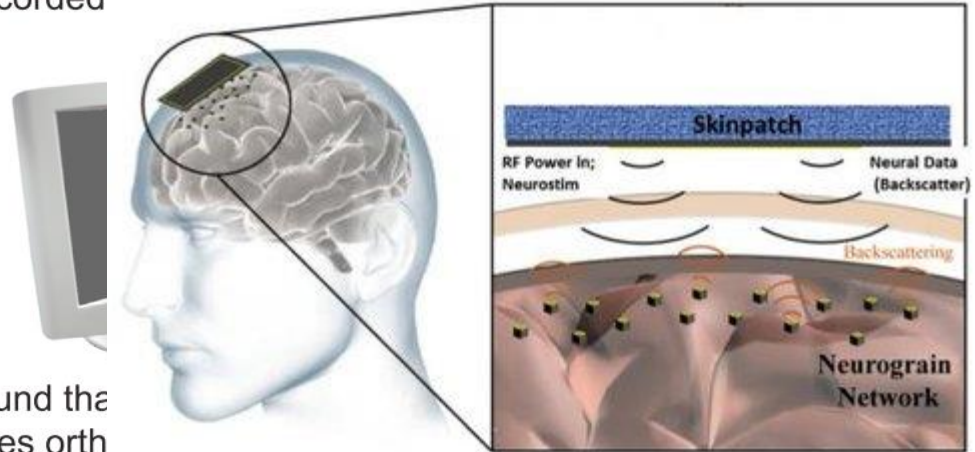
Neural screen

Features are discovered, and their combination remembered as face, but detailed recognition needs detailed recording from neurons – 205 neurons in various visual areas used.

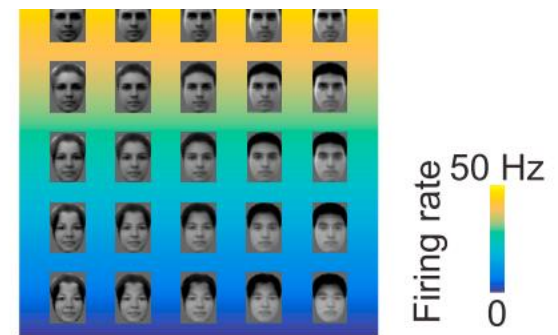
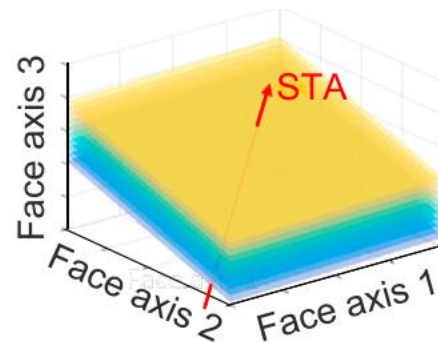
L. Chang and D.Y. Tsao, “The code for facial identity in the primate brain,” *Cell* 2017

DARPA (2016): put million nanowires in the brain!
Use them to read neural responses and 10% of them to activate neurons.

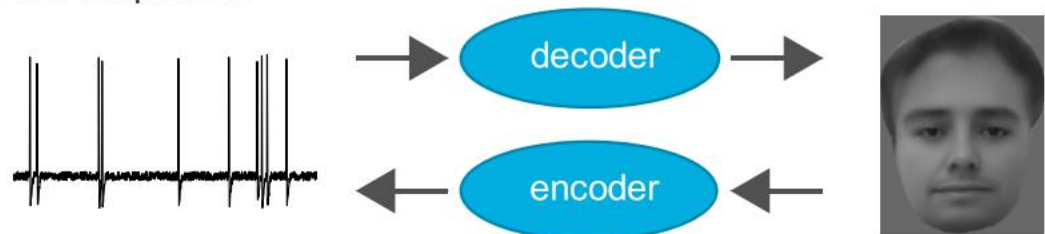
1. We recorded patches



2. We found the to changes orth

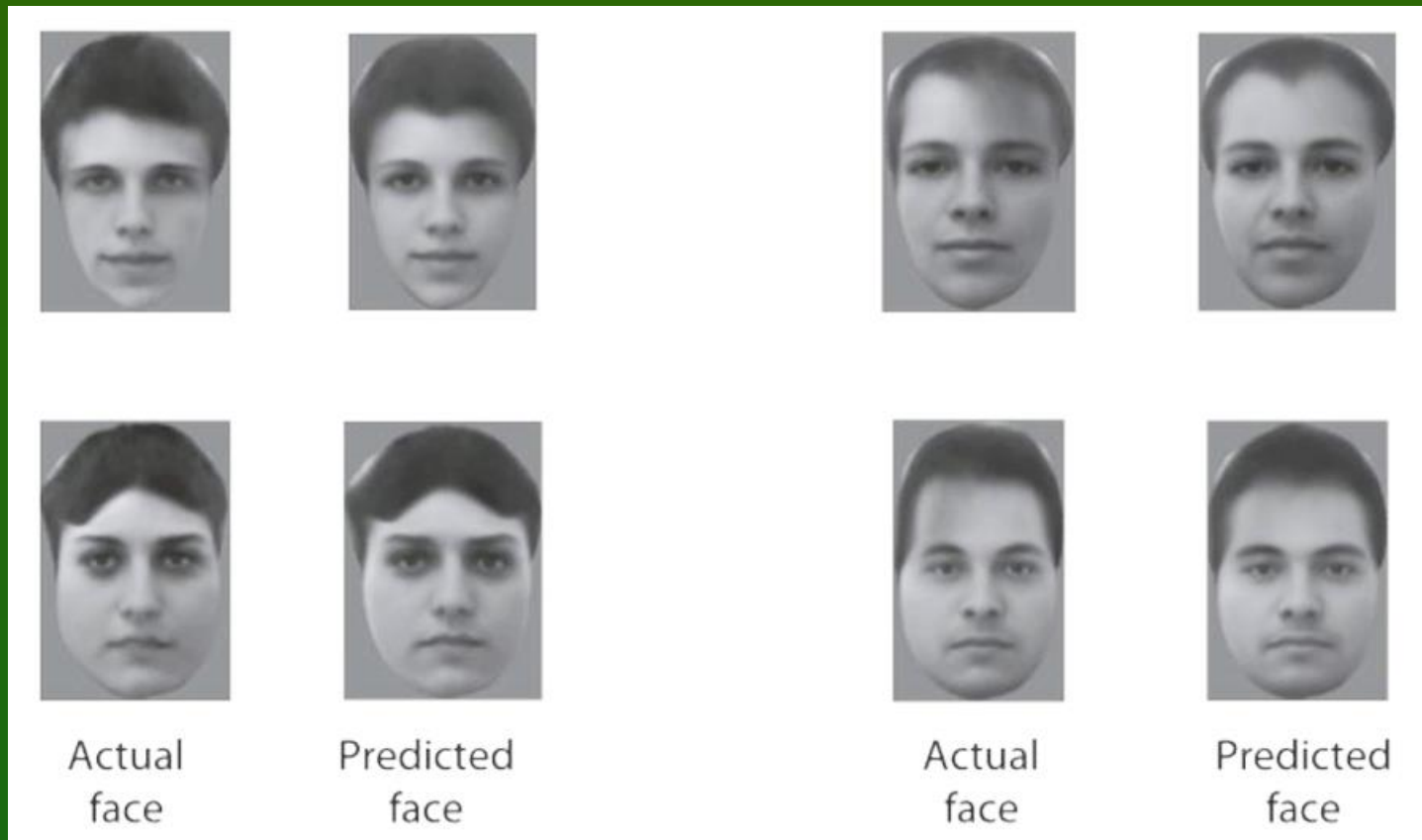


3. We found that an axis model allows precise encoding and decoding of neural responses

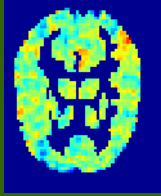


Mental images

Facial identity is encoded via a simple neural code that relies on the ability of neurons to distinguish facial features along specific axes in the face space.



Hidden concepts



Do we have conscious access of all brain states that influence thinking?

Language, symbols in the brain: phonological labels usually in the RH (right hemisphere) associated with prototypes of distributed activations of the brain.

Helps to structure the flow of brain states in the thinking process.

Right hemisphere activations just give us the feeling of something wrong.

- Right hemisphere is as busy as left – encoding concepts without verbal labels?
- Evidence: insight phenomena, intuitive understanding of grammar, etc.


Can we describe verbally natural categories?

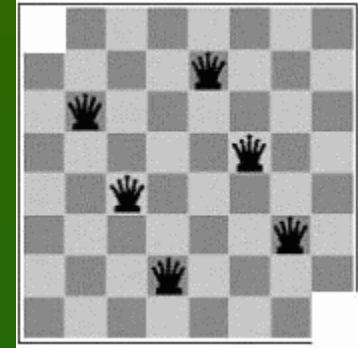
- Yes, if they are rather distinct: see 20 question game.
- Is object description in terms of properties sufficient and necessary?

Not always. Example: different animals and dog breeds.

- 20Q-game: weak questions (seemingly unrelated to the answer) may lead to precise identification! RH may contribute to activation enabling associations.

Problems requiring insights

Given 31 dominos  and a chessboard with 2 corners removed, can you cover all board with dominos?

