

Soft Computing, Iizuka, 30.09-5.10.1996



## Categorization, prototype theory and neural dynamics.

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://www.phys.uni.torun.pl/kmk>

FTP archive: <ftp.phys.uni.torun.pl/pub/papers/kmk>

1. Psychological theories of categorization.
2. Neurodynamics and connections between mind and brain.
3. Encoding categories in feature spaces - simplified neurodynamics.
4. Base rate effects.
5. Summary

# 1. Psychological theories of categorization.

*Categorization = creation of mental categories, cognitive process.*

*Category learning, concept formation - ignores constraints resulting from the neural plausibility of postulated mechanisms.*

*Connectionist models are at best loosely inspired by the idea that neural processes are at the basis of cognition.*

*Correlation does not imply causation - avoid superpower !*

*Classical approach to categorization:*

*defining feature approach, conceptual hierarchies*

*fuzzy prototype approach, typical features*

*exemplar model*

*Modern models of categorization (Cohen and Massaro 1992):*

*Fuzzy Logical Model of Perception (FLMP)*

*Gaussian Multidimensional Scaling Model (GMM)*

*Theory of Signal Detection (TSD)*

*Feedforward Connectionist Model (FCM)*

*Interactive Activation and Competition Model (IAC)*

*All these models predict probabilities of responses in a prototypical two and four-response situations in an almost equivalent way.*

## 2. Neurodynamics, mind/brain connections.

Chaotic neuroactivity of large groups of neurons - Freeman (1975)

SMNI, Statistical Mechanics of Neocortical Interactions, activity at mesoscopic scale - Ingberg (1982)

Darwinian selection, Edelman (1987), and Cerebral Code, Calvin (1995)

DFA, Deterministic Finite State Automata, HMM (Hidden Markov Models) - from recurrent neural networks -> logical rules from dynamical behavior, leading to behavioral description, without mind.

Synergetic approach to neurodynamics - Haken, Kelso (1988)

HyperBF networks for 3D recognition + motor control, Poggio (1990)

The Human Brain Project: PSYCHE (J. Taylor, 1995)

CALM and the memory model of Murre (1996)

- How are the internal representations formed?

Little solid neurophysiological evidence. Best: Miyashita (1989)

Neural cell assemblies (NCA), originally introduced by Hebb (1949), transcortical NCA; competition - most neurons are inhibitory.

Topographical maps: somatotopic, tonotopic, visual, motoric ...

Minicolumns: diameter 30  $\mu\text{m}$ , 110 neurons in column in most areas.

Microcolumns: diameter 0.4-1 mm, about 0.1-1  $\text{mm}^2$ ,  $10^3$ - $10^5$  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.

### 3. Encoding categories in feature spaces - simplified neurodynamics.

Neural dynamics during categorization: unknown.

Perceptual categorization for olfactory, auditory, visual stimuli partially known.

- **Category learning task experiment:**

Shepard, Hovland and Jenkins (1961), replicated by R.M. Nosofsky, M.A. Gluck, T.J. Palmeri, S.C. McKinley and P. Glauthier (1994)

6 types of classification problems, increasing complexity, results determined by logical rules. 8 stimuli constructed from 3 binary-valued dimensions, eg: shape (square vs. triangle), color (black vs. white), and size (large vs. small).

4 stimuli in each of the two categories presented.

Type I - categorization using one dimension only.

Type II - two dimensions are relevant (XOR problem).

Type VI - most complex, 3 dimensions relevant, logic = enumerate stimuli in each of the categories.

Types III, IV, and V - intermediate complexity between Type II - VI. All 3 dimensions relevant, "single dimension plus exception" type.

Difficulty (number of errors made): Type I < II < III ~ IV ~ V < VI

Many psychological categorization models predict wrong ordering.

Nosofsky et.al: for separable dimensions OK, for integral dimensions (color) gets Type I < III ~ IV ~ V < II < VI

- Dynamical system for XOR categorization:

