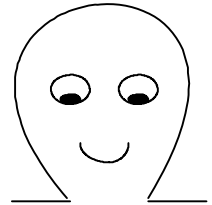


# Towards a plausible theory of mind



Włodzisław Duch

Department of Computer Methods,  
Nicholas Copernicus University,  
ul. Grudziądzka 5, 87-100 Toruń, Poland.  
duch @ phys.uni.torun.pl



**WWW home page:** <http://class1.phys.uni.torun.pl>

**FTP archive:** [class1.phys.uni.torun.pl/pub/papers/kmk](http://class1.phys.uni.torun.pl/pub/papers/kmk)

## Plan

1. Understanding the mind ?
2. The brain
3. Basic perception
4. Consciousness
5. Immediate behavior and intuition
6. Mind space - language for cognition
7. Higher cognition
8. Practical model and its applications
9. Philosophical implications

## Problems with understanding of the mind

Foundations of **cognitive sciences** are plagued with problems:

1. How can the non-material mind have influence on the material body and vice versa (mind-body problem) ?
2. The problem of qualia or qualitative character of experiences: “feeling” of redness or darkness or “cat-ness”.
3. The symbol grounding problem - how can symbols in a formal system ever acquire real meanings?
4. Mental representations require cognitive agents and cognitive agents require mental representations.

Many other philosophical objections (Searle, Penrose, ...) are variations on the above topics. Some philosophers of mind claim that we will never understand such problems and that they are the biggest obstacle to the progress of science!

5. The binding problem: why is our experience unified if there seems to be no central place in the brain where it all “comes together”?
6. The Central Paradox of Cognition: how can the structure and meaning, expressed in symbols and ideas at the mental level result from numerical processing at the brain level?
7. Basic concepts, such as “consciousness”, are not defined.
8. Productivity, systematicity, compositionality and inferential coherence are difficult for connectionist systems. Hard to find neurophysiological inspiration for classical systems. Pattern recognition, association, graceful degradation are difficult for physical symbol processing systems.

**Is a satisfactory understanding of the mind possible ?**

## Empirical Theory of Mind should explain:

- *Basic facts about perception, such as stereoscopic vision.*
- *Thousands of facts from cognitive psychology, such as the typing skills or the power law of learning.*
- *Stages of development, from infancy to adulthood. Learning to walk, learning basic perceptual categories and knowledge structures.*
- *Various types of memory, active memory, LTM and STM*
- *Conscious and subconscious perception, advantages of conscious perception, relation to brain events.*
- *Short time stimulation and conscious perception: Libet (1965-1993) direct stimulation of neurocortex, Walter (1963) “precognitive carousel”, cutaneous rabbit (Geldard 1977)*
- *Dynamic optical illusions: color phi (Kolars and von Grunau 1976), metacontrast, Stroop (1935) interference, tachistoscope*
- *First person experience, qualia, mental content, meaning of symbols*
- *Intuition and immediate response behavior*
- *Linguistic competence, thinking and reasoning*
- *Dream states (daydreaming)*
- *Hypnotic and other unusual states of mind*
- *Blindsight, hysterical blindness*
- *Disorders: from anxiety and dyslexia to schizophrenia*
- *Formation of ego, personality, Multiple Personality Disorder (MPD)*
- *Exceptional abilities, “idiot savant” syndrome*
- *“Free will”*
- *Esthetic perception in music and art*
- *Extrasensory perception ?*

Very few ambitious projects so far

Newell - unified theory of cognition (SOAR)

Edelman - Theory of Neuronal Group Selection (TNGS)