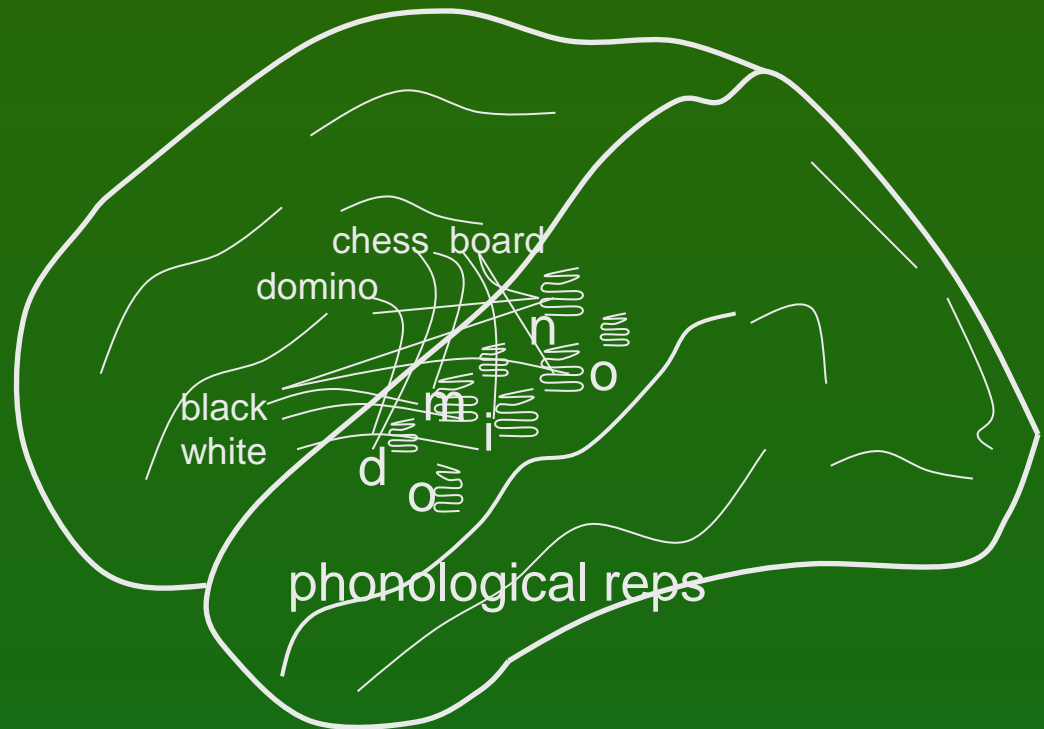


Analytical solution: try all combinations.

Does not work ... too many combinations to try.

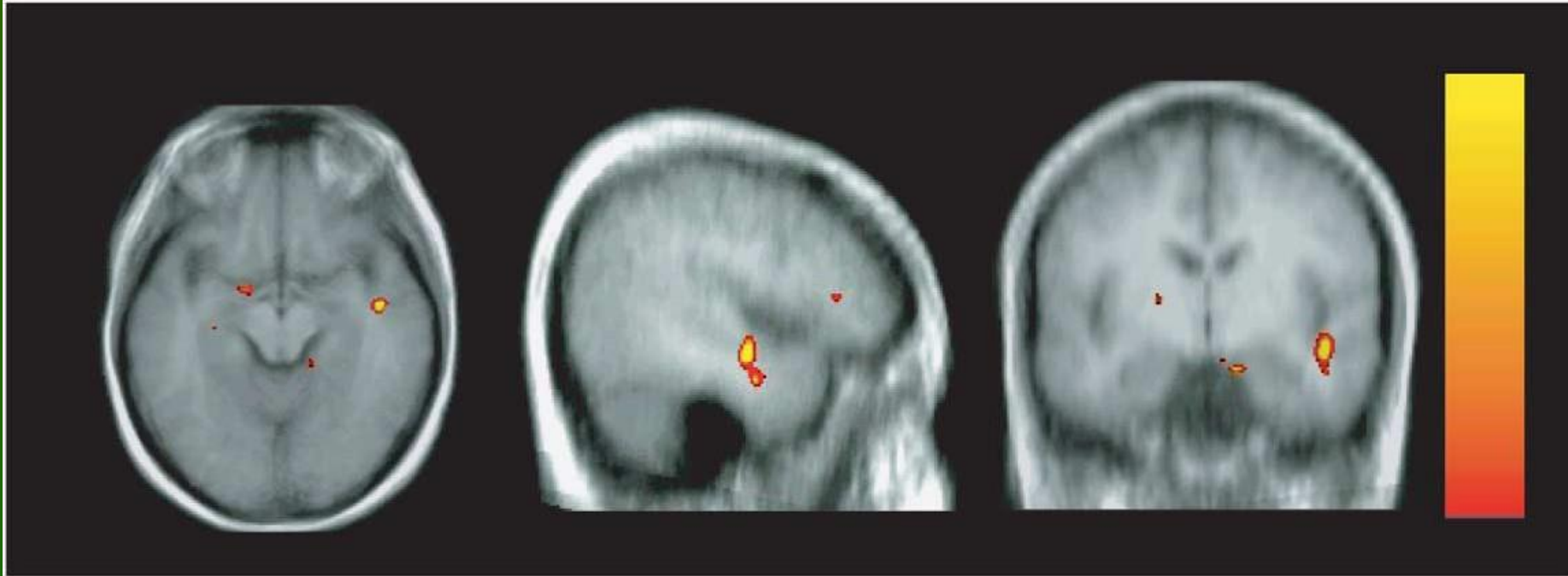
Logical, symbolic approach has little chance to create proper activations in the brain, linking new ideas: otherwise there will be too many associations, making thinking difficult.

Insight \leq right hemisphere, meta-level representations without phonological (symbolic) components ... counting?



Insights and brains

Activity of the brain while solving problems that required insight and that could be solved in schematic, sequential way has been investigated.

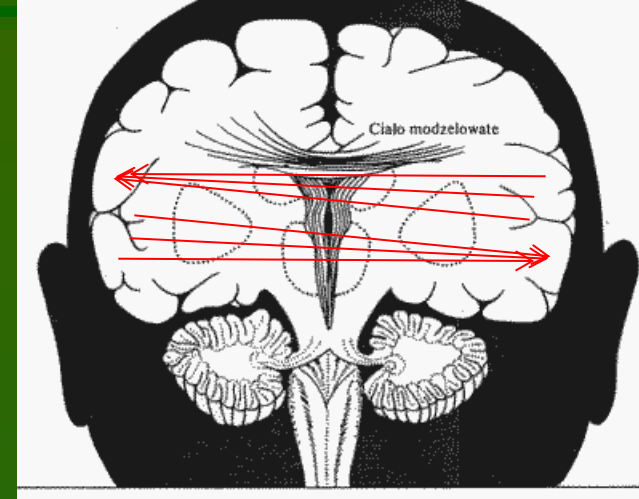


An increased activity of the right hemisphere anterior superior temporal gyrus (RH-aSTG) was observed during initial solving efforts and insights. About 300 ms before insight a burst of gamma activity was observed, interpreted by the authors as „making connections across distantly related information during comprehension ... that allow them to see connections that previously eluded them”.

