# Artificial intelligence and the limits of the humanities I

Wlodzislaw Duch

Neurocognitive Laboratory, Center for Modern Interdisciplinary Technologies,
Department of Informatics, Institute of Engineering and Technology,
Faculty of Physics, Astronomy & Informatics, Nicolaus Copernicus University, Toruń

Current Trends in Human Science, Jagiellonian University 24.04.24

# AI and brains at the warp speed

1. What makes us human?
2. Multi-level phenomics
3. AI minds and human brains
4. Foundation Models and Generative AI.
5. Large Multimodal Models.
6. AI for science.
7. Distributed artificial brains.
8. Conscious avatars.

AI news in my Flipboard.

ChatGPT << AI. We see a tip of the iceberg...

Every week arxiv cs.ai adds over 100 new papers.

# What/Who am I

Quis ego et qualis ego?

Who I am and what kind of self I am?

Św. Augustyn (~ 400 rok)

What is „I” ?

Pascal (~ 1670)

How can this question be answered?

In 1972, science was not interested in the study of consciousness, in the university library you could only find pre-war polish monthly magazines "Spiritual Knowledge“ or "Lotus", on occultism and theosophy.

They did not provided satisfactory answers.

# Mind is moving

Bodidharma (5-6th centrury), first Chan Patriarch, who came from India:

Mind, which is our true nature ... responding to circumstances, transforms situations into mental events.

Mumonkan (early 13th century) koan collection:

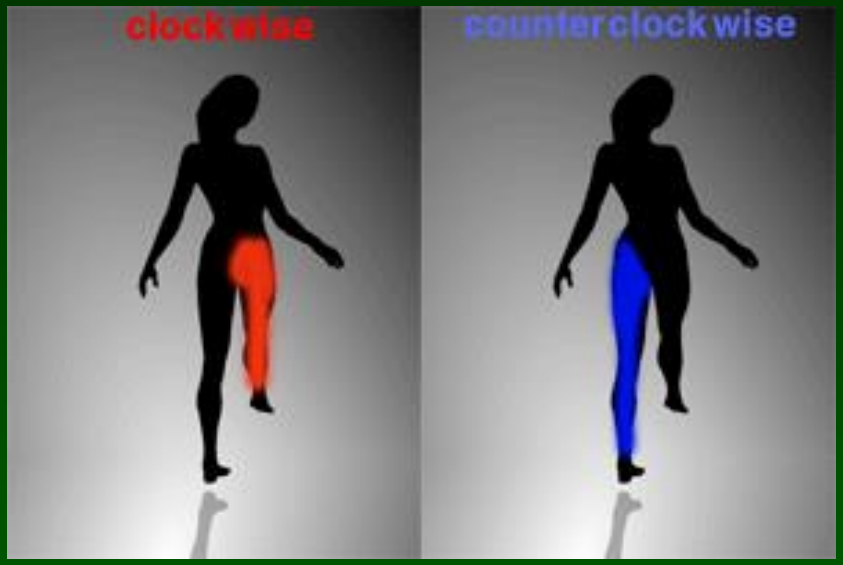
The Sixth Patriarch Hui-Neng arrived at the temple.

The wind fluttered the temple flag.
Two monks were arguing about it.

One said: the wind is moving.
The other said: the flag is moving.
In vain they try to convince each other.

The patriarch said: It's not the wind.
It's not the flag.
It is your mind that is what moves.

# *Māyā*, cosmic illusion that the phenomenal world is real




Whilst part of what we perceive comes through our senses
from the object before us, another part (and it may be the larger part)
always comes out of our own mind.

William James, The Principles of Psychology, 1890

# Neuronal space

Cortex activations => thoughts, experiences

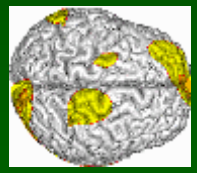
Streams of information sensory cortex ⇔ associative cortex
form quasi-stable states in the brain, can be distinguished from noise.

Neuronal activity in V1 cortex is weakly correlated (~10%) with retina image.

Most of V1 activations come from internal excitation:
you know what you see, you see what you know.

The perceived world is a product of our imagination! Maya …

It is only our mind that moves ... in each of us a little differently.



Our personal experiences are not testimony to objective events,
witnesses are rarely reliable.

# Can we know ourselves?

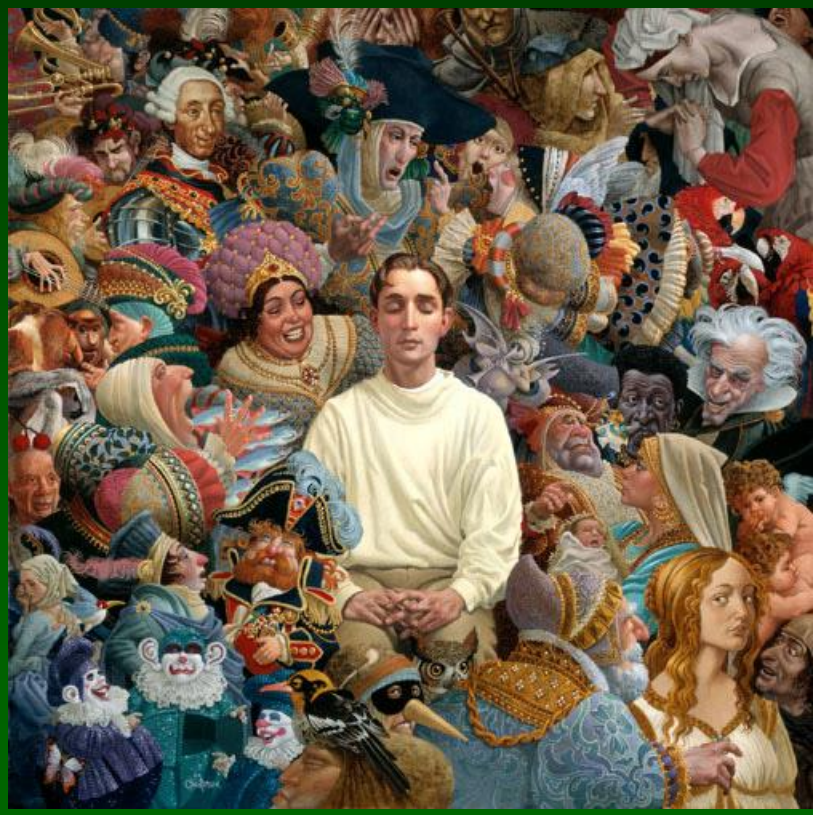
We are not aware of most processes that go on in “our” head.
Sigmund Freud wrote: “We are not the masters in our own house”.

In many cases we are unable to describe
our mental processes (E. Schwitzgebel, 2011).

Million voices in the brain compete
for conscious attention
(global network dynamics).

Can the **mind** understand the **brain**
that creates it? How deeply?

Can the **mind** control the **brain**?
What are the limits of such control?



James C. Christensen

# Introspection

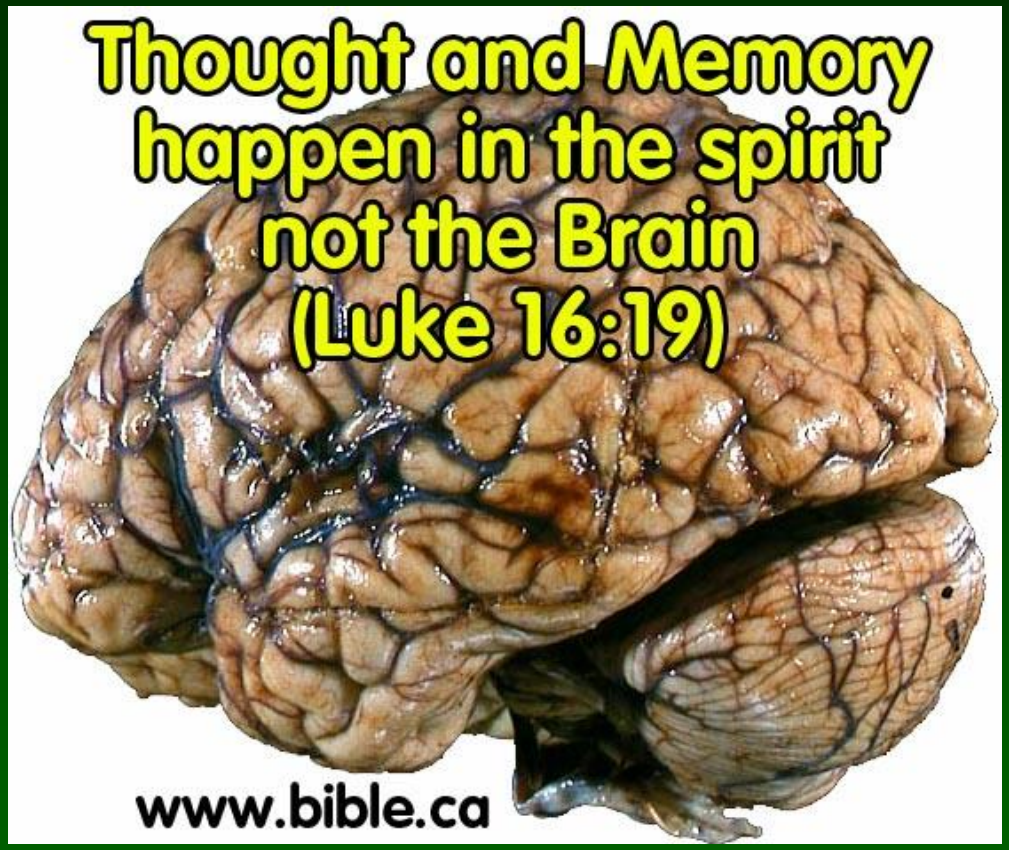
Zen Master Bassui (1327-1387) in "One Mind Sermon": ... you must first of all look into the source from which thoughts flow.
While sleeping or working, standing or sitting, deeply inquire of yourself:

What is it:
my own Mind?

Eventually the brain loses power over us ...

Phenomenology, embodied cognition, non-Cartesian cognitive science (Merleau-Ponty, Heidegger, Dreyfus, Varela, Lakoff, Johnson).

# Soul or brain?




Soul or brain: what makes us human? Interdisciplinary Workshop, Faculty of Theology, Nicolaus Copernicus University + Insbruck + Pampeluna (10/2016). Almost no dualists are left even among theologians.

Duch W. (2017) *Why minds cannot be received, but are created by brains*. Scientia et Fides 5(2), 171-198.

# Goal of Humanities

Humanities, “the human, or spiritual sciences”, in contrast with physical sciences, was introduced at European universities at the end od 19c.

**Bing:** humanities aim to deepen our understanding of the human experience, foster critical thought, promote a sense of understanding of others. Key goals:

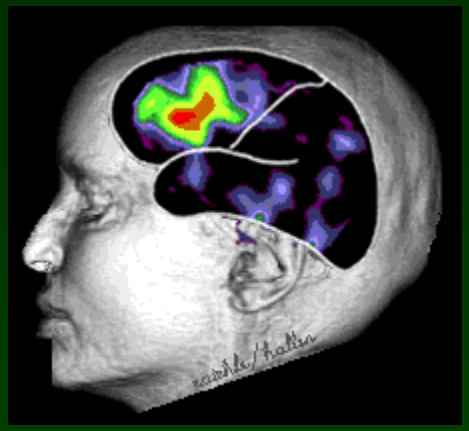
- **Understanding Human Culture**, use analytic and critical methods of inquiry derived from an appreciation of human values. This includes the study of languages, literatures, the arts, history, and philosophy (Britannica).
- **Fostering Intellectual Curiosity**, global knowledge, critical thinking, personal responsibility, and ethical and cultural awareness. Use language effectively and establish a framework for aesthetic appreciation for fine arts (The Humanities General Education Program, OIT).
- **Studying Creative Expressions** of humans that reflect our experiences, feelings, ideas about ourselves, other humans, the past, and the universe.
- **Promoting Collaboration and Social Justice** (Stanford Humanities Center).
- **Respecting and Understanding the World**: Humanities encourage students to respect and understand the world around them (Profoundtips.com).

# Brain-mind presentations 1

List of my presentations on brain-mind:

1. Towards a plausible theory of mind. Birmingham University, UK, 1995
2. How does the brain work? University of Tokyo, 2000
3. Neurophilosophical solution to the hard problem of consciousness. Konstanz, Germany, 2000
4. Mind from brain: psychological spaces and neuroscience. Starlab, Belgium, 2001
5. Attractor neural networks and concept formation in psychological spaces. Lejondal castle, Stockholm, Sweden 2002
6. Conscious mind as a limit of brain-like computing. Pretoria, S. Africa, 2003
7. Large-scale projects to build artificial brains: ABACCUS. Building Artificial Brain, ICANN'2005, Toruń, Poland.
8. Semantic Memory for Avatars in Cyberspace Singapore, 2005
9. Creativity, Intuition, Emotions and Perceptual Learning, School of Humanities and Social Sciences, Singapore, 2005
10. Neurocognitive science: mind from brain? SWPS, Warsaw, Poland, 2006
11. Neurocognitive approach to creativity and higher-level cognition, Poland, 2006

# Brains ⇔ Minds



Define mapping S(M)⇔S(B). BCI: intentions => actions.
How do we describe the state of mind?

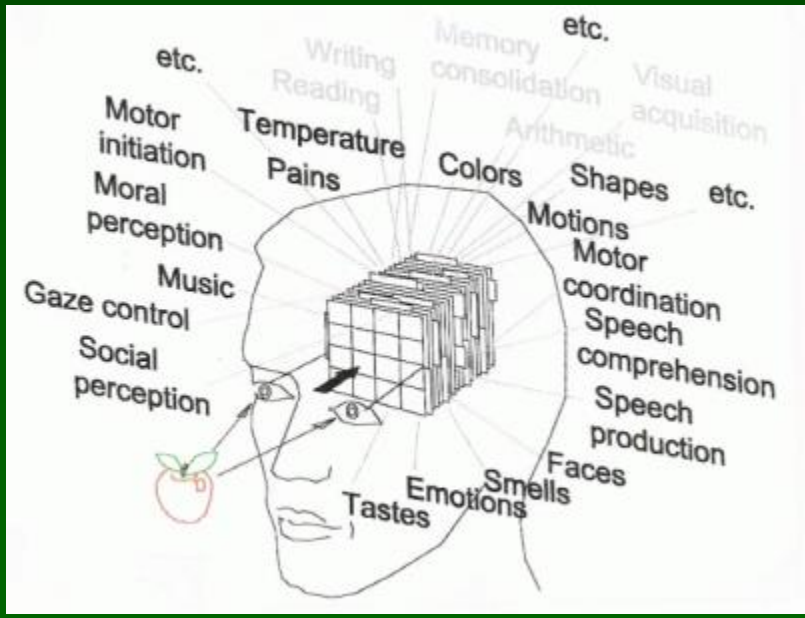
Verbal description is not sufficient, words should be represented in a space with dimensions that measure different aspects of our experience.

Stream of brain states ⇔ movement of thoughts in psychological spaces.

**Two problems**:

1. Discretization of continuous processes for explainable, symbolic models.
2. 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 ...




Eric Schwitzgabel, Perplexities of Consciousness. MIT Press 2011
R.T. Hurlburt, E. Schwitzgabel, Describing Inner Experience? Proponent Meets Skeptic 2007

# Brain-mind presentations 2

12. Neurocognitve NLP. Towards human-level competence, Mayo Clinic, USA, 2006
13. Brain, Mind, Consciousness and the Ghost in the Machine, Cincinnati, 2006
14. Neurocognitive approach to creativity, Yonsei University, Seoul, Korea, 2006
15. Are we automata? Self, Intersubjectivity and Social Neuroscience, Torun, 2007
16. Cognitive Architectures: Where do we go from here? First Conference on the Artificial General Intelligence, Memphis, TN, USA 2008
17. Brains, logic and computational models. Argumentation as a cognitive process, Torun, 2008
18. Consciousness, Imagery and Music, COST Action, Ghent, 2008
19. Free Will and the Brain: Are we Automata? Torun, UMK, 2008
20. Is embodiment necessary for natural language understanding? Torun, UMK, 2008
21. Mind from brain: physics & neuroscience Jagiellonian University, 2008
22. Neurocognitive approach to natural language understanding and creativity, Porvoo, Finland
23. Imagery, Creativity, Brains and Talent. Hitachi Advanced Research Lab. Japan, 2010
24. Consciousness and Creativity in Brain-Inspired Cognitive Architectures. First CHIST-ERA conference. Rome-EUR , 2010

# Brain-mind presentations 3

25. Social intelligence: what we need to understand, European Network for Social Intelligence, Warsaw 2012
26. Dariah and other digital humanities projects. Warsaw University, 2014
27. Grand challenge: Computational Neurophenomics for understanding people's behawior. Cyfronet Kraków 2014
28. Cognitive Science and Music. Cochlear Implants and Music, Warsaw 2015
29. Neurobiological foundations of ethics and law. Humboldt-Kolleg, 2015
30. Neurolinguistics: what do we want to achieve, UAM Poznań, 2015.
31. Signs of consciousness in humans and machines. X Congress of Polish Philosophy. UAM Poznań, 2015.
32. From brain to mind - challenges for mathematicians. Banacha Center, Warsaw 2016
33. What is digital world doing to our brains? Kraków, 2016
34. Brains and the limits of self-knowledge. Limits of knowledge. Kraków, 2017
35. Mapping psychological concepts on higher order brain dynamics, Zakopane 2017
36. Imagery agnosia, KU Leuven 2016
37. Self as a function of the brain, Theology Dept, NCU Toruń.
38. Communication as a resonance between cognitive systems. Wroclaw 2016

# Brain-mind presentations 4

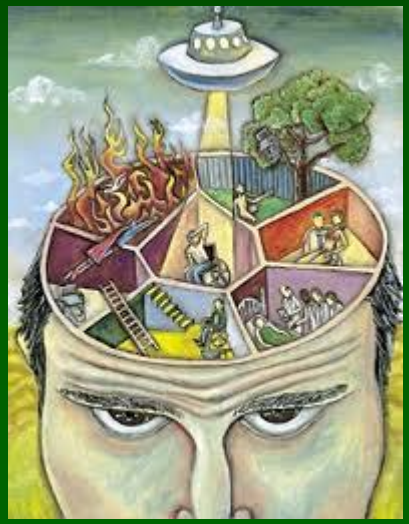
39. Brain and Machine Learning inspirations (3 lectures), Zakopane, 2018
40. Fingerprints of brain cognitive activity. ICAISC 2018, Zakopane, 2018
41. Optimizing brain processes. Neurocognitive technologies. University of Surrey, 2018
42. Representation of concepts in brain networks. NCU Toruń, 2018
43. Multi-level explanations in neuroscience (3 lectures), Val di Sole Monastery, 2019
44. Analysis of neurodynamics for diagnosis of mental states. Zakopane 2019
45. How to understand brain-mind. 10 years of Cognitive Science at NCU, Toruń 2019
46. AI for better brains. Artificial Intelligence: Art or Science? SISSA, Trieste, 2019
47. High performance computing and neuropsychiatry. Oak Ridge NL, TN USA, 2019
48. AI and Neurocognitive Technologies for Human Augmentation Thailand, 2020
49. Brain-inspired cognitive computing KIiA PAN-UMK-UTP seminar, Poland, 2020
50. Searching for fingerprints of brain activity. Huazhong UST, Wuhan, China, 2021
51. Human enhancement and the future of BCI, ACM ISS 2021
52. What is needed to fully understand mental processes? Virtual conference, 2021
53. Brain-inspired cognitive computing, ICAISC 2021, Zakopane, 2021
54. Brain-Computer-Brain Interfaces, CD DAMSI, Toruń, 2021
55. Brains and evolution of culture. UAM Poznań, 2022
56. BMIs for human enhancement: what has been done and what is coming? Prague 2022
57. Mental states in brains and computers. SANO Kraków, 2022.
58. Recipe for a conspiracy theory. Memes and neuroscience, NCU 2022

# Understanding human nature

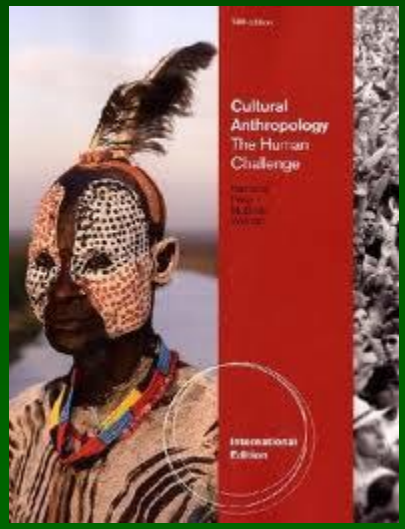
Psychophysics


Neuroscience

Psychology



Ethology
Sociocultural
Anthropology




+ Philosophy + Sociology … => The Science of Human Experience

# Brains

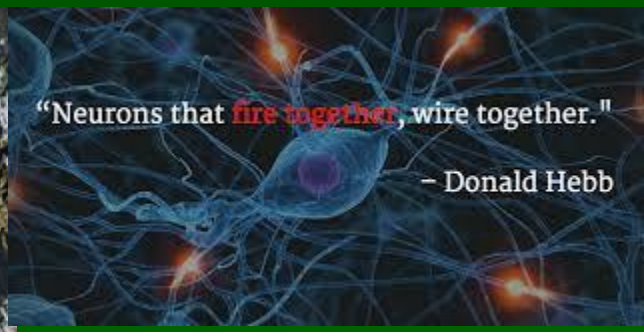
# Erosion

"Where do our inclinations come from?" - asked King Milinda the Buddhist sage Nagasena 1,600 years ago.

