

Consciousness and Creativity in Brain-Inspired Cognitive Architectures



Włodzisław Duch



Department of Informatics,
Nicolaus Copernicus University, Toruń, Poland

Google: W. Duch

CHIST-ERA, Rome, May 2010



A roadmap to human level intelligence



workshop organized by:

Włodzisław Duch (Google: W Duch)

Department of Informatics, Nicolaus Copernicus University, Torun, Poland
& School of Computer Engineering, Nanyang Technological Uni, Singapore

Nikola Kasabov (<http://www.kedri.info>)

KEDRI, Auckland, New Zealand

James Anderson, Paul Alloppenna, Robert Hecht-Nielsen, Andrew Coward,
Alexei Samsonovich, Giorgio Ascoli, Kenneth De Jong, Ben Goertzel

WCCI'2006, Vancouver, , British Columbia, Canada, July 17, 2006

Steps Toward an AGI Roadmap

Włodek Duch (Google: W. Duch)

Artificial General Intelligence (AGI):
architectures that can solve many problems and
transfer knowledge between the tasks.

Roadmaps:

- A Ten Year Roadmap to Machines with Common Sense (Push Singh, Marvin Minsky, 2002)
- Euron (EU Robotics) Research Roadmap (2004)
- Neuro-IT Roadmap (EU, A. Knoll, M de Kamps, 2006)

Challenges: Word games of increasing complexity:

- 20Q is the simplest, only object description.
- Yes/No game to understand situation.
- Logical entailment competitions.

Conference series, journal, movement.



In the year 1900 at the International Congress of Mathematicians in Paris David Hilbert delivered what is now considered the most important talk ever given in the history of mathematics, proposing 23 major problems worth working at in future. 100 years later the impact of this talk is still strong: some problems have been solved, new problems have been added, but the direction once set - identify the most important problems and focus on them - is still important.

It became quite obvious that this new field also requires a series of challenging problems that will give it a sense of direction.

Włodzisław Duch
Jacek Mańdziuk (Eds.)

Challenges for Computational Intelligence

Failures of AI



Many ambitious general AI projects failed, for example:

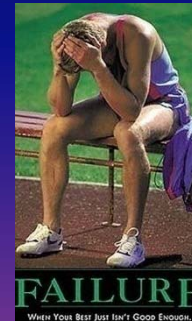
A. Newell, H. Simon, General Problem Solver (1957).

Eduardo Caianiello (1961) – mnemonic equations explain everything.

5th generation computer project 1982-1994.

AI has failed in many areas:

- problem solving, reasoning
- flexible control of behavior
- perception, computer vision
- language ...



Why?

- Too naive?
- Not focused on applications?
- Not addressing real challenges?



Ambitious approaches...



CYC, started by Douglas Lenat in 1984, commercial since 1995. Developed by CyCorp, with 2.5 millions of assertions linking over 150.000 concepts and using thousands of micro-theories (2004).

Cyc-NL is still a “potential application”, knowledge representation in frames is quite complicated and thus difficult to use.

Hall baby brain – developmental approach, www.a-i.com, failed.

Open Mind Common Sense Project (MIT): a WWW collaboration with over 14,000 authors, who contributed 710,000 sentences; used to generate **ConceptNet**, very large semantic network.

Some interesting projects are being developed now around this network but no systematic knowledge has been collected.

Other such projects:

HowNet (Chinese Academy of Science),
FrameNet (Berkeley), various large-scale ontologies,
MindNet (Microsoft) project, to improve translation.

Mostly focused on understanding of all relations the in text/dialogue.



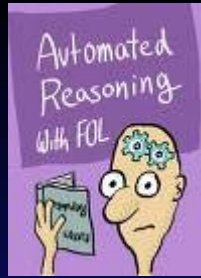
Challenges: language



- **Turing test** – original test is too difficult.
- **Loebner Prize** competition, for almost two decades played by chatterbots based on template or contextual pattern matching – cheating can get you quite far ...
- A “**personal Turing test**” (Carpenter and Freeman), with programs trying to impersonate real personally known individuals.
- **Question/answer systems**; Text Retrieval Conf. (TREC) competitions.
- **Word games**, 20-questions game - knowledge of objects/properties, but not about complex relations between objects. Success in learning language depends on automatic creation, maintenance and the ability to use large-scale knowledge bases.
- Intelligent tutoring systems? How to define milestones?



Challenges: reasoning



- **Super-expert** system in a narrow domain (Feigenbaum), needs a lot of general intelligence to communicate, should reason in math, bioscience or law, experts will pose problems, probe understanding.
- Same direction, but without NLP: **Automated Theorem Proving** (ATM) System Competitions (CASC) in many sub-categories.
- General AI in math: general theorem provers, perhaps using meta-learning techniques with specialized modules + NLP.
- Automatic curation of **genomic/pathways databases**, creation of models of genetic and metabolic processes for bioorganisms.
- Partners that **advise humans** in their work, evaluating their reasoning (theorem checking), adding creative ideas, interesting associations.

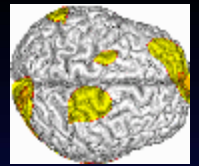
Real AI?



- General purpose systems that can be taught skills needed to perform human jobs, and to measure which fraction of these jobs can be done by AI systems (Nilsson, Turing's “child machine”).
- Knowledge-based information processing jobs – progress measured by passing a series of examinations, ex. accounting.
- Manual labor requires senso-motoric coordination, harder to do?
- DARPA Desert & Urban Challenge competitions (2005/07), old technology, integration of vision, signal processing, control, reasoning.
- Humanoid robotics: understanding of perception, attention, learning casual models from observations, hierarchical learning with different temp. scales.
- “Cognitive Assistant that Learns and Organizes” (CALO), part of DARPA Personalized Assistant that Learns (PAL) call, SRI+21 inst.

5-year project to create partners/personal assistants, rather than complete replacements for human workers (also CM RADAR).

Motivation & possibilities



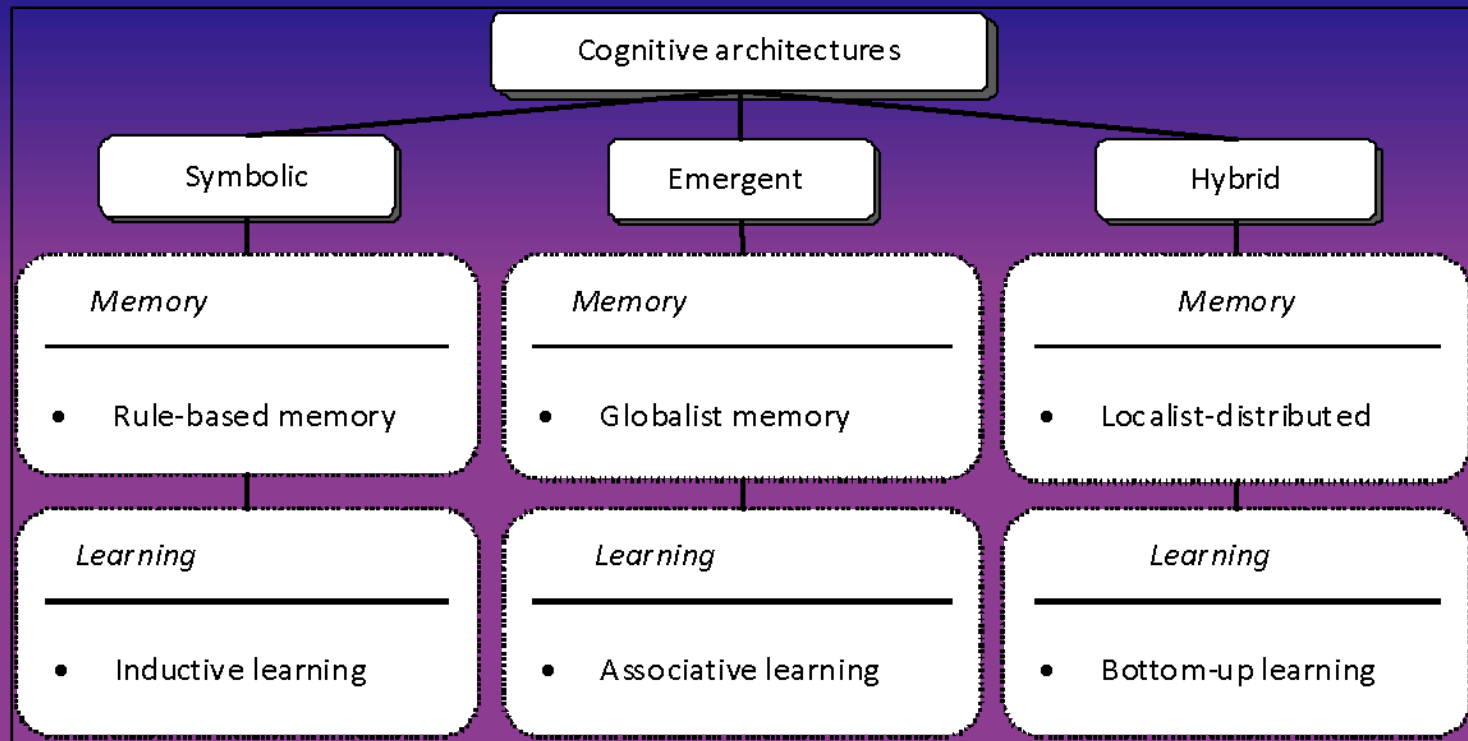
To reach human-level intelligence we need to go beyond pattern recognition, memory and control. How to reach this level?

- **Top down style**, inventing principles: laminar & complementary computing (Grossberg), chaotic attractors (Freeman & Kozma), AMD (John Weng), confabulation (Hecht-Nielsen), symbolic cognitive architectures (SOAR, ACT-R, CLARION ...).
- **Bottom up style**: systematic approximations to brain activity: neuromorphic systems, CCN (Izhikevich), Blue Brain, Emergent ...
- **Mind models**: “shadows of neurodynamics”: models of mind processes based on psychological spaces, providing inner perspective as an approximation to neurodynamics.
- Designs for artificial brains based on cognitive/affective architectures
Integration of perception, affect and cognition, large-scale semantic memory models, implementing control/attention mechanisms.

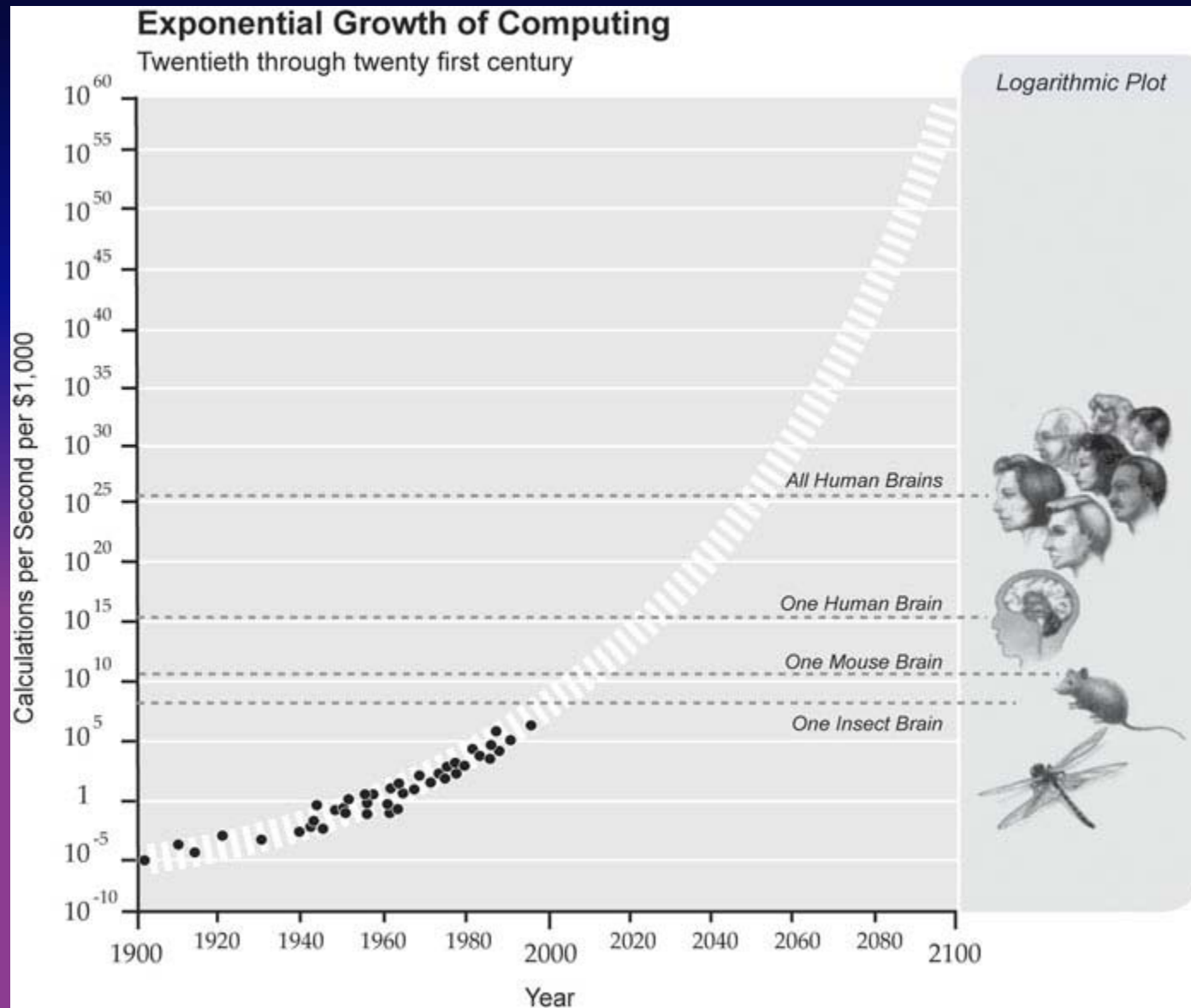
Cognitive architectures



- CA frequently created to model human performance in multimodal multiple task situations, rather than AGI.
- Newell, *Unified Theories of Cognition* (1990), 12 criteria for CS: behavioral: adaptive, dynamic, flexible; development, evolution, learning, knowledge integration, vast knowledge base, natural language, real-time performance, and brain realization.



Exponential growth of power



From
R. Kurzweil,

The Law of
Accelerating
Returns

By 2020 PC
computers
will match the
raw speed of
brain
operations!

What about
organization of
info flow?

Special hardware?

What is needed: elements performing like a spiking biological neurons connected in the layered 2-D structures of mammalian cerebral cortex.

- Nanoscale Memristor Device as Synapse in Neuromorphic Systems (Univ. of Michigan); neural plasticity in hardware.
- ALAVLSI, Attend-to-learn and learn-to-attend with analog VLSI, EU IST Consortium 2002-2005, Plymouth, ETH, Uni Berne, Siemens.
- A general architecture for perceptual attention and learning based on neuromorphic VLSI technology. Coherent motion + speech categorization, ended in 2005, new projects to create low-power analog spiking neuron chips along these lines are ongoing.
- RobotCub open-architecture for Cognition, Understanding, and Behavior, developing a humanoid robotic platform shaped as a 2.5 years old child, a lot of hardware for control.
- P-RAM neurons developed at KCL.
- ASSO-IC for Haikonen's associative architecture (VTT, planned).

Brain-like computing

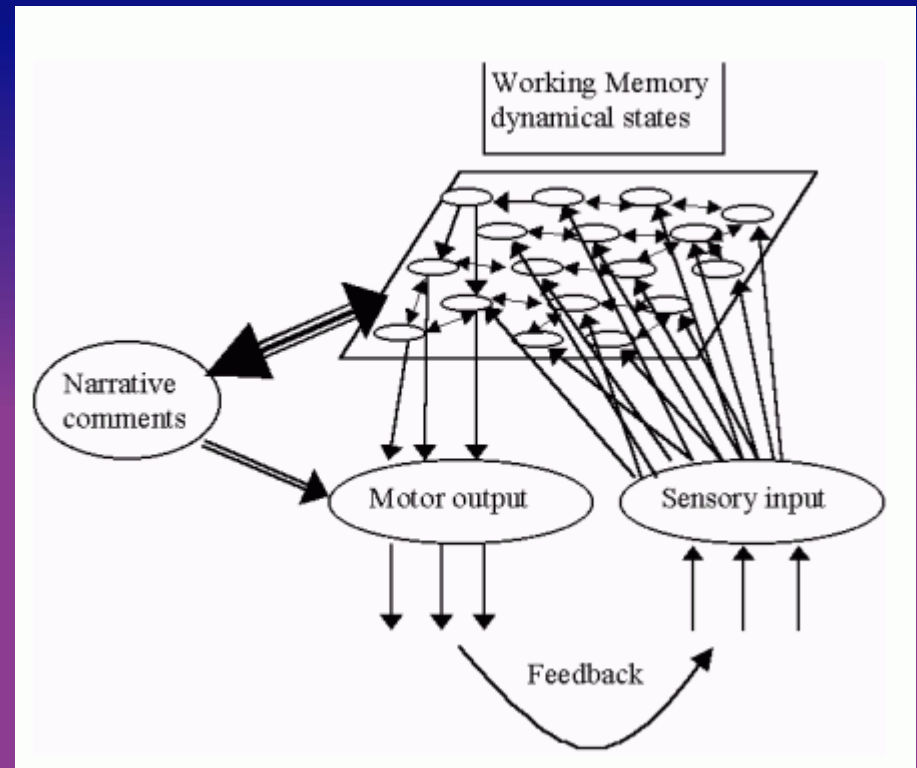
Brain states are physical, spatio-temporal states of neural tissue.

- I can see, hear and feel only my brain states! Ex: change blindness.
- Cognitive processes operate on highly processed sensory data.
- Redness, sweetness, itching, pain ... are all physical states of brain tissue.

In contrast to computer registers, brain states are dynamical, and thus contain in themselves many associations, relations.

Inner world is real! Mind is based on relations of brain's states.

Computers and robots do not have an equivalent of such WM.



Towards conscious robots



Few explicit attempts to build them so far.

Stan Franklin, "Conscious" Software Research Group, Institute of Intelligent Systems, University of Memphis, CMattie, LIDA projects:
an attempt to design and implement an intelligent agent under the framework of Bernard Baars' Global Workspace Theory.

My formulation: 1994, 1995 and: Brain-inspired conscious computing architecture. Journal of Mind and Behavior, Vol. 26(1-2), 1-22, 2005

Owen Holland, University of Essex: consciousness via increasingly intelligent behavior, robots with internal models, development of complex control systems, looking for "signs of consciousness", 0.5 M£ grant.

Pentti Haikonen (Nokia, Helsinki) + consortium led by VTT (Finland)
The cognitive approach to conscious machines (Imprint Academic 2003).
Simulations + microchips, ASSO-IC FET Proactive STERP, submitted.

Robot development

Nomad, DB, Cog, Kismet, Hal – develop robot mind in the same way as babies' minds, by social interactions.

Cog: saccadic eye movements, sound localization, motor coordination, balance, auditory/visual signal coordination, eye, hand and head movement coordination, face recognition, eye contact, haptic (tactile) object recognition ...

Interesting model of autism!

Kismet: sociable humanoid with emotional responses, that seems to be alive and aware.

DB: learning from demonstration, dance, pole balancing, tennis swing, juggling ...
complex eye movements, visuo-motor tasks, such as catching a ball.