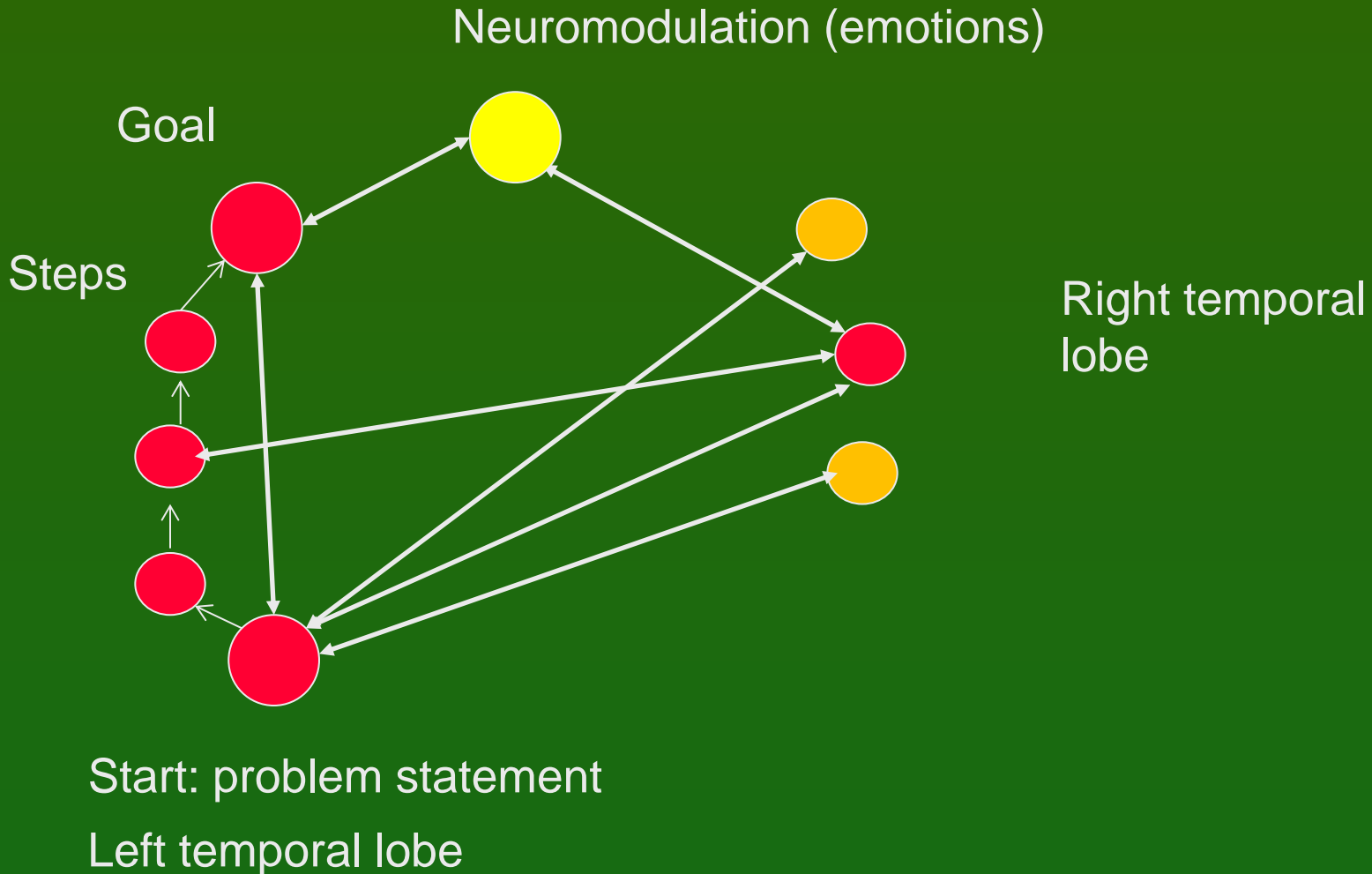
Insight interpreted

What really happens? My interpretation:

- LH-STG represents concepts, S=Start, F=final
- understanding, solving = transition, step by step, from S to F
- if no connection (transition) is found this leads to an impasse;
- RH-STG 'sees' LH activity on meta-level, clustering concepts into abstract categories (cosets, or constrained sets);
- connection between S to F is found in RH, leading to a feeling of vague understanding;
- gamma burst increases the activity of LH representations for S, F and intermediate configurations; feeling of imminent solution arises;
- stepwise transition between S and F is found;
- finding solution is rewarded by emotions during Aha! experience; they are necessary to increase plasticity and create permanent links.



Solving problems with insight



How to become an expert?

Textbook knowledge in medicine: detailed description of all possibilities.

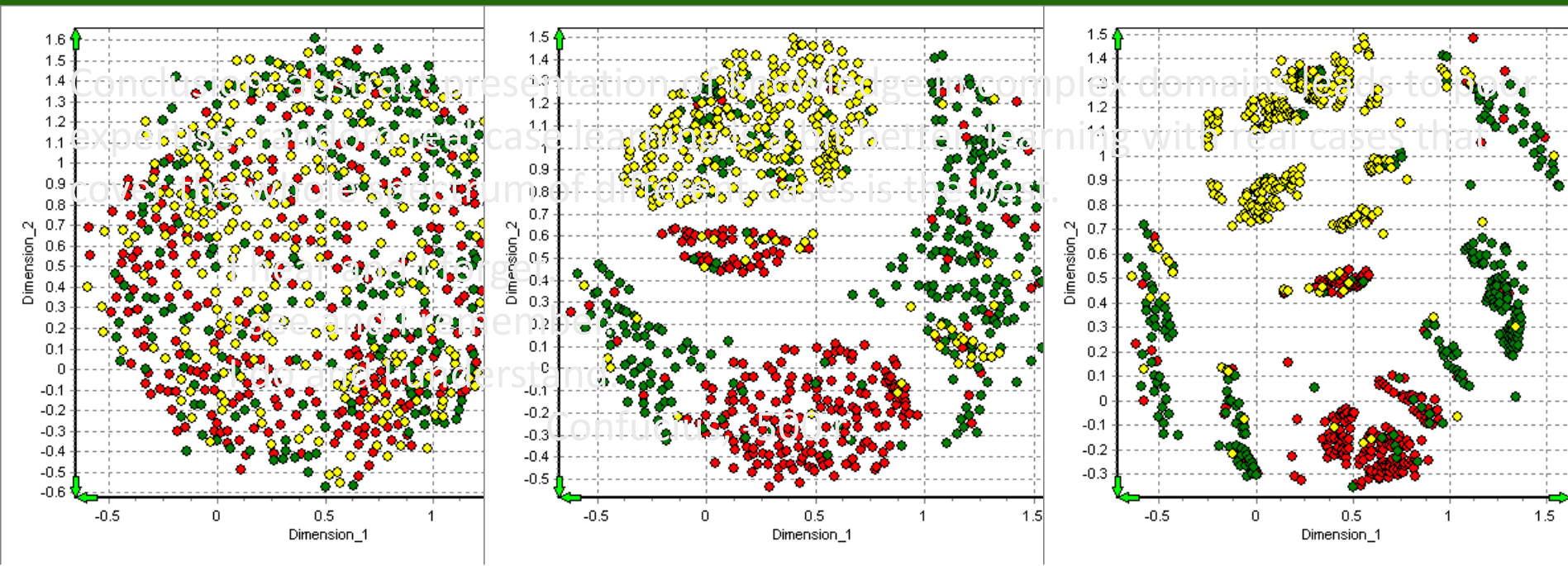
Effect: neural activation flows everywhere and correct diagnosis is impossible.

Correlations between observations forming prototypes are not firmly established.

Expert has correct associations.

Example: 3 diseases, clinical case description, MDS description.

- 1) System that has been trained on textbook knowledge.
- 2) Same system that has learned on real cases.
- 3) Experienced expert that has learned on real cases.



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 \sim angle, about $60^\circ/\text{sec}$.

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?

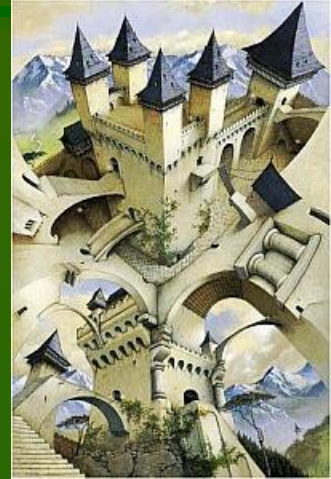
Systematic fallacies: a sort of cognitive illusions.

- 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 $A \Rightarrow B$; then $\sim B \Rightarrow \sim A$, 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.

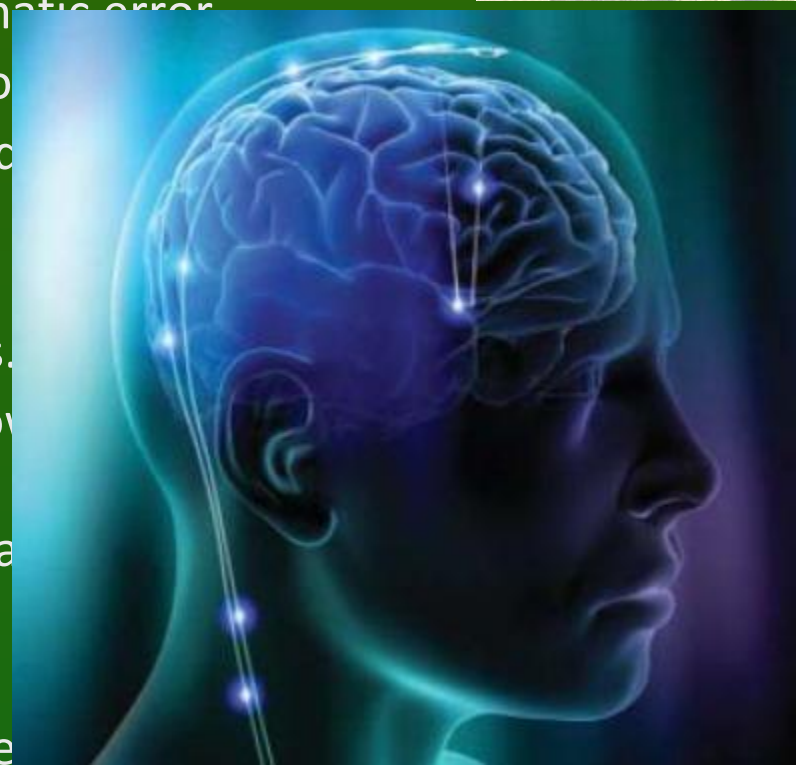
1. MM represent explicitly what is true, but not what is false; this may lead naive reasoner into systematic error
2. Large number of complex models => poor performance
3. Tendency to focus on a few possible models => irrational decisions.

Cognitive illusions are just like visual illusions.

M. Piattelli-Palmarini, *Inevitable Illusions: How Everyday Deceptions Shape the Human Mind* (1996)

R. Pohl, *Cognitive Illusions: A Handbook on Fallacious Reasoning, Judgment and Memory* (2005)

Amazing, but mental models theory ignores evidence for the role of learning in any form! How and why do we reason the way we do?
I'm innocent! My brain made me do it!