N - When it rains, where does the water go?
M - It will flow down the sloping ground.

N - And if the rain fell again, where would the water flow to?
M- It would flow in the same direction as the first water.

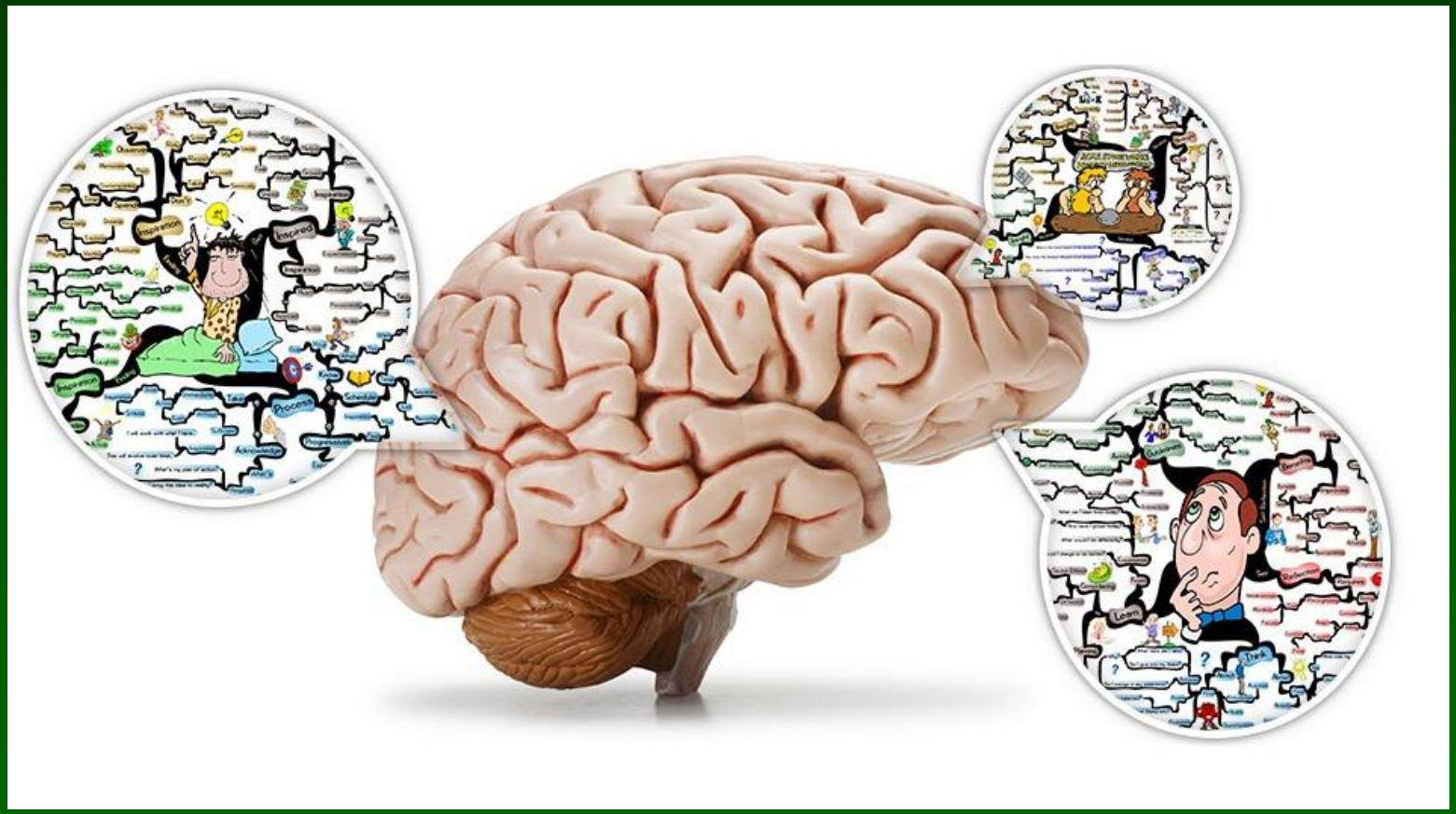
The new builds on the learned. This is essence of the Hebb's principle (1949): neurons that fire together wire together. The order of learning is important.



# Learning

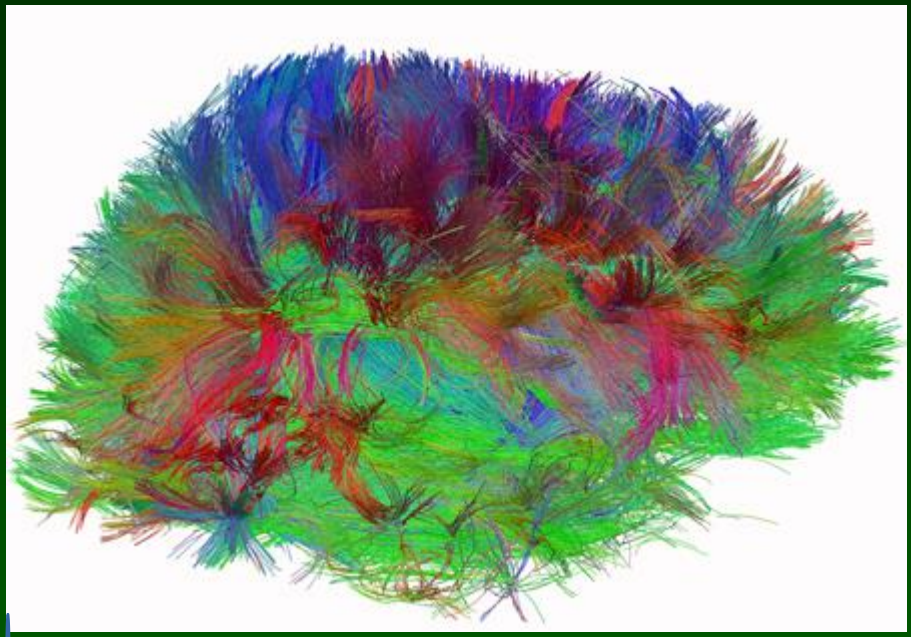
The order of learning is important.

First learn science, understand the world, than enjoy philosophy or religion, through the lens of science.



Iqmatrix.com, many mind maps".

# Neuronal determinism



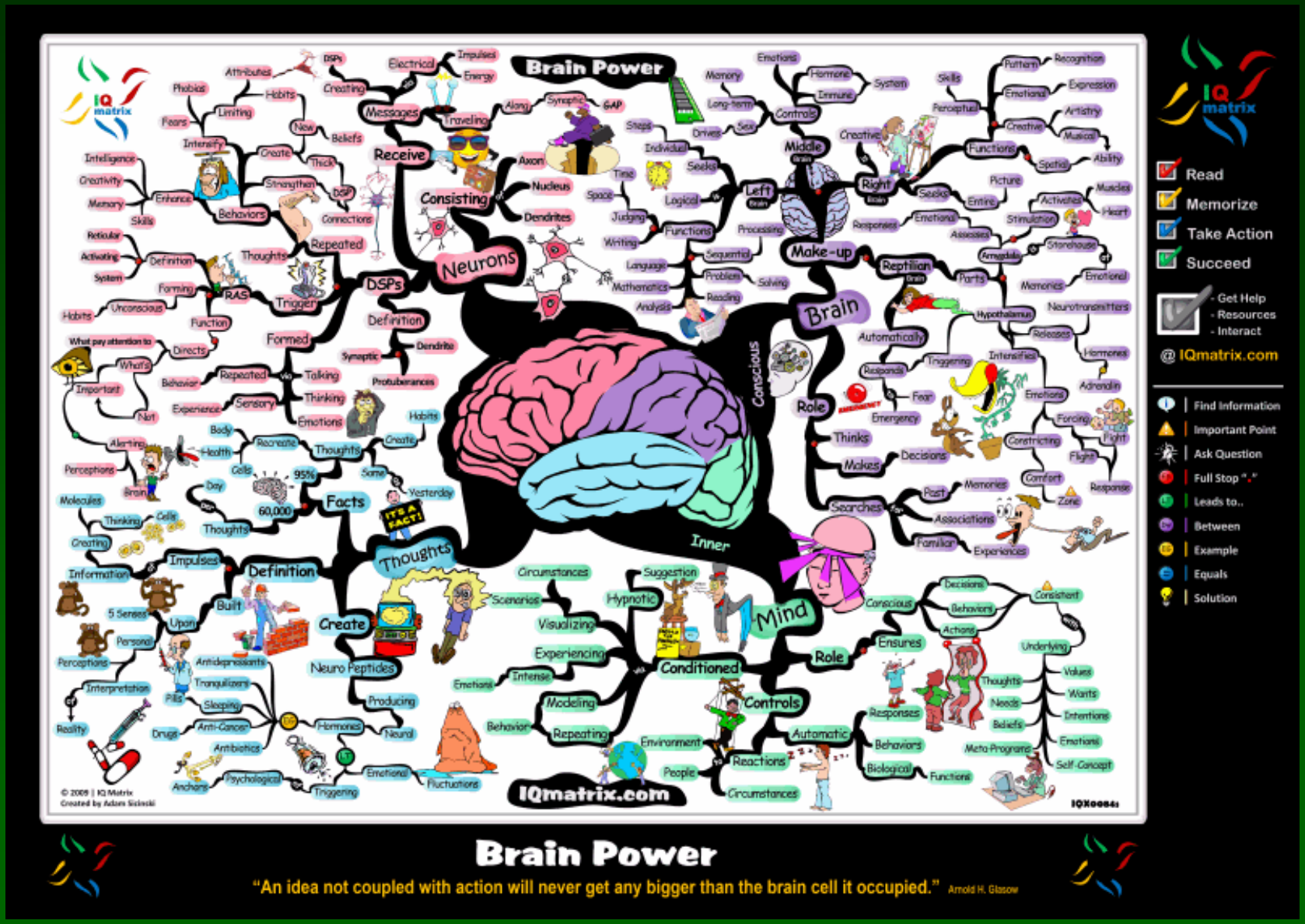
Genetic determinism imposes general constraints on the efficiency of brains, it is better to have numerous "wrinkles" and "hairy" brains than smooth and combed.

Neural activity (neurodynamics) determines what comes to mind.
Neuronal determinism: brain shaped by life experiences, upbringing, education, forms associations, thoughts, emotions. Connectome and neural properties are the key.

Genes ⇔ Brain/body
⇑⇔ Environment ⇔⇓

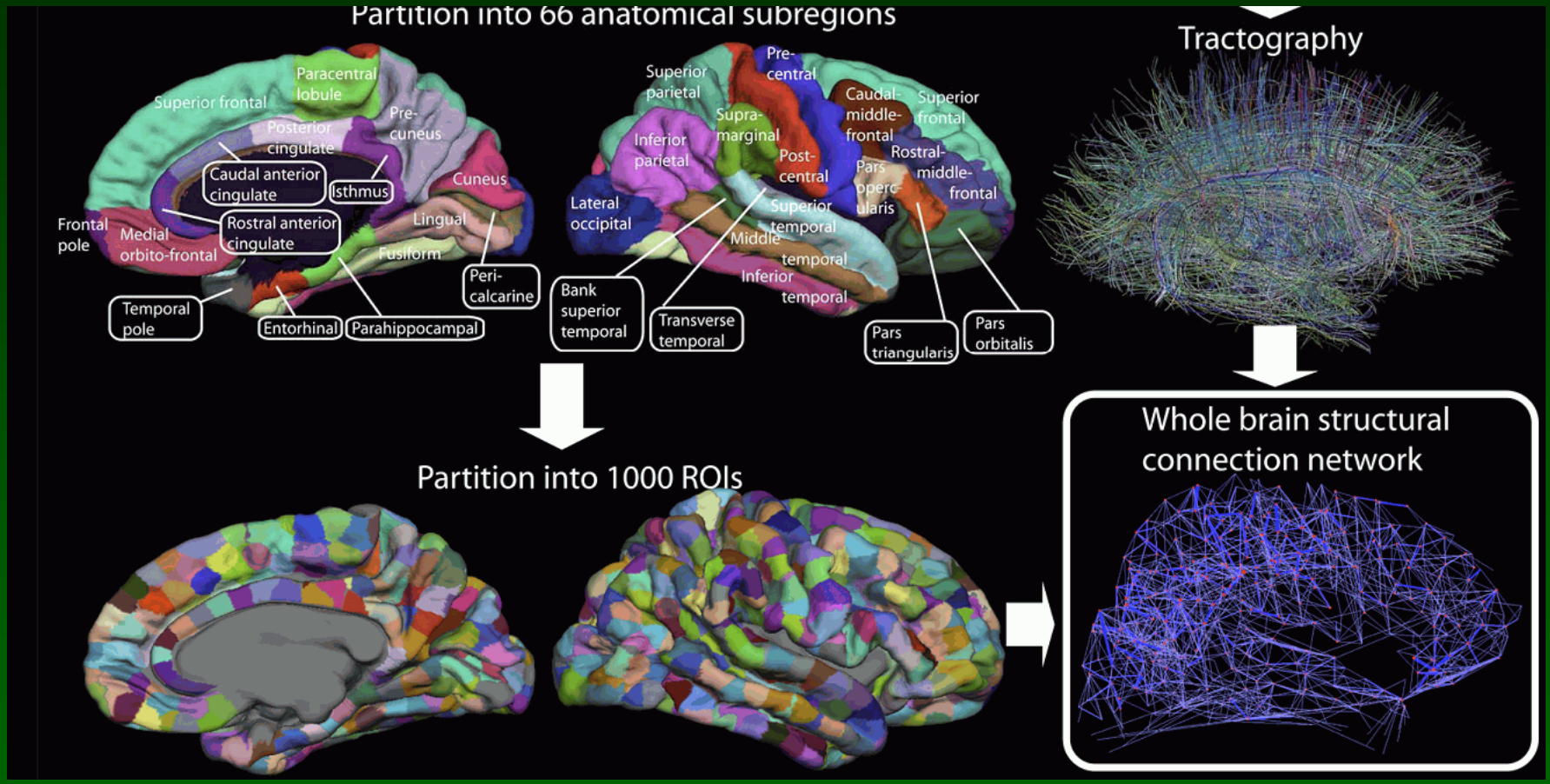
**Metaphor: mind is a shadow of brain activity (neurodynamics)**.

# Brain and mind



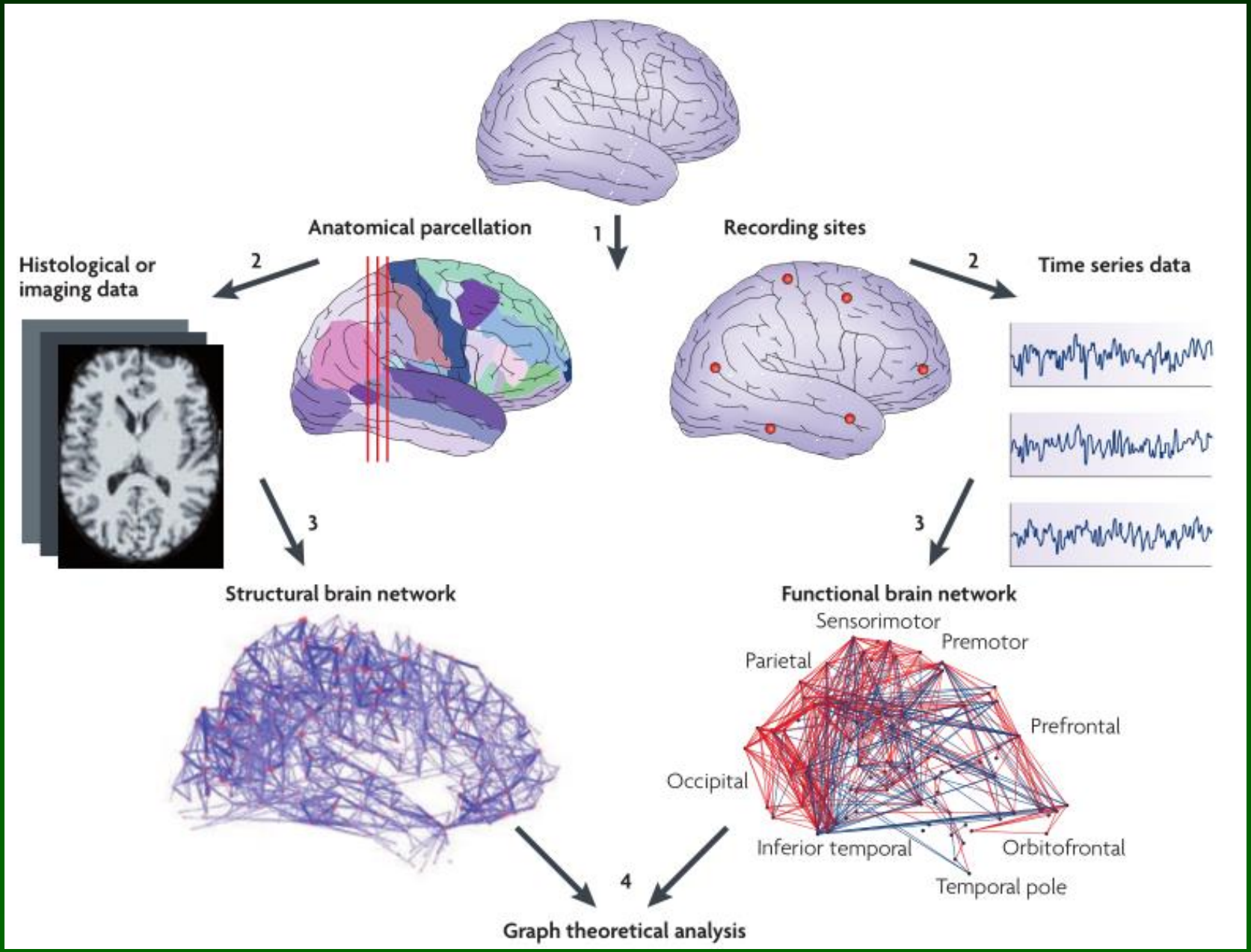

Iqmatrix.com, many mind maps".

# Connectome




Most functions require cooperation of different brain regions. We need to understand what these regions (ROI) do and how they are connected. Initially we had Brodmann areas with 52 regions. Now we have maps 100's. Relationship viewer.

# Structure and functions



# Conscious Perception

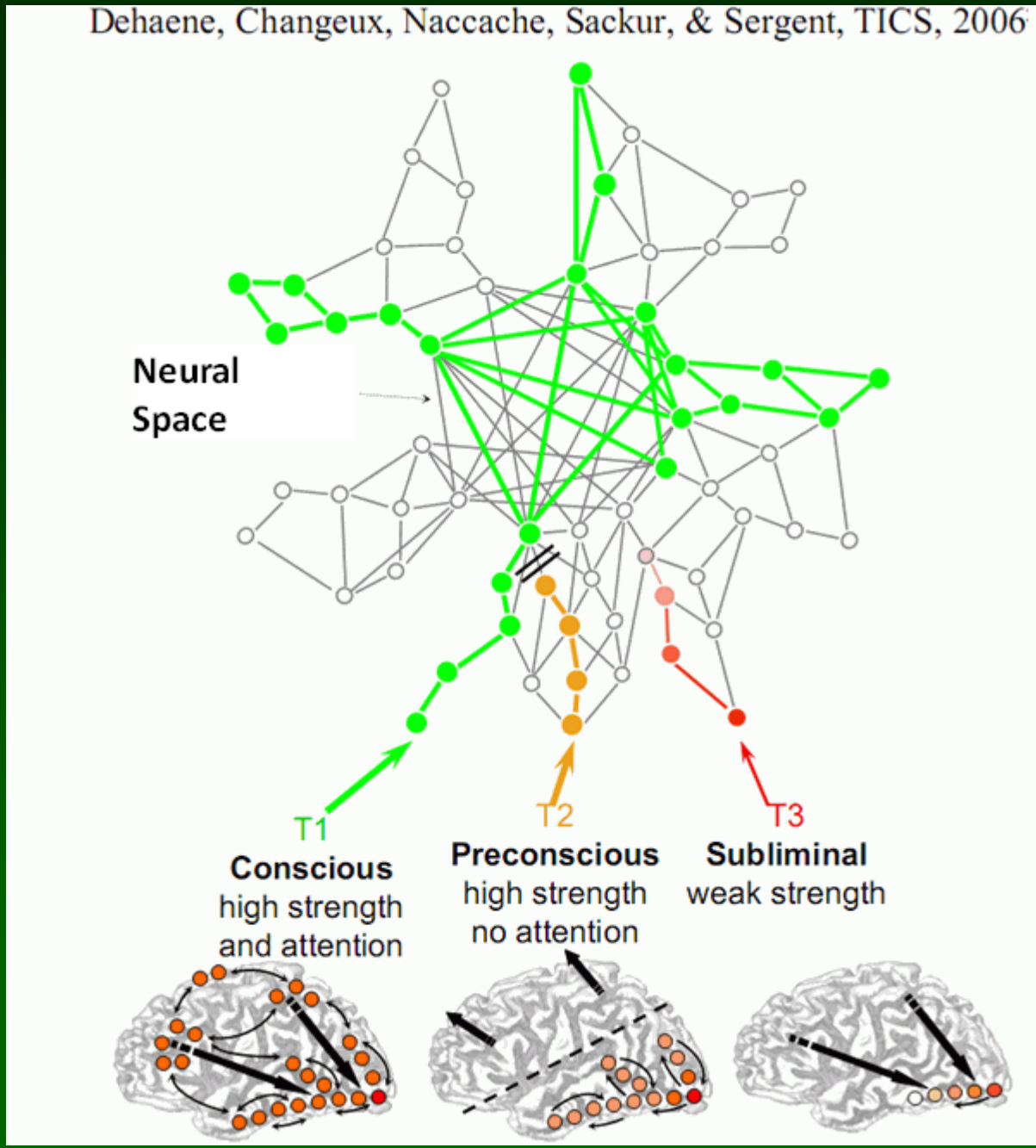
Very little of what passes In the brain is perceived.