Machine consciousness

Owen Holland (Essex Univ), Tom Troscianko and Ian Gilchrist (Bristol Univ) received 0.5 M£ from the EPSR Council (UK) for a project 'Machine consciousness through internal modeling', 2004-2007.

<http://www.machineconsciousness.org/>

To survive robots will plan actions, build a model of the world and a model of itself - its body, sensors, manipulators, preferences, history ...

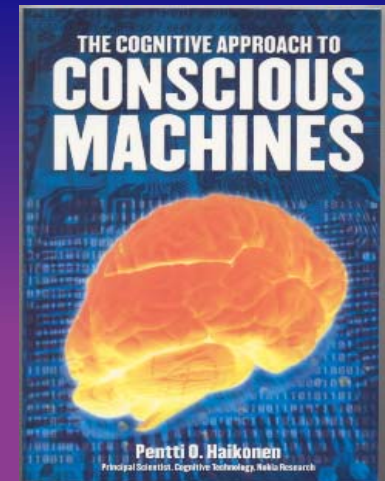
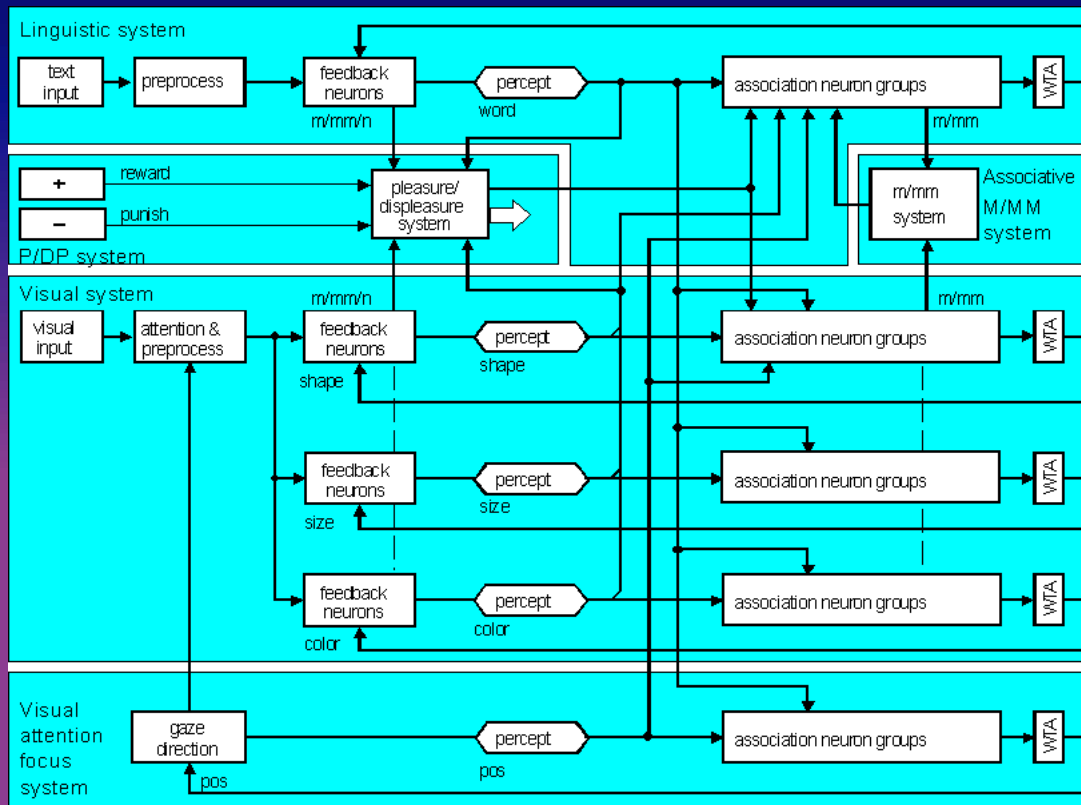
The main focus of interest will be the self-model; its characteristics and internal changes are expected to resemble those of the conscious self in humans, perhaps closely enough to enable some of the robots to be regarded as possessing a form of machine consciousness.

Increasingly complex biologically inspired autonomous mobile robots forced to survive in a series of progressively more difficult environments, and will then study the external and internal behavior of the robots, looking for signs and characteristics of consciousness.

Now Owen is more involved in mechanical design of robots ...

Conscious machines: Haikonen

Haikonen has done some simulations based on a rather straightforward design, with neural models feeding the sensory information (with WTA associative memory) into the associative “working memory” circuits.



What do we really want?

Do we really want conscious robots/systems that feel?

- Conscious in what sense? C-like cognitive behavior, in the sense of being aware - yes, robots should have it.
- Phenomenal consciousness with inner life, self, unreliable processes? Is this desired in machines?
- How reliable may machines with phenomenal C be?
- First, can we build them?
How to build a robot that feels, J.Kevin O'Regan at CogSys 2010 at ETH Zurich on 27/1/2010
- Sensorimotor account of action/perception shows that “hard problem” of consciousness is not such a problem after all.
Applications of this approach: sensory substitution, as long as the structure is right the signals are correctly interpreted.
- We want machines to be: human like, creative, intuitive, but also following our orders without psychological suffering.

Some failed attempts

Many have proposed the construction of brain-like computers, frequently using special hardware.

- **Connection Machines** from Thinking Machines, Inc. (D. Hills, 1987) was commercially almost successful, but never become massively parallel and the company went bankrupt.

CAM Brain (ATR Kyoto) – failed attempt to evolve the large-scale cellular neural network; based on a bad idea that one can evolve functions without knowing them. Evolutionary algorithms require supervision (fitness function) but it is not clear how to create fitness functions for particular brain structures without knowing their functions first; if we know the function it is easier to program it than evolve.



Attention-Based Artificial Cognitive Control Understanding System (ABACCUS)

Large EU integrated project with 9 participants (never started):

- King's College London (John G. Taylor, coordinator), UK
- Centre for Brain & Cognitive Development, Berkbeck College, University of London, UK
- Cognition and Brain Sciences Unit, Medical Research Council, UK
- Robotics and Embedded Systems, Technical University of Munich, G
- Institute of Neurophysiology and Pathophysiology, Universitätsklinikum Hamburg-Eppendorf, G
- Institute of Computer Science, Foundation for Research and Technology – Hellas, Heraklion, Crete, GR
- National Center for Scientific Research “Demokritos”, Athens, GR
- Dip. di Informatica, Sistemistica, Telematica, Università di Genova, I
- Dep. of Informatics, Nicholas Copernicus University, Torun, PL

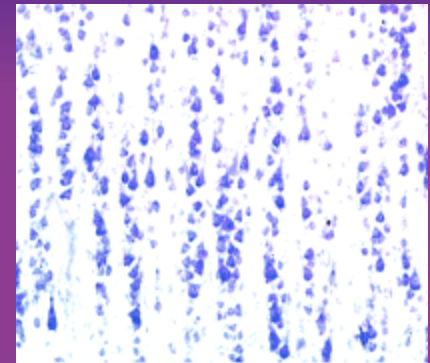
Other ambitious attempts

Artificial Development (www.ad.com) is building CCortex™, a complete 20-billion neuron simulation of the Human Cortex and peripheral systems, on a cluster of 500 computers - the largest neural network created to date.



Artificial Development plans to deliver a wide range of commercial products based on artificial versions of the human brain that will enhance business relationships globally. Rather unlikely?

The Ersatz Brain Project – James Anderson (Brown University), based on modeling of intermediate level cerebral cortex structures - cortical columns of various sizes (mini $\sim 10^2$, plain $\sim 10^4$, and hypercolumns $\sim 10^5$).



NofN, Network of Networks approximation, 2D BSB network.

<http://www.ersatzbrain.org/>

not much is happening with this project (2010).

Darwin/Nomad robots

G. Edelman (Neurosciences Institute) & collaborators, created a series of Darwin automata, brain-based devices, “physical devices whose behavior is controlled by a simulated nervous system”.

- (i) The device must engage in a behavioral task.
- (ii) The device’s behavior must be controlled by a simulated nervous system having a design that reflects the brain’s architecture and dynamics.
- (iii) The device’s behavior is modified by a reward or value system that signals the salience of environmental cues to its nervous system.
- (iv) The device must be situated in the real world.

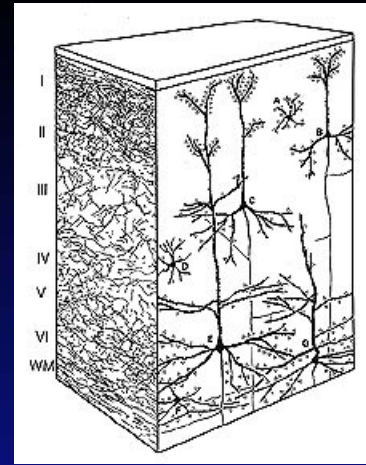


Darwin VII consists of: a mobile base equipped with a CCD camera and IR sensor for vision, microphones for hearing, conductivity sensors for taste, and effectors for movement of its base, of its head, and of a gripping manipulator having one degree-of-freedom; 53K mean firing +phase neurons, 1.7 M synapses, 28 brain areas.

Blue Brain

Models at different level of complexity:

<http://bluebrainproject.epfl.ch>



1. The Blue Synapse: A molecular level model of a single synapse.
2. The Blue Neuron: A molecular level model of a single neuron.
3. The Blue Column: A cellular level model of the Neocortical column with 10K neurons, later 50K, 100M connections (completed 2008).
4. The Blue Neocortex: A simplified Blue Column will be duplicated to produce Neocortical regions and eventually an entire Neocortex.
5. The Blue Brain Project will also build models of other Cortical and Subcortical models of the brain, and sensory + motor organs.

Blue Gene simulates ~100M minimal compartment neurons or 10-50'000 multi-compartmental neurons, with 10^3 - 10^4 x more synapses.

Next generation BG will simulate $>10^9$ neurons with significant complexity.

Great expectations but not new papers since 2008 ?

Ccortex & Brain Corp



Artificial Development (www.ad.com) is building CCortex™, a complete 20G neuron 20T connection simulation of the Human Cortex and peripheral systems, on a cluster of 500 computers - the largest neural network created to date.

Artificial Development plans to deliver a wide range of commercial products based on artificial versions of the human brain that will enhance business relationships globally.

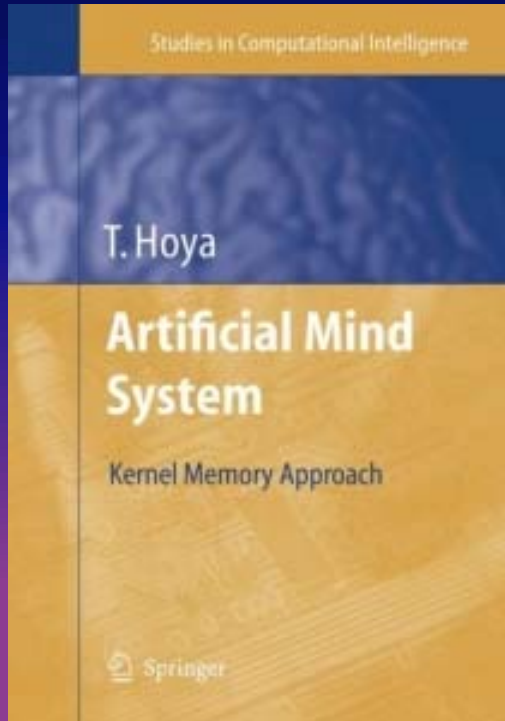
The project made bold statements in 2007, opened a lab in Kochi, Kerala, India, to “uncover relevant information on the functioning on the human brain, and help model and interpret the data.” Died in 2008?

The company is run by Marcos Guillen, who made money as ISP in Spain but has no experience in neuroscience or simulations.

More serious: **Brain Corporation** <http://braincorporation.com/> simulation of visual system and motor control using spiking neuron system. Eugene M. Izhikevich is CEO, already made whole 1 sec whole brain simulation.

Artificial Mind System (AMS)

Kernel Memory Approach



Series: **Studies in Computational Intelligence (SCI)**, Vol. 1 (270p)

Springer-Verlag: Heidelberg
Aug. 2005

available from:

<http://www.springeronline.com/>

by **Tetsuya Hoya**

BSI-RIKEN, Japan

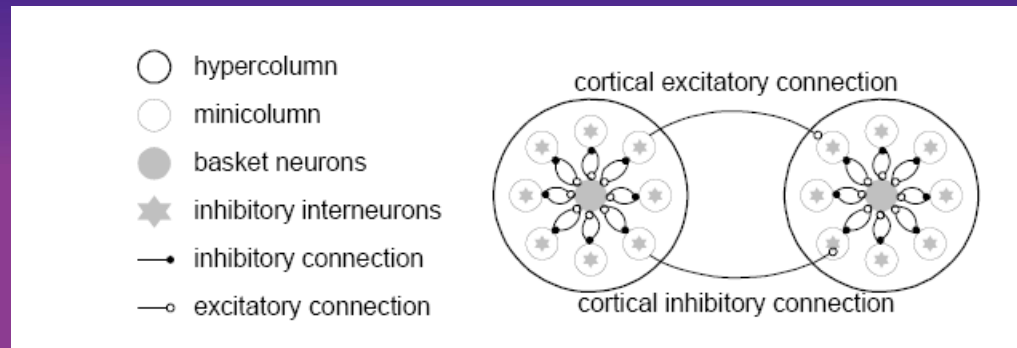
Lab. Advanced Brain Signal Processing

Hybrid approach, based on modularity of mind.

Bayesian Confidence Propagating NN.

Johansson/Lansner ideas:

- o Assumption: functional principles of cortex reside on a much higher level of abstraction than that of the single neuron i.e. closer to abstractions like ANN and connectionist models.
- o Target: artificial brain, compact, low-power, multi-network NN.
- o Mapping of cortical structure onto the BCPNN, an attractor network.
- o Implementation of BCPNN based on hyper columnar modules.

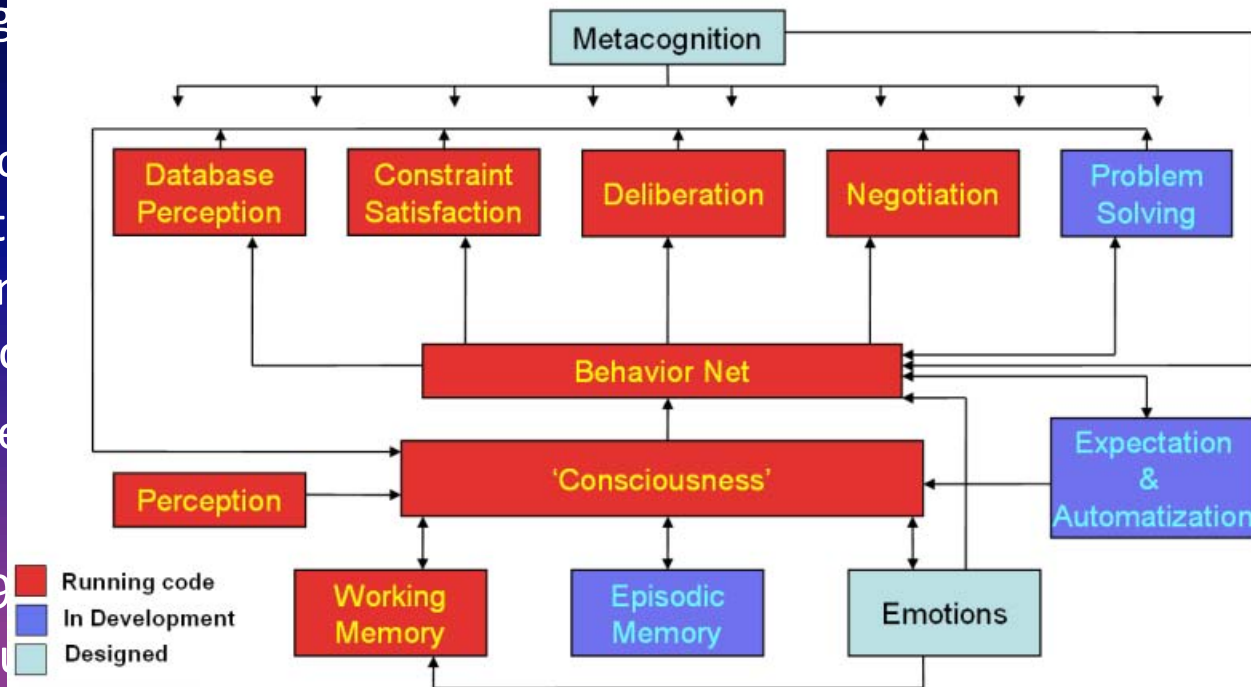


- o Hypercolumn needs $5 \cdot 10^9$ ops, with about $2 \cdot 10^6$ hypercolumns in human cortex, giving about 10^{16} ops.
- o No detailed structure proposed.

Hybrid CA: LIDA



IDA's Architecture



- **LIDA** (The *Learning In* framework for intelligent agents)

- LIDA: partly symbolic modules for perceptual episodic memory, action constraint satisfaction, metacognition, and cooperation

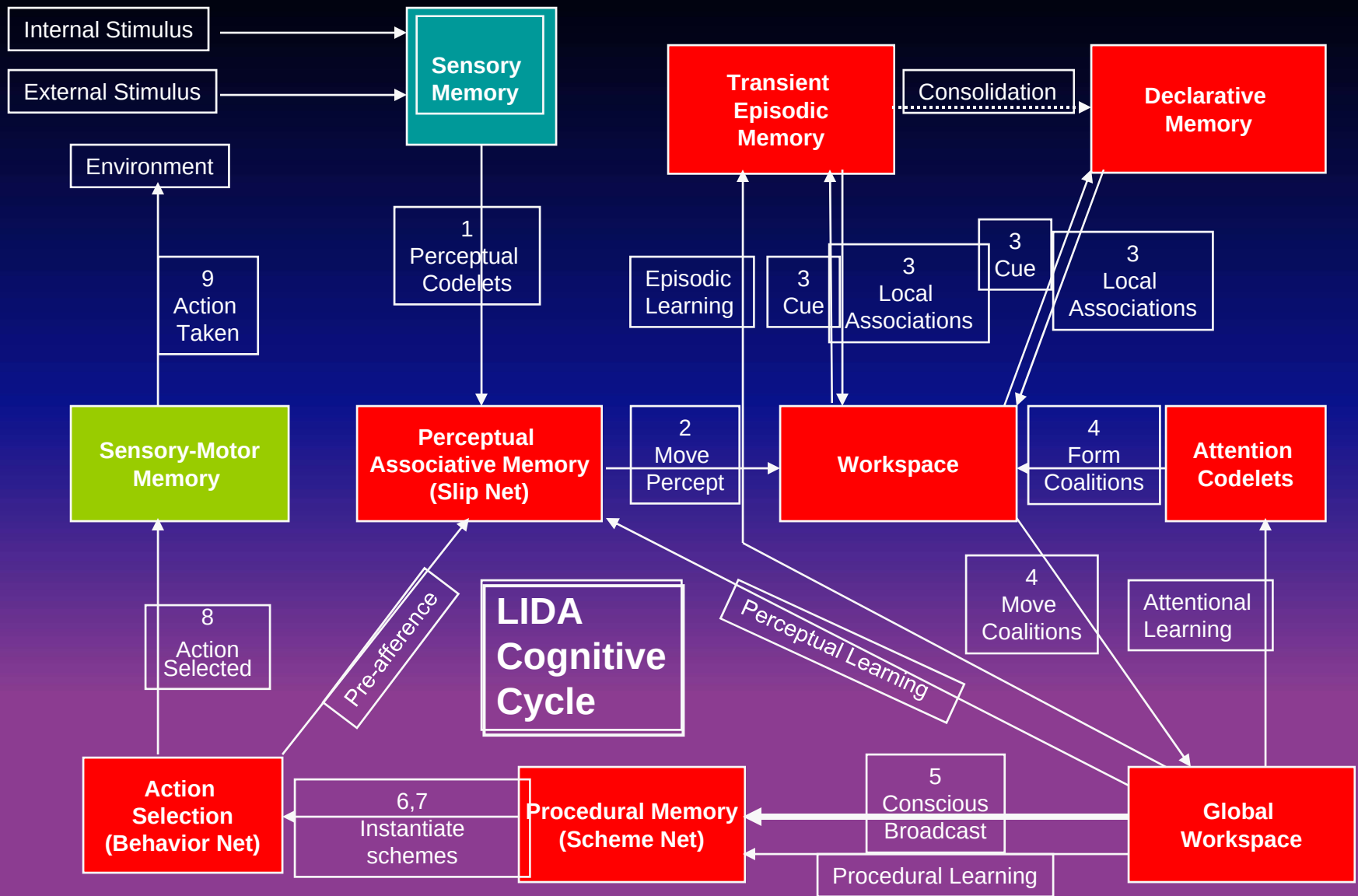
- Cooperation of codelets

- Perceptual, episodic,

- **DUAL** (B. Kokinov 1999) multi-agent architecture

micro-agents for memory and processing, agents form coalitions with emergent dynamics, at macrolevel psychological interpretations may be used to describe model properties.