Mental models

Easy reasoning $A \Rightarrow B$, $B \Rightarrow C$, so $A \Rightarrow C$

- All mammals suck milk.
- Humans are mammals.
- \Rightarrow 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?

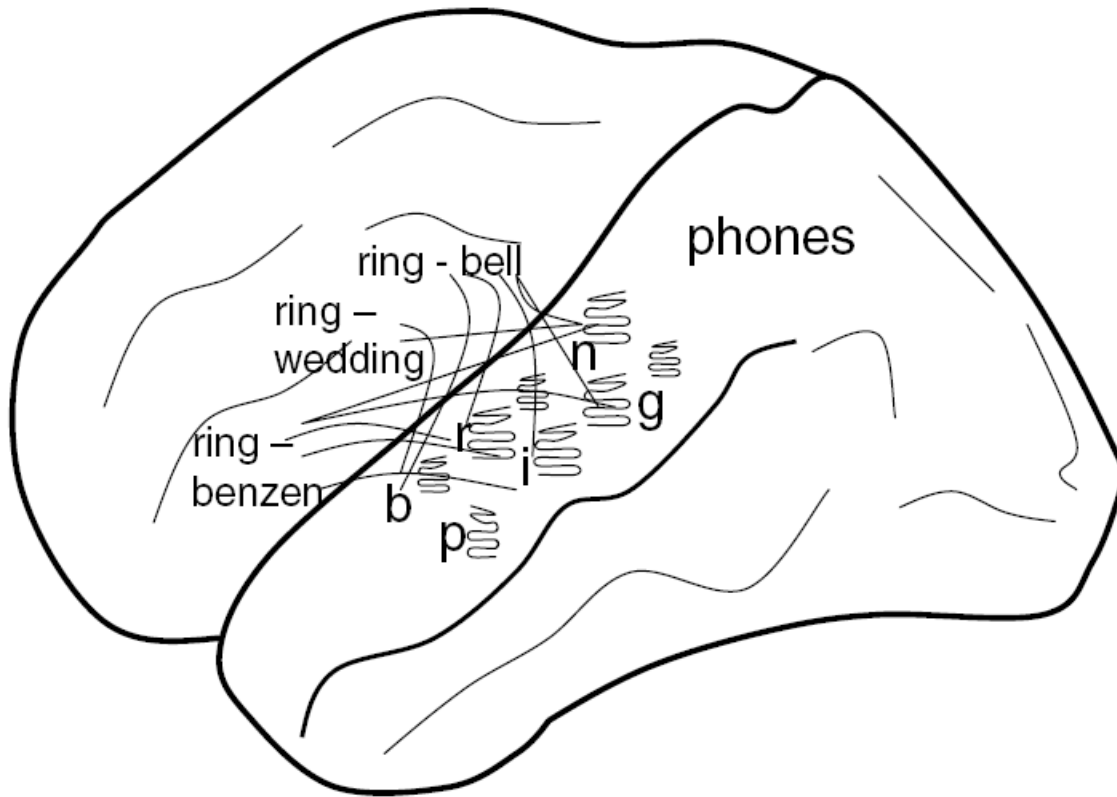
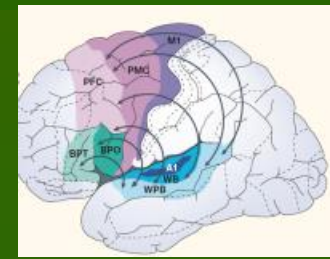
Surprisingly only $\sim 10\%$ of students get it right after days of thinking.

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.

ivity



nes, pay attention to

terns of activations.
parallel both words and
ptic connections.
antic density.

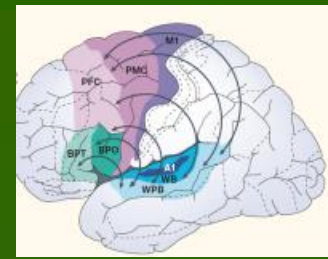
Start from keywords priming phonological representations in the auditory cortex; spread the activation to concepts that are strongly related.

Use inhibition in the winner-takes-most to avoid false associations.

Find fragments that are highly probable, estimate phonological probability.

Combine them, search for good morphemes, estimate semantic probability.

Creativity with words



The simplest testable model of creativity:

- create interesting novel words that capture some features of products;
- understand new words that cannot be found in the dictionary.

Model inspired by the putative brain processes when new words are being invented starting from some keywords priming auditory cortex.

Phonemes (allophones) are resonances, ordered activation of phonemes will activate both known words as well as their combinations; context + inhibition in the winner-takes-most leaves only a few candidate words.

Creativity = network+imagination (fluctuations)+filtering (competition)

Imagination: chains of phonemes activate both word and non-word representations, depending on the strength of the synaptic connections.

Filtering: based on associations, emotions, phonological/semantic density.

discoverity = {disc, disco, discover, verity} (discovery, creativity, verity)

digventure = {dig, digital, venture, adventure} new!

Server: <http://www-users.mat.uni.torun.pl/~macias/mambo/index.php>

Words: experiments

A real letter from a friend:

I am looking for a word that would capture the following qualities: portal to new worlds of imagination and creativity, a place where visitors embark on a journey discovering their inner selves, awakening the Peter Pan within. A place where we can travel through time and space (from the origin to the future and back), so, its about time, about space, infinite possibilities.

FAST!!! I need it soooooooooooooooooooooon.

creativital, creatival (creativity, portal), used in creatival.com

creativity (creativity, discovery), creativity.com (strategy+creativity)

discoverity = {disc, disco, discover, verity} (discovery, creativity, verity)

digventure = {dig, digital, venture, adventure} still new!

imativity (imagination, creativity); infinitime (infinitive, time)

infinition (infinitive, imagination), already a company name

portravel (portal, travel); sportal (space, sport, portal), taken

timagination (time, imagination); timativity (time, creativity)

tivity (time, discovery); trime (travel, time)

Server at: <http://www-users.mat.uni.torun.pl/~macias/mambo>

Conspiracy in the brain



Formation of deep beliefs, distorted memory, memetics, conspiracy ...
Slow and rapid scenarios are possible, here only rapid presented:

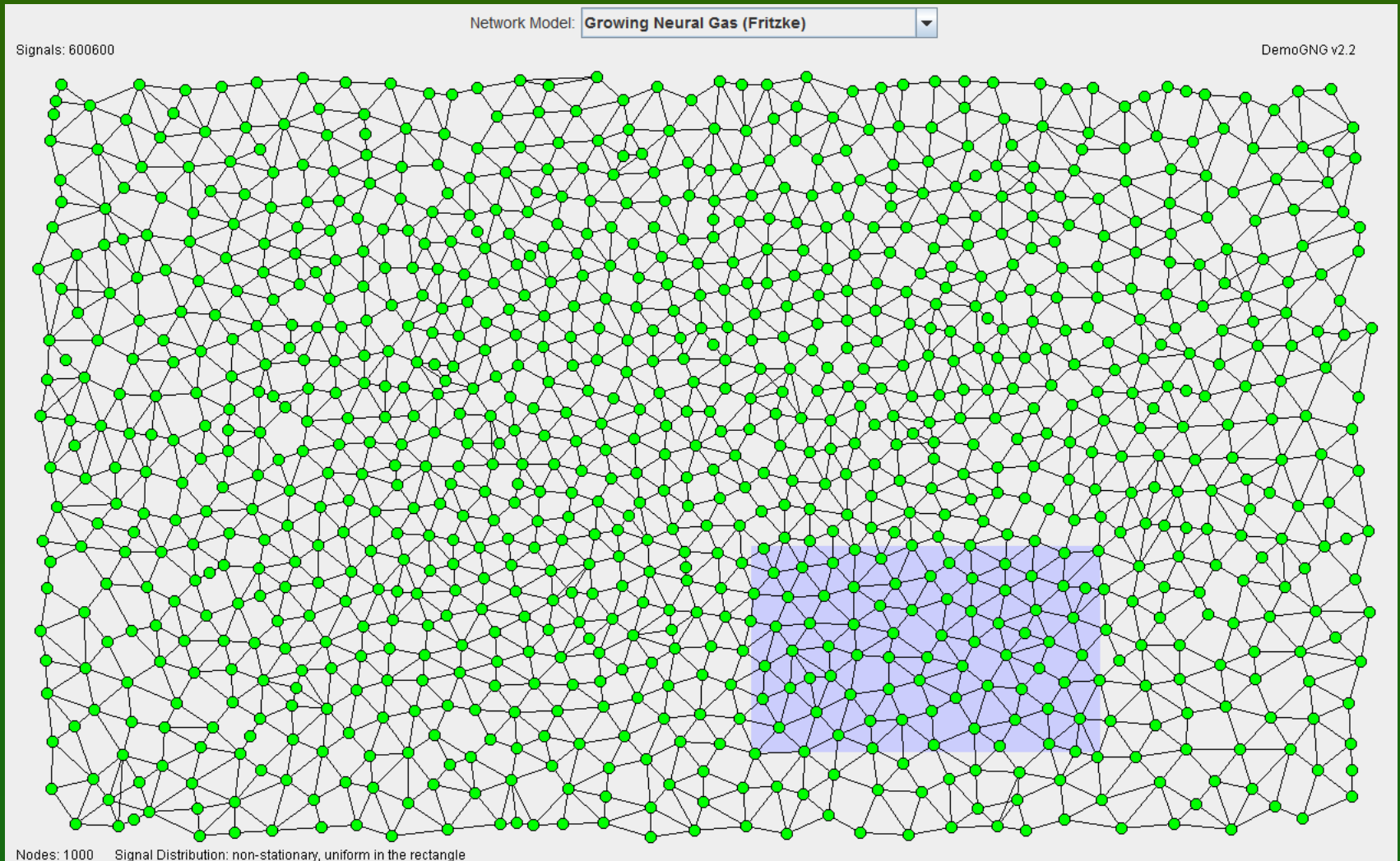
- Emotional situations => neurotransmitters => neuroplasticity => fast learning, must be important.
- Fast learning => high probability of wrong interpretation.
- Traumatic experiences, hopelessness, decrease brain plasticity and leave only strongest association – strongly connected pathways.
- Conspiracy theories form around such associations, “frozen” pathways lead to brain activations forming strong attractors, distorting rational thinking.
- Such strong associations save brain energy and cannot be changed by rational arguments, that influence weaker associations only.
- This explanation becomes so obviously obvious ...