Attention + stimulation is needed to create brain states that are persistent and can be distinguished from noise.

Attention: 20 Hz

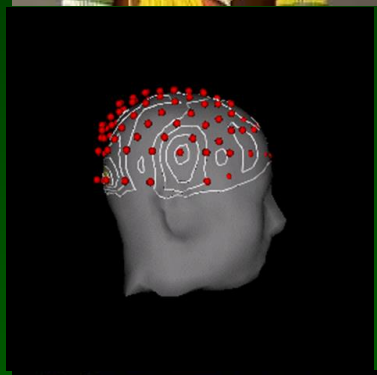
Perception: 40 Hz

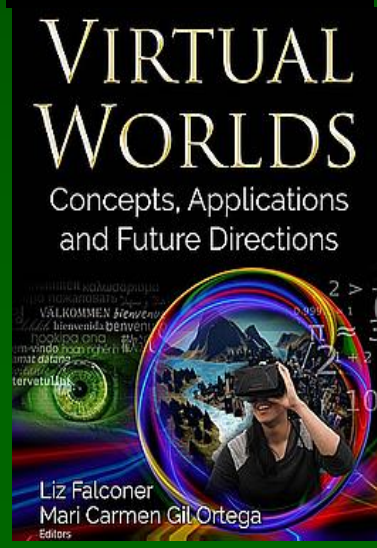
C. Gilbert, M. Sigman, Brain States: Top-Down Influences in Sensory Processing. Neuron 54(5), 677-696, 2007



# Understanding ourselves

- Understanding human mind and culture
  ⇔ deep understanding of human behavior
  ⇔ mental processes and their sources.
- Wonders of equilibria: from genes to culture.
- Learning how brains enable art, literature, culture.
- Phenomics: multi-level foundations of processes creating brain states, memory, cognition, mental life.
- Neuroplasticity and conspiracies in the brain.
- Future: brain-machine interfaces, neurotechnologies, augmentation of the brain, virtual worlds, artificial intelligence, cyborgization and transhumanism.






Duch (2024) Artificial intelligence and the limits of the humanities. Er(r)go - Humanities (in print).

Duch W. (2021). *Memetics and Neural Models of Conspiracy Theories*. Patterns 2(11), 2-13.

More papers on these topics.

# Explanations

Seconds

Minutes

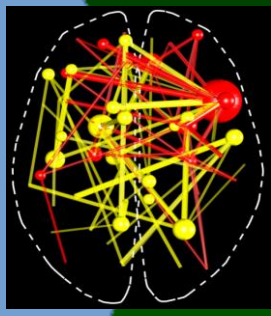
Days

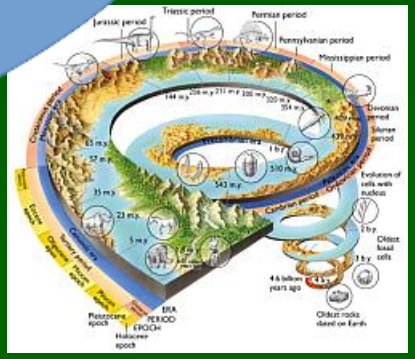
Month

Years

Millenia

Eras/Eons





Cognitive phenomics, fast/slow: neurodynamics, hormones, education, culture, infancy, gestation and evolution.

# Environment and culture

T. Talhelm et al., Large-Scale Psychological Differences Within China Explained by Rice Versus Wheat Agriculture. Science 344 (2014).

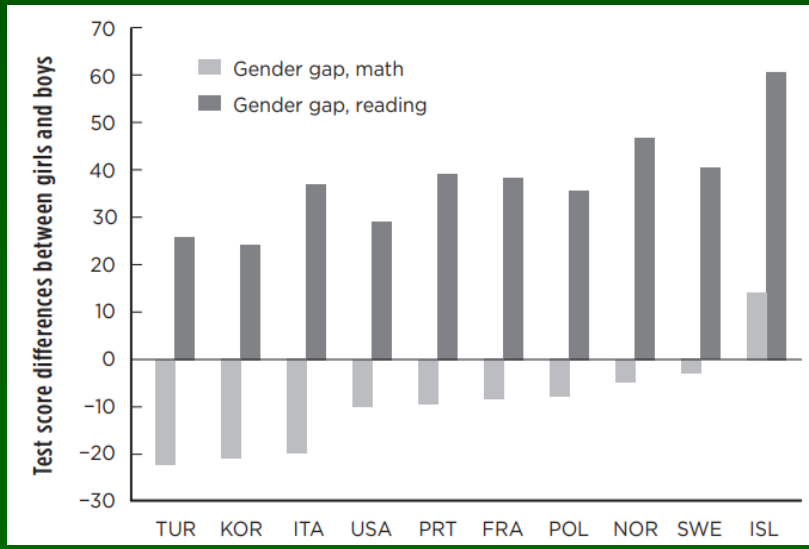
Individualist (USA) versus collectivist (China) cultures: wheat vs. rice. Divorce, inventiveness ⇔ 7R dopamine DRD4 receptor variants vs. 4R variants.

Behavior ⇔ ecosystem, climate.

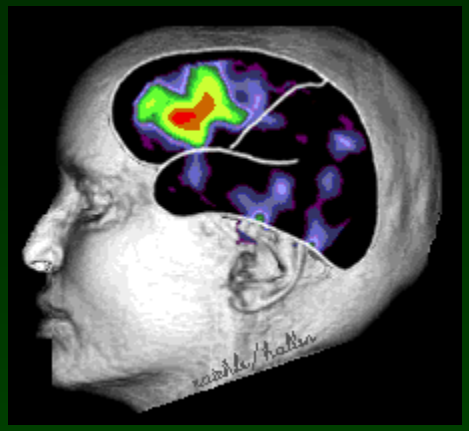
Brains ⇔ Culture ⇔ Genetics ⇔ Brains

Generalizations based on sex are **meaningless**. Scores of boys and girls in math and reading depend on culture.

Caucher Birkar, Kurdish refugee in UK, got 2018 Fields medal.



# Brains ⇔ Minds

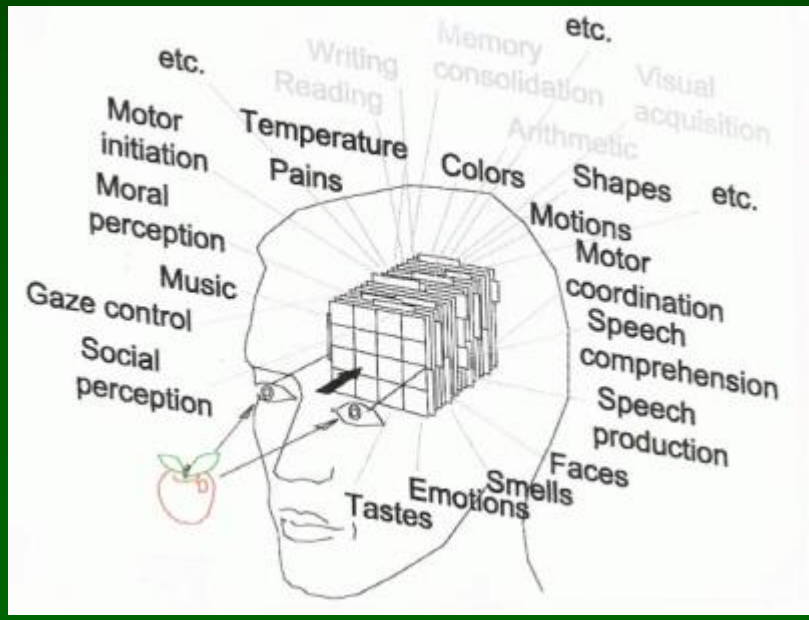


Define mapping S(M)⇔S(B). BCI: intentions => actions. How do we describe the state of mind?

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 explainable, symbolic models, and lack of good phenomenology – we are not able to describe our mental states.

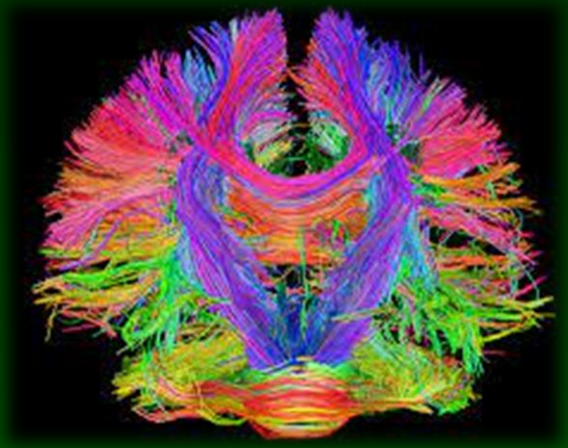


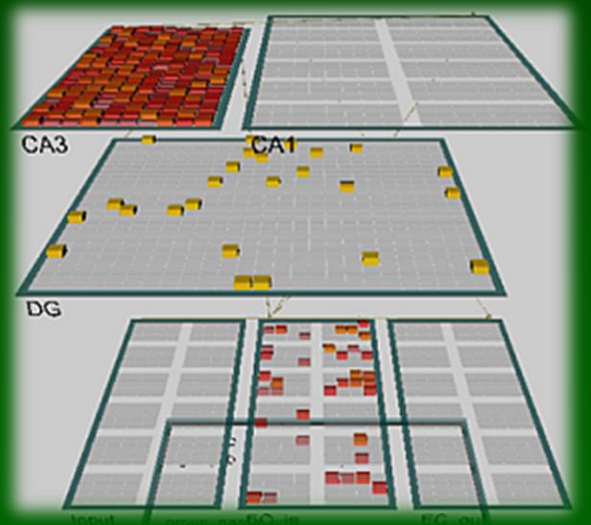

Where do my brain states come from?

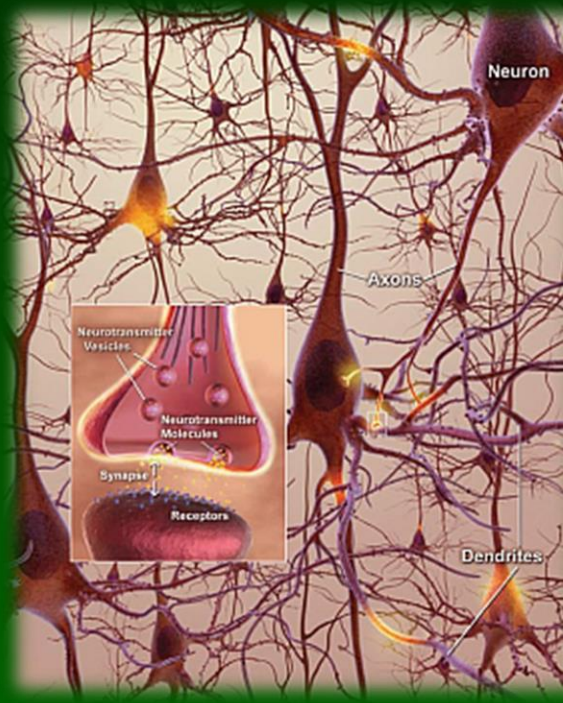
Understanding neurodynamics: bioelectrical activity of the brain, measured using EEG, MEG, NIRS-OT, PET, fMRI and behavioral techniques ...

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

# From Behavior to Neurons

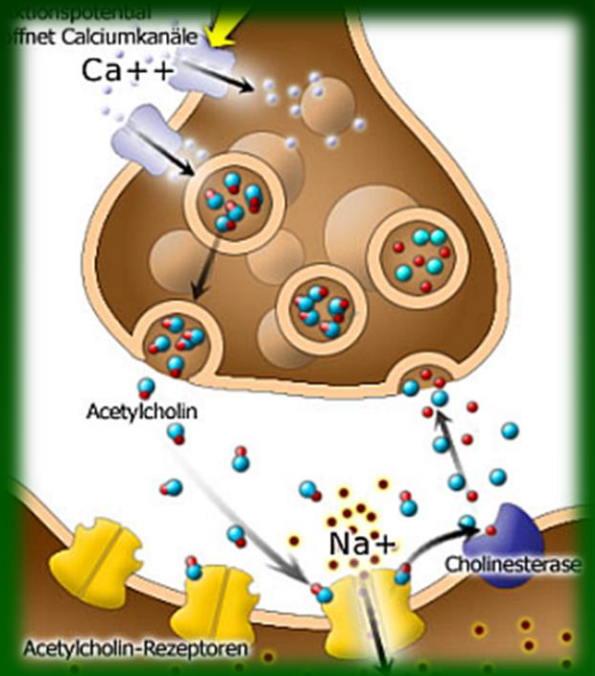


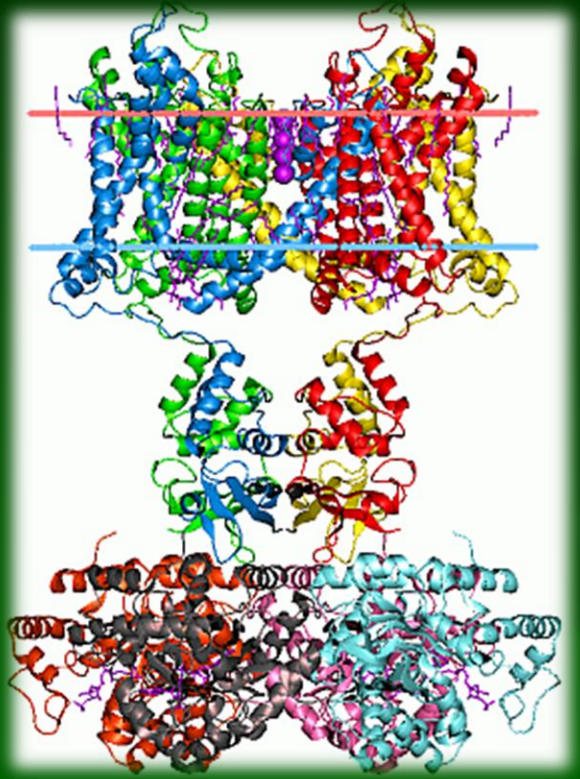


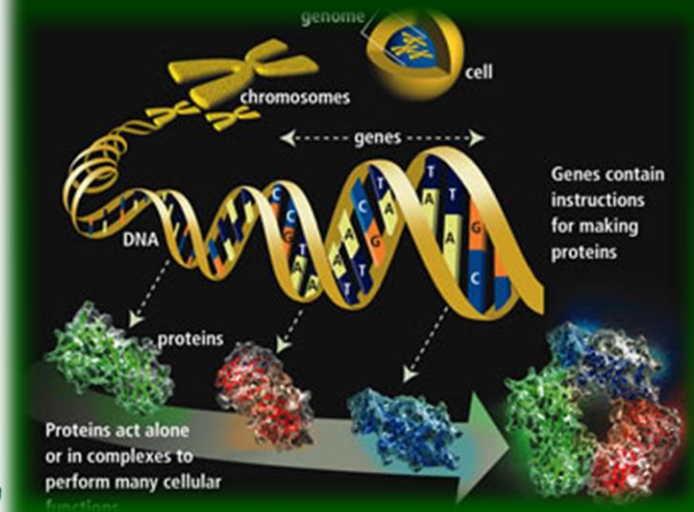




**=> neurodynamics**

# From Genes to Neurons






synapses, soma, receptors, ion channels => proteins => genes

# Neuropsychiatric phenomics

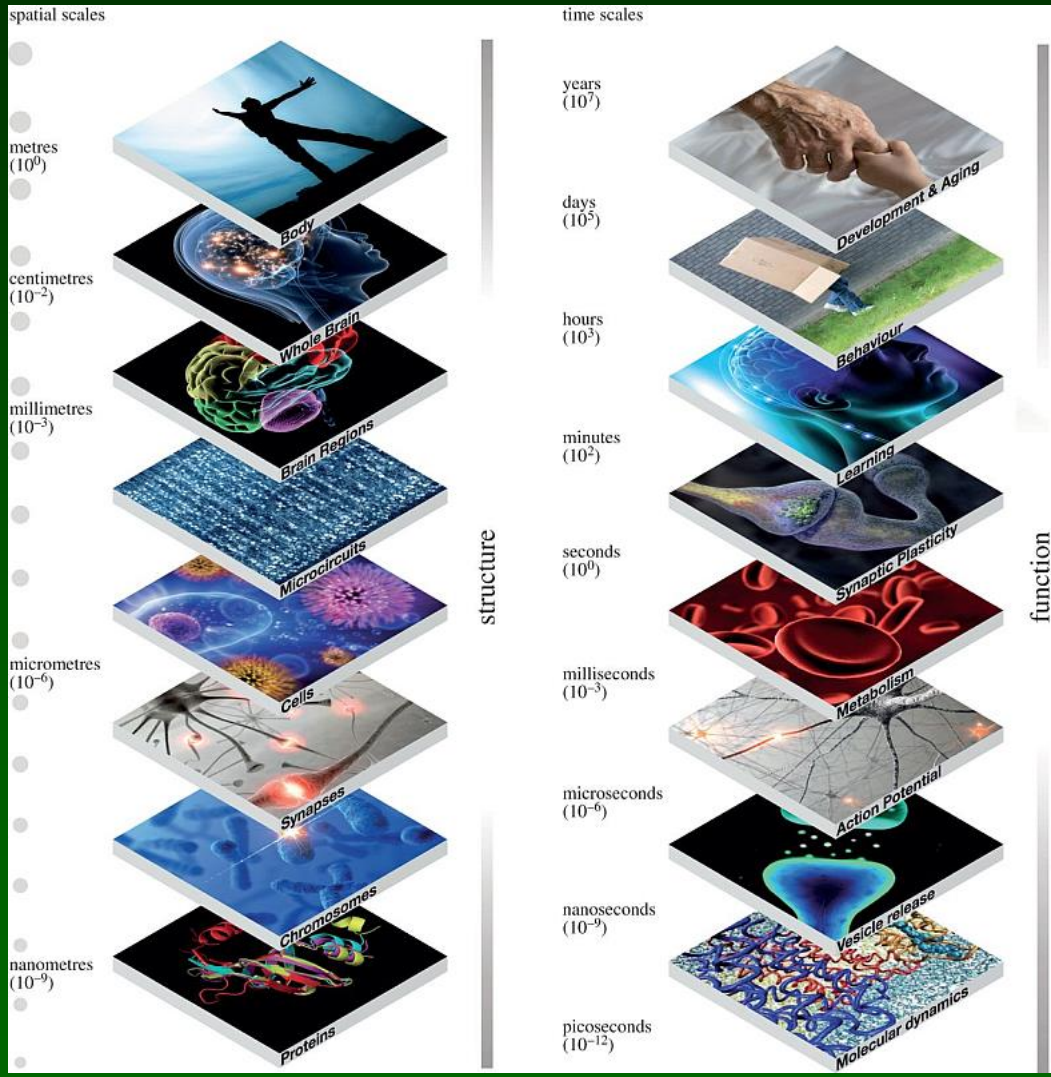
2008: The Consortium for Neuropsychiatric Phenomics

"… categories, based upon presenting signs and symptoms, may not capture fundamental underlying mechanisms of dysfunction" (Insel et al., 2010).

New approach: RDOC NIMH.

Description of organisms at different levels will help to answer different types of questions.

Network level is in the middle and can be connected to the mental level via computational models.