- Micro-frames used for symbolic representation of facts, relevance in a particular context is represented by network connections/activations.

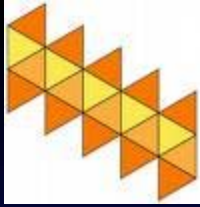


Hybrid CA: DUAL



- **DUAL** (B. Kokinov, since 1994), inspired by Minsky's "Society of Mind".
- Good example of hybrid, multi-agent architecture.
- Dynamic emergent computations: interacting micro-agents for memory and processing, agents form coalitions with emergent dynamics.
- At macrolevel psychological interpretations may be used to describe model properties.
- Micro-frames used for symbolic representation of facts.
- Relevance in a particular context is represented by network connections/activations.
- Used in a model of reasoning and psychophysics, model of active vision, attention, eye movement simulation.
- Combined with AMBR, Associative Memory-Based Reasoning, analogy making and reasoning.
- Scaling? No larger scale attempts so far.

Hybrid CA: Polyscheme



- **Polyscheme** (N.L. Cassimatis, 2002) integrates multiple methods of representation, reasoning and inference schemes in problem solving. Polyscheme “specialist” models some aspects of the world.
- Scripts, frames, logical propositions, neural networks and constraint graphs represent knowledge, interacting & learning from other specialists; attention is guided by a reflective specialist, focus schemes implement inferences via script matching, backtracking search, reason maintenance, stochastic simulation and counterfactual reasoning.
- High-order reasoning is guided by policies for focusing attention. Operations handled by specialists include forward inference, subgoaling, grounding, with different representations but same focus, may integrate lower-level perceptual and motor processes.
- Both for abstract and common sense physical reasoning in robots.
- Used to model infant reasoning including object identity, events, causality, spatial relations. Meta-learning, different approaches to problem solving.
- No ambitious larger-scale applications yet.

Hybrid CA: 4CAPS



- **4CAPS** (M.A. Just 1992), designed for complex tasks, language comprehension, problem solving or spatial reasoning.
- Operating principle: “Thinking is the product of the concurrent activity of multiple centers that collaborate in a large scale cortical network”.
- Used to model human behavioral data (response times and error rates) for analogical problem solving, human–computer interaction, problem solving, discourse comprehension and other complex tasks solved by normal and mentally impaired people.
- Has number of centers (corresponding to particular brain areas) that have different processing styles; ex. Wernicke’s area is constructing and selectively accessing structured sequential & hierarchical reps. Each center can perform/be a part of **multiple cognitive functions**, but has a limited computational capacity constraining its activity. Functions are assigned to centers depending on the resource availability, therefore the topology of the whole large-scale network is not fixed.
- Activity of 4CAPS modules **correlates with fMRI** and other neuroimaging data.
- M. Anderson: neural reuse, important concept (BBS, forthcoming 2010).

Hybrid CA: ACT-R



- **ACT-R** (*Adaptive Components of Thought-Rational*) (Anderson, >20 y), aims at simulations of full range of human cognitive tasks.
- Perceptual-motor modules, memory modules, pattern matcher.
- Symbolic-connectionist structures for declarative memory (DM), chunks for facts; procedural memory (PM), production rules.
Buffers - WM for inter-module communications and pattern matcher searching for production that matches the present state of buffers.
- Top-down learning approach, sub-symbolic parameters of most useful chunks or productions are tuned using Bayesian approach.
- Rough mapping of ACT-R architecture on the brain structures.
- Used in a large number of psychological studies, intelligent tutoring systems, no ambitious applications to problem solving and reasoning.
- **SAIL**: combining symbolic ACT-R with O'Reilley's Emergent architecture. Combines advantages of both approaches, but will it scale up?

Hybrid CA: others 2



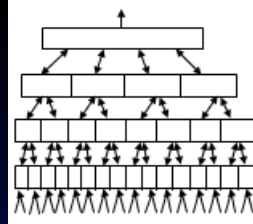
- **Shruti** (Shastri 1993), biologically-inspired model of human reflexive inference, represents in connectionist architecture relations, types, entities and causal rules using focal-clusters. These clusters encode universal/existential quantification, degree of belief, and the query status. The synchronous firing of nodes represents dynamic binding, allowing for representations of complex knowledge and inferences. Has great potential, but development is slow .
- **The Novamente AI Engine** (B. Goertzel, 1993), *psynet* model and “patternist philosophy of mind”: self-organizing goal-oriented interactions between patterns are responsible for mental states.
- Emergent properties of network activations lead to hierarchical and relational (heterarchical) pattern organization.
- Probabilistic term logic (PTL), and the Bayesian Optimization Algorithm (BOA) algorithms are used for flexible inference.
- Actions, perceptions, internal states represented by tree-like structures.
- New architecture, scaling properties are not yet known.

Emergent CA: others



- **NOMAD** (Neurally Organized Mobile Adaptive Device) (Edelman >20y) based on “neural Darwinism” theory, emergent architectures for pattern recognition task in real time. Large ($\sim 10^5$ neurons with $\sim 10^7$ synapses) simulated nervous system, development through behavioral tasks, value systems based on reward mechanisms in adaptation and learning, importance of self-generated movement in development of perception, the role of hippocampus in spatial navigation and episodic memory, invariant visual object recognition, binding of visual features by neural synchrony, concurrent, real-time control. Higher-level cognition?
- **Cortronics** (Hecht-Nielsen 2006), thalamocortical brain functions.
- Lexicons based on localist cortical patches with reciprocal connections create symbols, with some neurons in patches overlapping.
- Items of knowledge = linked symbols, with learning and information retrieval via confabulation, a competitive activation of symbols.
- Confabulation is involved in anticipation, imagination and creativity, on a shorter time scale than reasoning processes.

Emergent CA: directions



- The **NuPIC** (*Numenta Platform for Intelligent Computing*) (J. Hawking 2004), Hierarchical Temporal Memory (HTM) technology, each node implementing learning and memory functions. Specific connectivity between layers leads to invariant object representation. Stresses temporal aspects of perception, memory for sequences, anticipation.
- Autonomous mental development (J. Weng, ~10 y).
- M.P. Shanahan, internal simulation with a global workspace (2006) weightless neural network, control of simulated robot, very simple.
- **COLAMN** (M. Denham, 2006), and Grossberg “laminar computing”.
- E. Korner & G. Matsumoto: CA controls constraints used to select a proper algorithm from existing repertoire to solve a specific problem.
- DARPA Biologically-Inspired Cognitive Architectures (**BICA**) program (2006), several consortia formed, for example:
- “**TOSCA**: Comprehensive brain-based model of human mind”, Michigan, MIT Media Lab, AlgoTek, Dartmouth, Johns Hopkins, Harvard and Rutgers.

Where will it go?



- Many architectures, some developed over ~ 30 y, others new.
- Used in very few real-world applications.
- **Grand challenges** + smaller steps that lead to human and super-human levels of competence should be formulated to focus the research.
- Extend small demonstrations in which a cognitive system reasons in a trivial domain to results that may be of interest to experts, or acting as an assistant to human expert.
- What type of intelligence do we want?
H. Gardner (1993), at least seven kinds of intelligence: logical-mathematical, linguistic, spatial, musical, bodily-kinesthetic, interpersonal and intrapersonal intelligence, perhaps extended by emotional intelligence and a few others.
- To some degree they are independent!
Perhaps AGI does not have to be very general ... just sufficiently broad to achieve human-level competence in some areas and lower in others.

Trends

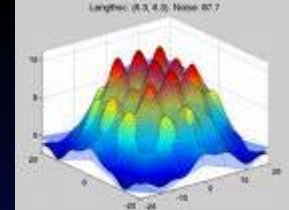


- Hybrid architectures dominate, but biological inspirations are very important, expect domination of BICA architectures.
- Focus is mainly on the role of thalamo-cortical and limbic systems, identified with cognitive and emotional aspects.
- Several key brain-inspired features should be preserved in all BICA:
hierarchical organization of information processing at all levels;
specific spatial localization of functions, flexible use of resources, time scales;
attention; role of different types of memory, imagination, intuition, creativity.

Missing so far:

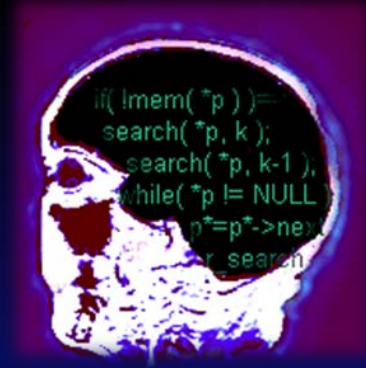
- Specific role of left and right hemisphere, brain stem etc.
- Many specific functions, ex. various aspects of self, fear vs. apprehension, processed by different amygdala structures.
- Regulatory role of the brain stem which may provide overall meta-control selecting different types of behavior is completely neglected.

BICA as approximation



- Significant progress has been made in drawing inspirations from neuroscience in analysis of perception, less in higher cognition.
- For example, neurocognitive approach to linguistics has been used only to analyze linguistic phenomena, but has no influence on NLP.
- “Brain pattern calculus” to approximate spreading neural activation in higher cognitive functions is urgently needed! How to do it?
Neural template matching? Network-constrained quasi-stationary waves describing global brain states $\Psi(w, Cont)$?
- Practical algorithms to discover “pathways of the brain” has been introduced recently (Duch et al, in print) to approximate symbolic knowledge & associations stored in human brain.
- Efforts to build concept descriptions from electronic dictionaries, ontologies, encyclopedias, results of collaborative projects and active searches in unstructured sources are under way.
- Architecture that uses large semantic memory to control an avatar playing word games has been demonstrated.

Neurocognitive informatics



Use inspirations from the brain, derive practical algorithms!

My own attempts - see my webpage, Google: W. Duch

1. Mind as a shadow of neurodynamics: geometrical model of mind processes, psychological spaces providing inner perspective as an approximation to neurodynamics.
2. Intuition: learning from partial observations, solving problems without explicit reasoning (and combinatorial complexity) in an intuitive way.
3. Neurocognitive linguistics: how to find neural pathways in the brain.
4. Creativity in limited domains + word games, good fields for testing.

Duch W, Intuition, Insight, Imagination and Creativity,
IEEE Computational Intelligence Magazine 2(3), August 2007, pp. 40-52

Intuition

Intuition is a concept difficult to grasp, but commonly believed to play important role in business and other decision making; „knowing without being able to explain how we know”.

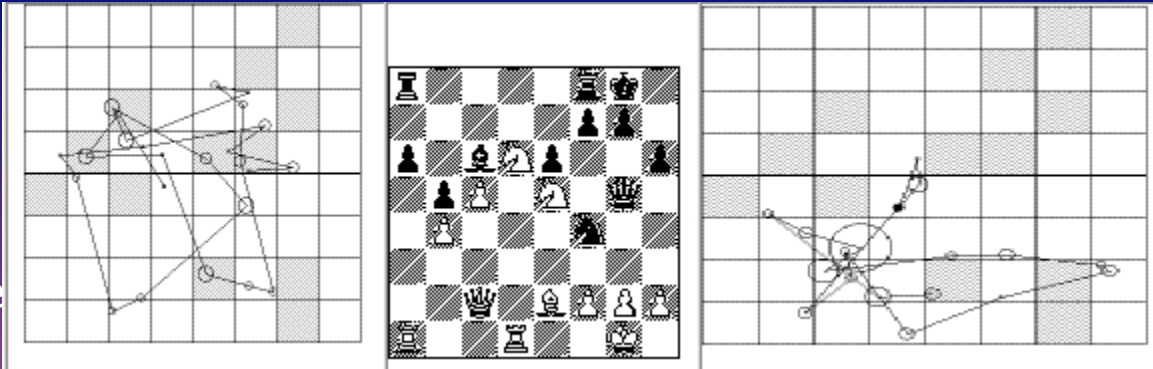


Sinclair Ashkanasy (2005): intuition is a „non-sequential information-processing mode, with cognitive & affective elements, resulting in direct knowing without any use of conscious reasoning”.

3 tests measuring intuition: Ratio
Type Inventory (MBTI) and Accu

Different intuition measures are
theoretical concept of intuition.

intuition scale and some measures of creativity.

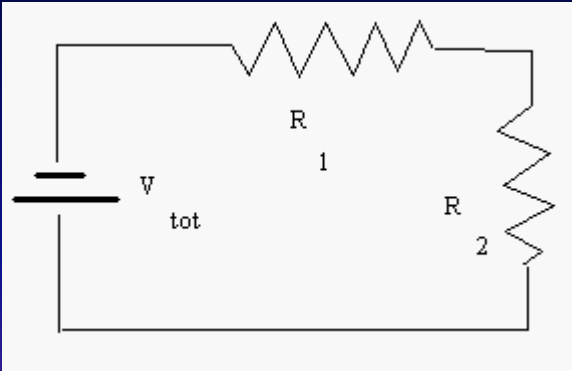


ANNs evaluate intuitively? Yes, although intuition is used also in reasoning.

Intuition in chess has been studied in details (Newell, Simon 1975).

Intuition may result from implicit learning of complex similarity-based evaluation that are difficult to express in symbolic (logical) way.

Intuitive thinking



Question in qualitative physics (PDP book):
if R_2 increases, R_1 and V_t are constant, what will happen with current and V_1 , V_2 ?

Learning from partial observations:

Ohm's law $V=I \times R$; Kirhoff's $V=V_1+V_2$.

Geometric representation of qualitative facts:

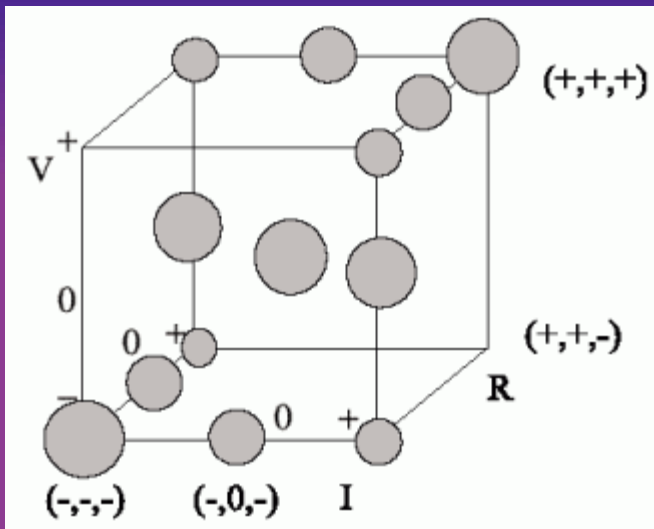
+ increasing, 0 constant, - decreasing.

True (I_-, V_-, R_0) , (I_+, V_+, R_0) , false (I_+, V_-, R_0) .

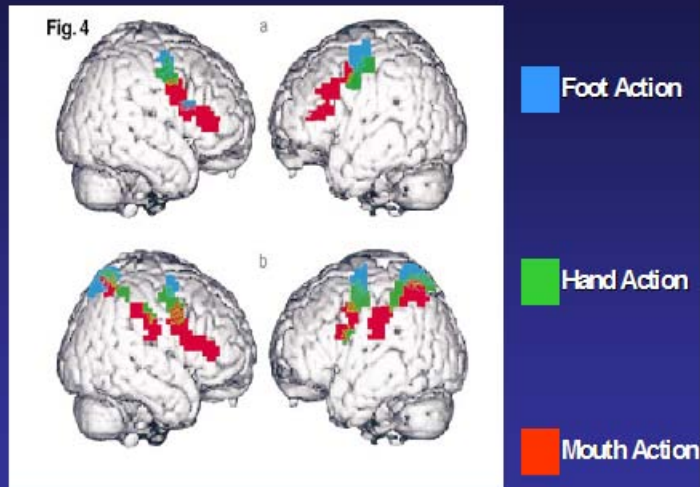
5 laws: 3 Ohm's 2 Kirhoff's laws.

All laws $A=B+C$, $A=B \times C$, $A^{-1}=B^{-1}+C^{-1}$, have identical geometric interpretation!

13 true, 14 false facts; simple P-space, but complex neurodynamics.



Somatotopy of Action Observation



Buccino et al. Eur J Neurosci 2001

s in the brain



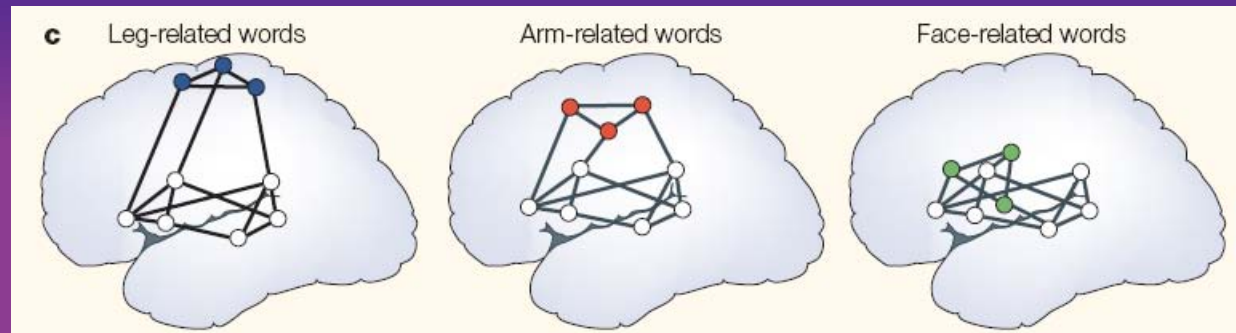
show that most likely categorical, used, not the acoustic input.

words => semantic concepts.

semantic by 90 ms (from N200 ERPs).

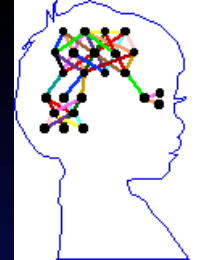
Science of Language. On Brain Circuits of Words University Press.