Model: concept vectors derived from a corpus + MDS or Growing Neural Gas visualization (Martinetz & Schulten, 1991).

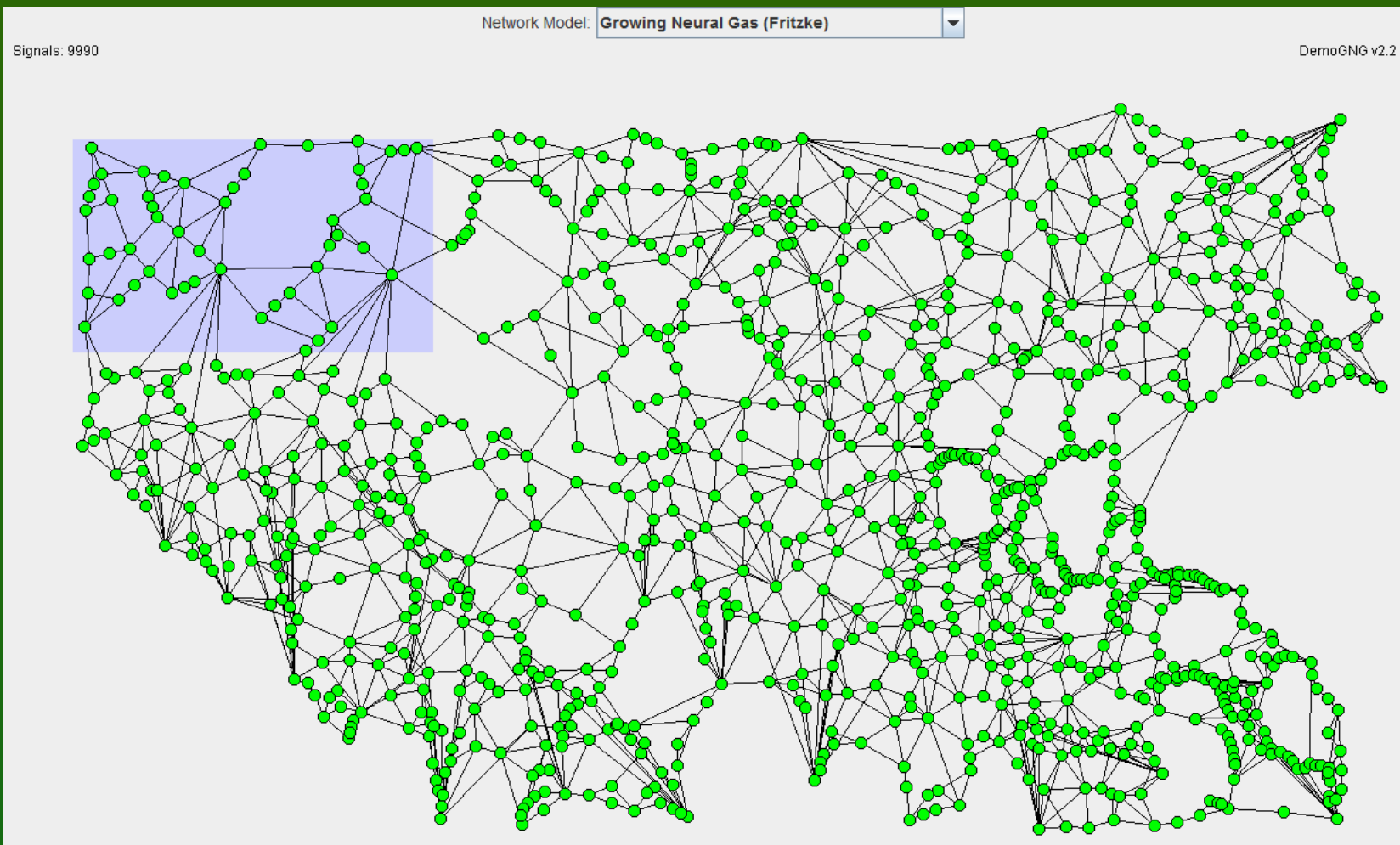
Internalization of environment

Episodes are remembered and serve as reference points, if observations are unbiased they reflect reality.



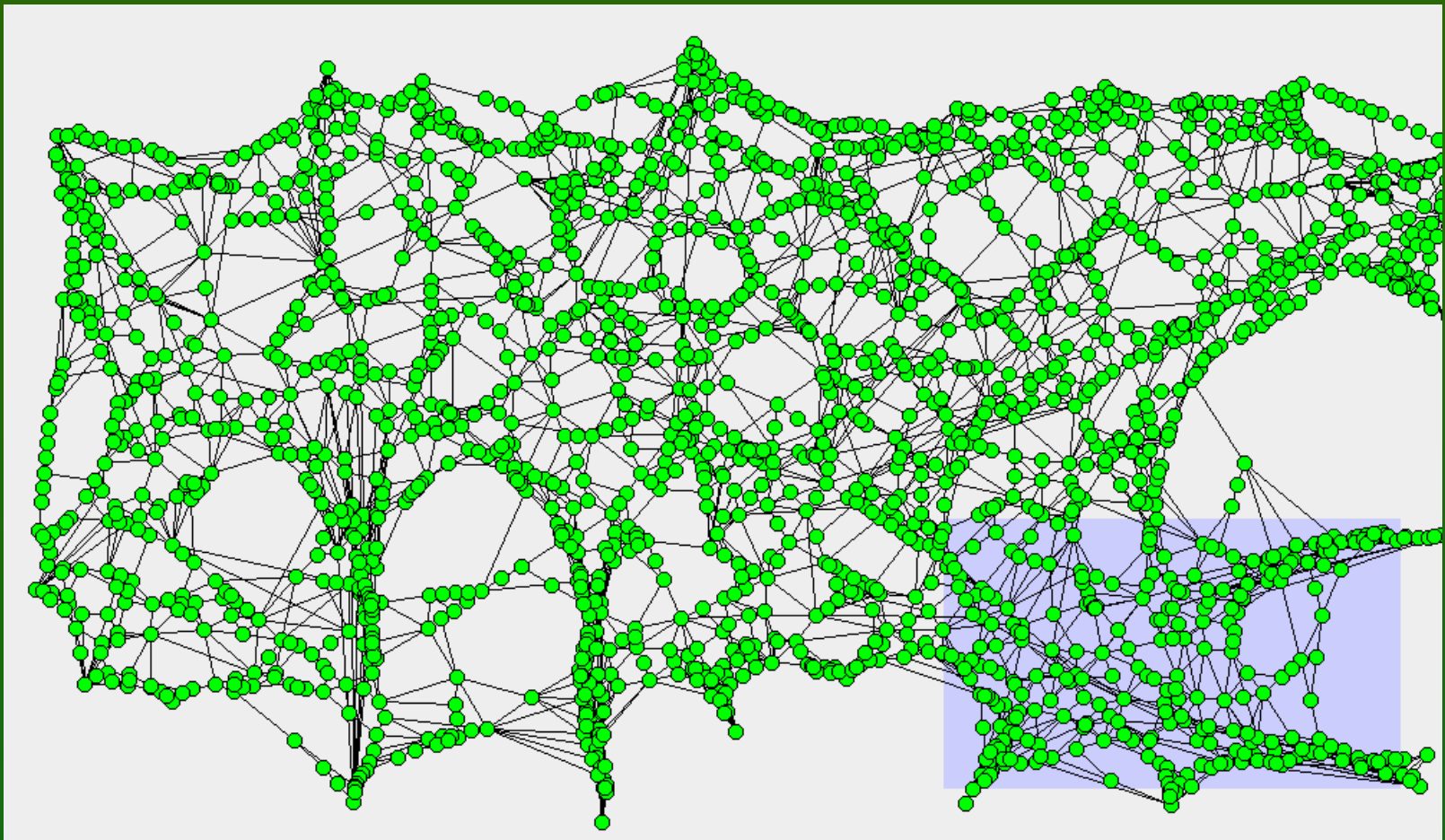
Extreme plasticity

Brain plasticity (learning) is increased if long, Slow strong emotions are involved. Followed by depressive mood it leads to severe distortions, false associations, simplistic understanding.