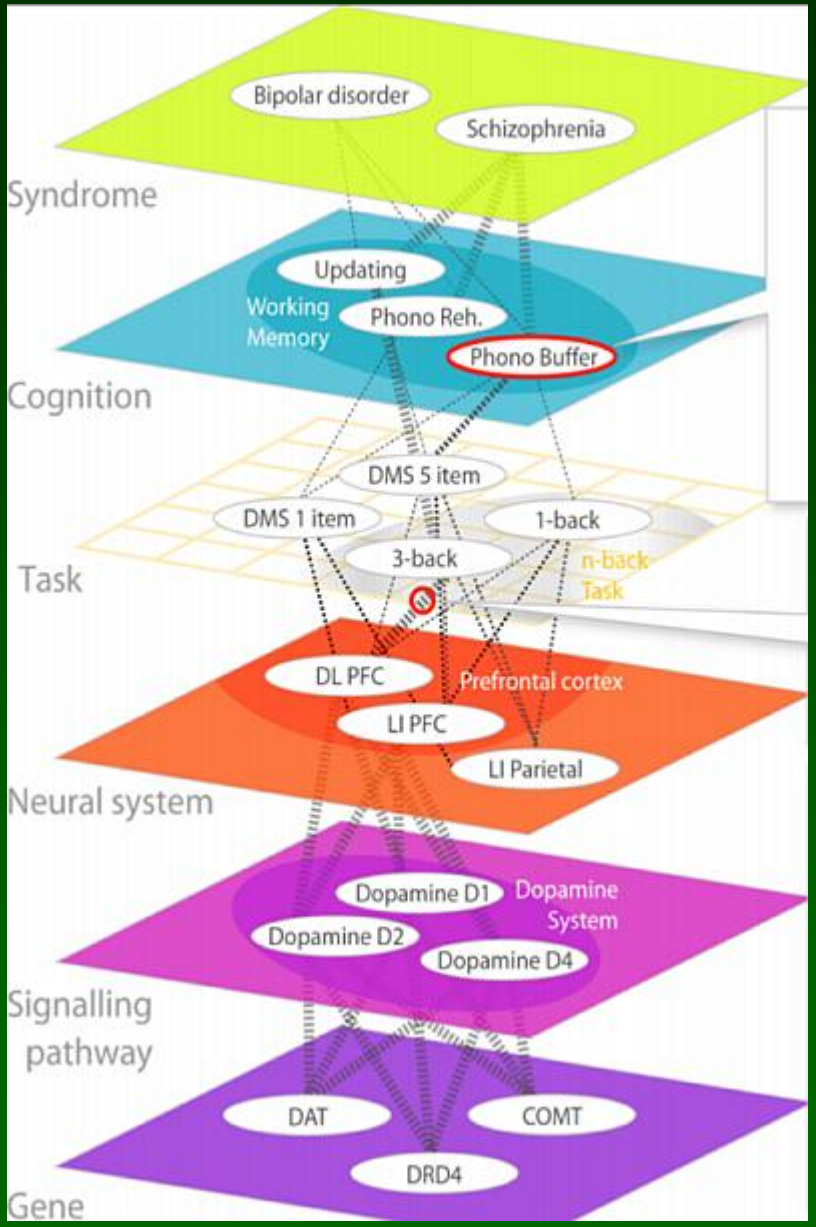


# Neuropsychiatric Phenomics Levels

According to
The Consortium for Neuropsychiatric Phenomics (CNP)
http://www.phenomics.ucla.edu

From genes to molecules to neurons and their systems to tasks, cognitive subsystems and syndromes.

Neurons and networks are right in the middle of this hierarchy.



RDoC
Research Domain Criteria Initiative

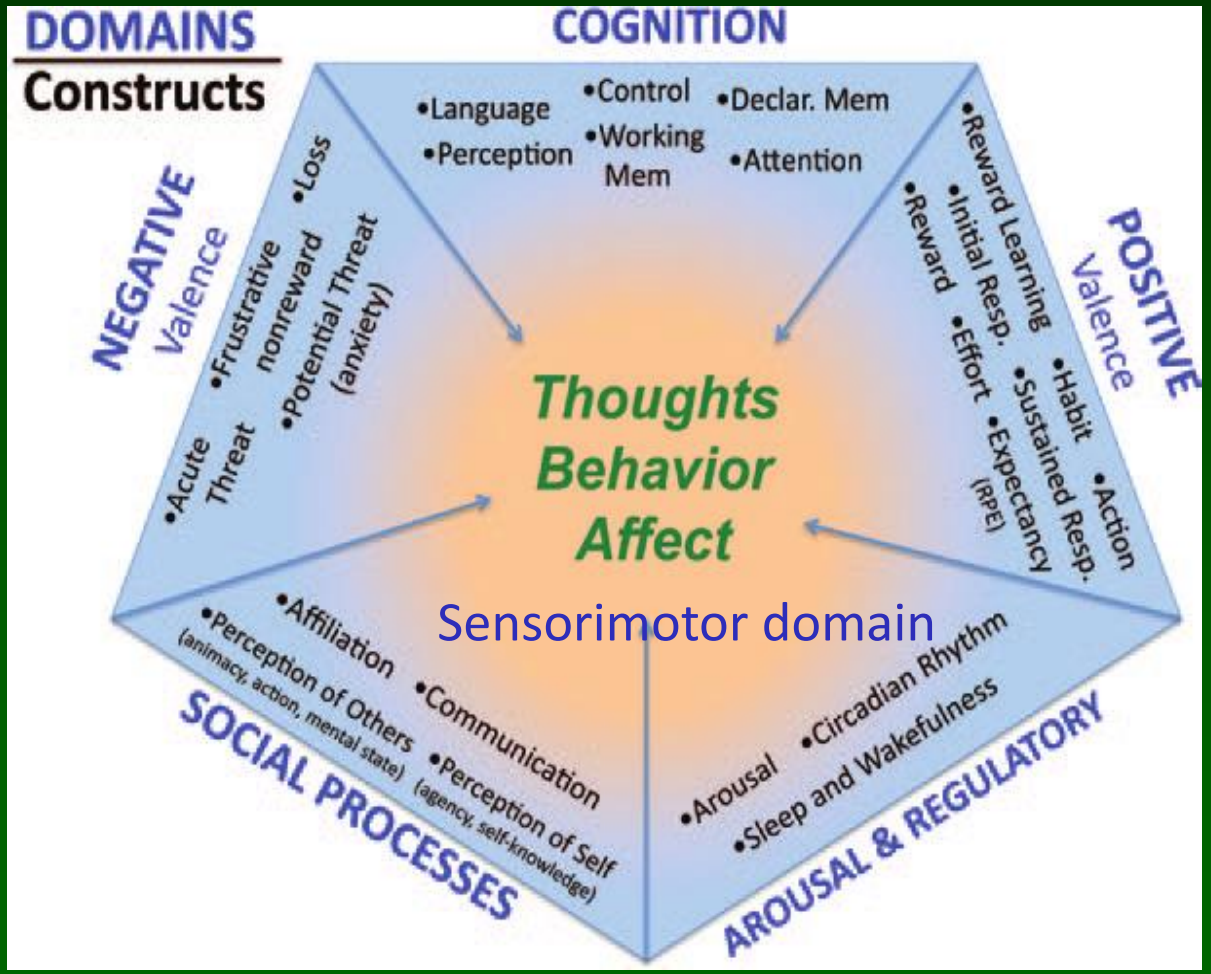
NIMH RDoC Matrix for analysis of (de)regulation of 6 large brain systems.

Psychological constructs are necessary to talk about mental states.

Regulation of these 6 large networks forms the basis of all human behavior.

**Sensorimotor systems** added in Jan. 2019 as the sixth brain system.

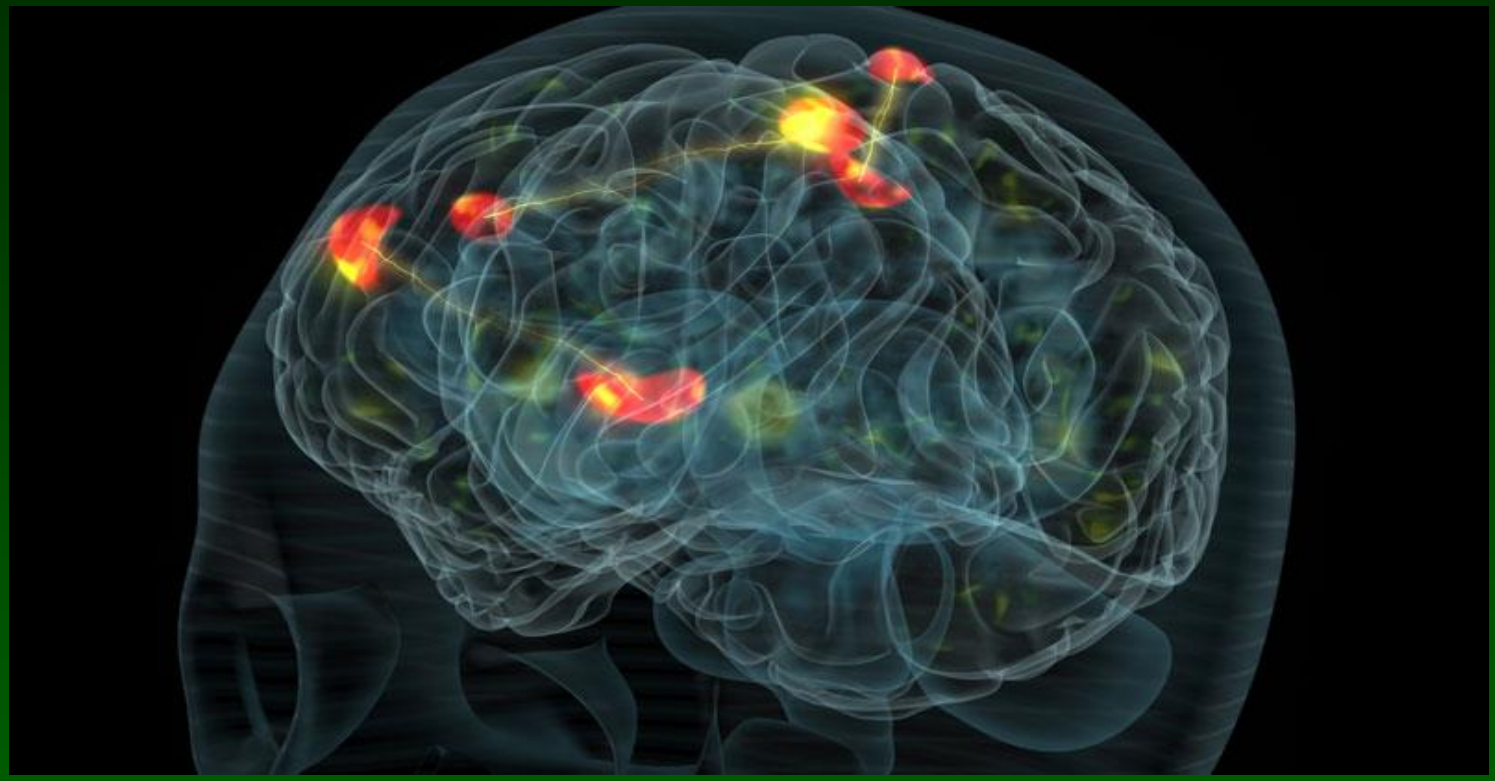
How are these functions implemented in the brain and what contributes to their activity?



# RDoC Matrix for „cognitive domain”

| Construct/Subconstruct | | Genes | Molecules | Cells | Circuits | Physiology | Behavior | Self-Report | Paradigms |
|---|---|---|---|---|---|---|---|---|---|
| Attention | | Elements | Elements | Elements | Elements | Elements | Elements | | Elements |
| Perception | Visual Perception | Elements | Elements | Elements | Elements | Elements | Elements | Elements | Elements |
| | Auditory Perception | Elements | Elements | Elements | Elements | Elements | Elements | Elements | Elements |
| | Olfactory/Somatosensory/Multimodal/Perception | | | | | | | | Elements |
| Declarative Memory | | Elements | Elements | Elements | Elements | Elements | Elements | Elements | Elements |
| Language | | Elements | | | Elements | Elements | Elements | Elements | Elements |
| Cognitive Control | Goal Selection; Updating, Representation, and Maintenance ⇒ Focus 1 of 2 ⇒ Goal Selection | | | | Elements | | | Elements | Elements |
| | Goal Selection; Updating, Representation, and Maintenance ⇒ Focus 2 of 2 ⇒ Updating, Representation, and Maintenance | Elements | Elements | Elements | Elements | Elements | Elements | Elements | Elements |
| | Response Selection; Inhibition/Suppression ⇒ Focus 1 of 2 ⇒ Response Selection | Elements | Elements | Elements | Elements | Elements | Elements | Elements | Elements |
| | Response Selection; Inhibition/Suppression ⇒ Focus 2 of 2 ⇒ Inhibition/Suppression | Elements | Elements | Elements | Elements | Elements | Elements | Elements | Elements |
| | Performance Monitoring | Elements | Elements | | Elements | Elements | Elements | Elements | Elements |
| Working Memory | Active Maintenance | Elements | Elements | Elements | Elements | Elements | | | Elements |
| | Flexible Updating | Elements | Elements | Elements | Elements | Elements | | | Elements |
| | Limited Capacity | Elements | Elements | | Elements | Elements | | | Elements |
| | Interference Control | Elements | Elements | Elements | Elements | Elements | | | Elements |

# Mental state: strong coherent activation



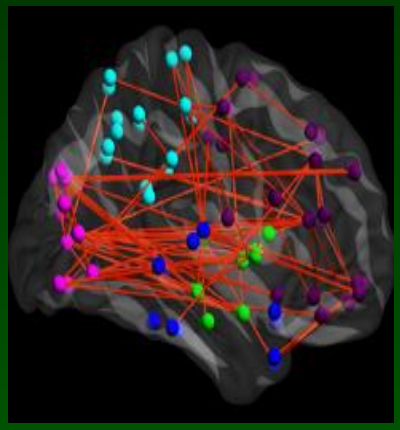
Many processes go on in parallel, controlling homeostasis and behavior. Most are automatic, hidden from our Self. What goes on in my head?

Various subnetworks influenced by hundreds of factors compete for access to the highest level of control – consciousness.
The winner-takes-most mechanism leaves only the strongest.
How to extract stable intentions from such chaos? BCI is never easy.

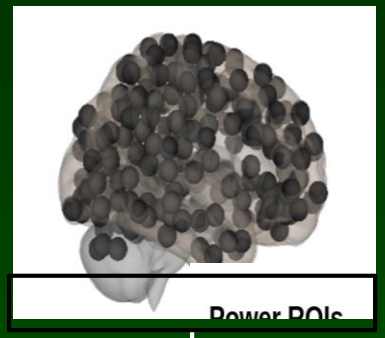
# Human connectome and MRI/fMRI

Structural connectivity

Functional connectivity

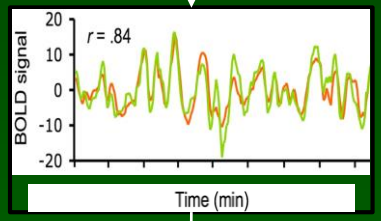


Node definition (parcelation)

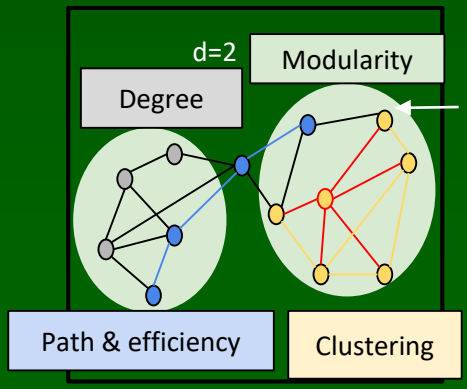



Signal extraction

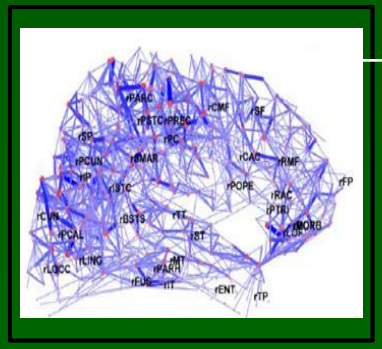
Correlation calculation



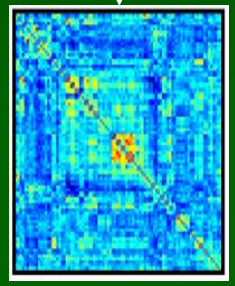

Graph theory




Whole-brain graph



Binary matrix



Correlation matrix

Many toolboxes are available for such analysis.

Bullmore & Sporns (2009)

# Brains vs. computers

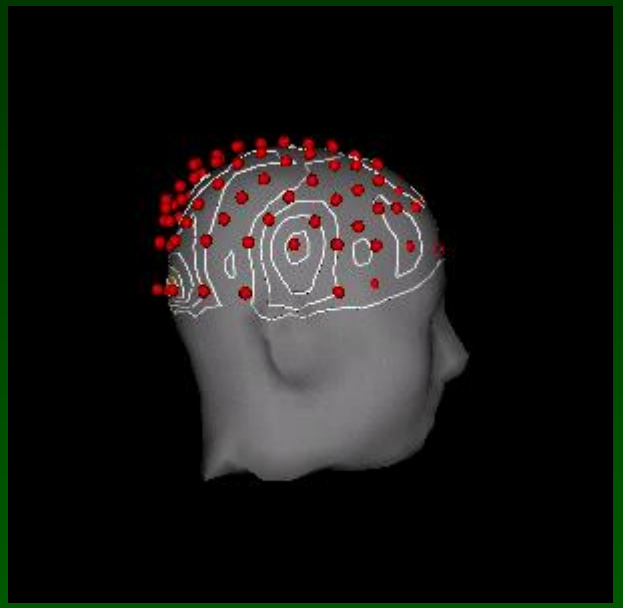
Brains: neurodynamics, continuously changing activation of the brain in space and time.

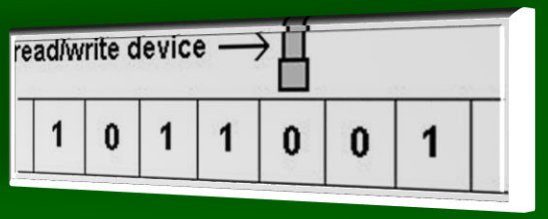
Computer registers: no space, time irrelevant, counting bits in central processors.

Brain states: distributed neurodynamics, each brain state partially contains in itself many associations, relations, other states.

Mind states are internal interpretations of attractor states.

Computer programs and robots do not work like neurodynamics, arising of attractor states.
Large language model can simulate it.
Analog neurochips may form such dynamics.






W. Duch, Brain-inspired conscious computing architecture.
Journal of Mind and Behavior 26, 1-22, 2005

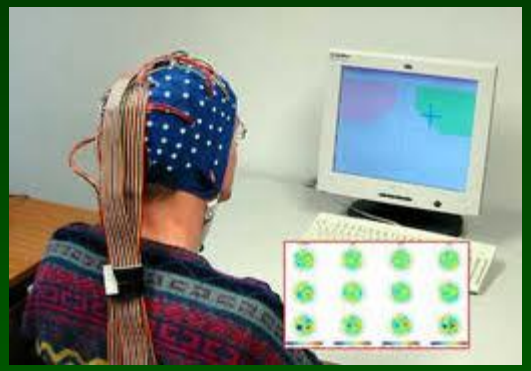
# Brain-computer interfaces

Mind reading is an exciting and rapidly developing field.

Brain-computer interfaces (BCI) read and interpret the activity of the brain.

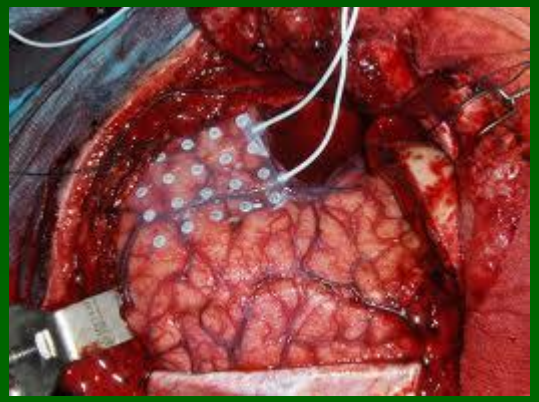
Conscious, intentional activity is detected.

BCI development is motivated by the desire to communicate with people in locked-in or minimal consciousness states (and games -;).



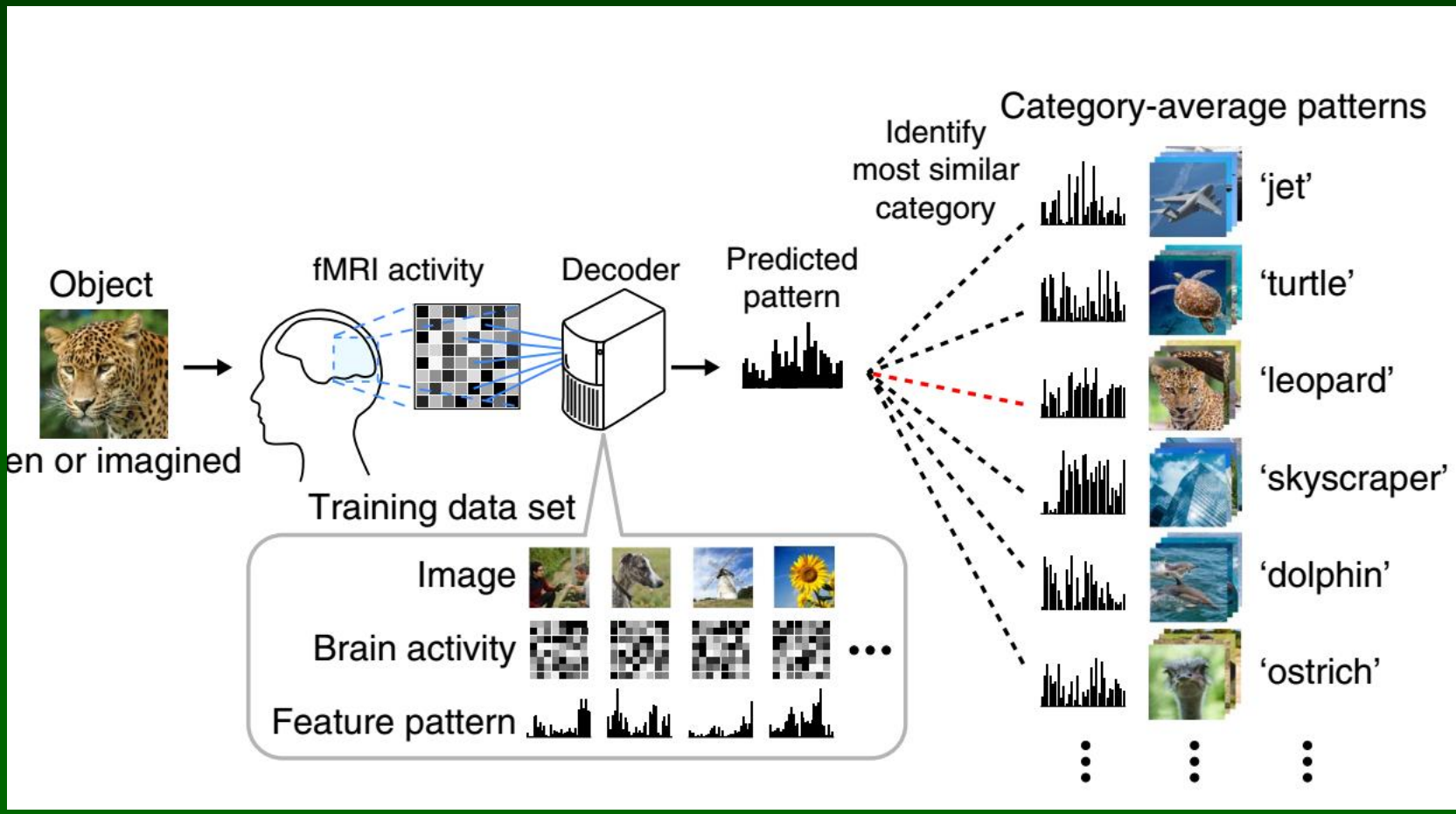
Can we detect signs of consciousness in the same way in artificial brains?