Action-perception networks inferred from ERP and fMRI



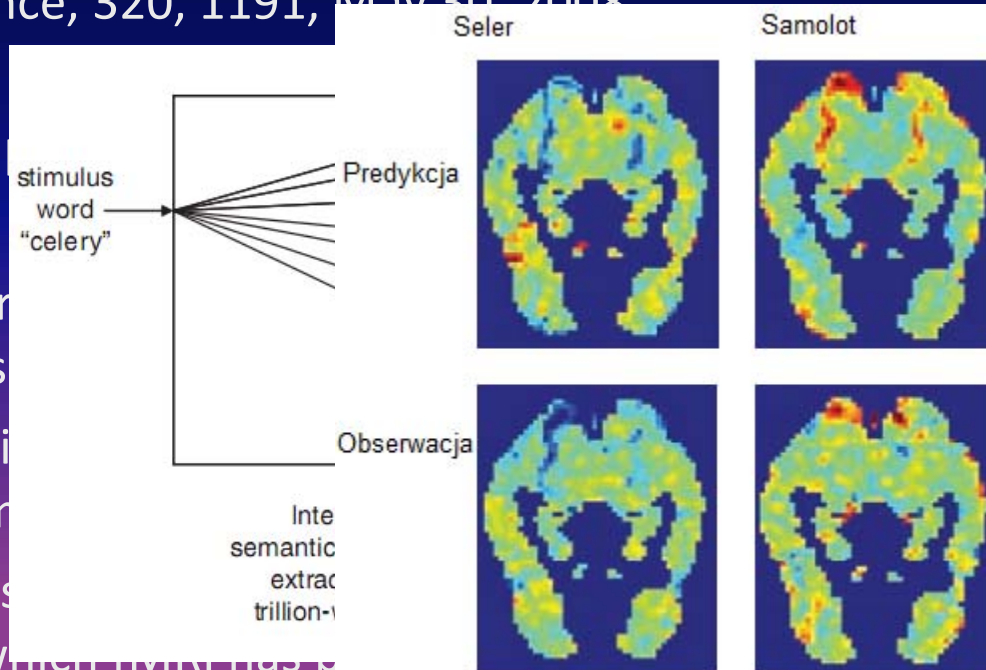
Left hemisphere: precise representations of symbols, including phonological components; right hemisphere? Sees clusters of concepts.

Neuroimaging words



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

- Clear differences between fMRI scans about different nouns.
- Reading words and seeing the drawings presumably reflecting semantics.
- Although individual variance is significant of different people, a classifier model can predict the meaning of words.
- Model trained on ~10 fMRI scans and brain activity for over 100 nouns for which fMRI has been recorded.



Overlaps between activation of the brain for different words may serve as expansion coefficients for word-activation basis set.

In future: I may know what you'll think before you will know it yourself!
Intentions may be known seconds before they become conscious!

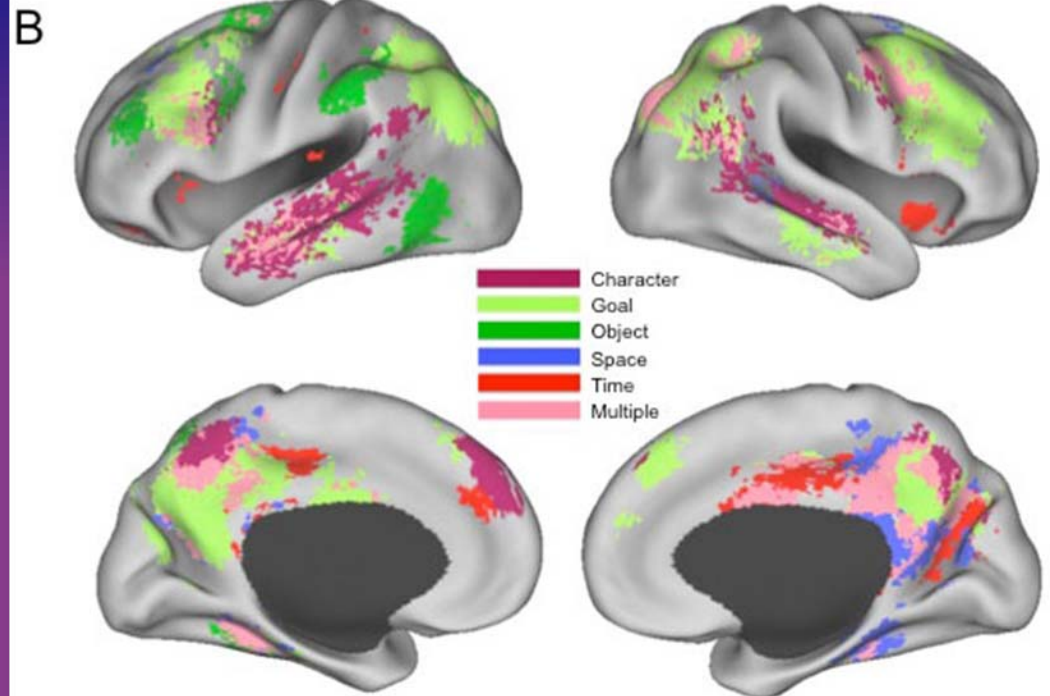
Nicole Speer et al.
Reading Stories Activates
Neural Representations of
Visual and Motor
Experiences.

Psychological Science
(2010, in print).

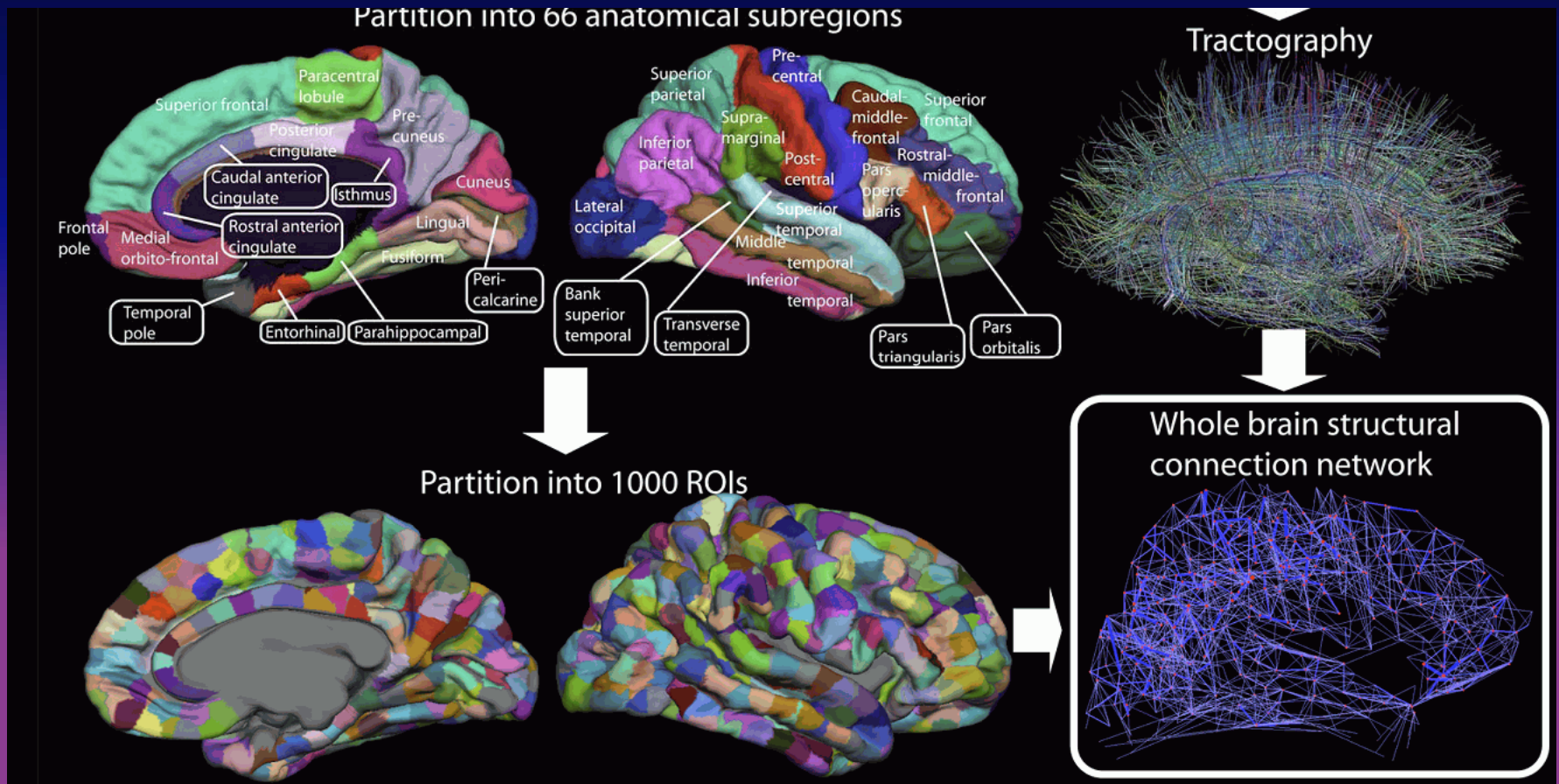
Meaning: always slightly
different, depending on the
context, but still may be
clusterized into relatively
small number of distinct
meanings.

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.						

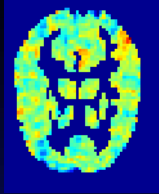


Connectome Project



Brain-based representations of concepts should be possible.

Hidden concepts



- **Language**, symbols in the brain: phonological labels associated with prototypes of distributed activations of the brain.

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

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

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

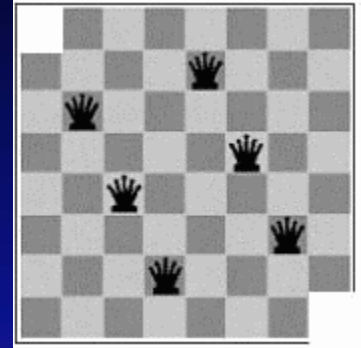
- Right hemisphere is as busy as left – concepts without verbal labels?
- Evidence: insight phenomena, intuitive understanding of grammar, etc.

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 question (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 ... to many combinations to try.

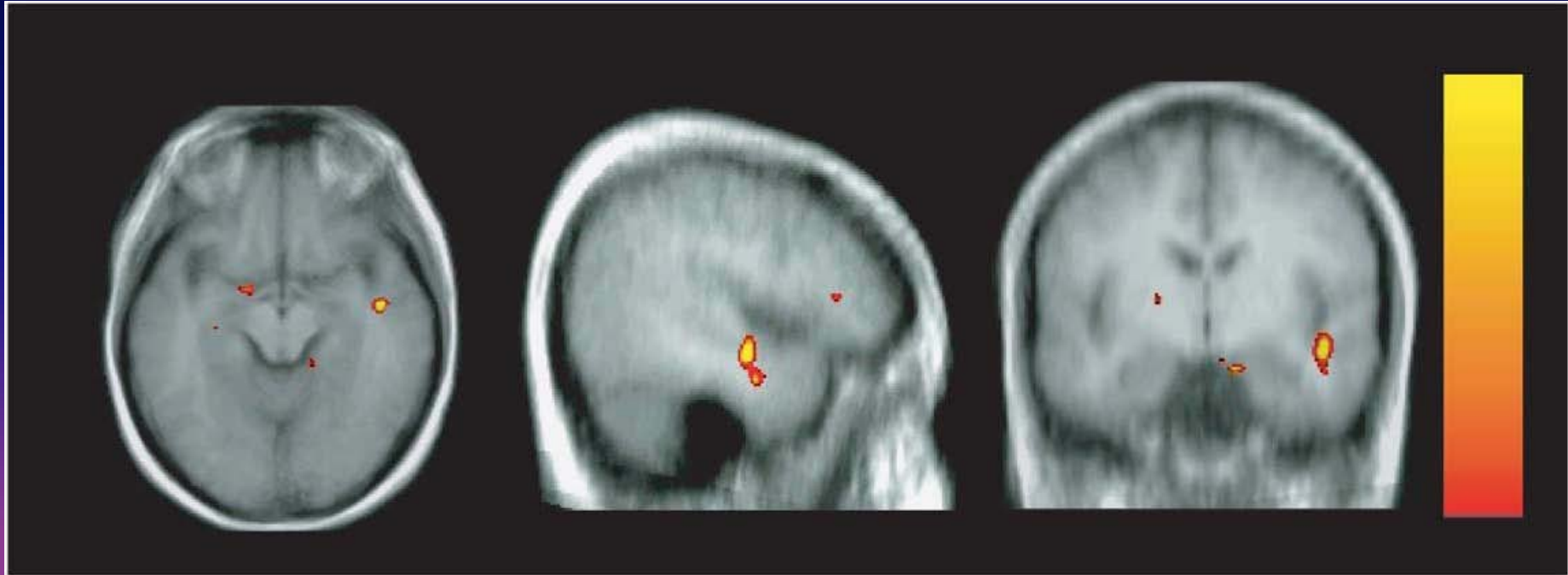
Logical, symbolic approach has little chance to create proper activations in the brain, linking new ideas: otherwise there will be too many associations, making thinking difficult.



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.

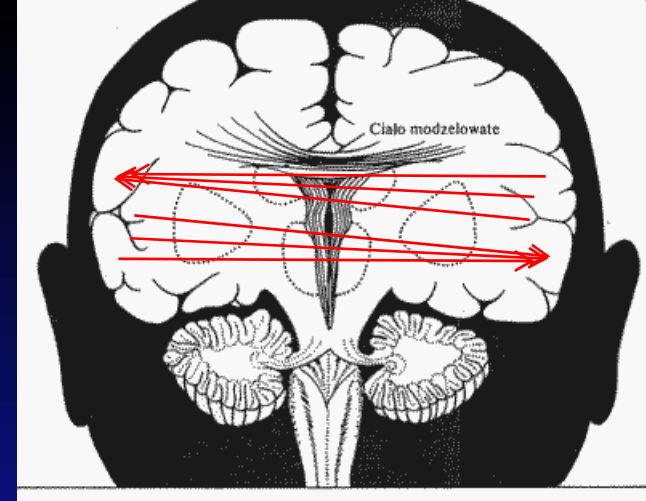


An increased activity of the right hemisphere anterior superior temporal gyrus (RH-aSTG) was observed during initial solving efforts and insights. About 300 ms before insight a burst of gamma activity was observed, interpreted by the authors as „making connections across distantly related information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

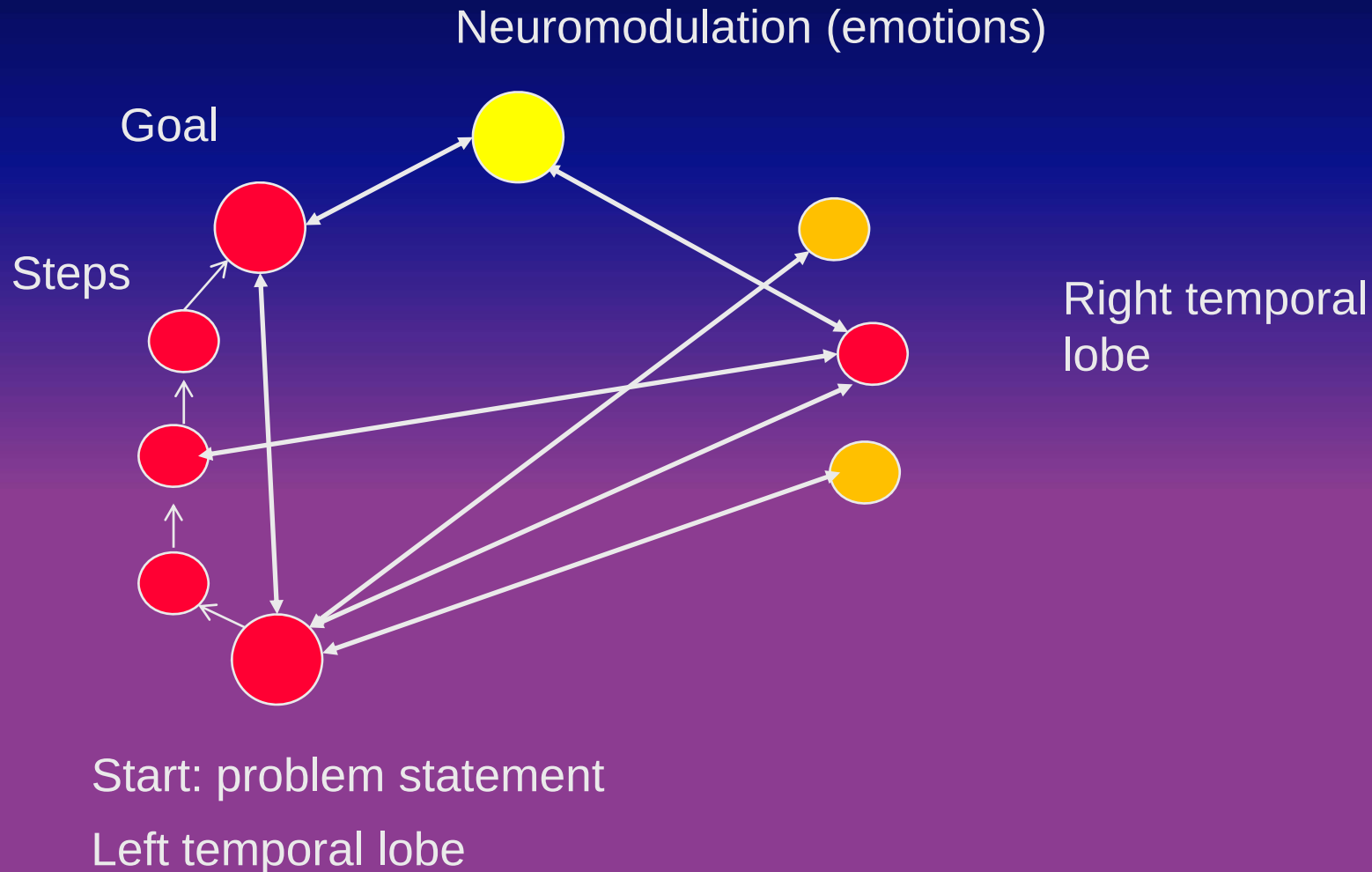
Insight interpreted

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



Dog breeds

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, all traditional categorizations are based on behaviour and features that are not easy to observe.

- Ontologies do not agree with visual similarity.
- Brains discover it easily => not all brain states have linguistic labels.



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!

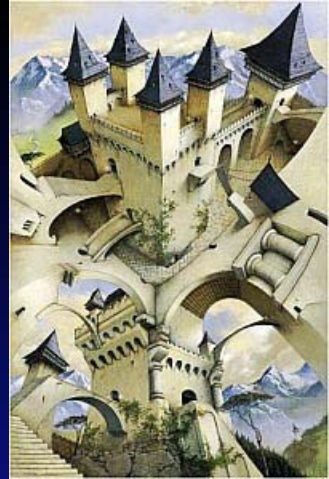
Reasoning: imaging relations, “seeing” mental picture, semantic?

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
3. Tendency to focus on a few possible models => irrational decisions.

Cognitive illusions are just like visual illusions.