What is mind? *No matter.*

What is matter? *Never mind.*

## The brain

- **Complexity:**  $10^{10}$ - $10^{11}$  neurons,  $10^{14}$ - $10^{15}$  synapses,  $\sim 10^3$  - $10^5$  Tflops
- **Levels of modeling:** from molecular  $\sim 10^{-10}$ m to macroscopic  $\sim 1$ m
- **Structures:** few hundred large, anatomically different structures
- **Topographical maps:** somatotopic, tonotopic, visual, motoric ...
- **Recurrent feedbacks:** backprojection re-entry loops
- **Subcortical multimodal integration:** superior colliculus, claustrum, thalamus, intralaminar complex nuclei
- **Neural cell assemblies (NCA),** originally introduced by Hebb (1949), transcortical NCA
- **Six layers of neurocortex** in 2-4 mm
- **Microcolumns:** about  $0.1$ - $1$  mm<sup>2</sup>,  $10^3$ - $10^5$  neurons, few percent internal connectivity, tangential spread  $\sim 8$  mm
- **Collective oscillations**  $\sim 40$  Hz, synchronization leads to transient TNCA bindings in networks of inhibitory neurons.
- **Experimental techniques:** multielectrode EEG correlations, in live brains and slices of neural tissue.

## Basic perception

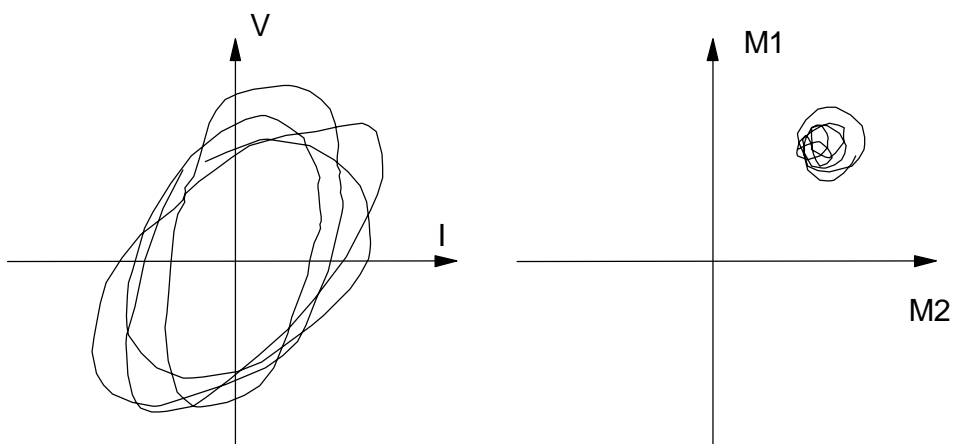
- **Basic processing of sensory signals: topographical maps**

Formation of these maps is elucidated by self-organizing neural network models learning in an unsupervised way.

- Short-term memory (STM), working memory, sensory buffers.  
Half-life about 7 seconds, capacity  $7 \pm 2$  chunks.
- Local and global reverberations and active memory.
- Perception as wave process - global pattern of excitations of transcortical neural cell assemblies (TNCAs) or attractor of dynamical system composed of transcortical columns.

Symbolic approach to dynamical systems - simplified description. Instead of the NCAs and TNCAs phase space description feature space representation of each attractor - mind object. Evidence from EEG observations for visual and auditory stimuli.

M1, M2 - average intrasassembly signals invoking attractor on the left.