Can we communicate creating resonance states coupling human-robot brains?



# Brain activity ⇔ 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 categorization (~20 categories).

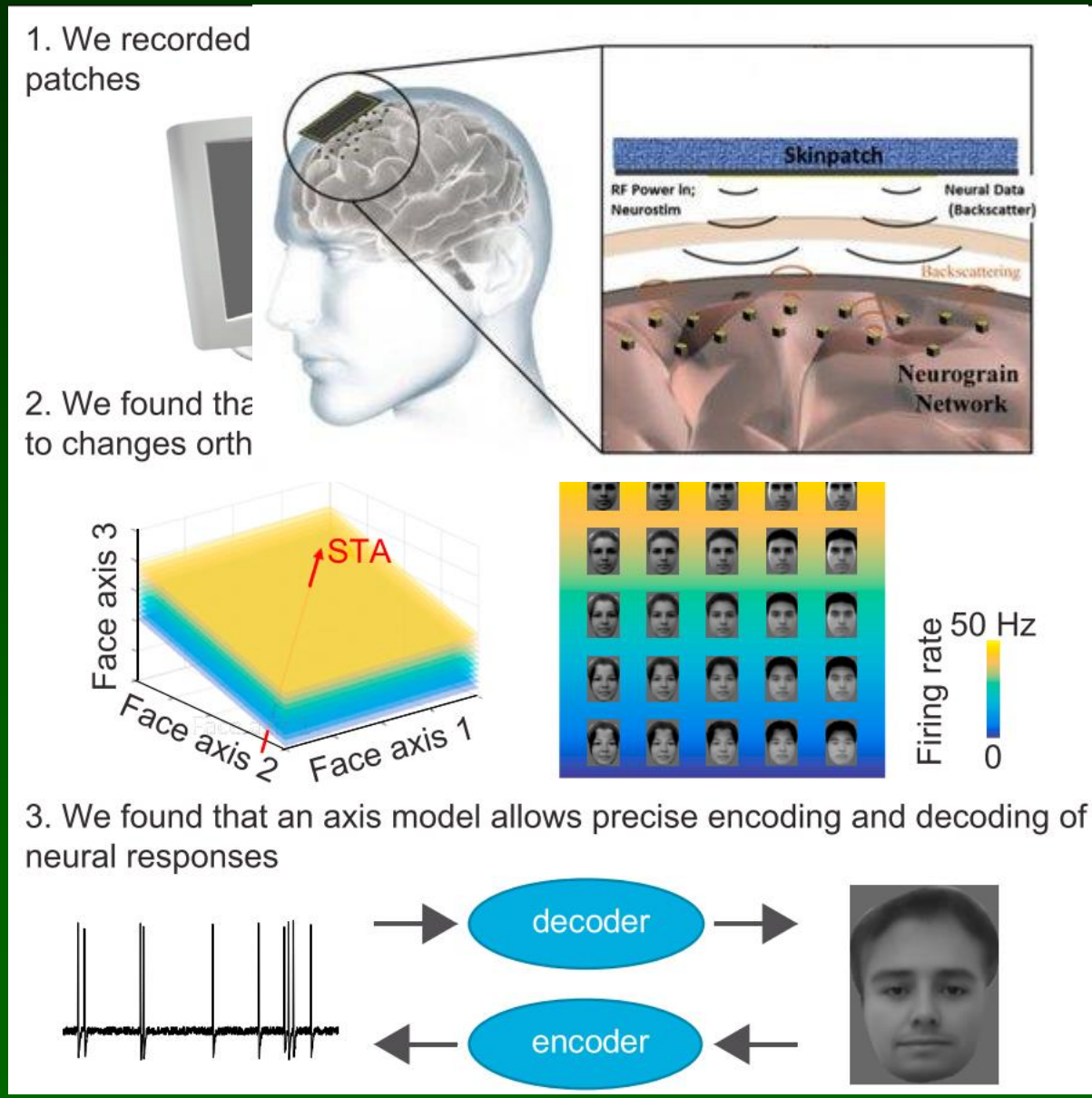
Dreams, thoughts … soon we will use AI to make movies of our dreams?

# 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.

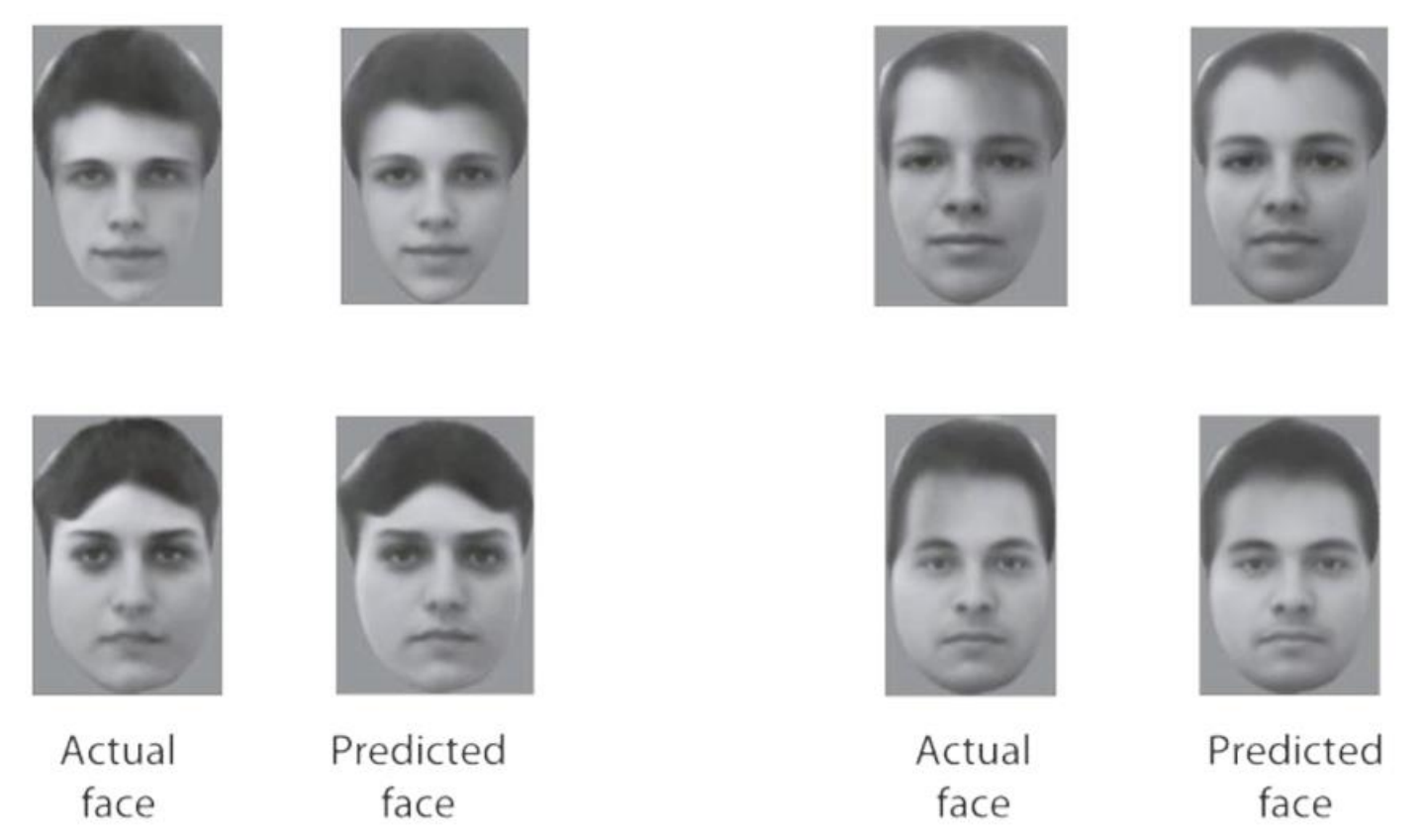
L. Chang and D.Y. Tsao, "The code for facial identity in the primate brain," *Cell* 2017

Elon Musk Neurolace, DARPA projects: put million nanowires in the brain! Use all to read neural responses, 10% to activate specific neurons.

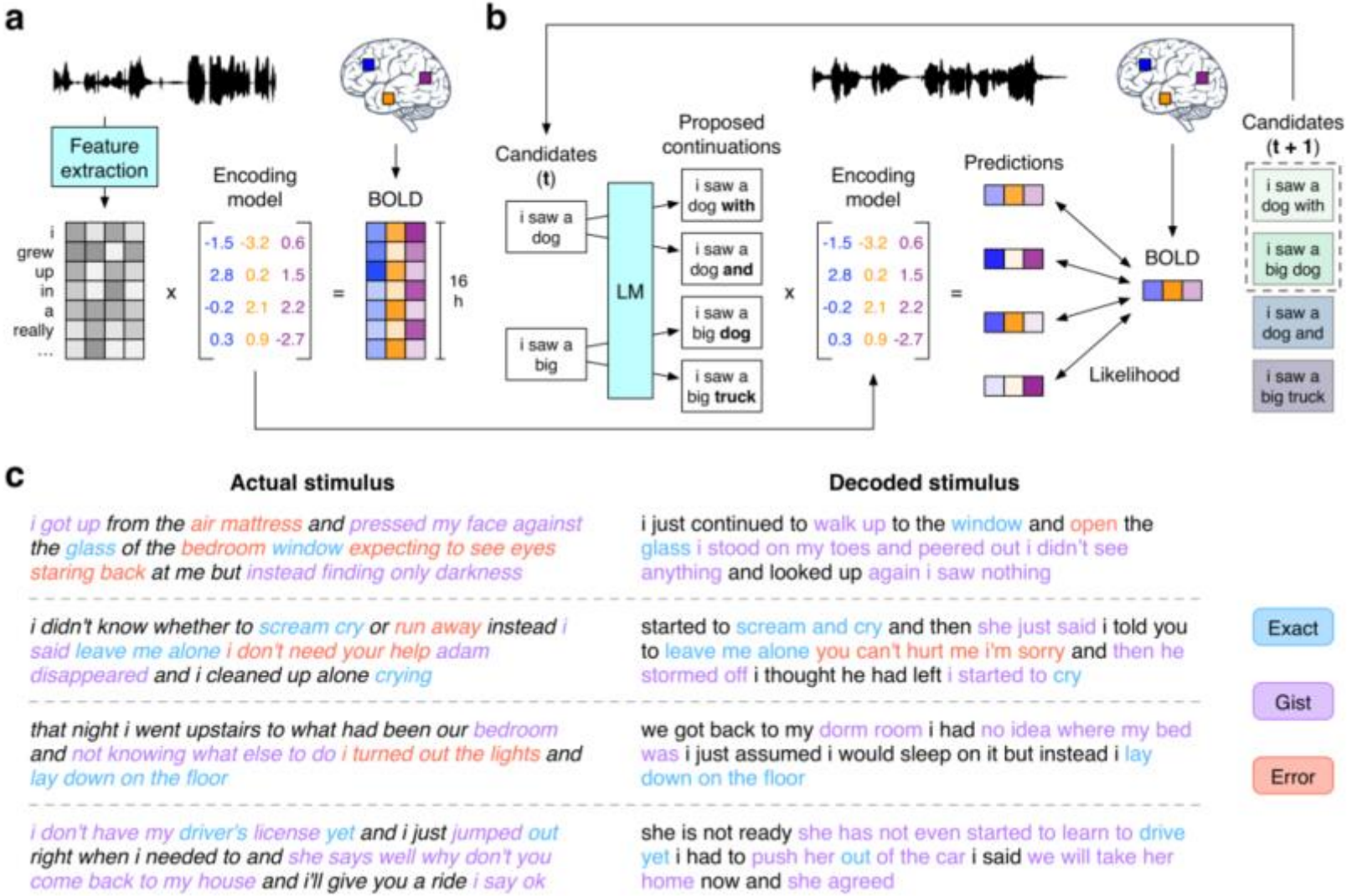


# Mental images

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.



# Brain reading

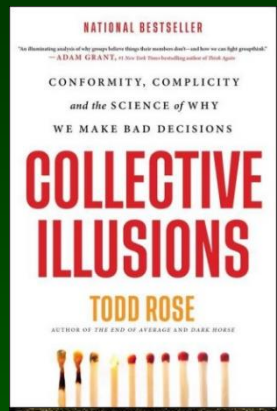


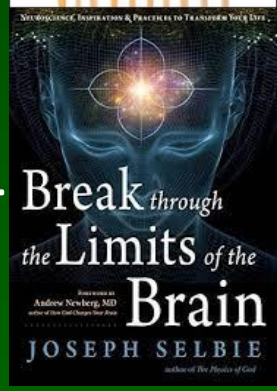

Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. Nature Neuroscience, 26(5)

# Humans have limitations

We have many problems with emotional and cognitive functions:

- Separating emotional responses from rational conclusions.
- Limitation of our senses, memory problems.
- Collective illusions, or wrong assumptions about the thoughts or intentions of others, leading to bad decisions.
- Understanding our real needs, leading to happiness in a longer time, resisting addictions, seeking immediate rewards.
- Making decisions based on limited available data, in complex situations/problems, holding multiple perspectives simultaneously.
- Use of educated guessing and verification to find new answers.
- Limited ability to predict future and construct what-if scenarios.




The good Lord has already done what He could, now its time to call experts.
We need to consciously shape our minds/brains, empowering ourselves.
Neuroethics, neuropolitcs, neuroeconomics, neurolaw, neurotheology … AI?

# The Cognitive Bias Codex

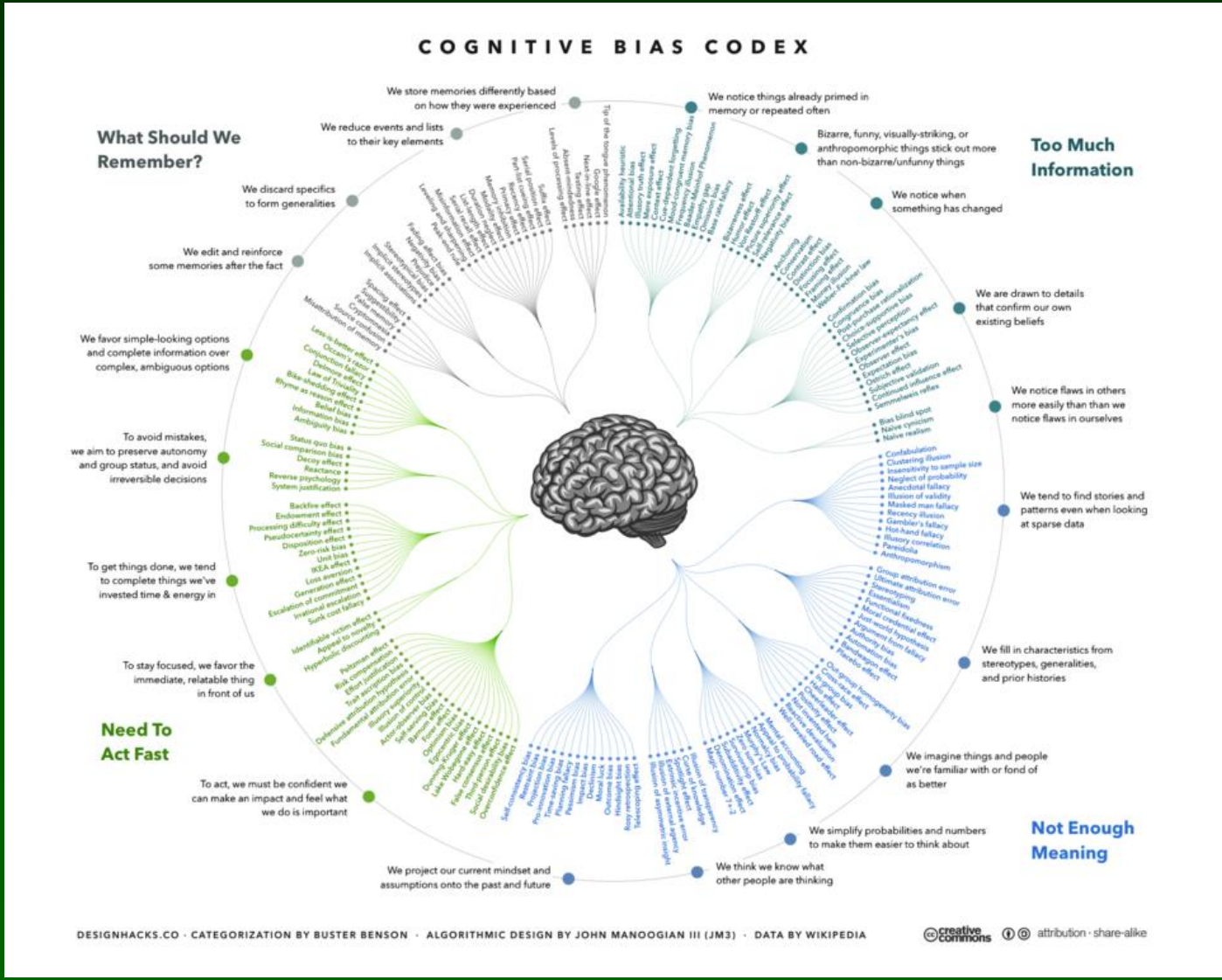
Cognitive biases are systematic patterns of deviation from rationality in judgment, studied in psychology, sociology, behavioral economics.

Over 180 biases …

Nobel 2002 in the Economic Sciences

**Daniel Kahneman**, Bounded Rationality

Nobel 2017, **Richard H. Thaler**, decision making, nudging, .




Source: Wikimedia commons

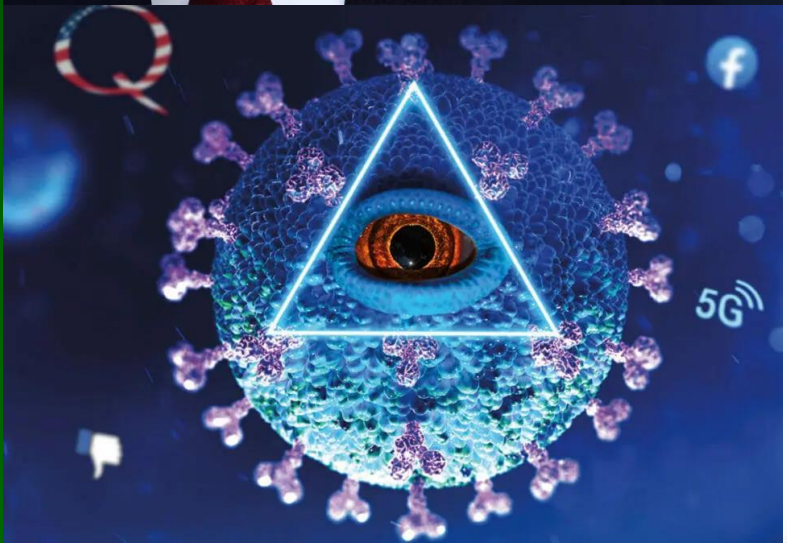

5G
DOES NOT
SPREAD
COVID-19
STOP
5G

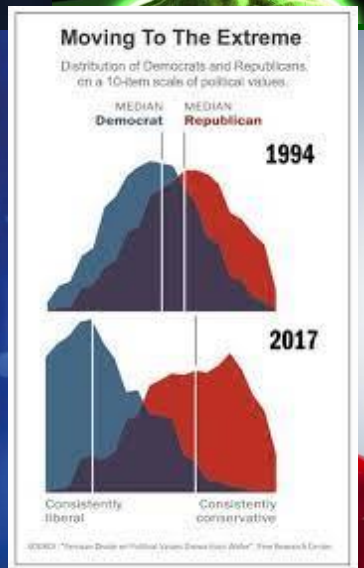

INVESTIGATE
COVID
AND 5G!