Conspiracy views

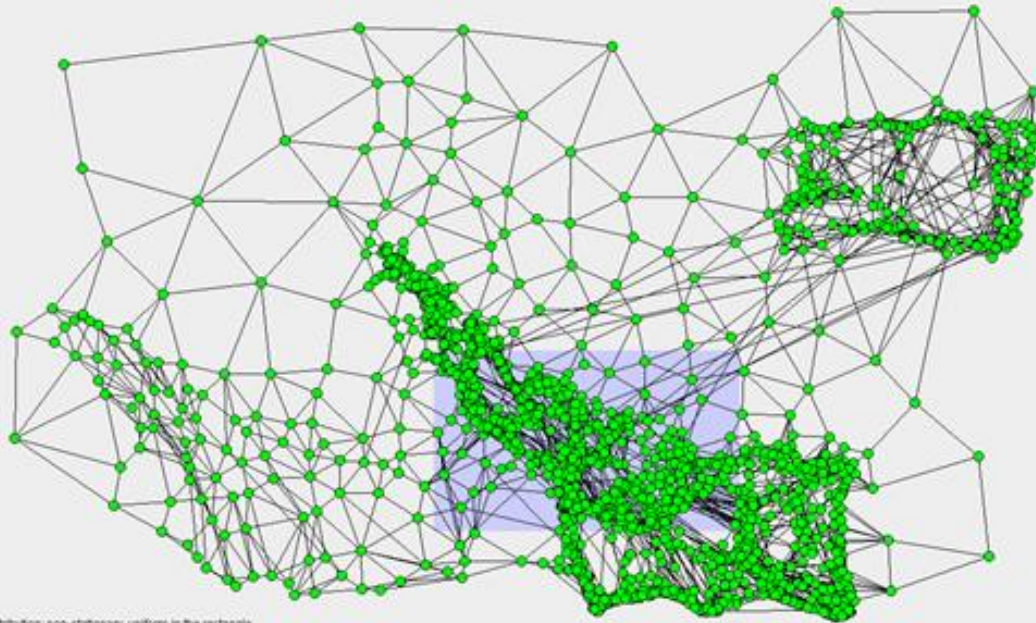
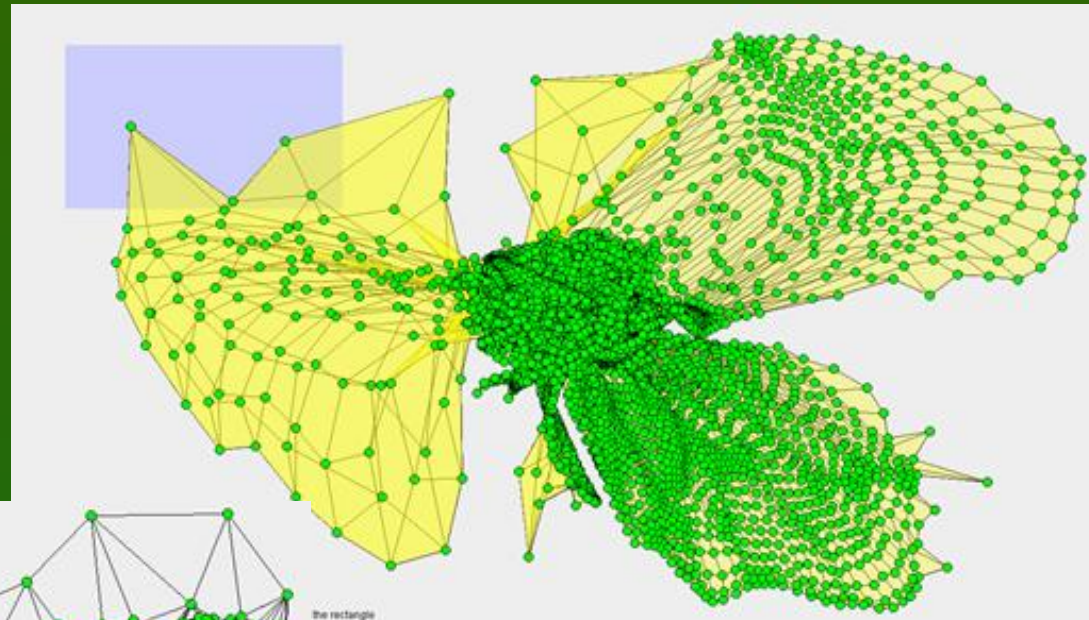
Illuminati, masons, Jews, UFOs, or twisted view of the world leaves big holes and admits simple explanations that save mental energy, creating „sinks” that attract many unrelated episodes.



Memoids ...

Totally distorted world view, mind changed into a memplex.

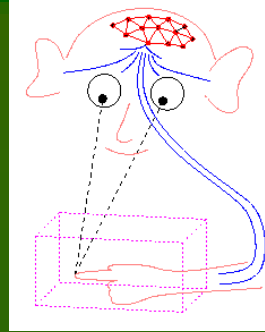
Ready for sacrifice.



WD: Memetics and Neural Models of Conspiracy Theories

[arXiv:1508.04561](https://arxiv.org/abs/1508.04561)

Conclusions



- Brain reading has made impressive progress in recent years. New techniques based on nanowires will bring much more info.
- Connectomics and network science has shown how global brain states give rise to mental functions, connecting different brain areas.
- So far only brains are capable of understanding language and use complex reasoning. Formal methods used for real-world problems have limitations.
- Words have relatively stable and unique distributions of activity in the brain, semantic representation partially recreates direct experience, or in case of abstract concepts and metaphors relations to other concepts.
- Agents are functional subnetworks performing specialized functions and encoding specific information.
- Computational simulations and analysis of neuroimaging converge on useful models for natural language processing. Psychological constructions and models do not provide correct conceptualization of brain processes.
- Insight and intuition are functions of the right hemisphere.

Project „Symfonia”, NCN, Kraków, 18 July 2016

In search of the sources
of brain's cognitive activity



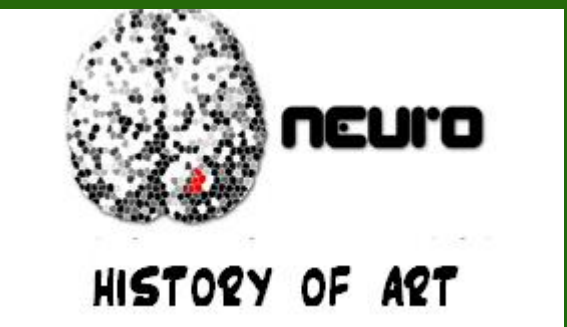
Soul or brain: what makes us human?
Interdisciplinary Workshop with theologians,
Toruń 19-21.10.2016



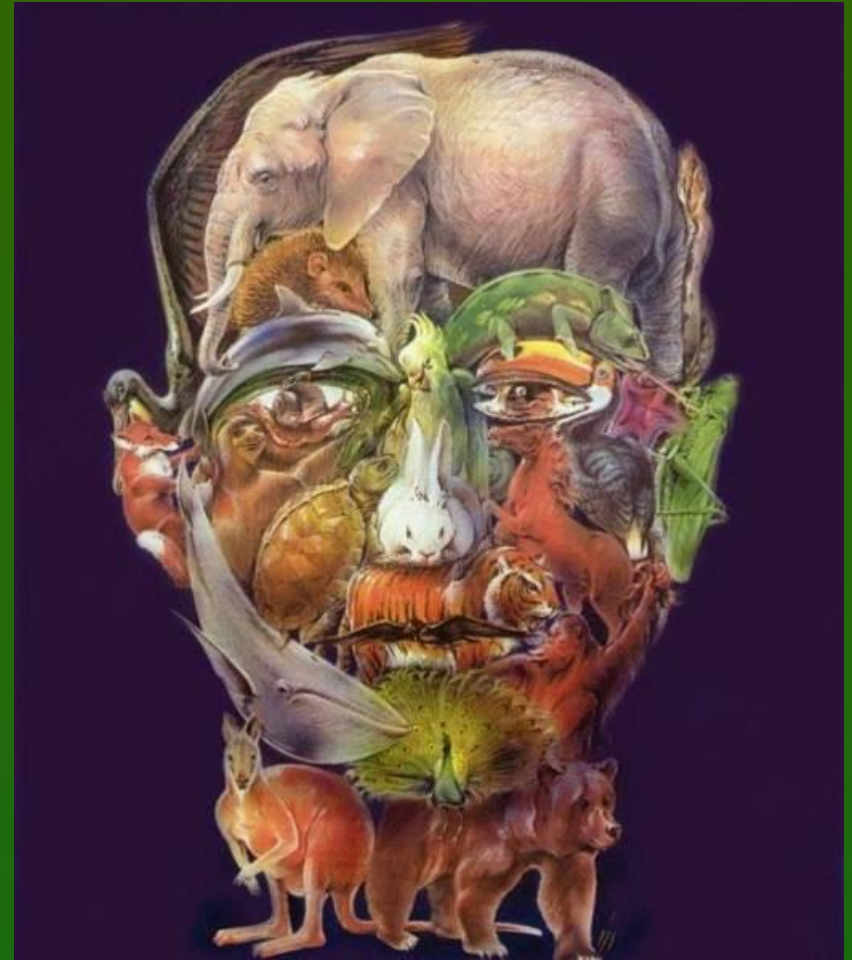
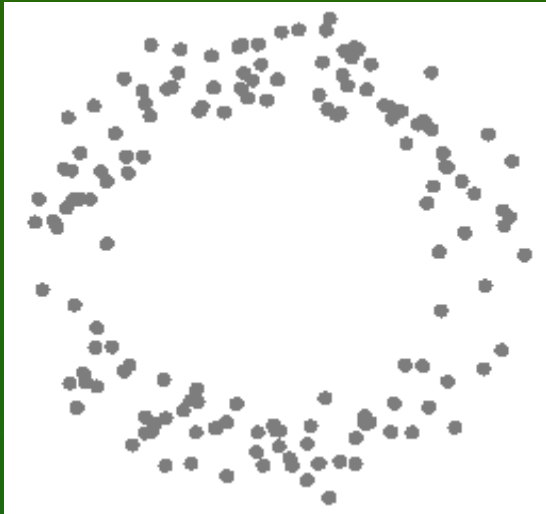
Monthly international
developmental seminars
(2017): Infants, learning,
and cognitive development

Disorders of consciousness
17-21.09.2017

Autism: science, therapies
23.05.2017



Thank you for
synchronization
of your neurons



Google: W. Duch
=> talks, papers, lectures, Flipboard ...