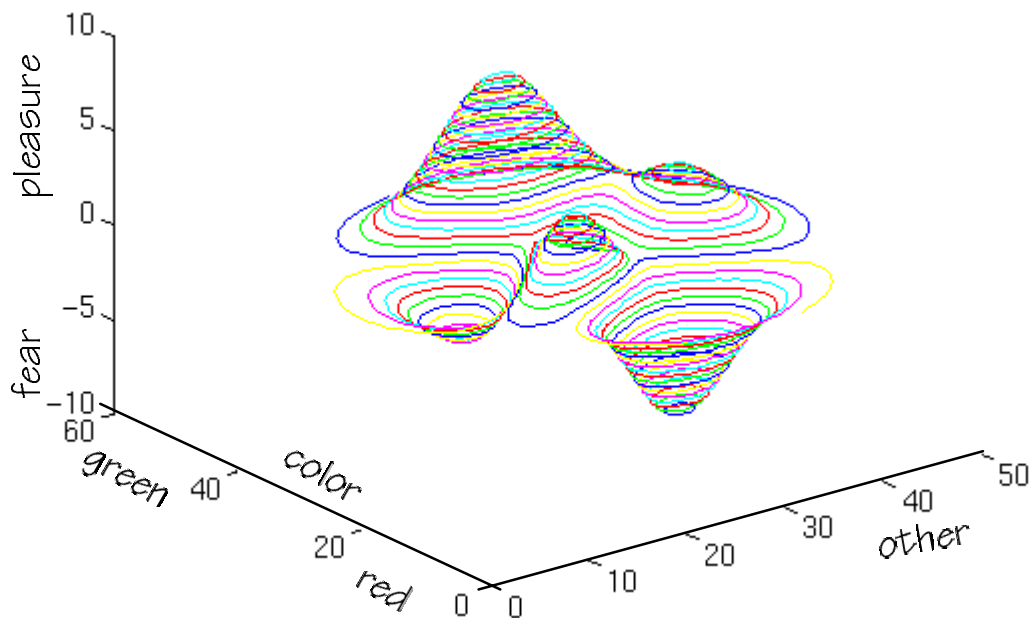


## Consciousness

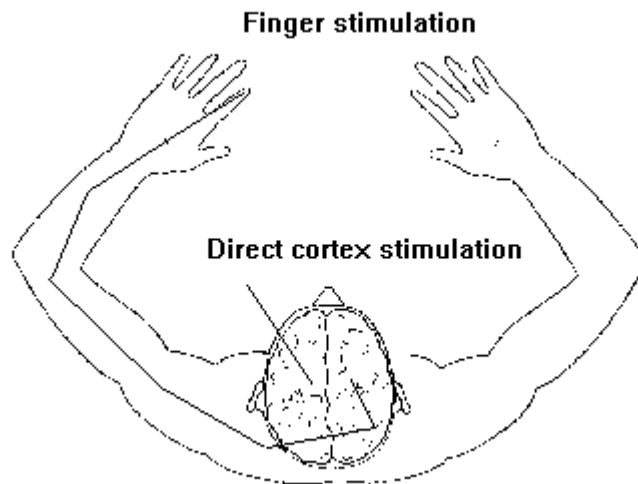
- *Consciousness is not a thing, it is an experience.*
- *Experience is a relaxation of brain and body states.*
- *Empirical approach to consciousness: which brain processes are labeled as conscious and which not conscious?*
- *Conscious experiences: only STM processes, global brain states*
- *Consciousness emerged together with the growing sophistication of animal brains.*
- *Primary consciousness requires global TNCA stable states.*
- *Reflective consciousness - only in brains sophisticated enough to form a representation of self. Strong qualia are connected with the representation of self.*

**Brain states** are labeled “conscious” when short term memory patterns are created by TNCA excitations. For humans required time for reaching a global attractor in noisy brain is about 0.5 seconds .

Consider an animal conditioned to associate the red color with pain and green color with pleasure. We can represent the stable states of the brain dynamics in coordinate system (fear-pleasure, color, other). Values on the pain -pleasure axis are defined by projections from visual neurocortex to subcortical structures such as amygdala.



*Libet's experiments: if a hand is stimulated by a brief electrical pulse and at the same time somatosensory cortex is directly stimulated by an electric current conscious perception of the hand stimulation precedes the perception of directly induced tactile sensation.*



- *Precognitive carousel (Walter 1963)*
- *Dynamical illusions: metacontrast, color Phi experiment, cutaneous rabbit (Geldard and Sherrick 1972), Stroop effect (1935)*
- *Blindsight*