CNN
CNN SPECIAL REPORT
DONALD TRUMP'S
CONSPIRACY
THEORIES
MONDAY 9P ET/PT

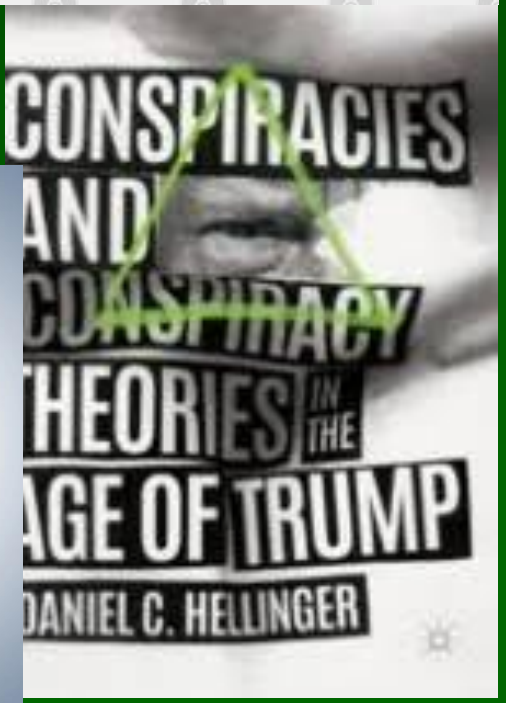

I WANT TO
WE COME IN PEACE
Believe

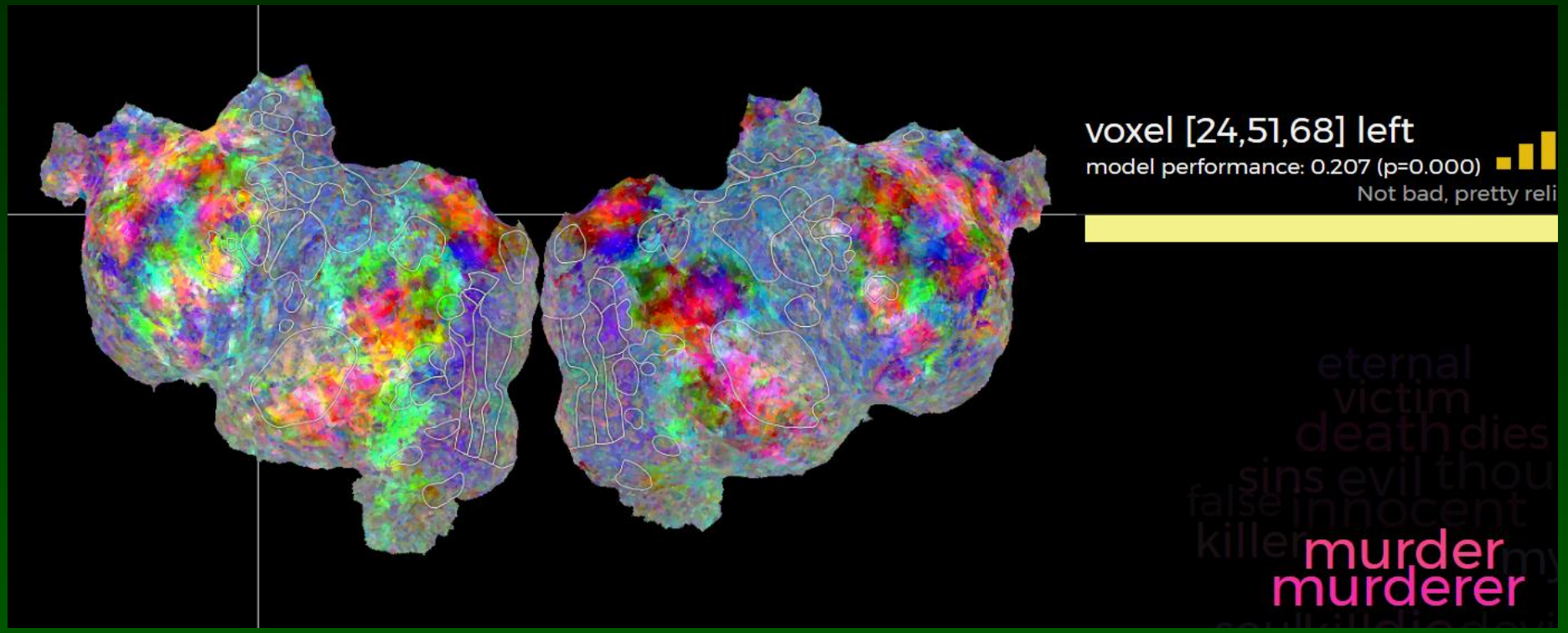

THE TRUTH
BEHIND
AREA 51


5G


Moving To The Extreme
Distribution of Democrats and Republicans on a 10-item scale of political values
MEDIAN Democrat
MEDIAN Republican
1994
2017
Consistently liberal
Consistently conservative


COVID-19


CONSPIRACIES
AND
CONSPIRACY
THEORIES IN THE
AGE OF TRUMP
DANIEL C. HELLINGER




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

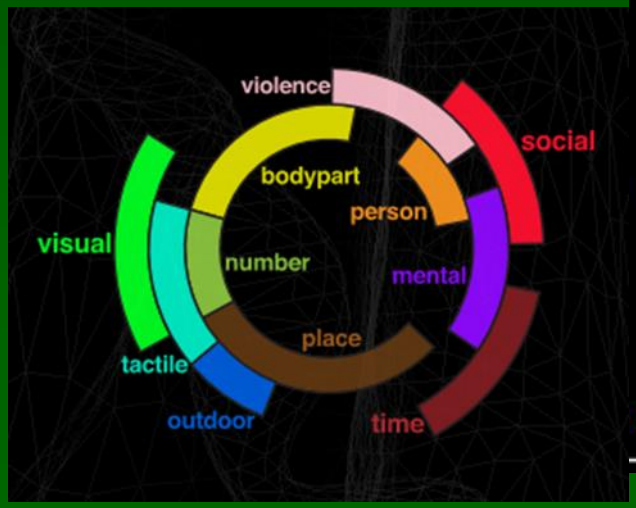
Each word activates a whole map of activity in the brain, depending on sensory features, motor actions and affective components associated with this word. Why such activity patterns arise? Brain subnetworks connect active areas.

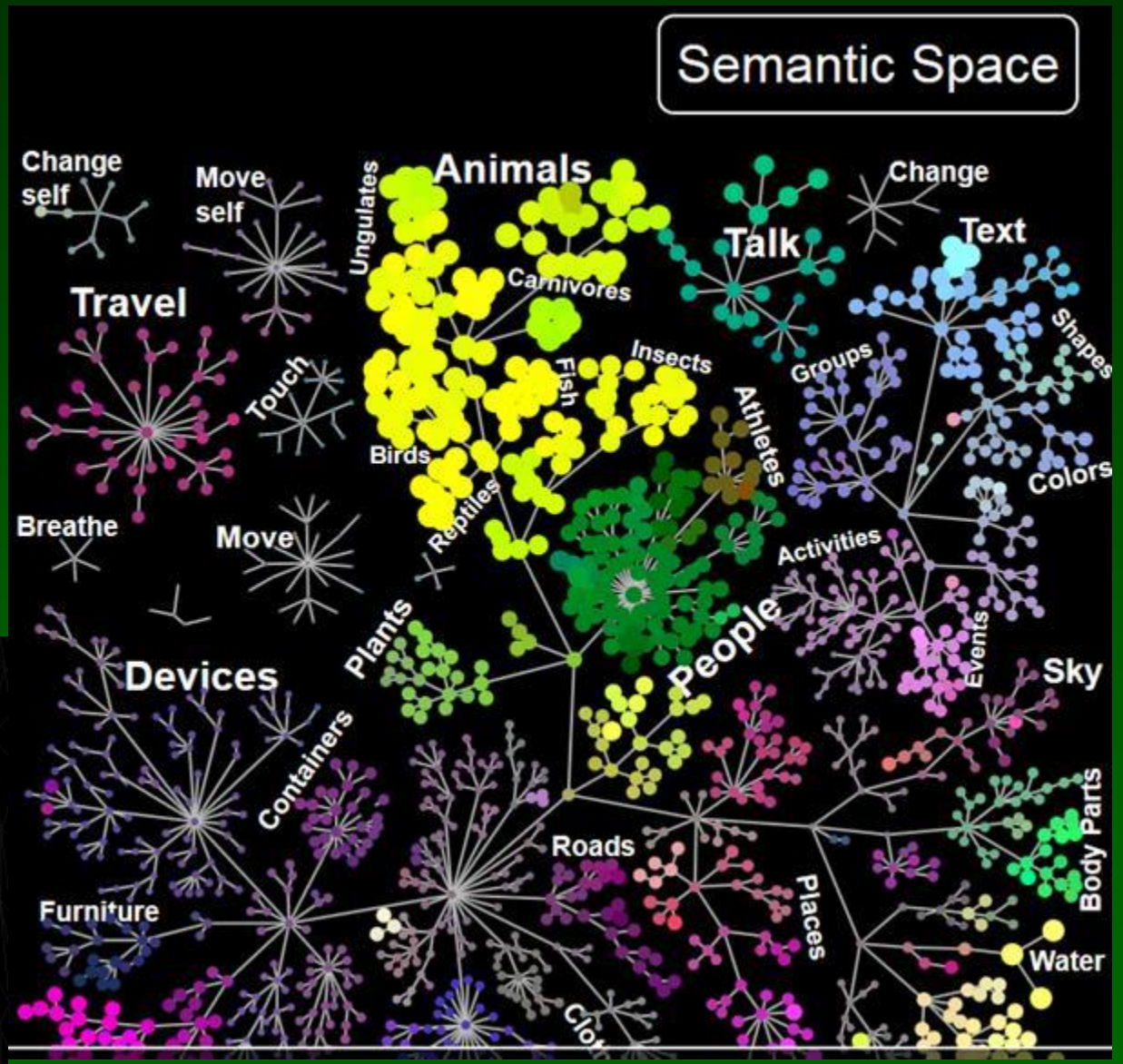
http://gallantlab.org/huth2016/ and short movie intro.

Can one do something like that with EEG or MEG? Brain dictionary - Nature video

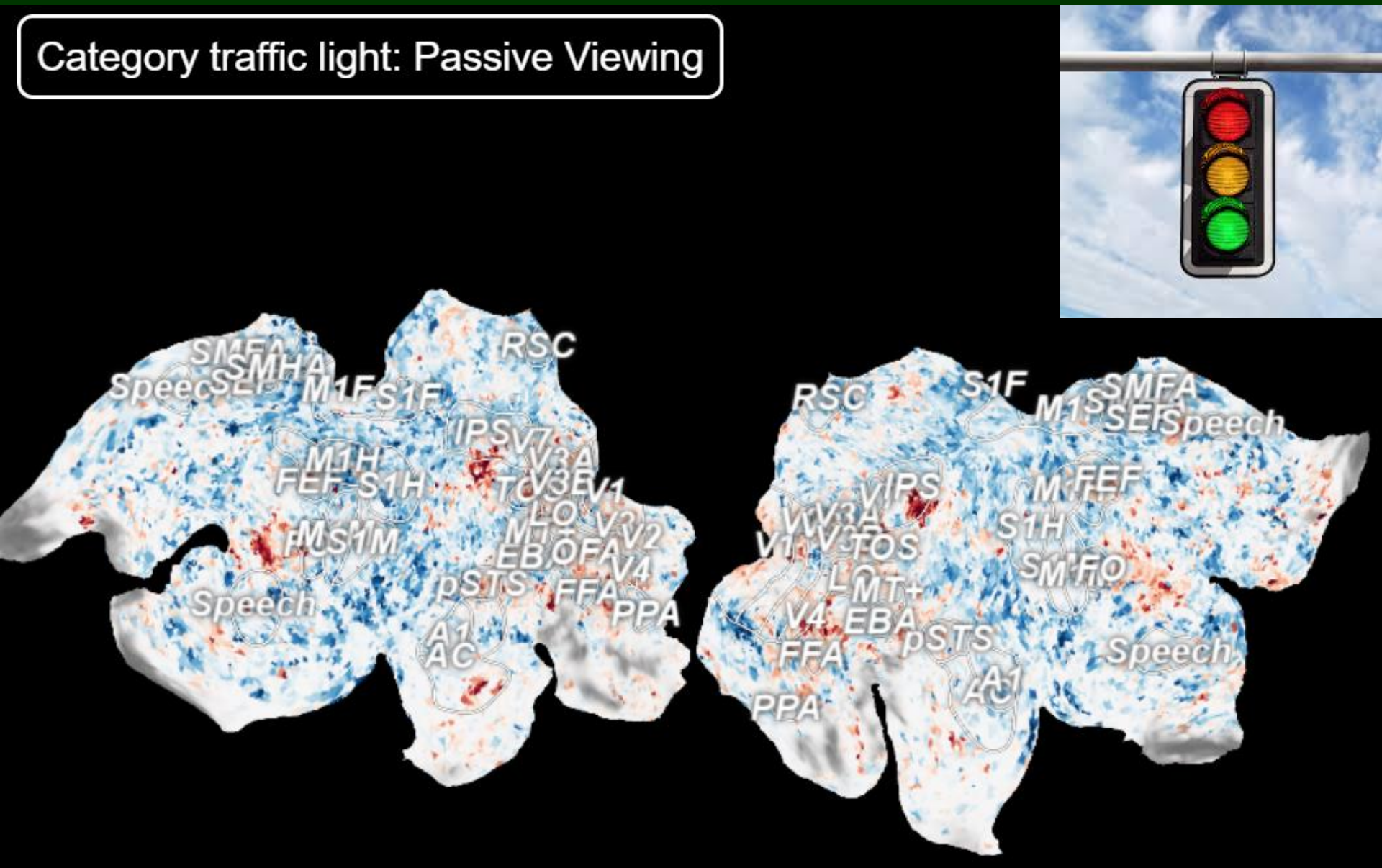
# Semantic neuronal space

1700 words in the semantic space are grouped by similarity. Words activate specific ROIs, similar words create similar maps of brain activity. Video or audio stimuli, fMRI 60.000 voxel). Gallant lab, Berkeley.



Interpretation for simple objects is easy: IPS – visual attention, V4 – color, AC – object recogniton.



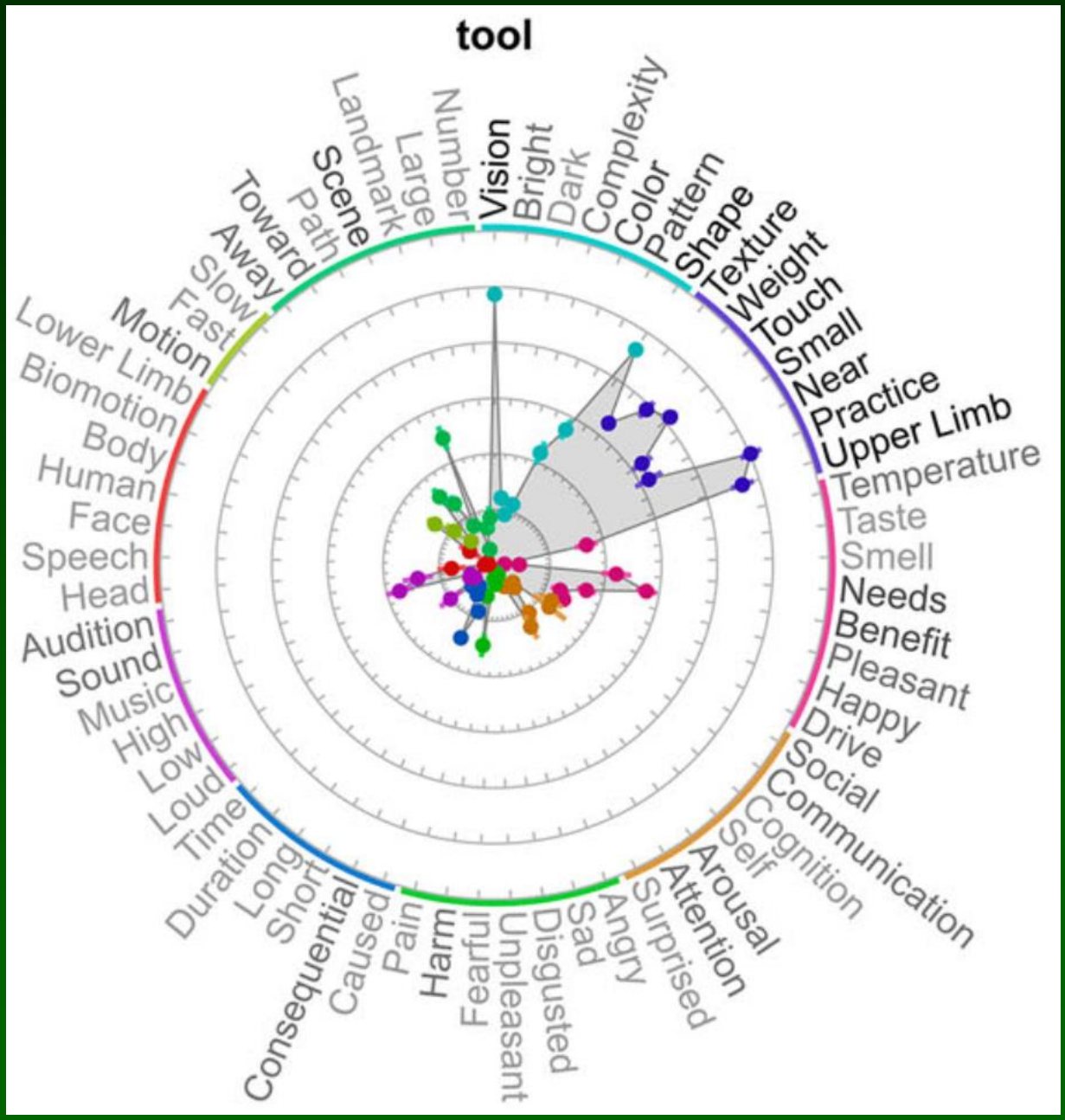
65 attributes related to neural processes.

Brain-Based Representation of tools.

J.R. Binder et al

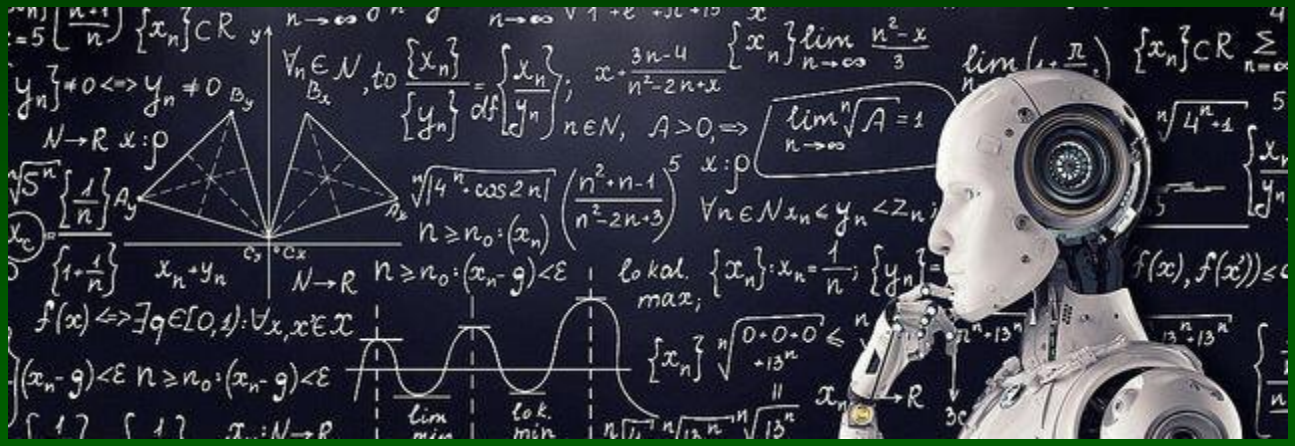
Toward a Brain-Based Componential Semantic Representation

Cognitive Neuropsychology 2016



# AI intro

# AI intro: why, what, where?



# Superhuman AI?

**New AI**: predicts words, but shows no understanding, requires small adjustments,
or: new superhuman form of intelligence, will lead to **radical changes**.

# Imitation may take you quite far …


Is AI only imitating understanding?

Are our brains magically doing something else?

Neurons are just counting spikes!

Brains predict next word.

Cogni
Cognitive sciences

Biohybrids

Bio
Neuroscience
Organoids

Nano
Quantum
Technologies

Neurocognitive
Informatics

Exaflop speed $10^{18}$ – $10^{21}$ op/sec, GPU, TPU, NU
Nano LLMs in phones.
Stargate 100B project?