M. Piattelli-Palmarini, Inevitable Illusions: How Reason Deceives the Mind (1996)

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

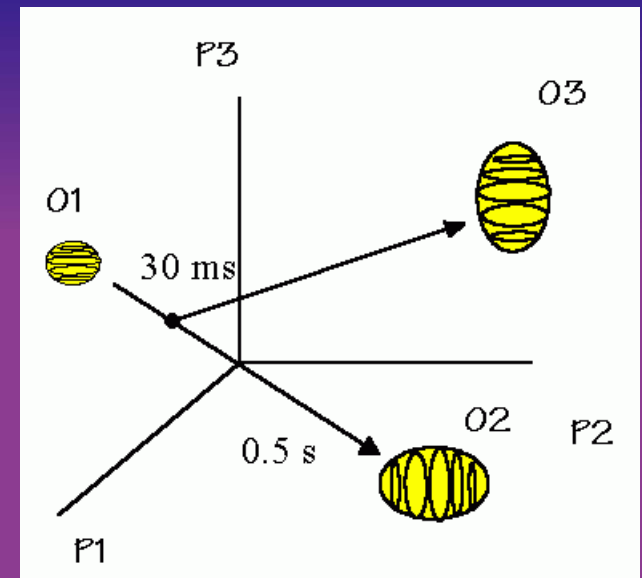
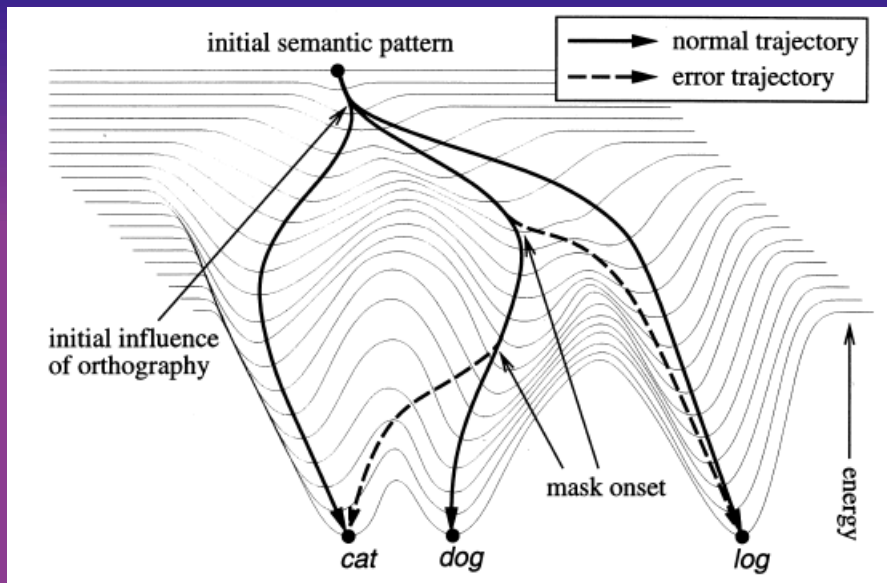


Amazing, but mental models theory ignores evidence for learning in any form! How and why do we reason the way we do?
I'm innocent! My brain made me do it!

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.



Model of reading & dyslexia



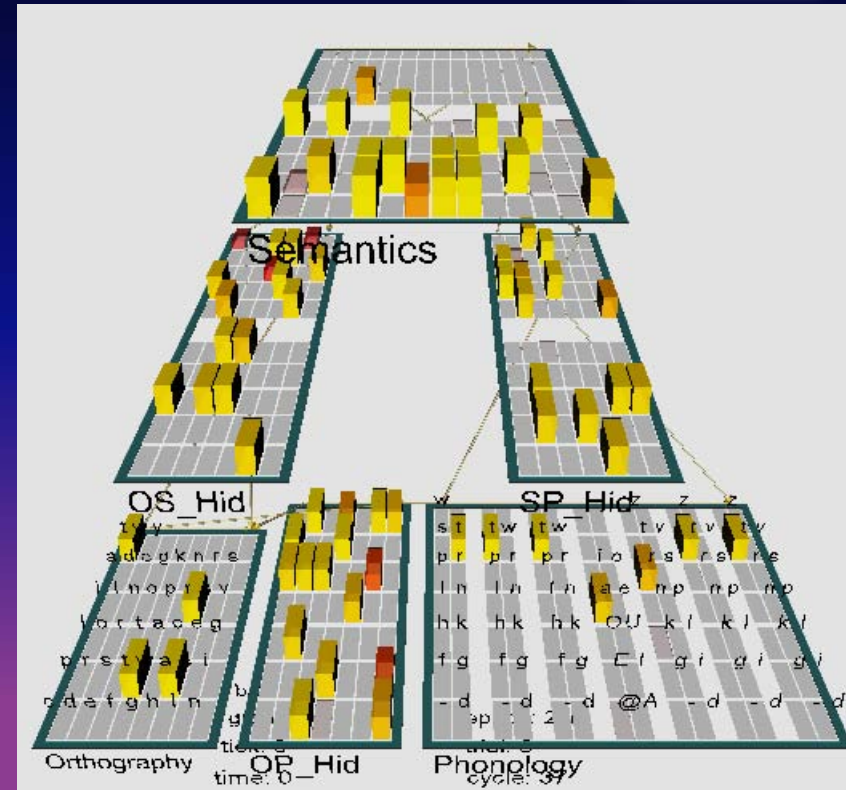
Emergent neural simulator:

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.

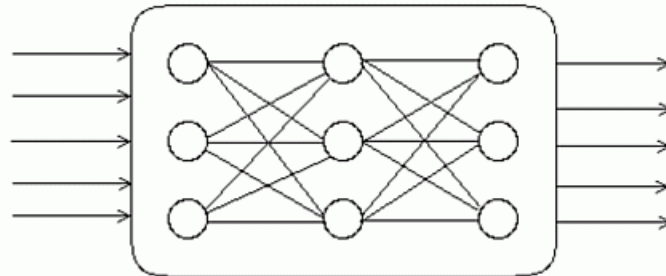
Hidden layers in between.



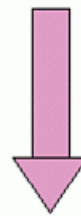
Learning: mapping one of the 3 layers to the other two.

Fluctuations around final configuration = attractors representing concepts.

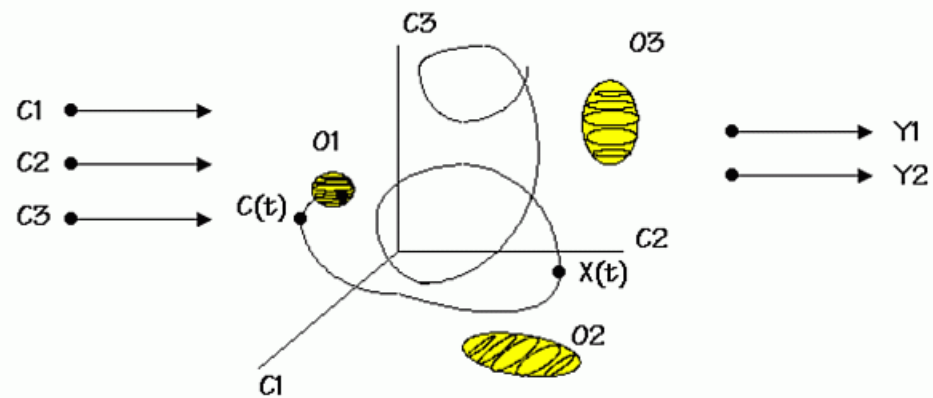
How to see properties of their basins, their relations?



Neurodynamics



Psychological
space



P-spaces

Psychological spaces: how to visualize inner life?

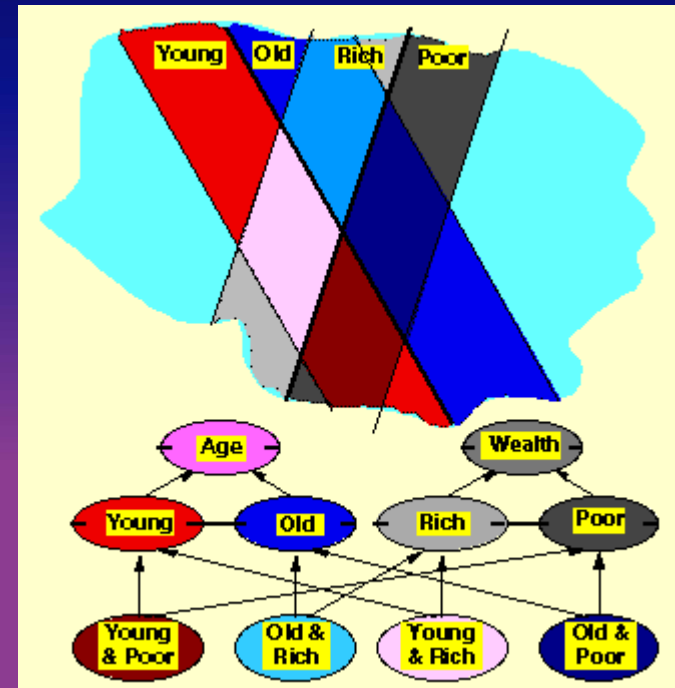
K. Lewin, The conceptual representation and the measurement of psychological forces (1938), cognitive dynamic movement in phenomenological space.

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.

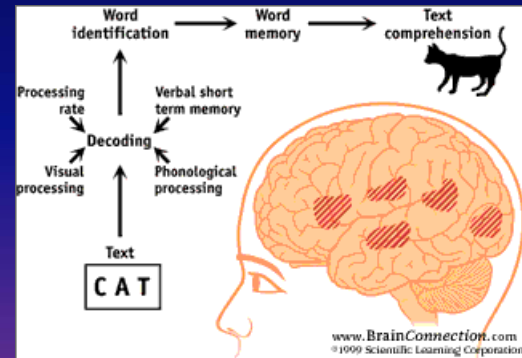


P-space definition

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.

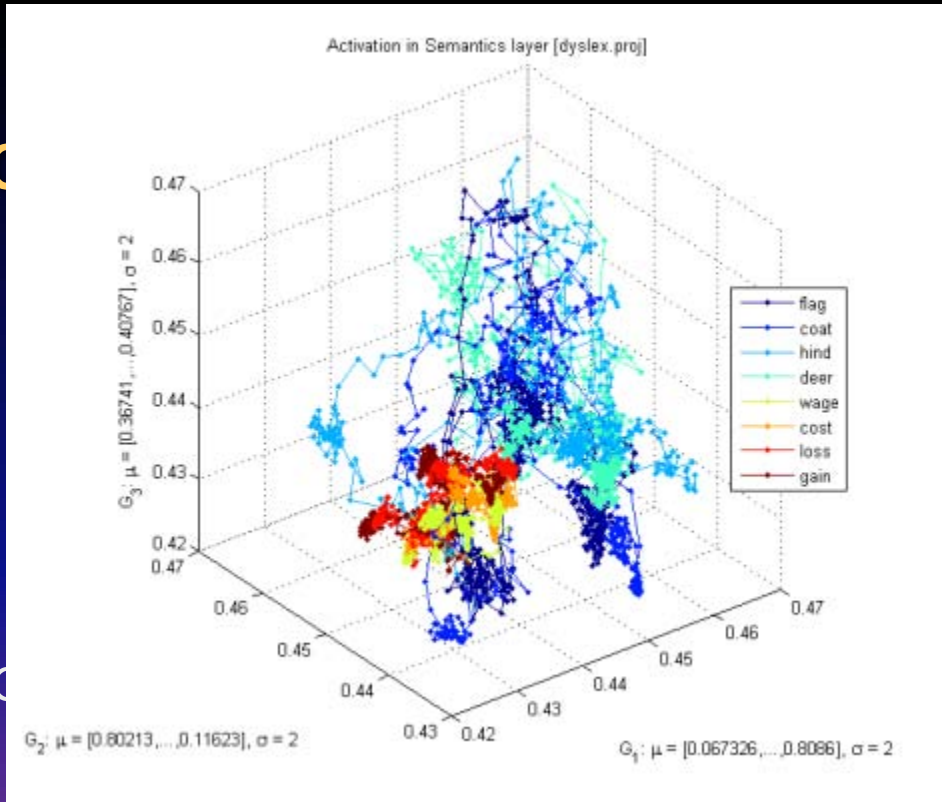
Can one describe the state of mind in similar way?

Attractor

Attention results from:

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

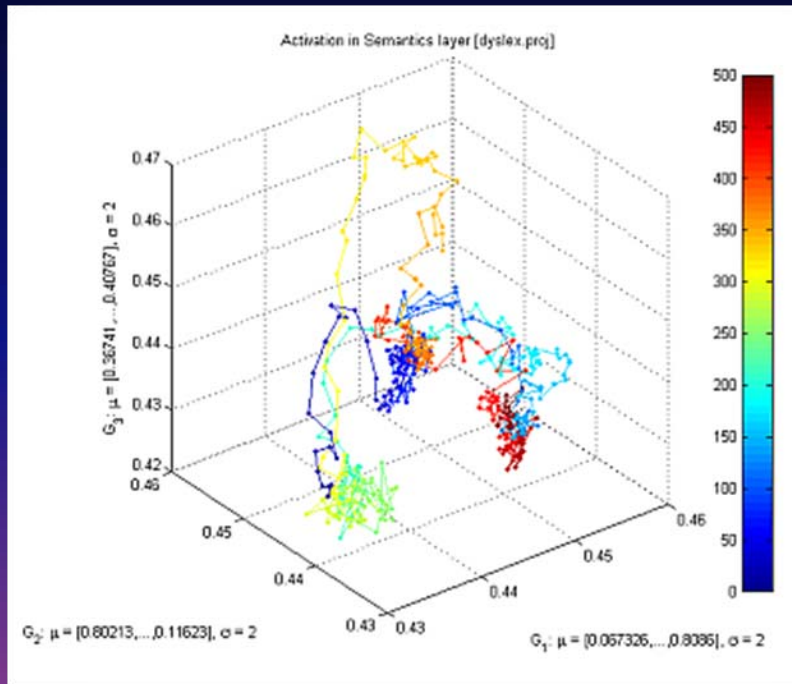
Basins of attractors: input activations {LO



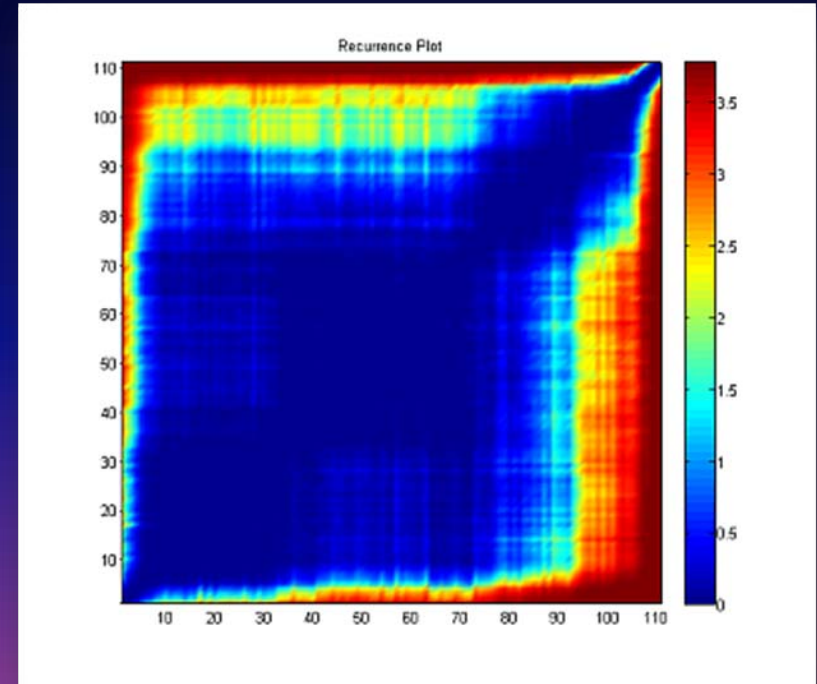
- Normal case: relatively large, easy associations, moving from one basin of attraction to another, exploring the activation space.
- Without accommodation (voltage-dependent K^+ channels): deep, narrow basins, hard to move out of the basin, associations are weak.

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).

Recurrence plots

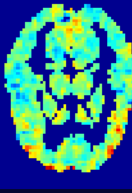


Starting from the word “flag”, with small synaptic noise (var=0.02), the network starts from reaching an attractor and moves to another one (frequently quite distant), creating a “chain of thoughts”.



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

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).

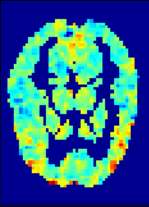
$\Psi(w, Cont, t)$ evolving prob. distribution (pdf) of brain activations.

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? Spreading activation in neural spaces, 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 .

QM-like formalism is useful for any probability waves.

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.

How to become an expert?

Textbook knowledge in medicine: detailed description of all possibilities.

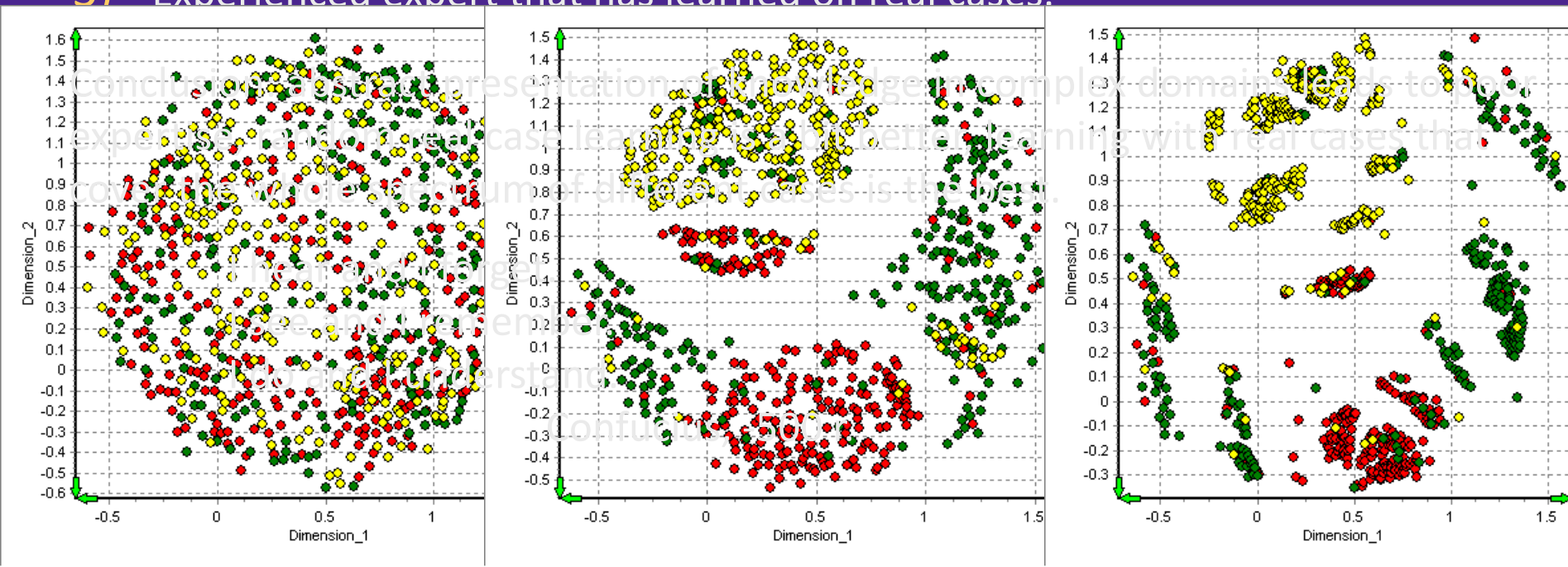
Effect: neural activation flows everywhere and correct diagnosis is impossible.