*Dynamical models predict the content of consciousness !*

- *Recent theories of consciousness:*

*Baars and Newman - Global Workspace Theory*

*Crick and Koch - visual awareness*

*Eccles and Margenau - dualist theory, mind = probability field*

*Gray - subicular comparator*

*Wilber - spectrum of consciousness*

*Quantum phenomena in microtubules.*

## Immediate behavior and intuition

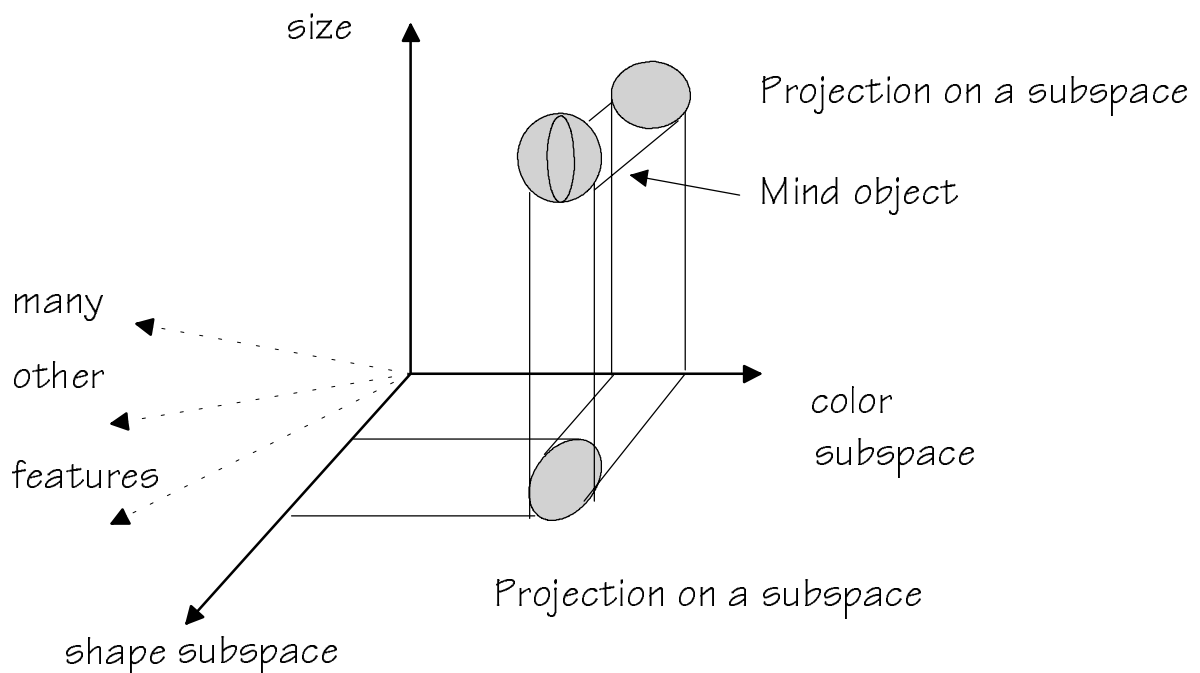
Immediate response - few seconds, no reasoning

The *Great Psychology Data Puzzle*: Newell's estimates that about 3000 regularities are known.

**Neurolinguistics:** Experiments with recognition of sounds/words: real words lead to stronger and longer activity than meaningless sounds.