Garagnani et al. conclusions

“Finally, the present results provide evidence in support of the hypothesis that words, similar to other units of cognitive processing (e.g. objects, faces), are represented in the human brain as distributed and anatomically distinct action-perception circuits.”

“The present results suggest that anatomically distinct and distributed action-perception circuits can emerge spontaneously in the cortex as a result of synaptic plasticity. Our model predicts and explains the formation of lexical representations consisting of strongly interconnected, anatomically distinct cortical circuits distributed across multiple cortical areas, allowing two or more lexical items to be active at the same time. Crucially, our simulations provide a principled, mechanistic explanation of where and why such representations should emerge in the brain, making predictions about the spreading of activity in large neuronal assemblies distributed over precisely defined areas, thus paving the way for an investigation of the physiology of language and memory guided by neurocomputational and brain theory.”

P-spaces

Psychological spaces: how to visualize inner life?

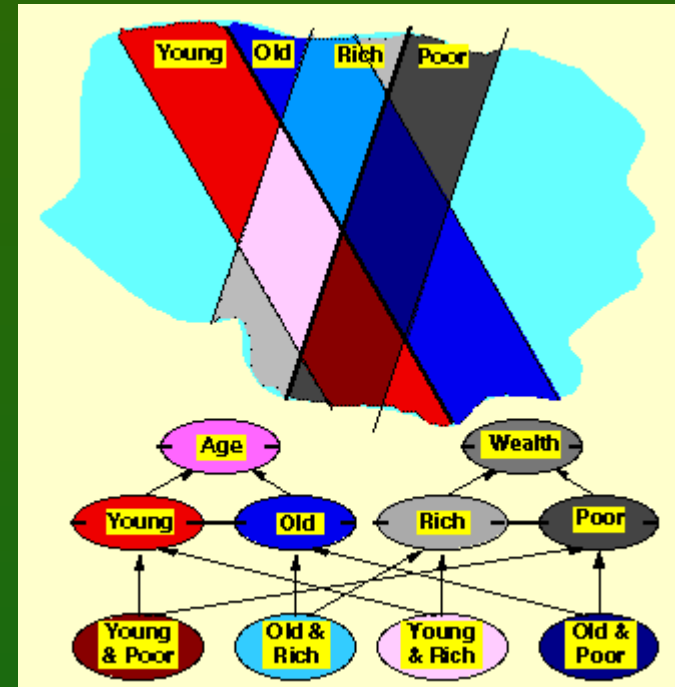
K. Lewin, The conceptual representation and the measurement of psychological forces (1938), cognitive dynamic movement in phenomenological space.

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.

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

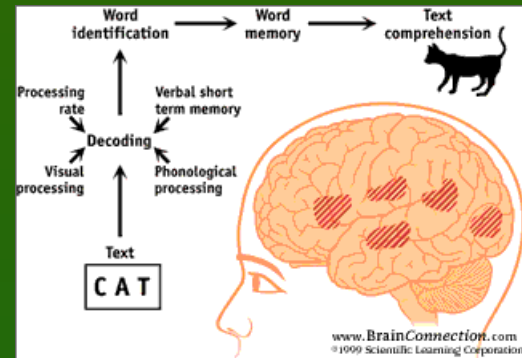


P-space definition

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 (Shepard 1957-2001):

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



This may be achieved using multi-dimensional non-metric scaling (MDS), reproducing similarity relations in low-dimensional spaces.

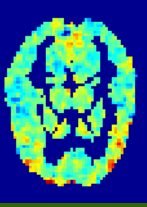
Many Object Recognition and Perceptual Categorization models assume that objects are represented in a multidimensional psychological space; similarity between objects $\sim 1/\text{distance}$ in this space.

Can one describe the state of mind in similar way?

- Nishida, S., & Nishimoto, S. (2018). Decoding naturalistic experiences from human brain activity via distributed representations of words. *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2017.08.017>
- Natural visual scenes induce rich perceptual experiences that are highly diverse from scene to scene and from person to person. Here, we propose a new framework for decoding such experiences using a distributed representation of words. We used functional magnetic resonance imaging (fMRI) to measure brain activity evoked by natural movie scenes. Then, we constructed a high-dimensional feature space of perceptual experiences using skip-gram, a state-of-the-art distributed word embedding model. We built a decoder that associates brain activity with perceptual experiences via the distributed word representation. The decoder successfully estimated perceptual contents consistent with the scene descriptions by multiple annotators. Our results illustrate three advantages of our decoding framework: (1) three types of perceptual contents could be decoded in the form of nouns (objects), verbs (actions), and

- Nishida, S., & Nishimoto, S. (2017). Decoding naturalistic experiences from human brain activity via distributed representations of words. *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2017.08.017>
-
- Natural visual scenes induce rich perceptual experiences that are highly diverse from scene to scene and from person to person. Here, we propose a new framework for decoding such experiences using a distributed representation of words. We used functional magnetic resonance imaging (fMRI) to measure brain activity evoked by natural movie scenes. Then, we constructed a high-dimensional feature space of perceptual experiences using skip-gram, a state-of-the-art distributed word embedding model. We built a decoder that associates brain activity with perceptual experiences via the distributed word representation. The decoder successfully estimated perceptual contents consistent with the scene descriptions by multiple annotators. Our results illustrate three advantages of our decoding framework: (1) three types of perceptual contents could be decoded in the form of nouns (objects), verbs (actions), and

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.

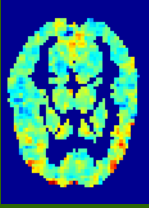
Hearing or thinking a word w , or seeing an object labeled as w adds to the overall brain activation in a non-linear way.

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

Result: almost continuous variation of this meaning.

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

Approximate reps.



States $\Psi(w, Cont) \leftrightarrow$ 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.

$\Psi(w, Cont) \sim$ 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.

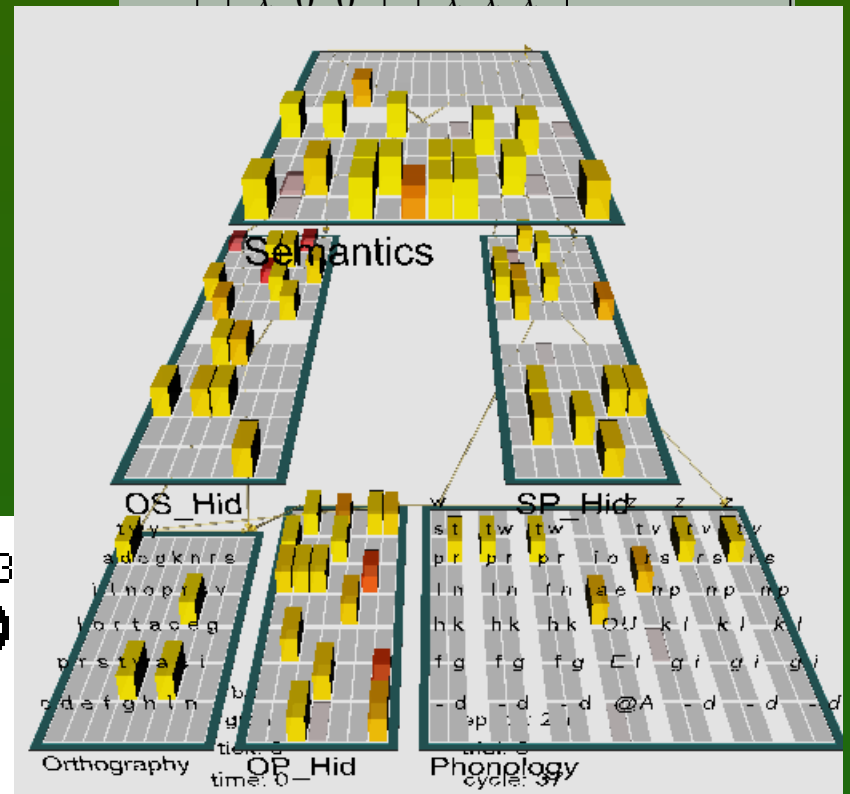
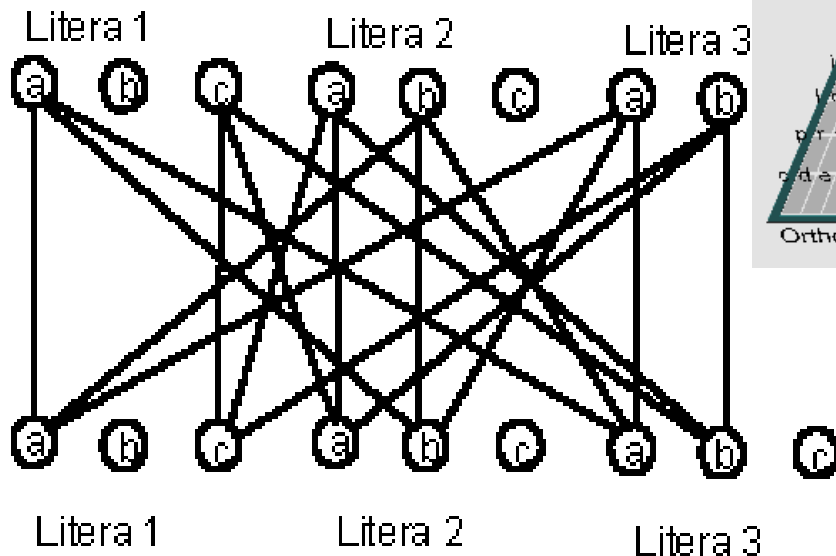
Result: a set of **X** vectors representing concepts in enhanced spaces, partially including effects of spreading activation.