Correlations between observations forming prototypes are not firmly established.

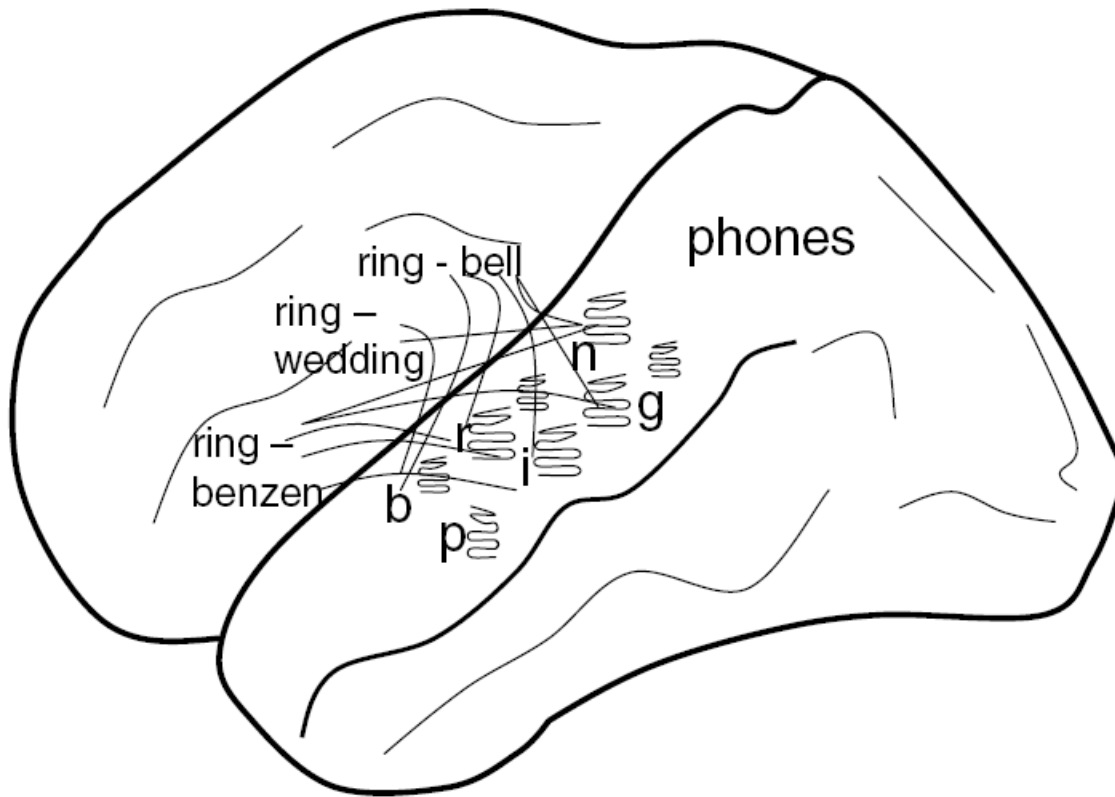
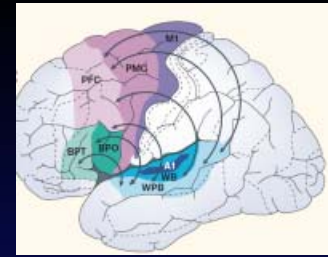
Expert has correct associations.

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.



ds



features of products;
dictionary.

new words are being
y cortex.

ation of phonemes will
ns; context + inhibition in
rds.

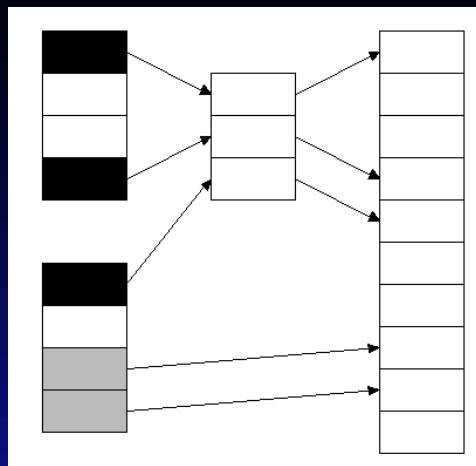
filtering (competition)

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.

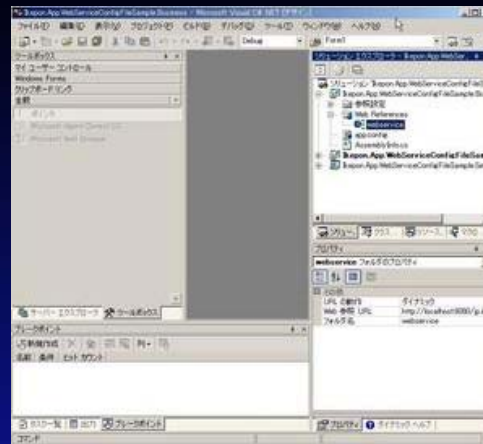
discovery = {disc, disco, discover, verity} (discovery, creativity, verity)

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

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



Semantic memory



Applications, search,
20 questions game.



Humanized interface

Query

Store

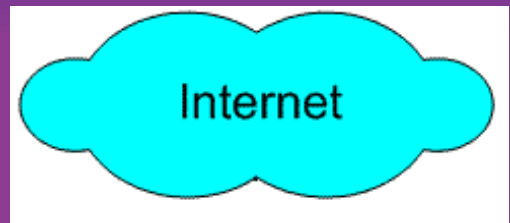


Part of speech tagger
& phrase extractor

verification

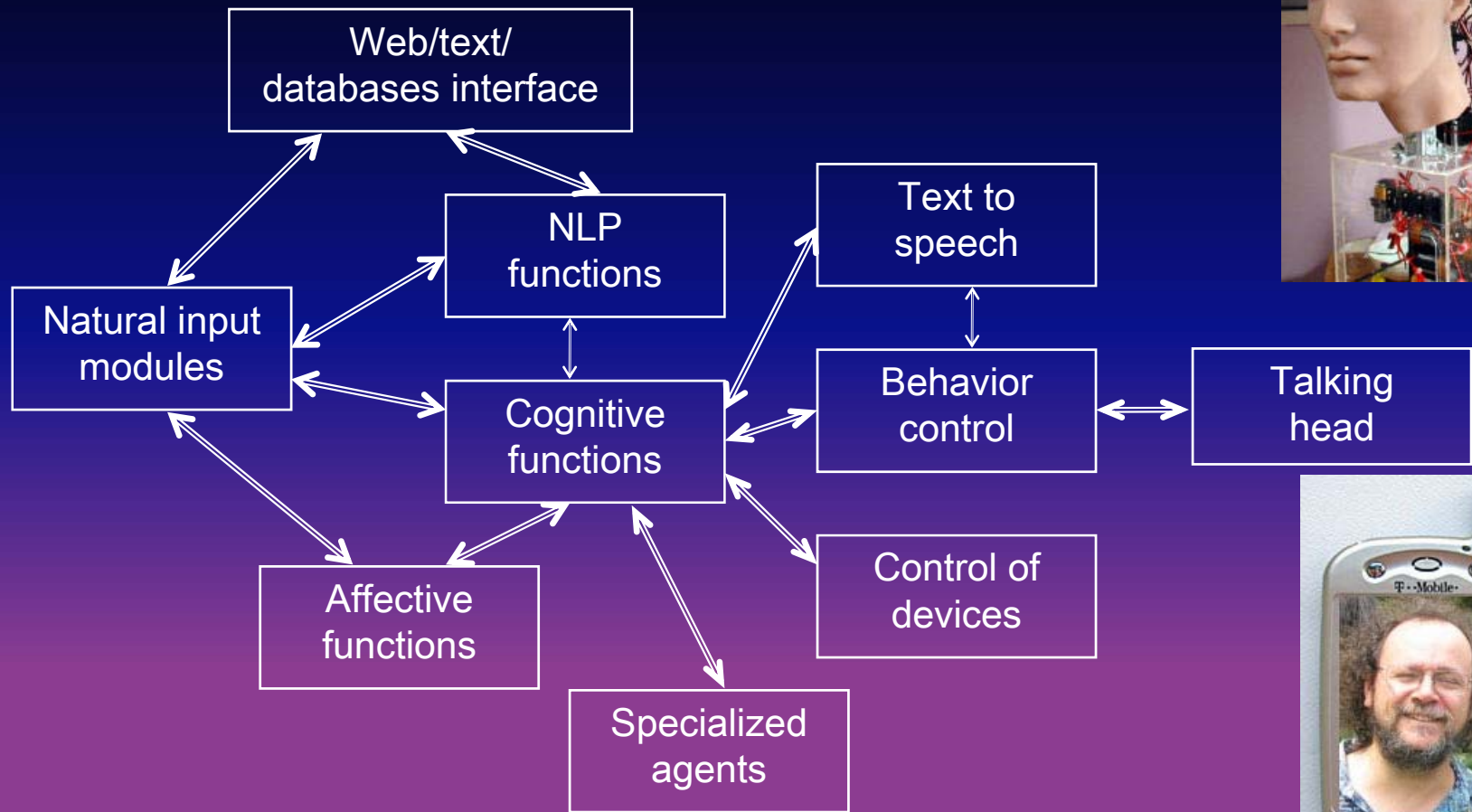
Manual

Parser



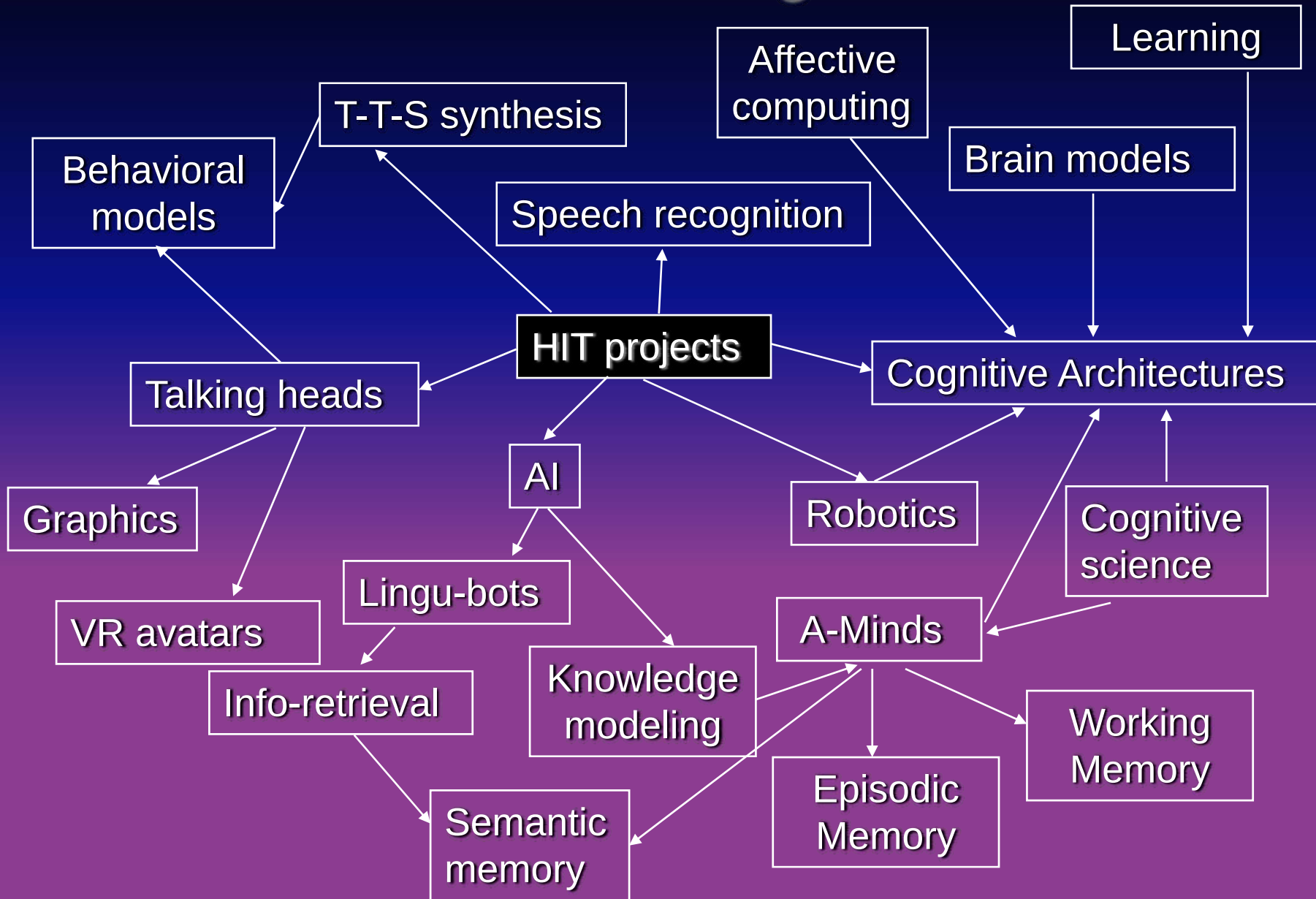
On line dictionaries
Active search and
dialogues with users

DREAM top-level architecture



DREAM project is focused on perception (visual, auditory, text inputs), cognitive functions (reasoning based on perceptions), natural language communication in well defined contexts, real time control of the simulated/physical head.

DREAM/HIT – larger view ...



Connections

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

Neuroscience:

D. Marr (1970) “probabilistic landscape”.

C.H. Anderson, D.C. van Essen (1994): Superior Colliculus PDF maps

S. Edelman: “neural spaces”, object recognition, global representation space approximates the Cartesian product of spaces that code object fragments, representation of similarities is sufficient.

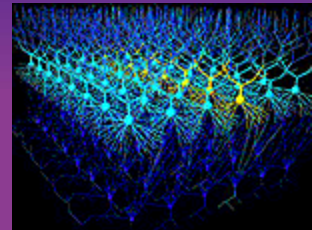
Psychology:

K. Levin, psychological forces.

G. Kelly, Personal Construct Psychology.

R. Shepard, universal invariant laws.

P. Johnson-Laird, mind models.



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

More connections



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

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

Philosophy:

P. Gärdenfors, Conceptual spaces

R.F. Port, T. van Gelder, ed. Mind as motion (MIT Press 1995)

Linguistics:

G. Fauconnier, Mental Spaces (Cambridge U.P. 1994).

Mental spaces and non-classical feature spaces.

J. Elman, Language as a dynamical system; J. Feldman neural basis;

Stream of thoughts, sentence as a trajectory in P-space.

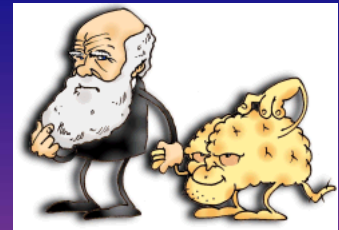
Psycholinguistics: T. Landauer, S. Dumais, Latent Semantic Analysis, Psych. Rev.
(1997) Semantic for 60 k words corpus requires about 300 dim.

Conclusions

Robots and avatars will make a steady progress towards realistic human-like behavior – think about progress in computer graphics.



- Artificial minds of brain-like systems will claim qualia; they will be as real in artificial systems as they are in our brains.
- There are no good arguments against convergence of the neural modeling process to conscious artifacts.
- Achieving human-level competence in perception, language and problem-solving may take longer than creation of basic consciousness.



Creation of conscious artifacts will open Pandora's box

What should be their status?

Will it degrade our own dignity?

Is switching off a conscious robot a form of killing?

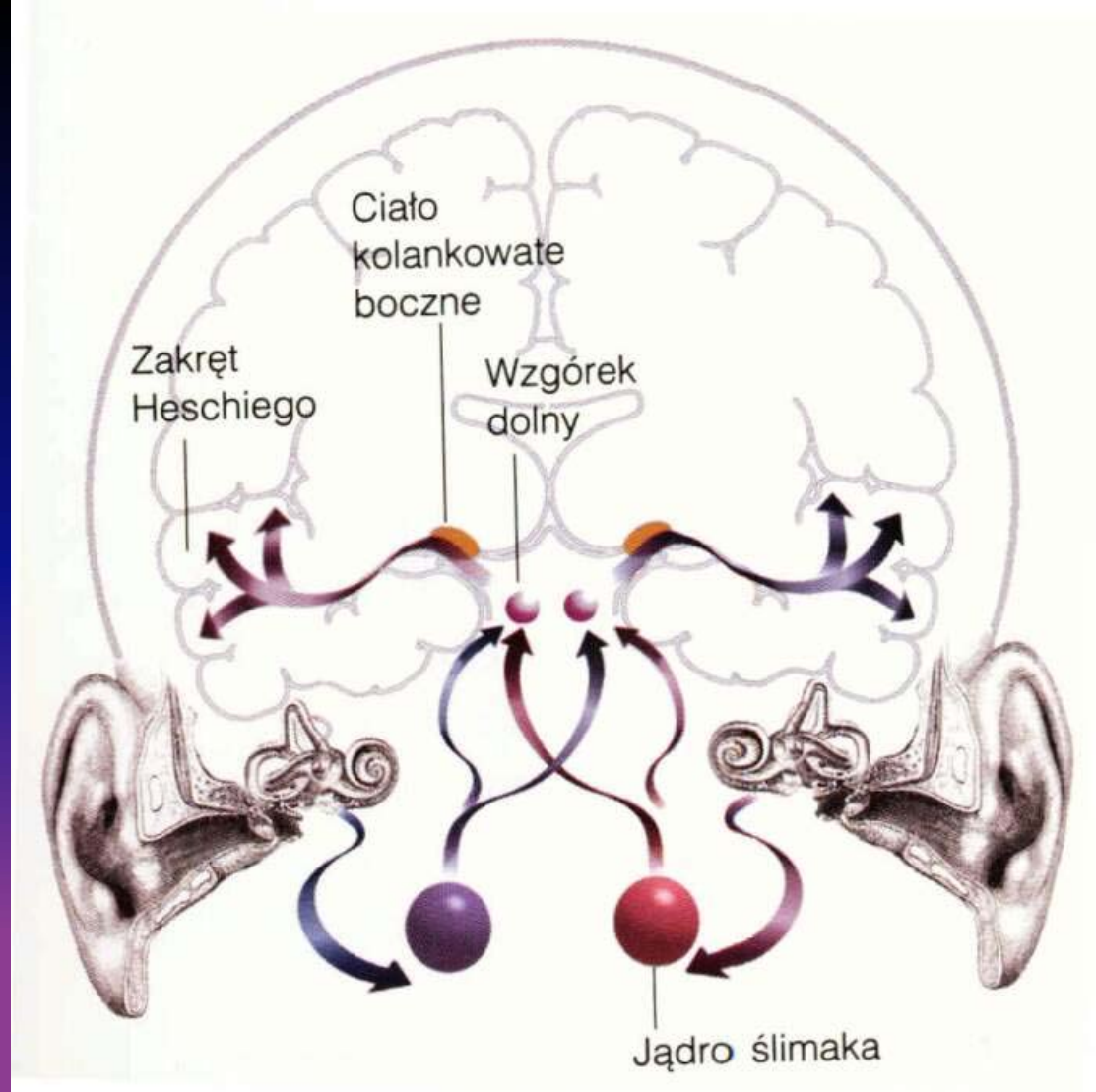
...

Will they ever turn against us ... or is the governor of California already one of them ?