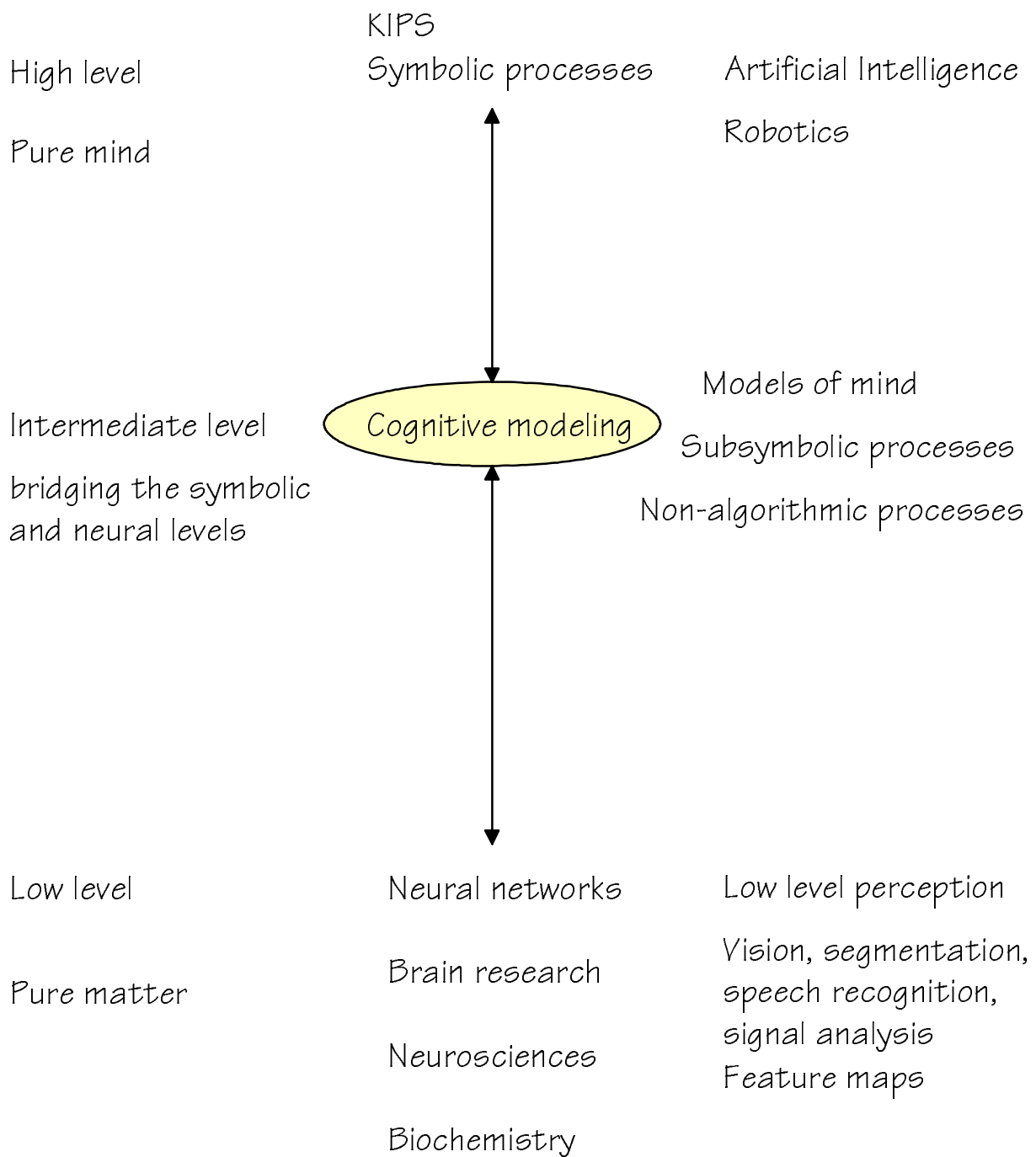
**Miyashita experiments** with monkeys performing delayed (up to 16 seconds) image matching. Stimulus selective 1 mm<sup>2</sup> region of neurocortex was identified. Representations are of prototypes, temporal sequence of pictures during training is preserved in correlations of neuron activities persisting up to 5 pictures.

**Feature spaces** and similarity-based intuitive recall: results obtained from psychological experiments with associative recall presented using multidimensional scaling (MDS) techniques are similar to the results from neural models.





Cognitive modeling - modeling at intermediate level, using concepts that may be reduced to ideas in neuroscience and extended to symbolic processes.



## Elements of a language useful for cognitive modeling.

- *The need for a language to speak about mind events.*

*Concept of physics:* matter, space, time, energy, momentum... are metaphorical descriptions of the “real world”.