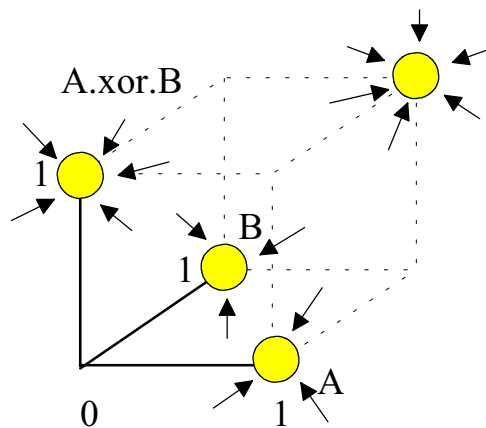
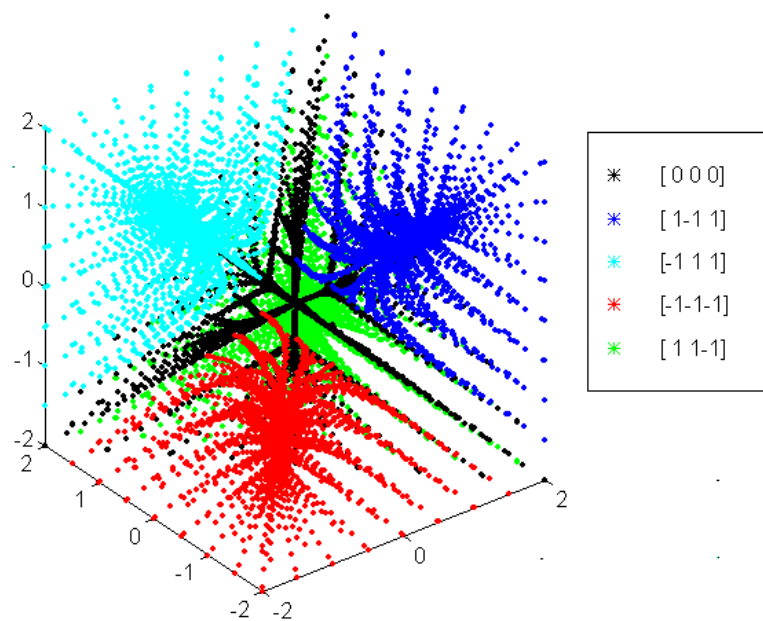
$$V(x,y,z) = 3xyz + \frac{1}{2}(x^2 + y^2 + z^2)^2$$

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

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

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

## Trajectories



## 4. Base rate effects

Neurodynamics -> Feature spaces

Analysis: symbolic dynamics, cell mapping (Hsu), neural networks

State trajectory using gradient dynamics:

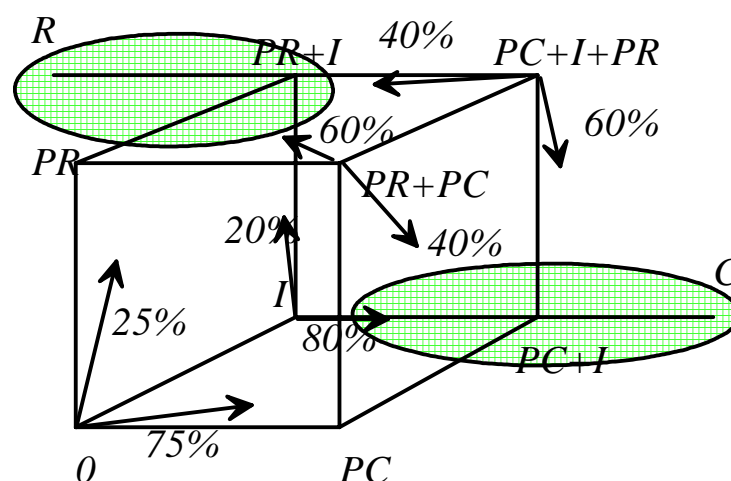
$$S(0) = X_{inp}; \dot{S}(t) = \beta \nabla_S M(S; t) / (1 + g(M(S; t))) + \eta(t)$$

$g(x)=0$  for small  $x$ , large around local maxima of the memory function.

Energy (distance) to go from one attractor (object) to another.

- **Base rate effect and inverse base rate effect:**

Relative frequencies (base rates) of categories used for classification but in some cases predictions contrary to the base rates are made.



# Interpretation

Neurodynamical point of view	Psychological point of view
Learning	
I+PC more frequent -> stronger synaptic connections, larger and deeper basins of attractors.	Symptoms I, PC typical for C because they appear more often.
To avoid attractor around I+PC leading to C deeper, localized attractor around PR is created.	Rare disease R - symptom I is misleading, attention shifted to PR associated with R.
Probing	
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 more frequently to C because I+PC puts the system in the middle of C basin.	I+PC+PR -> C because all symptoms are present and C is more frequent (base rates again).
Activation by PR and PC leads more frequently to R because the basin of attractor for R is deeper.	PC+PR -> R because R is distinct symptom, although PC is more common.

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

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.