**Exciting times
are coming!**

**Thank you
for lending
your ears**



Google: W Duch => Papers

Towards Full Cognitive Autonomy in Technical Systems

Ricardo Sanz

Autonomous Systems Laboratory
Universidad Politécnica de Madrid



1st CHIST-ERA Conference, Rome 27-28 May 2010



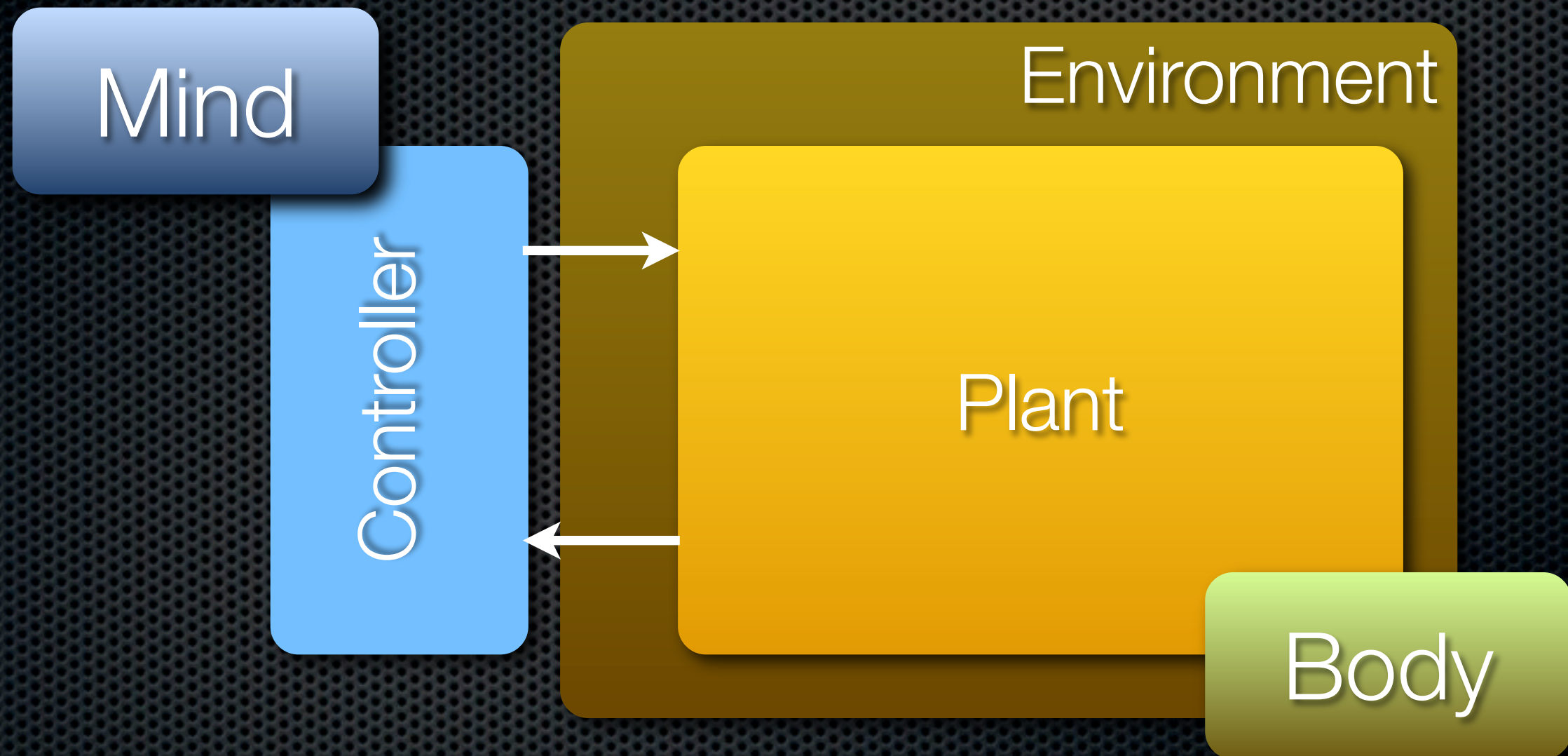
Two up-front issues

- ✦ Say thanks
- ✦ Apologise
 - ✦ Topic: **Self-awareness and self-consciousness**
 - ✦ *“A fuzzy topic which sprouts into cognitive psychology, neuroscience and philosophy.”*
 - ✦ I feel myself more **incompetent** than **expert**.

A bit of context

ICT in control systems

Control Engineering



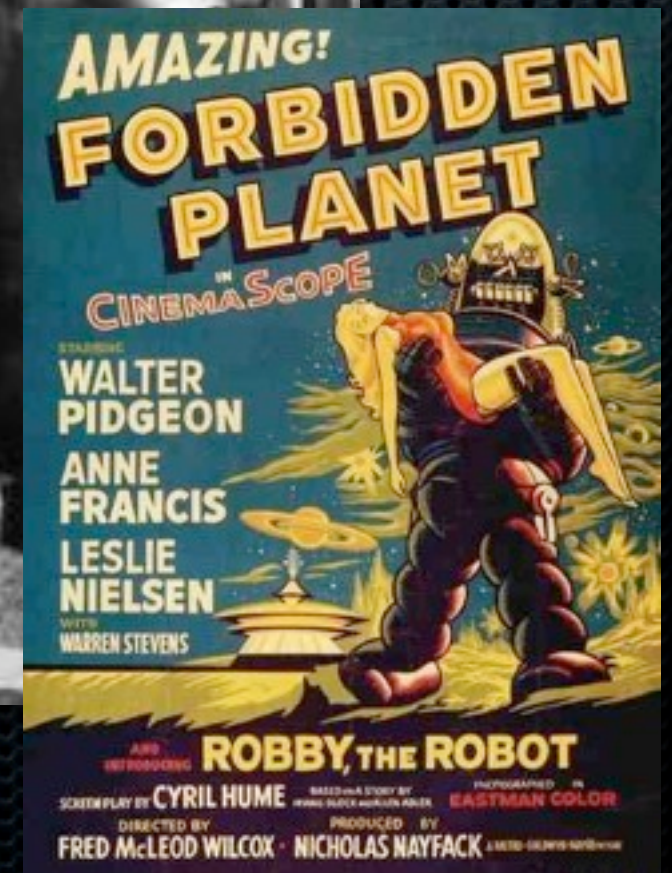
From dreams ...

Forbidden Planet

Robbie

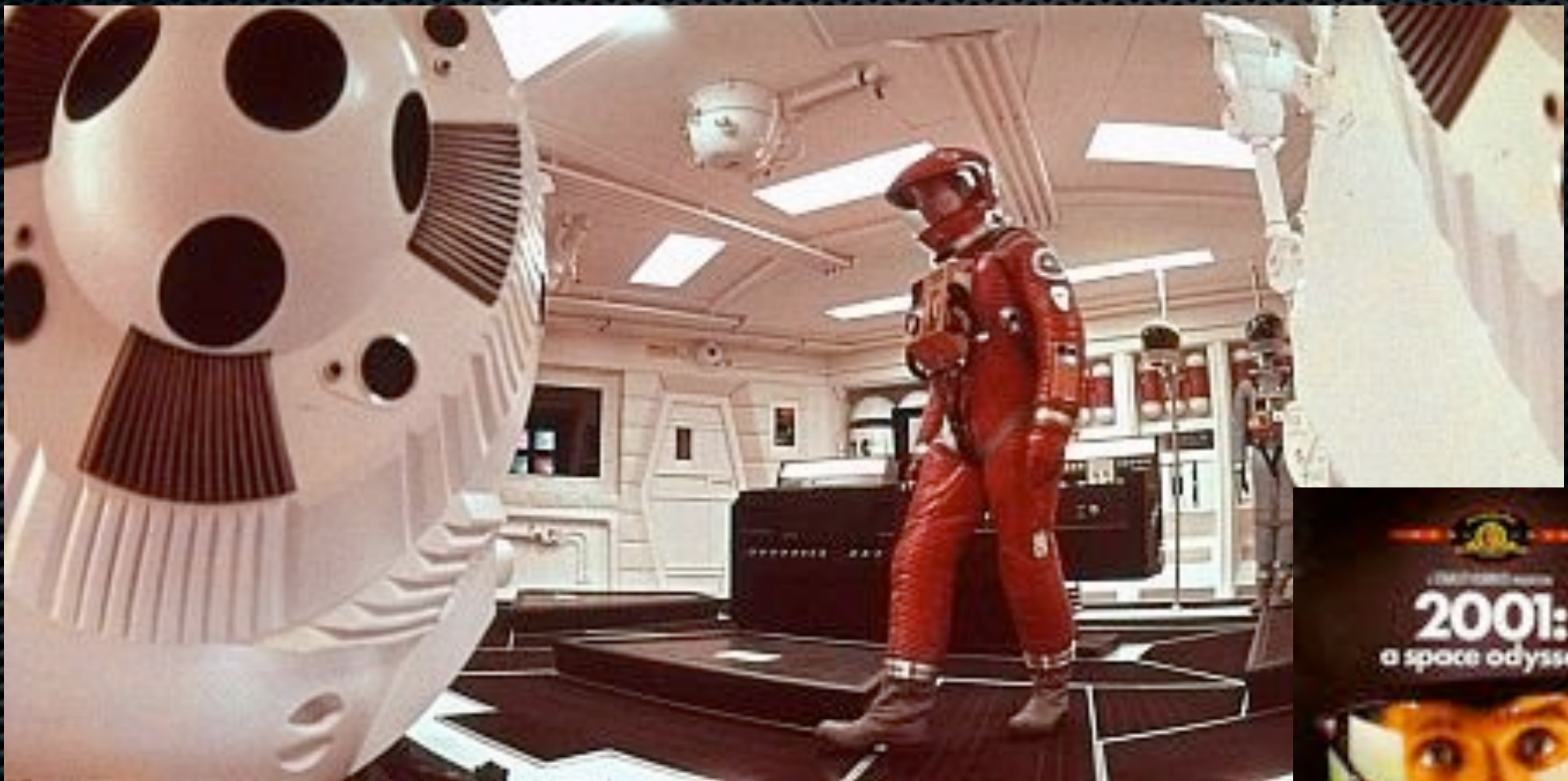


(1956)



2001: A Space Odyssey

HAL

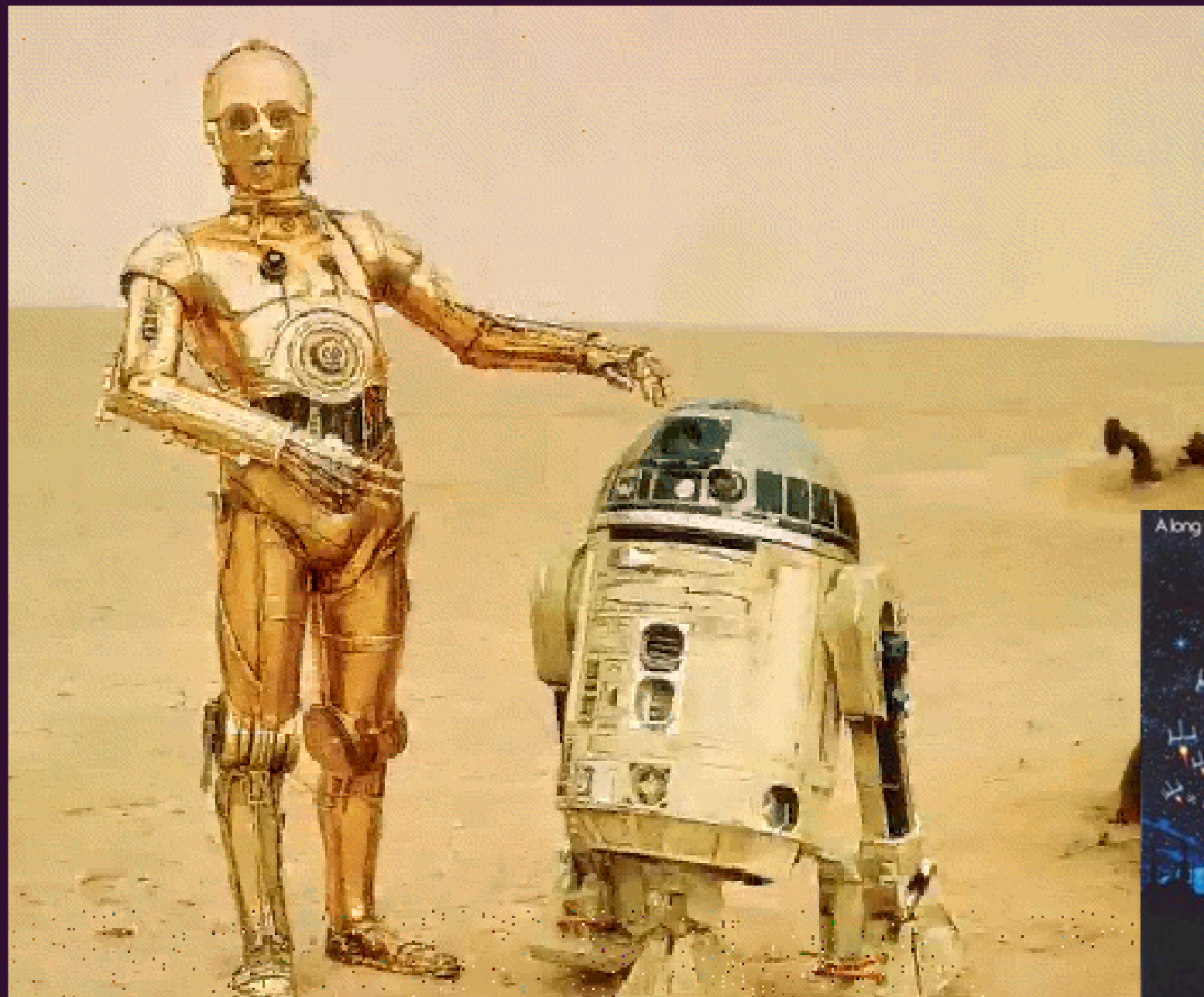


(1968)

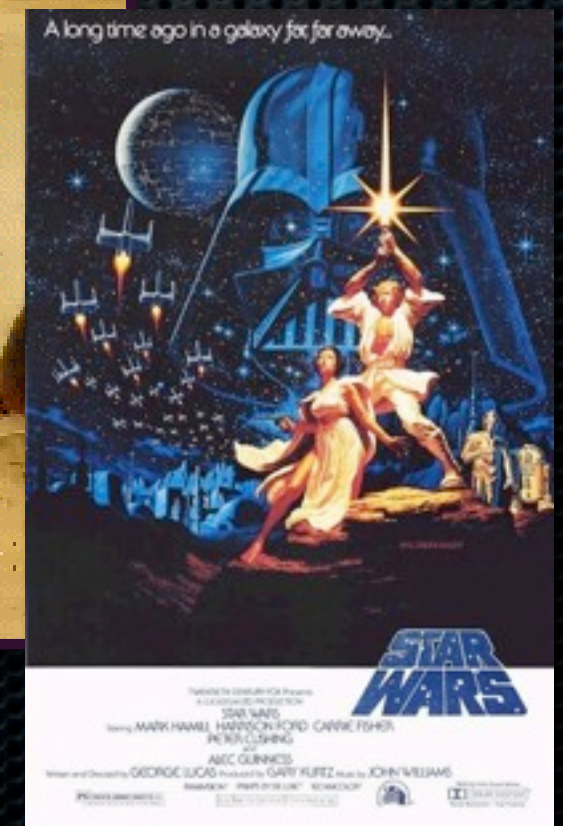


Star Wars

C3PO
R2D2

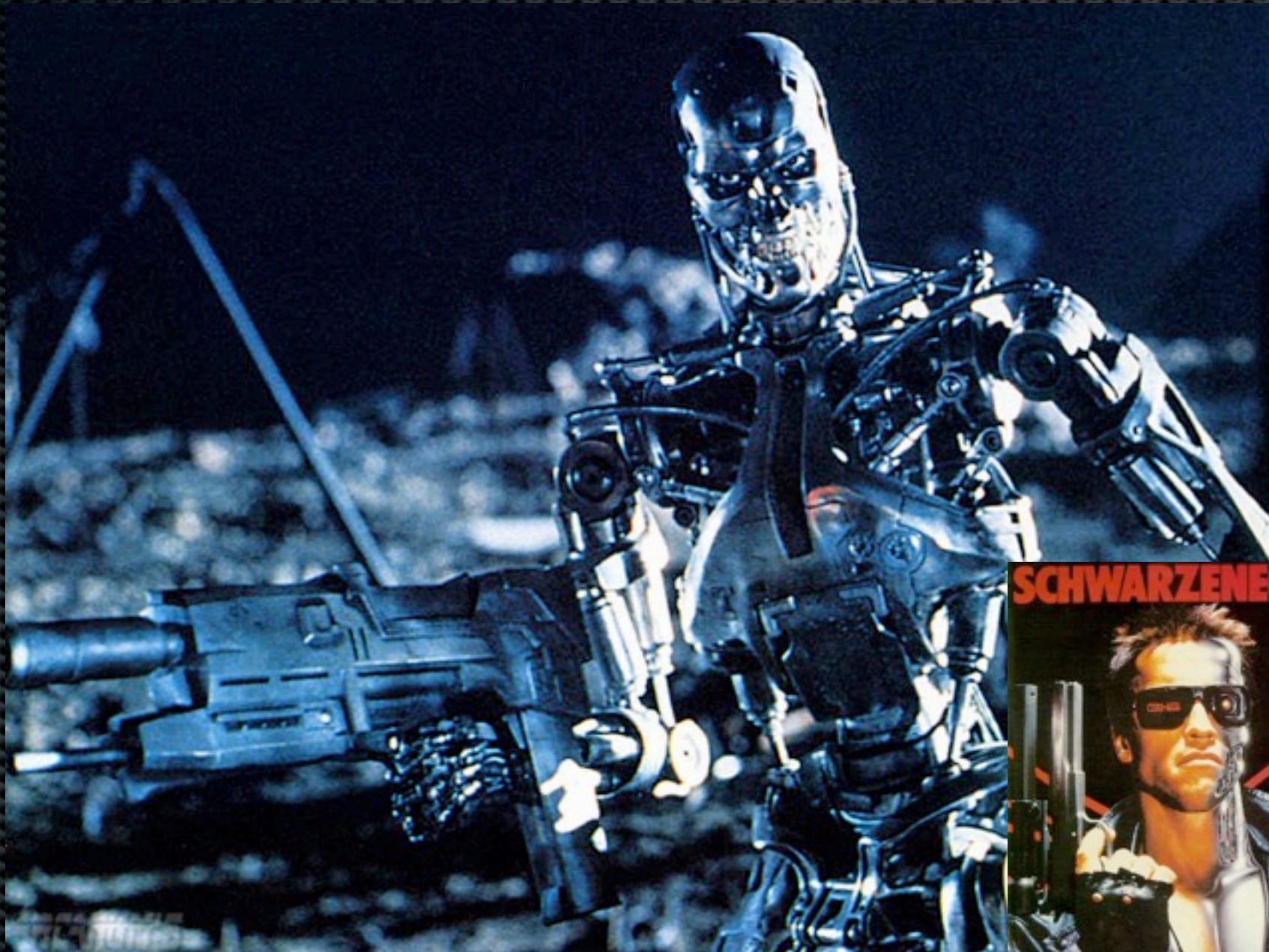


(1977)