*Cognitive concepts:* mind space, objects, events, dynamics, states, energy, momentum etc. are not more metaphorical.

Even in physics *phenomenological concepts* are still used, the flow of water is not described by atomic interactions.

In cognitive science *mind events* are hard to describe directly by dynamic neural activity, *phenomenological concepts* are needed.

- *What are the real objects of mind?*

*Platonic world:* abstract concepts, represented by abstract symbols, intellectual thought constructions.

*Real mind objects:* primarily preprocessed sensory data, iconic representations, perception-action multidimensional objects.

*Concepts* - the real grounding is in the sensory experiences.

*Perception:* low-level features are recognized - phonemes, colors, shapes, movements - primitive features of internal representations. Combinations of these features define *primary mind objects*.

*Mind space* = space spanned by the axes corresponding to the primitive features of representations, containing *mind objects*.

*Common expressions:* to put in *mind*, to have in *mind*, to keep in *mind*, to make up one's *mind*, be of *one mind* ... (*space*) are quite literally understood here.

- **Abstract mathematical theory** -> particular realizations.

All features of representations may define independent variables = new dimensions in the mind space.

- **Objects** in the mind space = **fuzzy areas** of non-vanishing density (other possibility: local coordinate systems).

Formation of mind objects takes many months.

Mind space is **almost empty**, only a very small fraction of all possible mind objects are realized by real minds.

- **Words:** some mind objects have **symbolic labels**, less fuzzy than other features, facilitating faster identification.
- Mind objects are defined by the "**mind function**" with the value  $M(X_i)$  proportional to the confidence of recognition of these objects.
- **Mind-maps** project relations from mind-space to 2-D surface.
- **Learning** = creating new objects and changing topographical relations among existing objects of the mind space.
- **Dynamics:** a collection of time dependent features of internal representation  $X_i(t)$  is called "**the mind state**". Object  $O$  is **activated** when the mind state  $X_i(t) \in O$  points at or belongs to it.

Simple recognition and learning processes activate only one object at a time. **Complex objects**, composed from simpler ones, are analyzed in the recognition process and created in the learning process.

- **Knowledge**=structure of mind function and evolution of mind state.
- **Short-term memory:** new objects appear for a short time and decay quickly. Many objects in the mind space may be active at the same time but only a few influence global activity, enable feedback (reflection) and are remembered as "an experience" by the system.
- **Evolution of the mind state** = **dynamics** of activation of objects, with energy and momentum in the mind space.
- **Reasoning and logic** = approximations to this evolution.

Transition probability  $p(A \rightarrow B)$  from the mind object  $A$  to the mind object  $B$  is given by the conditional expectancy (Sommerhoff 1990):

- “The brain’s internal representations of the world consist of linked sets of conditional expectancies of the what-leads-to-what kind, which are in the main based on the past experiences”.

Mathematical formalism useful for description of actual mind dynamics resembles quantum-mechanical formalism, with

$$p(A \rightarrow B) = \langle M_A | P | M_B \rangle$$

$M_A$  = rep. of the mind object  $A$  and  $P$  = momentum of the mind state.