Autoassociative networks

Simplest networks:

- binary correlation matrix,
- probabilistic $p(a_i, b_j | w)$

Major issue: rep. of symbols, morphemes, phonology ...



Static Platonic model

Newton introduced space-time, arena for physical events.

Mind events need psychological spaces.

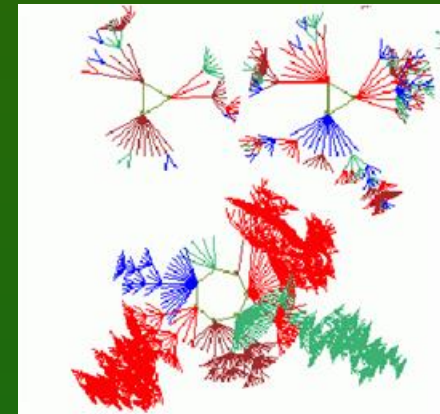
Goal: integrate neural and behavioral information in one model, create model of mental processes at intermediate level between psychology and neuroscience.

Static version: short-term response properties of the brain, behavioral (sensomotoric) or memory-based (cognitive).

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.



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.

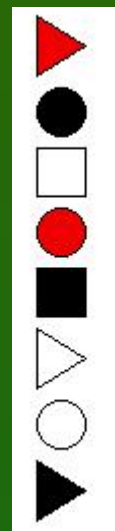
Types III, IV, and V - intermediate complexity between Type II - VI.

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 \Leftrightarrow simplest, canonical dynamics.

For all logical functions one may write corresponding equations.

For XOR (type II problems) equations are:

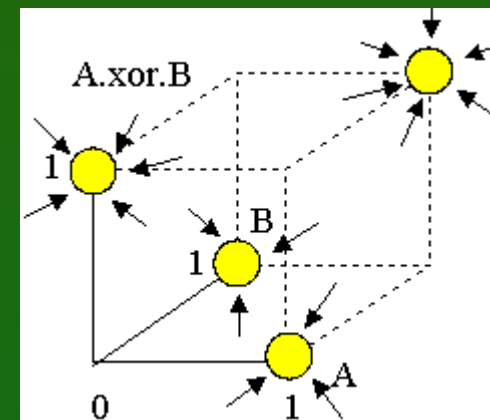
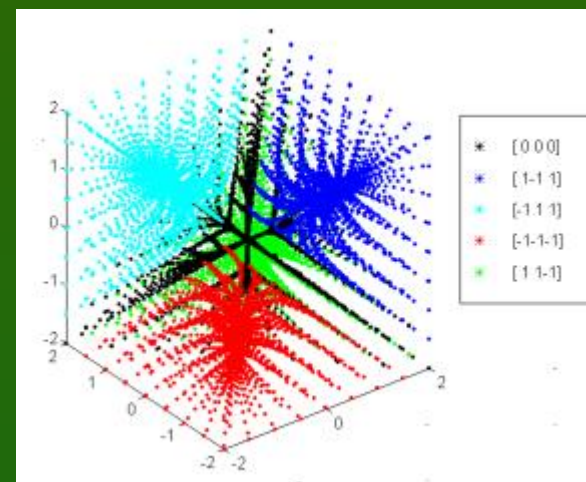
$$V(x, y, z) = 3xyz + \frac{1}{4}(x^2 + y^2 + z^2)^2$$

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

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

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

Corresponding feature space for relevant dimensions A, B



Inverse based

Relative frequencies (base rates) of categories

if on a list of disease and symptoms disease C is 3 times more common as R, then symptoms $PC \Rightarrow C$, $I \Rightarrow C$ (base rate effect)

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

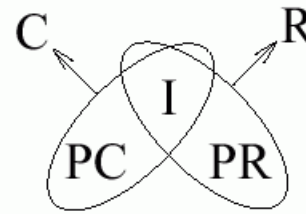
Although $PC + I + PR \Rightarrow C$ (60% answers)
 $PC + PR \Rightarrow R$ (60% answers)

Why such answers?
Psychological explanations are not convincing.

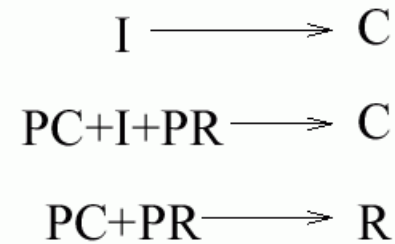
Effects due to the neurodynamics of learning?

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

Training:



Transfer:



Legend:

C = Common disease

R = Rare disease

I = Imperfect predictor

PC = Perfect predictor of
Common disease

PR = Perfect predictor of
Rare disease

Legend:

C = Common disease

R = Rare disease

I = Imperfect predictor

PC = Perfect predictor of
Common disease

PR = Perfect predictor of
Rare disease

IBR neurocognitive explanation

Psychological explanation:

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

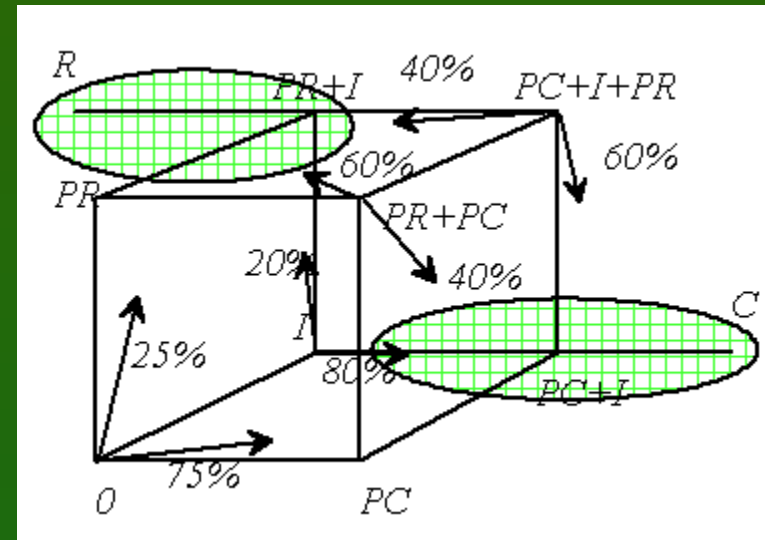
PR is attended to because it is a distinct symptom, although PC is more common.

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

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

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 gradients 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, $\sim W$ is A; relation $S \Leftrightarrow W$?

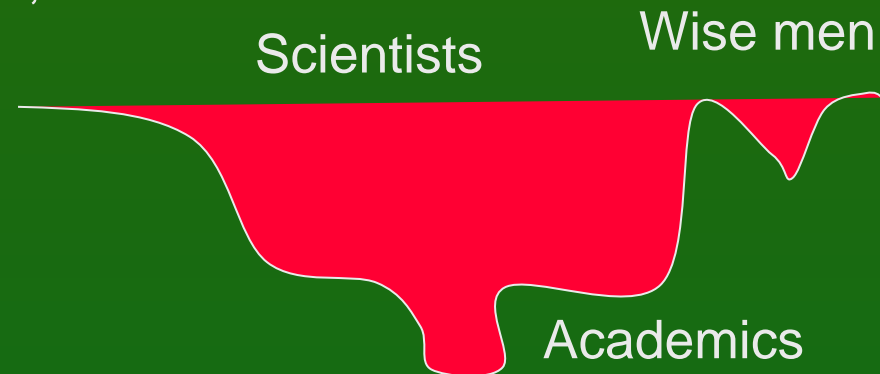
What happens with neural dynamics?

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.



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.

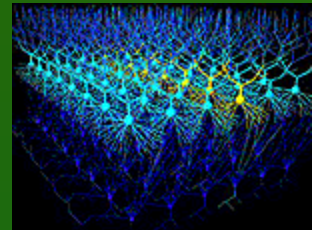
Psychology:

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.