Info
Artificial/Computational Intelligence,
Machine Learning, Neural Networks

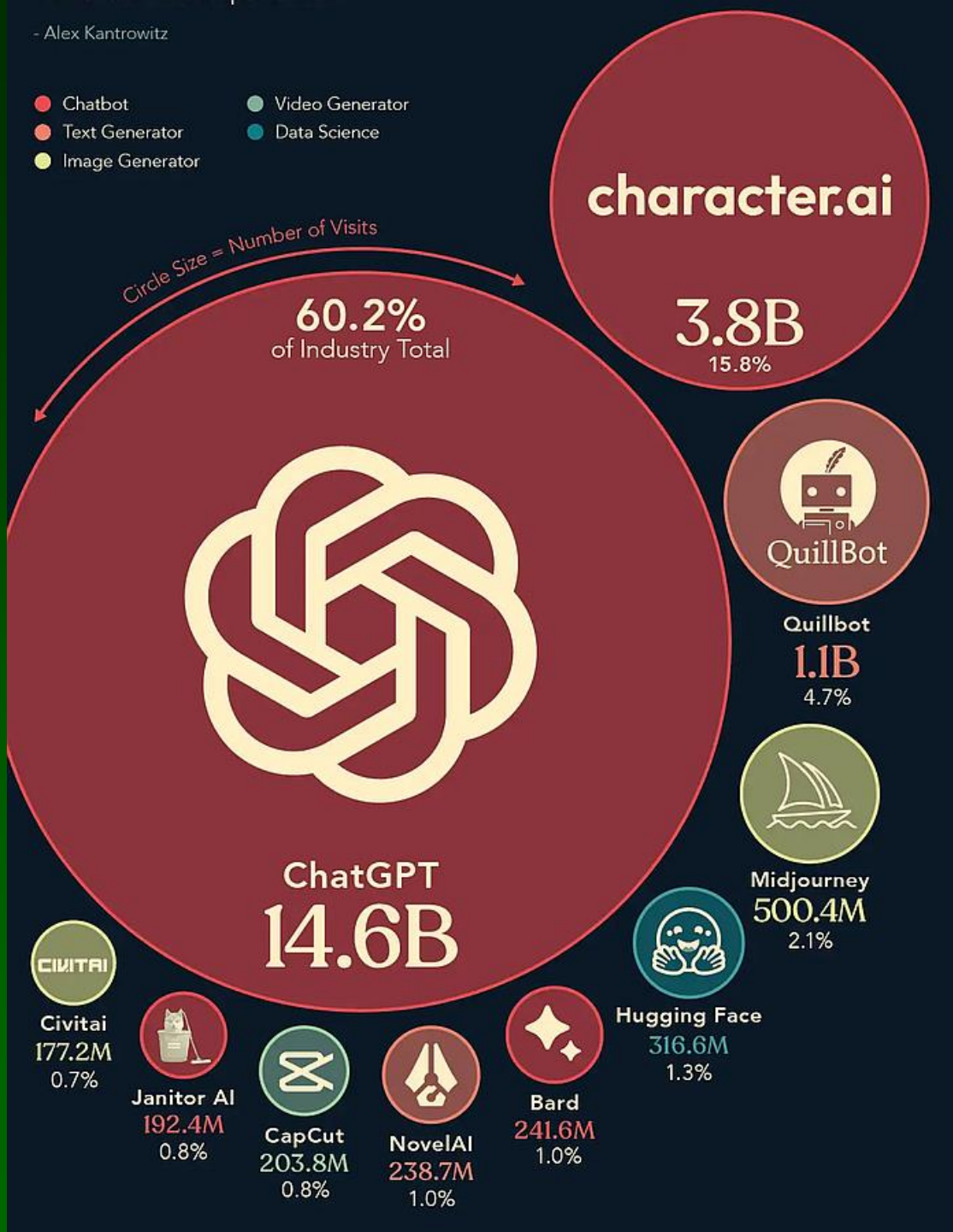
# AI Tools 2023

Billions of users.
We like to talk (chat), write, and create images.

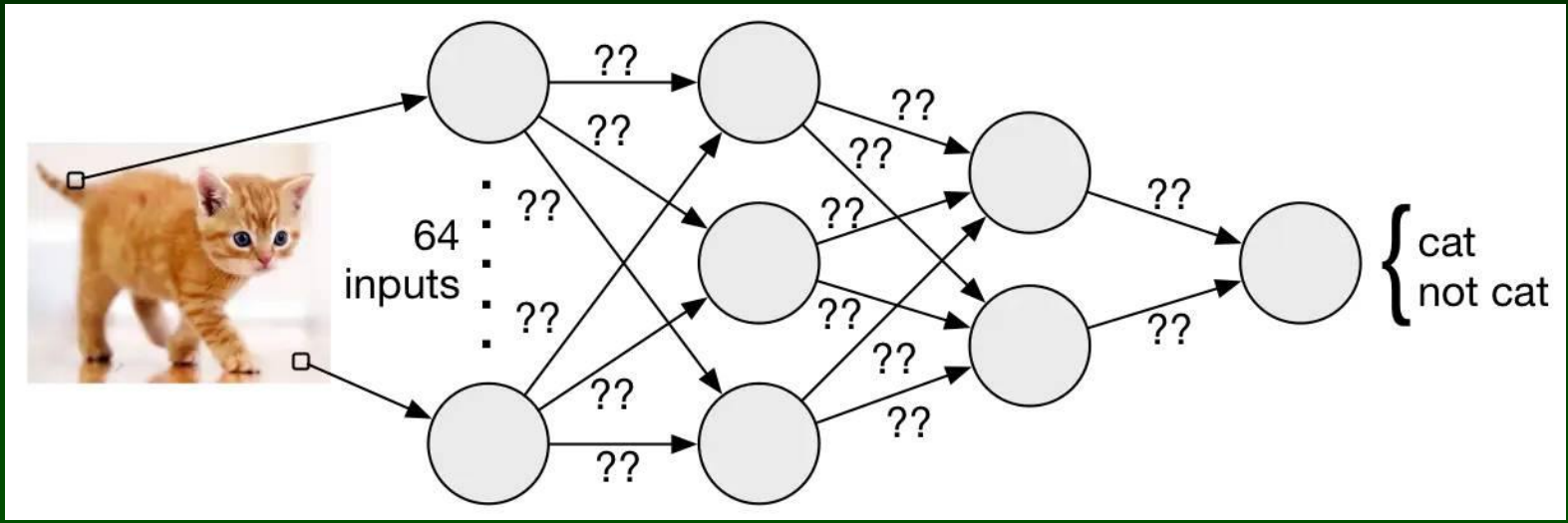
My favorite philosophers.

More useful for science:

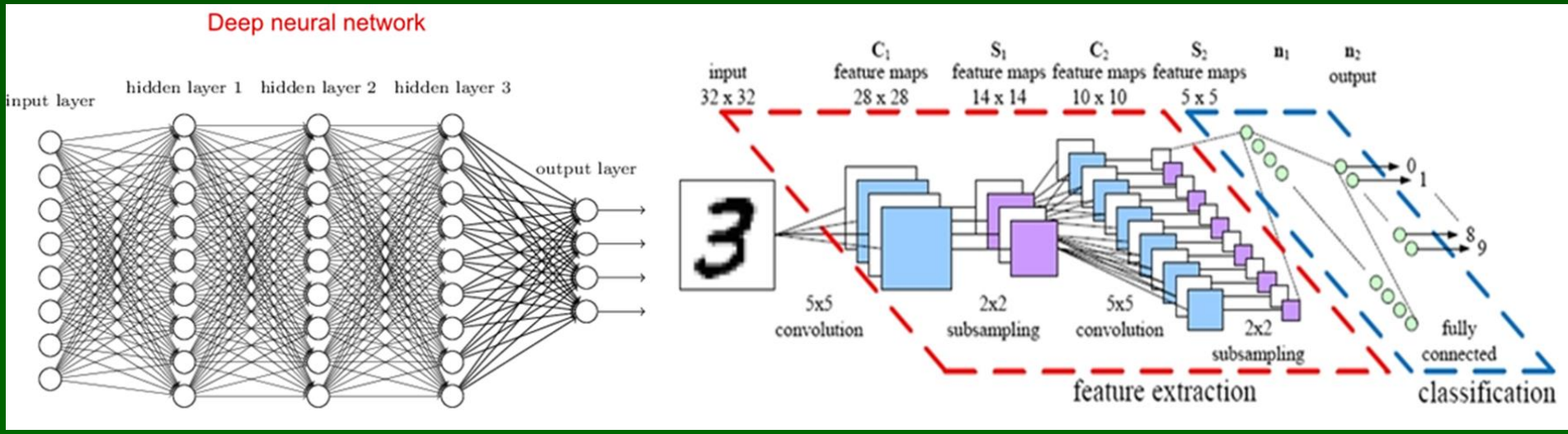
Perplexity, Elicit, Consensus, SciSpace, Iris.ai, Insightful, Open knowledge maps, Litmaps, Explainpaper, ScienceOpen, X-mol, SciMat, InfraNodus, ChatPDF, TAPoR.

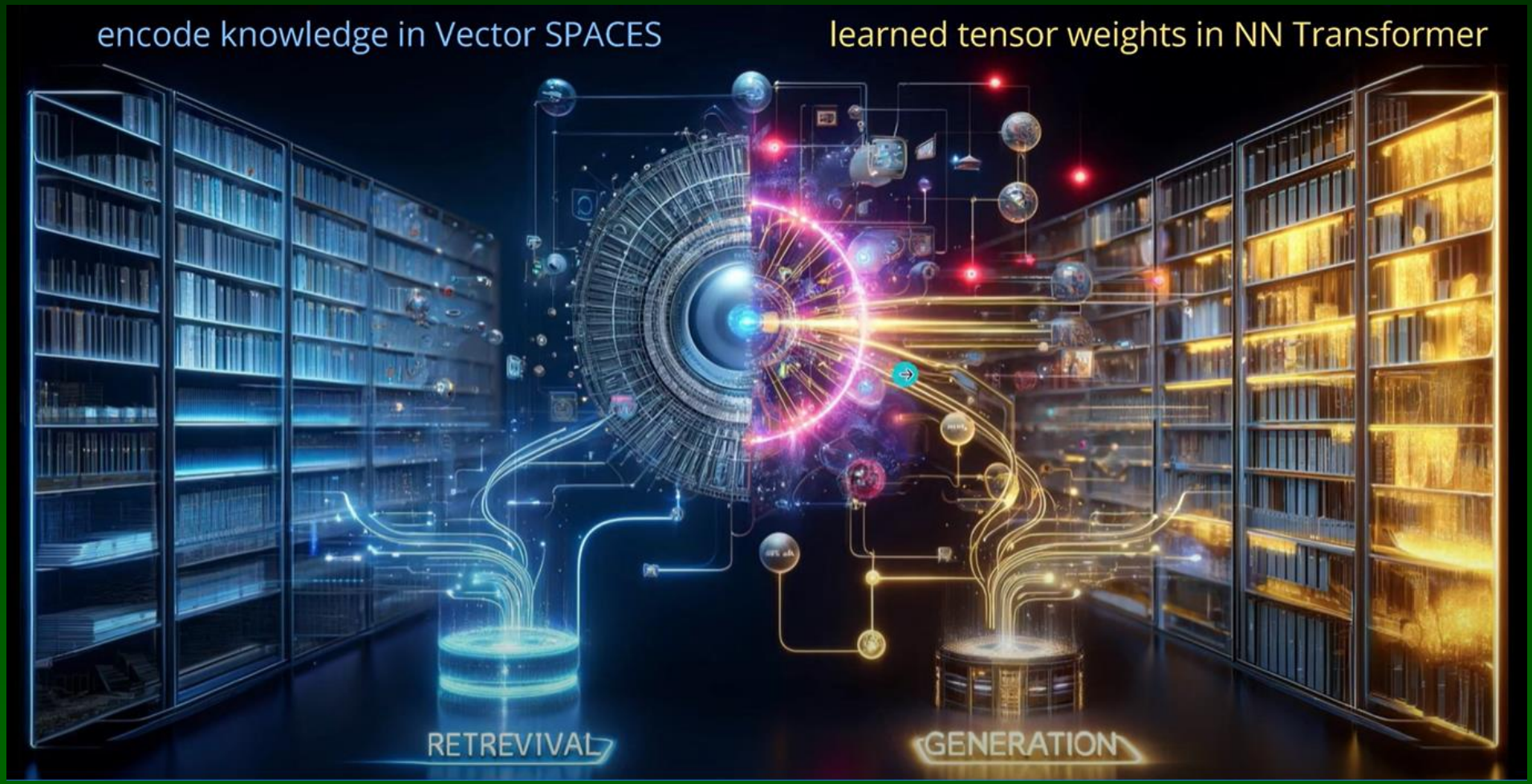


# Neural classifiers




Data, words, image patches => networks with adjustable parameters
=> training to recognize patterns => object classification, diagnosis.




encode knowledge in Vector SPACES
learned tensor weights in NN Transformer
RETREVIVAL
GENERATION

Think of it like compressing the internet.

Chunk of the internet,
~10TB of text

6,000 GPUs for 12 days, ~$2M
~1e24 FLOPS


parameters.zip

~140GB file

*numbers for Llama 2 70B

Llama is a family of open models, much smaller than large commercial models. Compression 10.000/140=71 times is the key, helps to form associations and be creative. List of open-source language models: 33169 (24/3/2024). https://llm.extractum.io

# Cognition as Compression

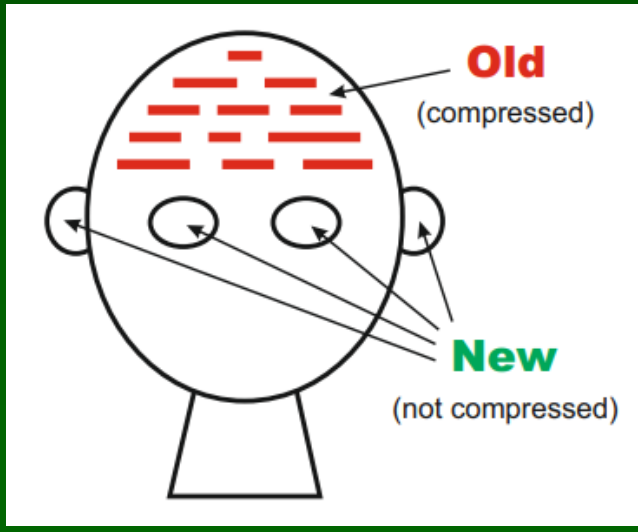
Computing ↔ Cognition,
artificial ↔ natural systems.

J.G. Wolff, SP theory of intelligence: computing as compression.
http://www.cognitionresearch.org (mostly for 1D sequences)

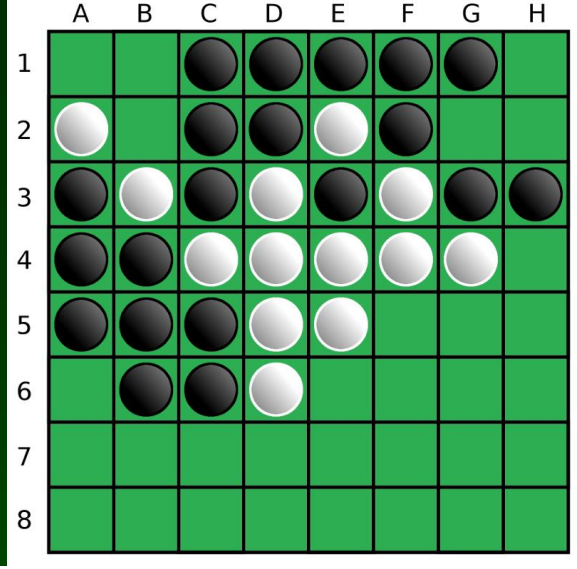
Related to model selection in ML, Algorithmic Information Theory (AIT): Minimum Length Encoding (MLE), Minimum Description Length (MDL), Minimum Run Length Encoding, Minimum Message Length Encoding, etc.

- Cognition as Compression, SP theory
- Language Learning as Compression
- Natural Language Processing as Compression

The **Hutter Large Text Compression Benchmark Prize (**2006) for data compression on *enwik9* English Wikipedia 1 GB text file. 5000 euros for each 1% improvement in the compressed size. 2023 best result: 113 746 218 bytes.

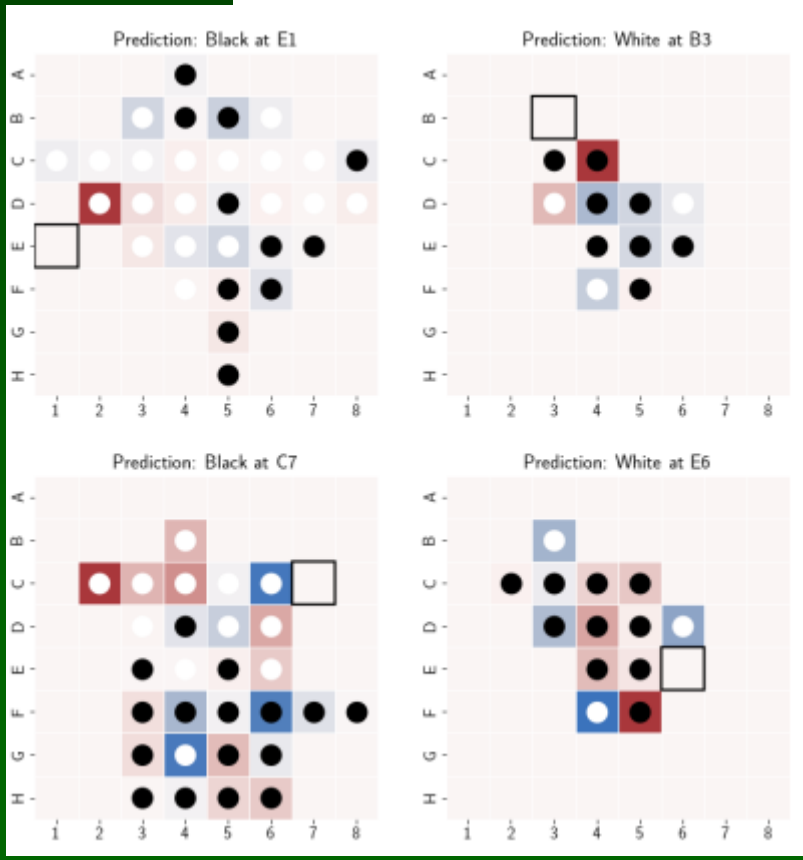


# Othello-GPT




A GPT variant Othello-GPT was trained to extend a list of moves with legal moves. The model has no a priori knowledge of the game or its rules, it only predicts the next move. Internal board representation emerged.

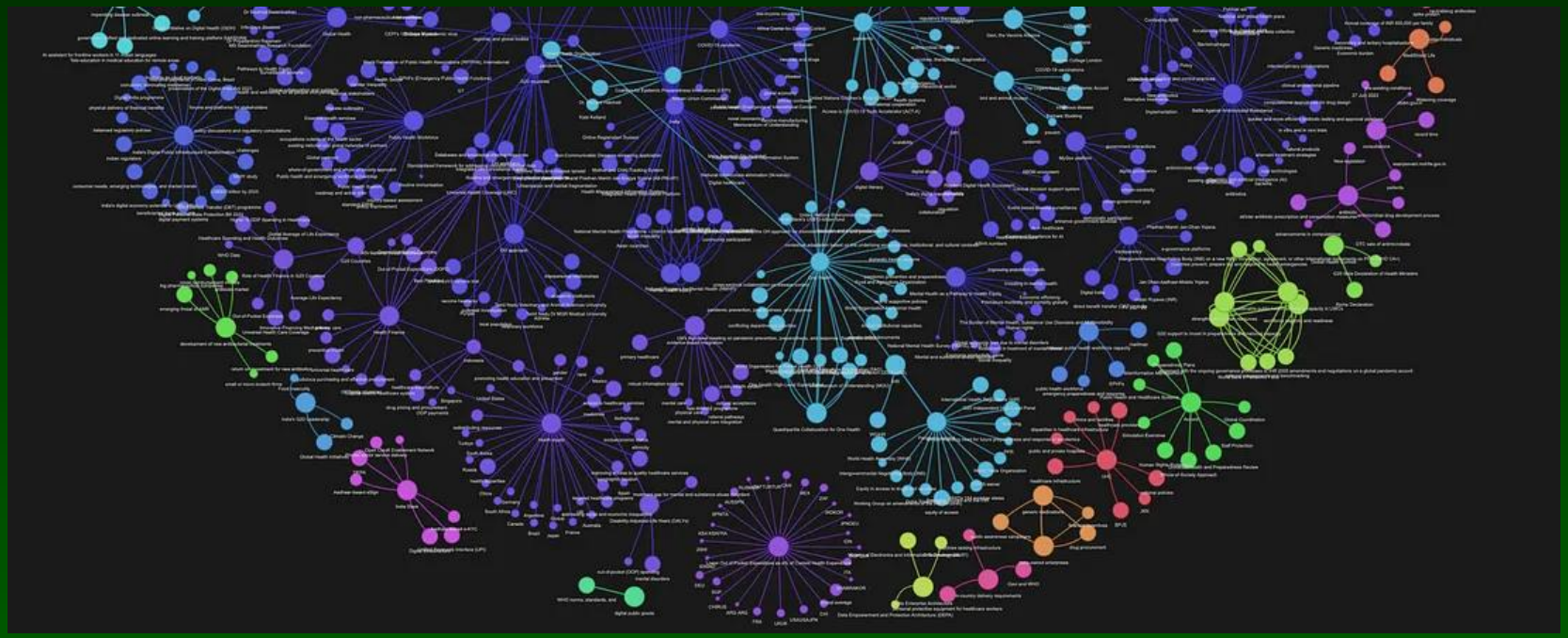
A 2-layer MLP classifier takes as input internal activations of a network, outputs next position. Activation-intervention technique is used to create latent saliency maps. Black box: Top-1 prediction by the model. Contribution to prediction: red=high, blue=low




GPT perceives what appears in its imagery.