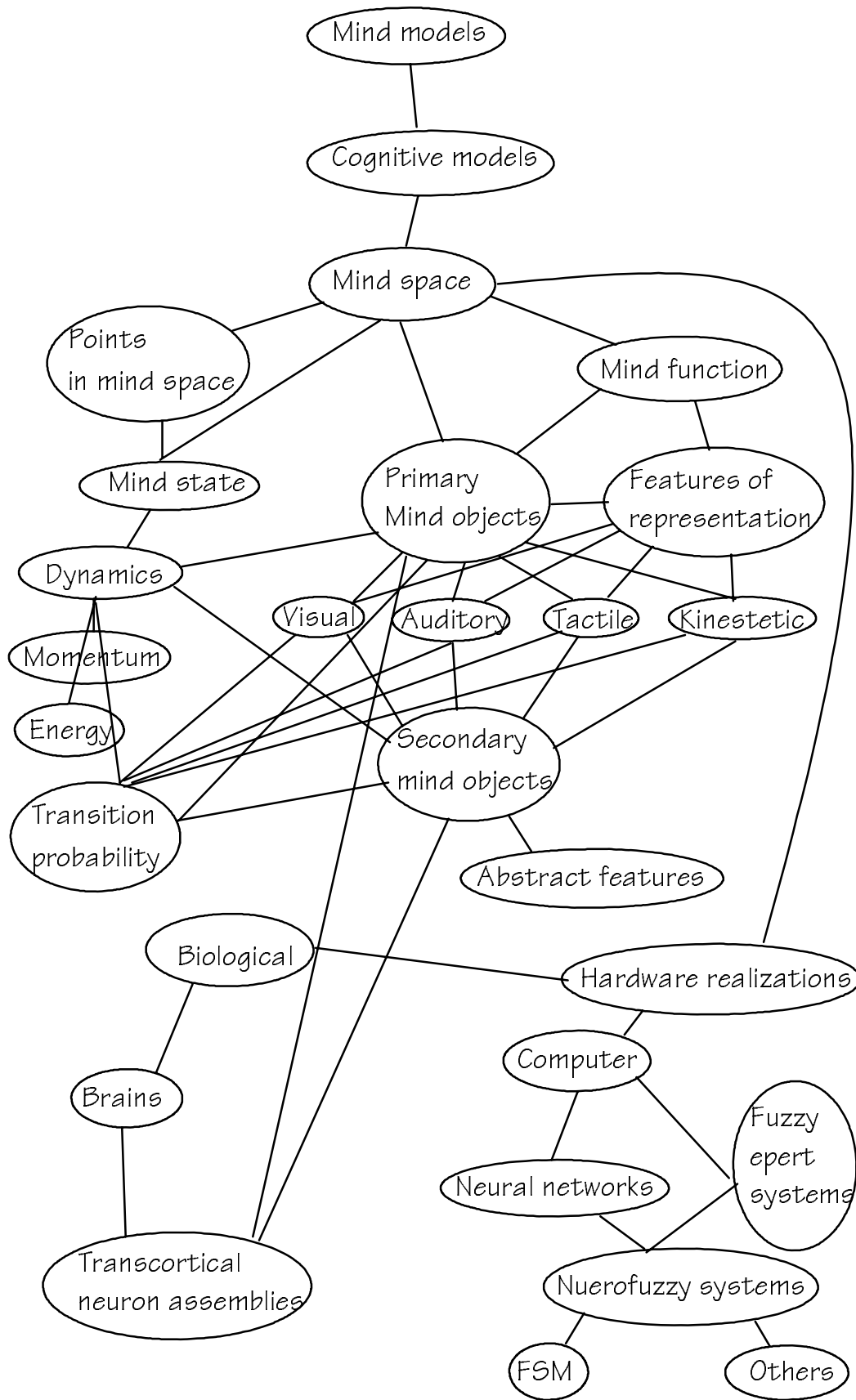
- “Intuition” is based on the topography of the mind space, depends on how well the mind mirrors reality.
- Mind space is an **open system**. The number of dimensions is not fixed and the topography of mind space is time dependent.

Mind function describing topography of the mind space is created via **unsupervised learning** and **genetic evolution**. Low level (sensory data processing)  $\Leftarrow$  dynamical attractor neural networks, feature maps (tonotopic, somatosensory, visual); high level  $\Leftarrow$  **TNCAs** specialize in representation of some objects of the mind space.

The forward cortical projections are accompanied by prominent projections back to the original sites making the activation by an internal dynamics indistinguishable from the external activation. The state of mind is constantly changing due to the sensory stimuli (including the proprioceptive) and the internal noise of the system.

- **Dreaming**, day-dreaming, hallucinations, granularity of experience and detailed descriptions of past experiences during hypnosis should be expected in mind-space systems.
- Mind function  $M(X_i(t))$  must be realized by some **hardware**.

**Natural realization** of  $M(X_i(t))$ : a modular neural network.



## Higher cognition

- Fodor and Pylyshyn criticism of connectionism: productivity, systematicity, compositionality and inferential coherence. Connectionist systems admit only causal connectedness, not true constituency relations.
- In neural systems **data** is represented by **prototypes** => data compression.

Relations should also be represented by prototypes => productivity.

- *A few recent connectionist models of higher cognition.*
- **Mental models** (Johnson-Laird 1983)
- DISCERN (Distributed SCRIPT processing and Episodic memory Network, Miikkulainen 1993) - modular connectionist system performing at the level of symbolic natural language systems.
- **Arithmetic learning** (J.A. Anderson et al. 1994) - beyond simple pattern association, showing how sensory information added to abstract representation lead to human-like errors/timing.
- **Logics and Neural Assemblies** (Ron Sun 1995) - implementation of complex schema structures in connectionist networks.
- **Knowledge-based neural networks** (Towell and Shavlik 1992).
- **Model of intuitionistic logic emerging from dynamical models** (Amsterdam and Lund), in particular from neural networks defined formally as 4-tuple  $\langle S, F, C, G \rangle$ ;  $S$ =states of the network,  $F$ =activation functions,  $C$ =connectivity (weight) matrix and  $G$  is a set of learning functions.

## Philosophical implications

- *Mind space and mind events* - purely mental, nonphysical.
- *Hardware realization* = modular neural networks.
- *Mind-body problem* arises only when linguistic labels are separated from other dimensions of the objects of mind.

Mental beliefs, ex. "low gas" observation, lead to intentional actions due to the **entrainment** of mind objects.

- **Meaning:** mind states are grounded in the kinesthetic image and sensoro-motoric schemes. Mind is embodied and creative rather than simply representing external reality
- **First-person experience:** States of mind are about something, since mind objects are non-decomposable mixtures of many features of internal representation.

Mind objects are activated using a subset of all features, bringing a state of mind into a specific region of mind space. **Qualia** are immediately accessible and have **observable** consequences: the probability of the next mind state depends on them.

No qualia are present in the Chinese Room example of Searle.

Words are particularly effective in activation of mind objects providing labels uniquely identifying the regions of mind space where "chunks of our experiences" are present.

Mind develops as a result of interaction with the environment, therefore the non-algorithmic nature of mind is a reflection of the non-algorithmic nature of the environment. Randomized computing, with chaotic component, allows to solve problems of high complexity in suboptimal way.

- Chomsky's "poverty of the stimulus" argument
- Systems processing mind objects, grounded in sensory experiences, are physical symbol processing systems.
- Artificial Intelligence is just a shadow of Artificial Mind.

An important task for the cognitive science is to elucidate the topography of the mind space and the internal representations of the mind objects in this space.

Minds of very different types should be possible, depending on the way the mind space is formed and the way objects are grounded in real experiences.



### Properties of FSM:

1. Symbolic interpretation, neural realization
2. Direct modeling of the mind (feature) space with fuzzy facts, direct association and controlled generalization.
3. Supervised and unsupervised learning, fine tuning of knowledge for pattern recognition.
4. Learns from **examples** and **general laws**, adding more nodes or removing existing nodes of the network if necessary.
5. Implementation of typical **expert system** production rules in the form IF (FACT 1.AND.FACT2.OR.FACT3...) THAN (FACT4) is straightforward. Facts are fuzzy.
6. Reasoning via one-dimensional searches, focusing on single variable, with the depth=no. of unknown features; gradient techniques also applicable.
7. Formation of **categories** and **metaconcepts** is straightforward
8. **Multi-scale** approach (focusing and defocusing) to concentration on relevant parts of conceptual space by changing dispersions of facts in the mind space.
9. Scaling linearly with the number of facts, ideal for parallel processing.

Many other systems may be based on the general idea of the mind space.