Li, K. … H., & Wattenberg, M. (2023). Emergent World Representations: Exploring a Sequence Model Trained on a Synthetic Task (arXiv:2210.13382).

# LLM => semantic network



Concept = activity of LLM network nodes => word sense in a context.
LLM creates semantic network, compressing information, leaving essentials for pragmatic use. This facilitates associations via spreading of activation.

W. Duch et al. Processing clinical text with domain-specific spreading activation methods. US Patent Application No. 12/006.813 (4/2008).

# Horses (Aria chatbot)

1. **Physical Characteristics**: Horses are large mammals known for their strong, muscular bodies, long legs, and flowing mane and tail.
2. **Domestication**: Horses have been domesticated for thousands of years and have played crucial roles in transportation, agriculture, and leisure activities.
3. **Breeds**: There are numerous horse breeds, each with unique characteristics such as size, color, and temperament. Examples include the Arabian ...
4. **Behavior**: Horses are social animals that form strong bonds with other horses and with humans, communicate through body language, vocalizations …
5. **Diet**: Horses are herbivores and graze on grass, hay, and other plant materials. They have a complex digestive system processing fibrous plant matter.
6. **Uses**: Horses are utilized for various purposes, including riding, racing, working on farms, therapy, and as companions.
7. **Anatomy**: Horses have a unique hoof structure that allows them to run at high speeds, a strong skeletal system, and a keen sense of hearing and vision.
8. **Healthcare**: Proper care for horses includes regular veterinary check-ups, vaccinations, dental care, hoof trimming, and appropriate nutrition.
9. **Gaits**: Horses have distinct gaits including walk, trot, canter, and gallop. Each gait serves a different purpose and requires specific training.
10. **Symbolism and Cultural Significance**: Throughout history, horses have symbolized strength, freedom, and beauty in various cultures. …

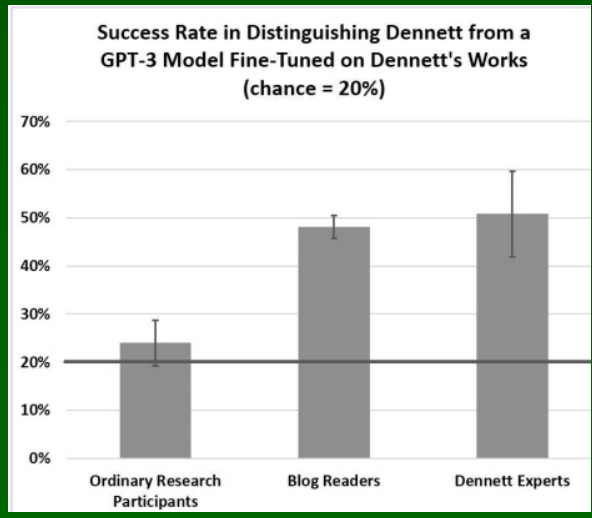
# GPT-3 as philosopher

Eric Schwitzgebel, David Schwitzgebel, Anna Strasser,
Creating a Large Language Model of a Philosopher, arXiv:2302.01339

"Can large language models be trained to produce philosophical texts that are difficult to distinguish from texts produced by human philosophers?"

We asked prof. Dennett ten philosophical questions, posing the same questions to the ChatGPT-3, fine-tuned on his books/papers, collecting 4 responses for each question, without any cherry-picking.

425 participants tried to distinguish Dennett's answer from ChatGPT.
25 experts on Dennett's work succeeded 51% of the time. Philosophy blog readers (N = 302) performed similarly to the experts.
Ordinary participants (N = 98) were near chance (24%).




So, is Dennett intelligent? If we agree, then GPT-3 must also be intelligent.

Duch W. (2023), Artificial intelligence and the limits of the humanities. Er(r)go 47 (2/2023) - Humanities.

# From calculator to superhuman AI



Reasoning: 1997–Deep Blue wins in chess;
2016 –AlphaGo wins in Go; 2017 Alpha GoZero 100:0.

Open Games: 2017–Poker, Dota 2; 2019-Starcraft II,
2022 Stratego, Diplomacy, Bridge – what is left?

Perception: speech, vision, recognition of faces,
personality traits, political and other preferences ...

Robotics: 2020 Atlas robot (Boston Dynamics) backflip
and parkour, autonomous vehicles, 2023 Tesla Optimus.

Science: 2020 AlphaFold 2, now 620 M 3D proteins,
2023-GNoME (Deep Mind) 2.2 mln structures; math.

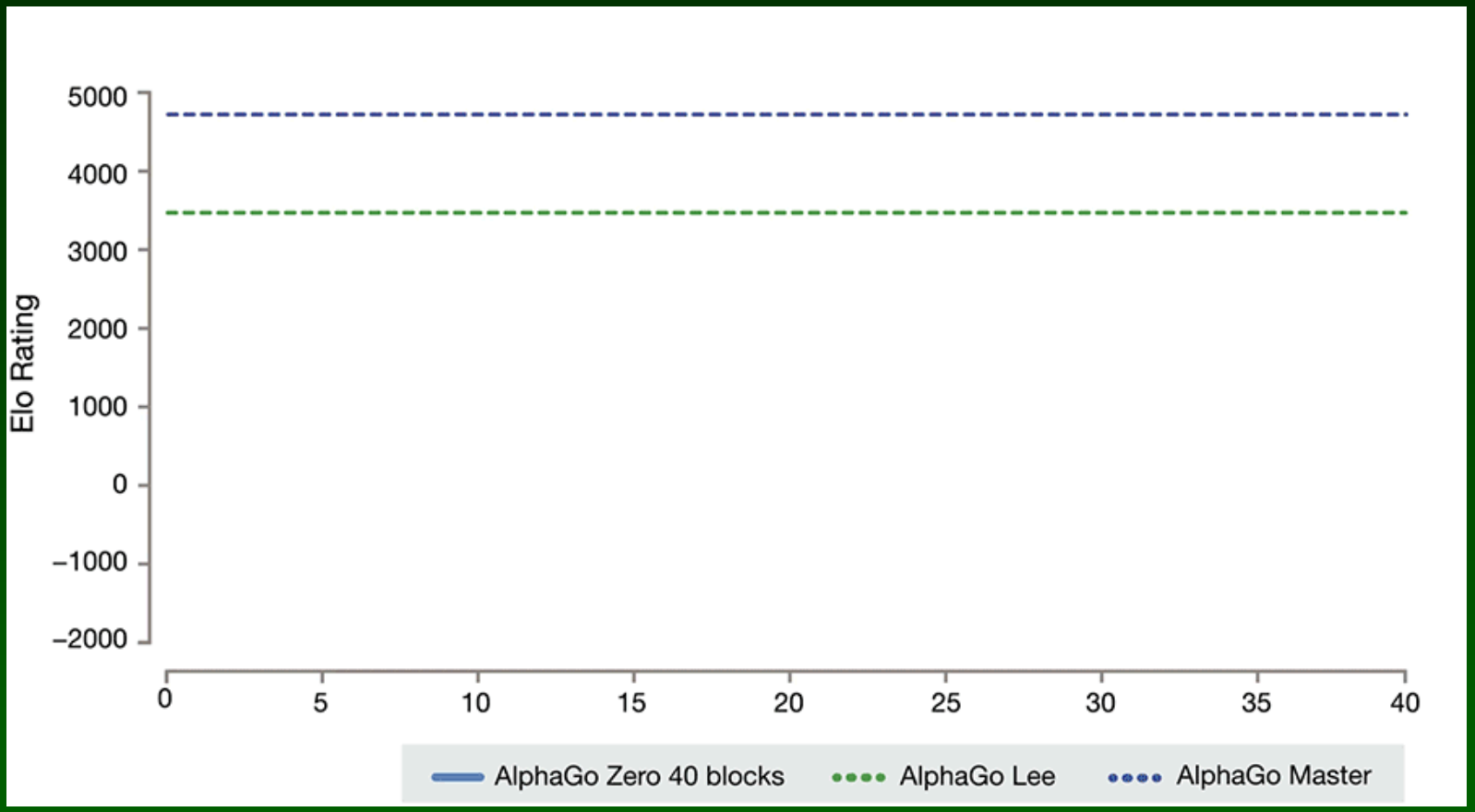
Creativity and imagination: GAN revolution, Dall-E,
Midjourney, Stable Diffusion, AIVA, music composers.

Language: 2011–IBM Watson wins in Jeopardy;
2018–Watson Debater wins with professionals.
2020: BERT answers questions from SQuAD database.

Cyborgization: BCI, brain-computer symbiosis, soon?

What are we better in comparison to AI? For how long?

# AlphaGo Zero learns Go from 0!




Superhuman level in the strategic game of Go. Human experience surpassed by software playing against its own copy. Search + NN as heuristics.
Human knowledge becomes irrelevant … **Shocking news:**
Ruoss … & Genewein, T. (2024). *Grandmaster-Level Chess Without Search* (arXiv:2402.04494) 270M parameter transformer model, 1-step search!

# Spreading activation

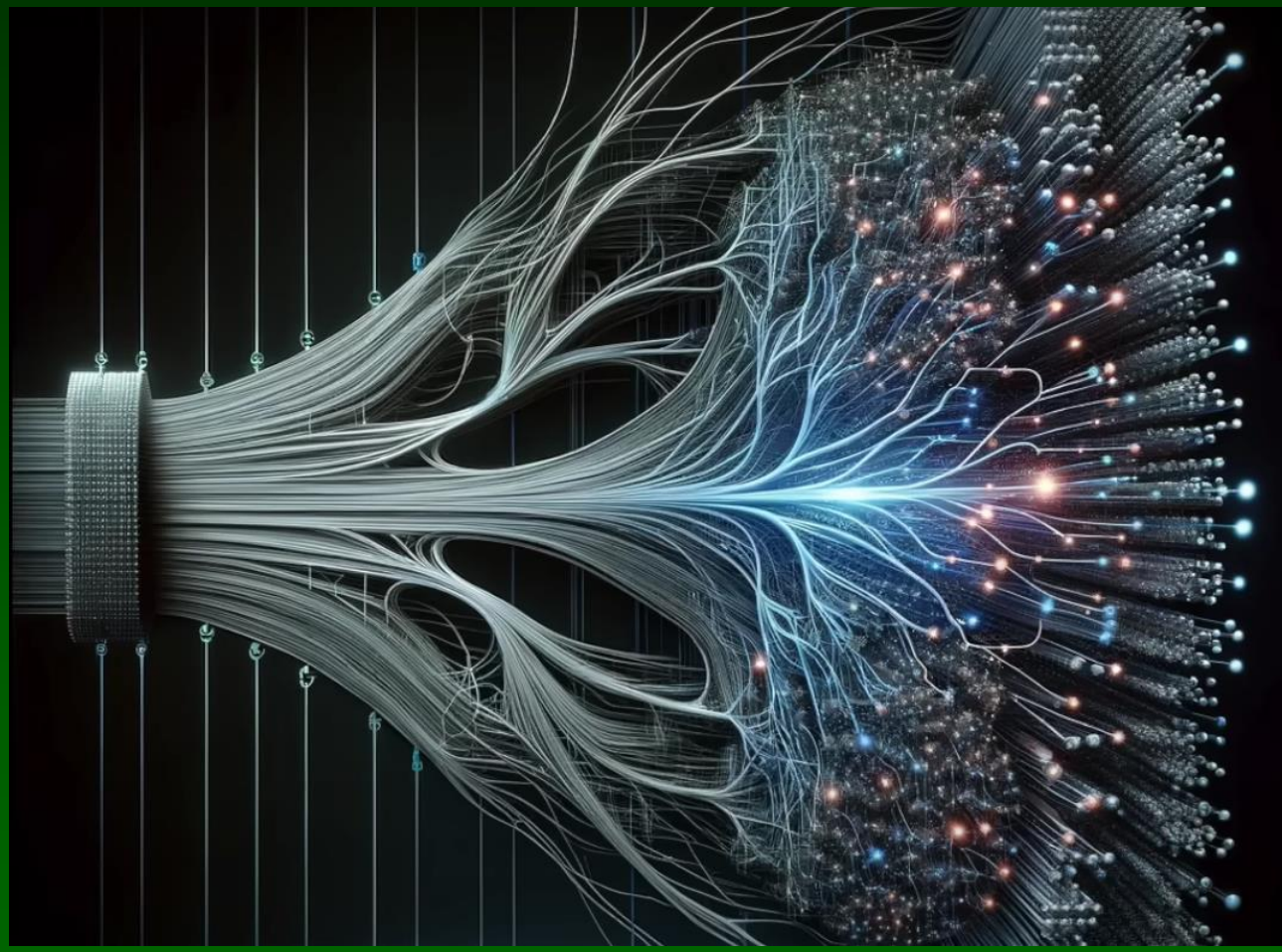
Spreading activation networks.
Duch et al., Towards Understanding of Natural Language: Neurocognitive Inspirations. LNCS 4668, 953–962, 2007

GPT = Generative Pre-trained Transformer

How transformers work
Financial Times + visual storytelling.

LLM visualization

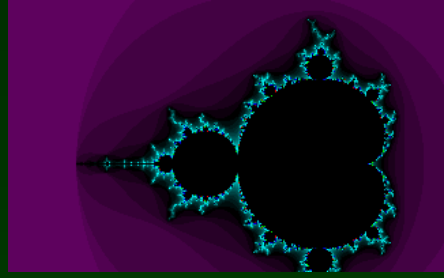
https://bbycroft.net/llm



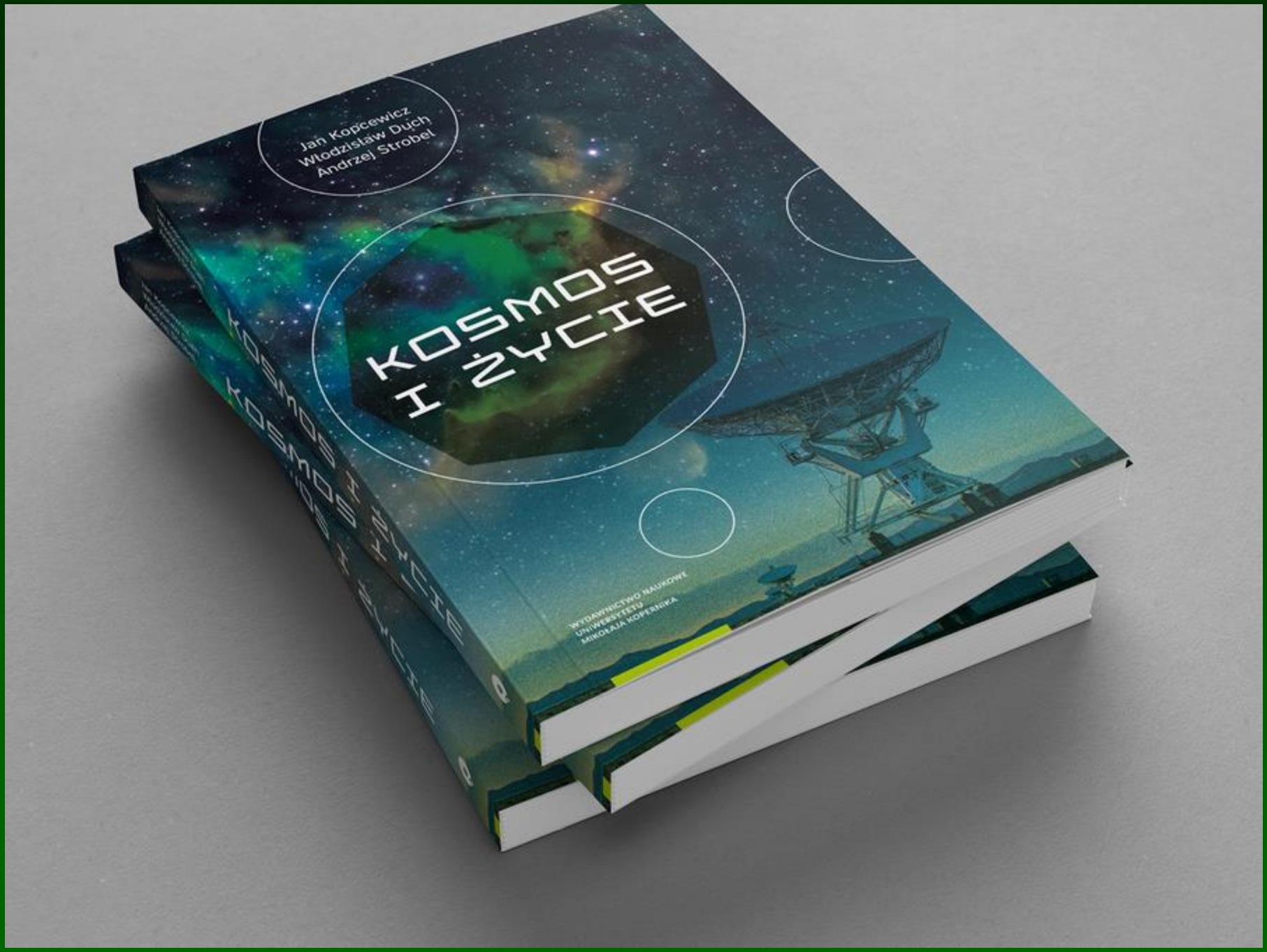
# Secrete geometry of language


Universal algorithm: identify tokens, create high-dimensional embeddings in many contexts, use self-attention in transformer architecture.
Structure and relations are similar in all thousands of languages.
Even animal communication can be analyzed in this way.

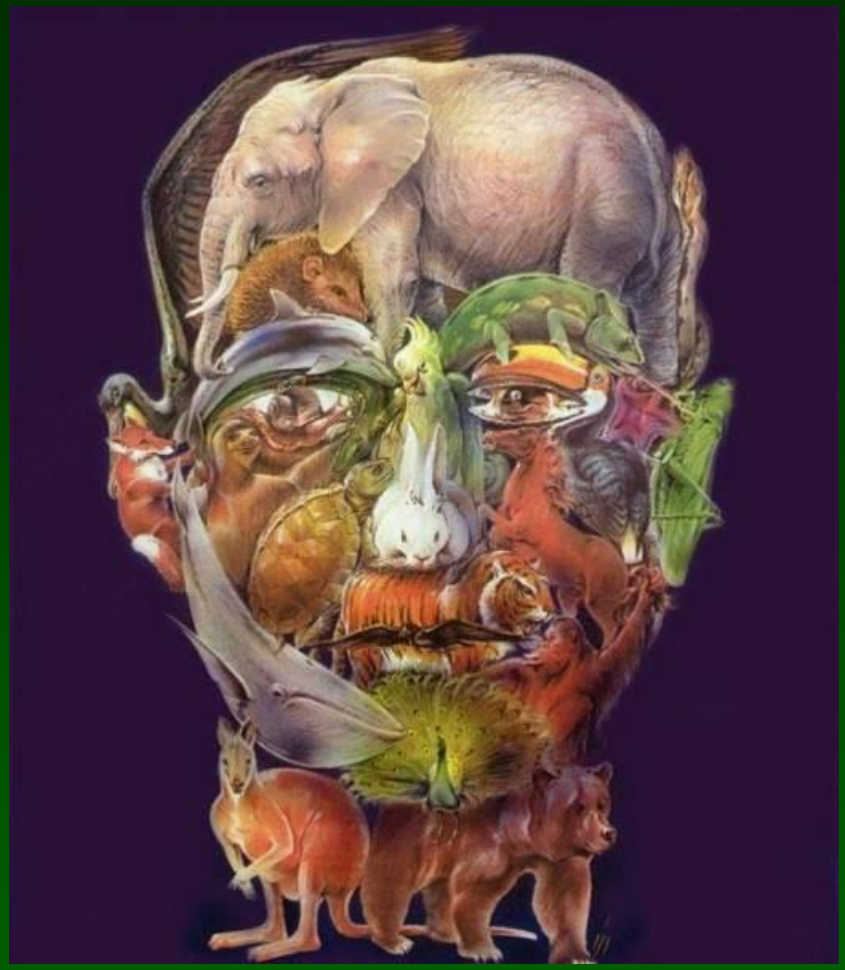
# Preliminary conclusions



- Coevolution: brains shape culture shape environment shape brains … loops within loops. Mandelbrot fractal shows emerging complexity.
- Many brain states are now linked to specific mental states, and can be transformed into signals that we can understand using BCI techniques: motor intentions, plans, images, inner voices …
- Neurodynamics is the key to understanding immediate causes of mental states; it creates dynamical forms. Neurodynamics itself on hormone levels, brain structure formed by genetics and developmental processes, environmental factors, ecosystems, social interactions, culture ...
- Our understanding allows for development of neurocognitive technologies, helping to diagnose, repair and optimize brain processes.
- AI does not need perfect replication of human brains to be intelligent, compression is the key for associative thinking.
- We need detailed comparison of large AI models and brains to understand the difference. Sentience and consciousness included.


Jan Kopcewicz
Włodzisław Duch
Andrzej Strobel
KOSMOS I ŻYCIE
WYDAWNICTWO NAUKOWE
UNIWERSYTETU
MIKOŁAJA KOPERNIKA

# Artificial ?

Search: Wlodzislaw Duch

=> talks, papers, lectures, Flipboard, YouTube