Terminator

T800



(1984)



I Robot

Sonny



(2004)



... to realities

Industrial Robots



Automatic Planes



Chemical Plants



Distribution Grids



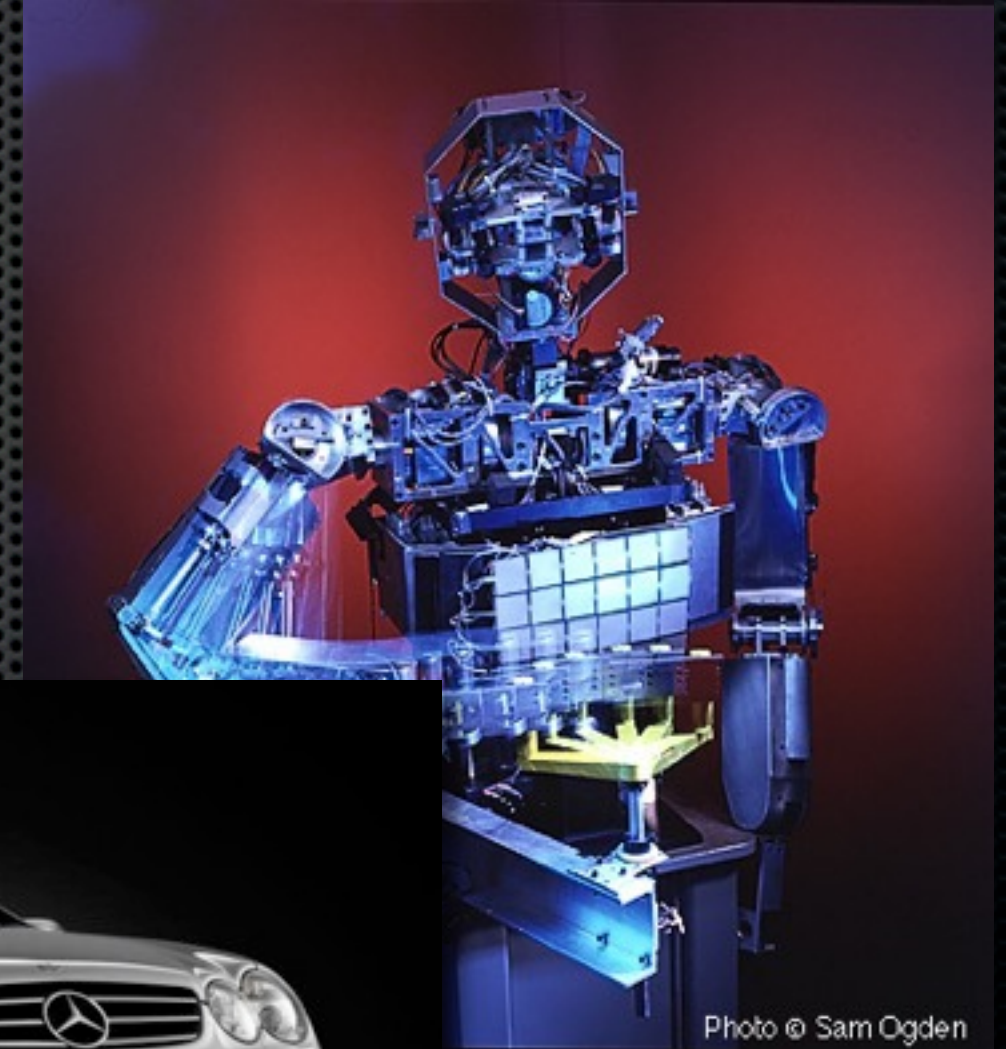
Computerised cars



Mars Rovers



Control ... for what ?



Pieces in a controlled system

- ✦ The **Plant**

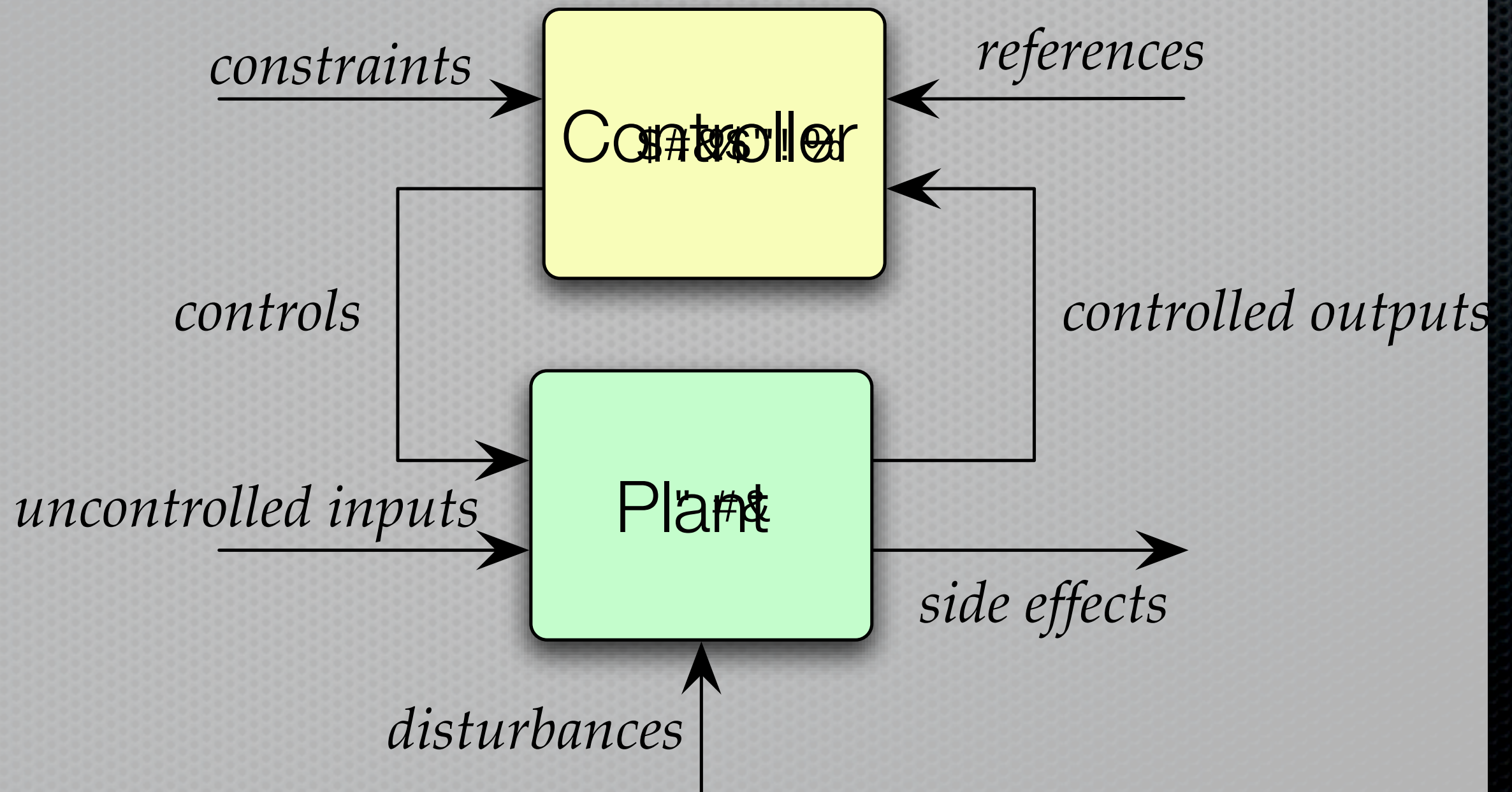
- ✦ The system that performs the task we're interested in
 - ✦ A plane, a chemical reactor, a router, etc.

- ✦ The **Environment**

- ✦ The **Controller**

- ✦ The system that forces the plant to follow desirable trajectories
 - ✦ The FCS, the PID, etc.

Control ... for what ?



Control engineering process

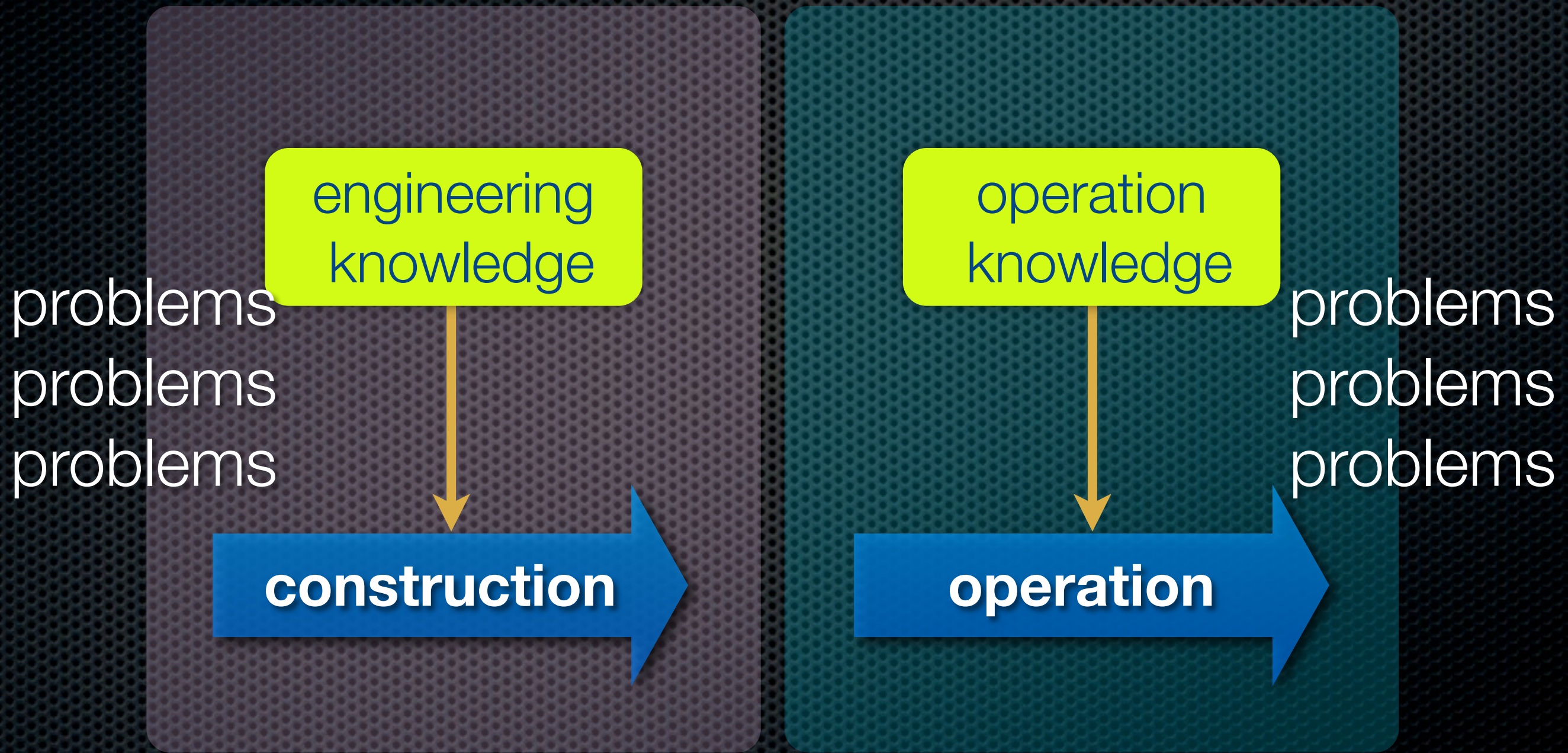
1. **Specify** desired plant behaviour
2. **Model** body and environment
3. **Design** controller
4. **Build** (normally using ICT)
5. **Deploy**
6. **Tune** (continuously !)
7. **Exploit**
8. **Decommission** when obsolete

Two classes of problems

- ✦ **Run-time** problems
 - ✦ What shall the controller do to handle disturbances?
- ✦ **Construction-time** problems
 - ✦ How to design and build such controller?



Two knowledge batteries



The basic desideratum is achieving

Robust autonomy

Make a system able to **solve its own problems**

Self-X

Self-Control

Controllers that Control Themselves

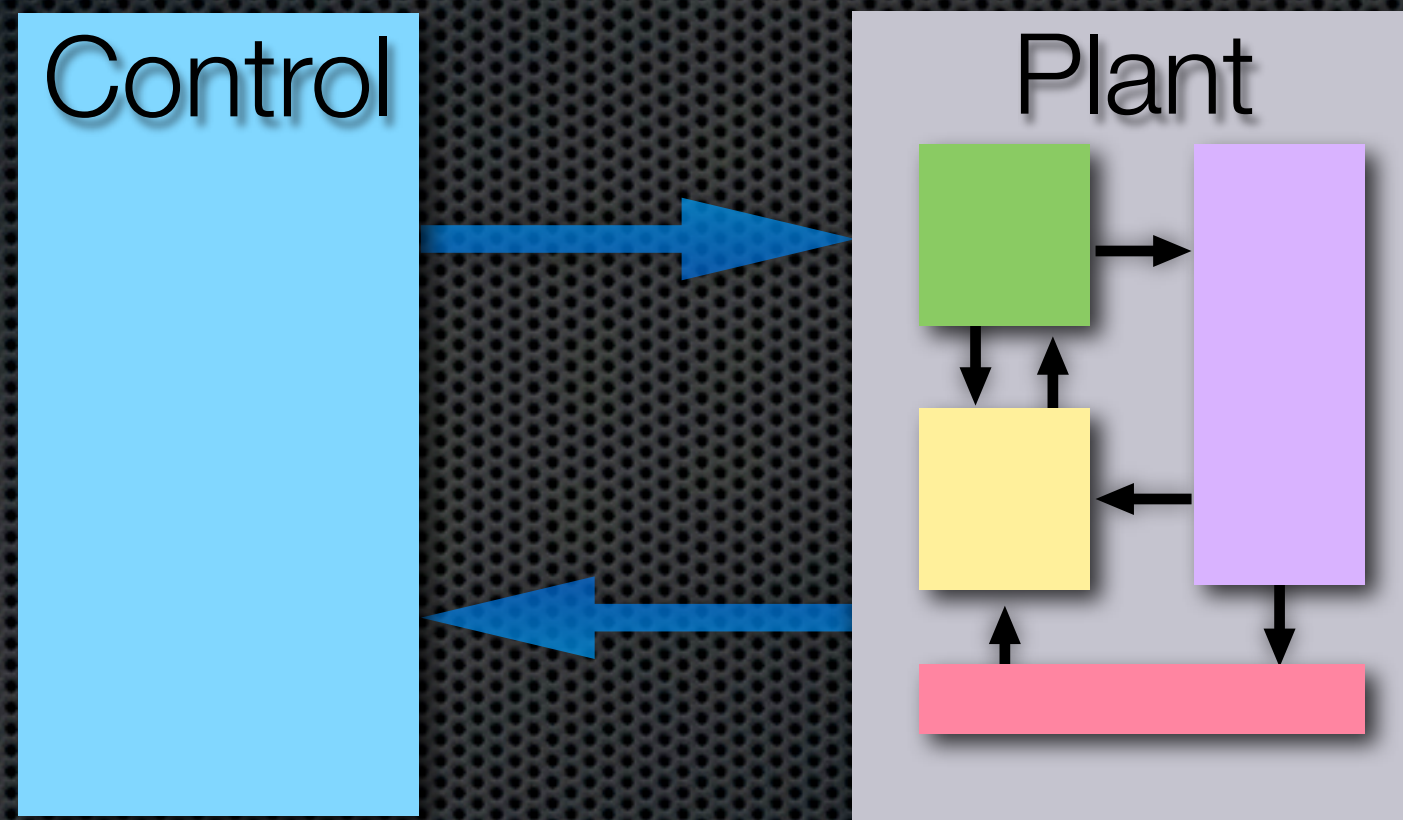
Increasing resilience

- ✦ Make **better** things (problem avoidance)
- ✦ Make **robust** things (problem tolerance)
 - ✦ Passive (Masking)
 - ✦ Transparent, fast, cost, deteriorating
 - ✦ Active (Compensating)
 - ✦ Cheaper, flexible, overhead, adaptable
- ✦ Controllers do **also suffer faults**, errors and failures

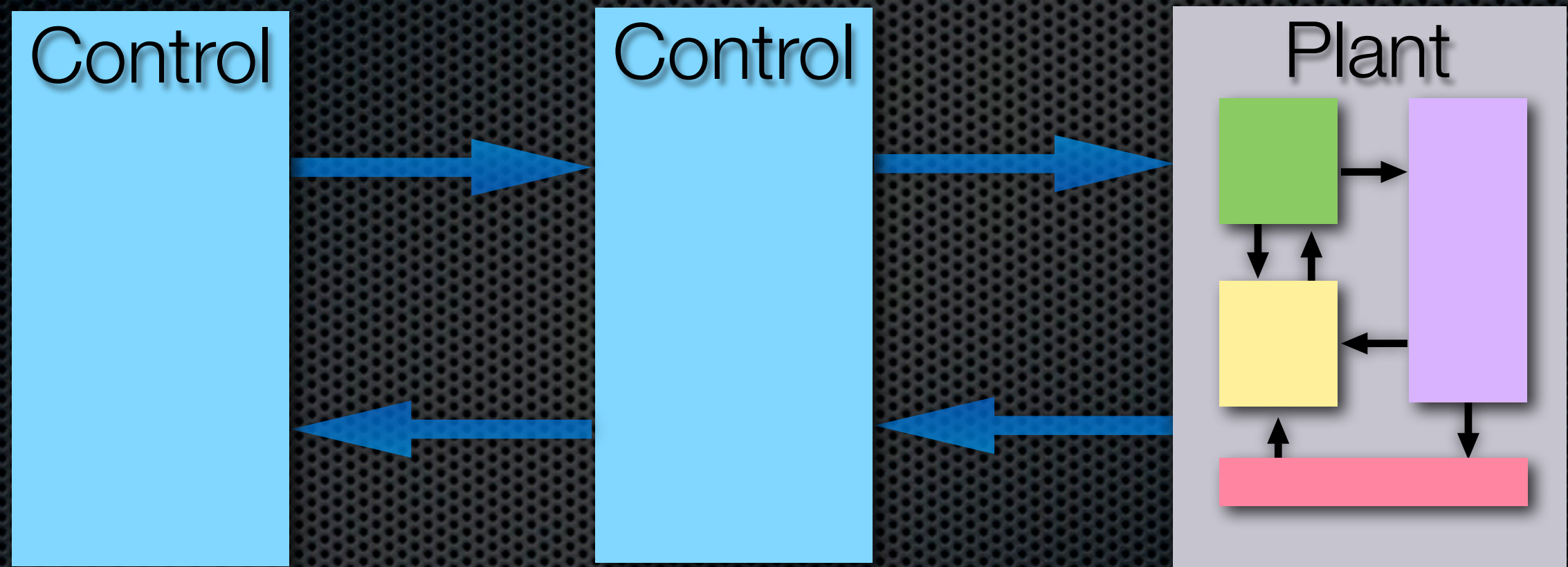
Controller resilience

- ✦ One way: Make passive-robust controllers
- ✦ Other way: Control the controller
 - ✦ Monitor, diagnose, modify the controller
- ✦ Second level controllers
 - ✦ Nested controllers
 - ✦ Adaptive controllers
 - ✦ Redundant fault-tolerant controllers
 - ✦ ... **conscious** controllers?

Control a plant

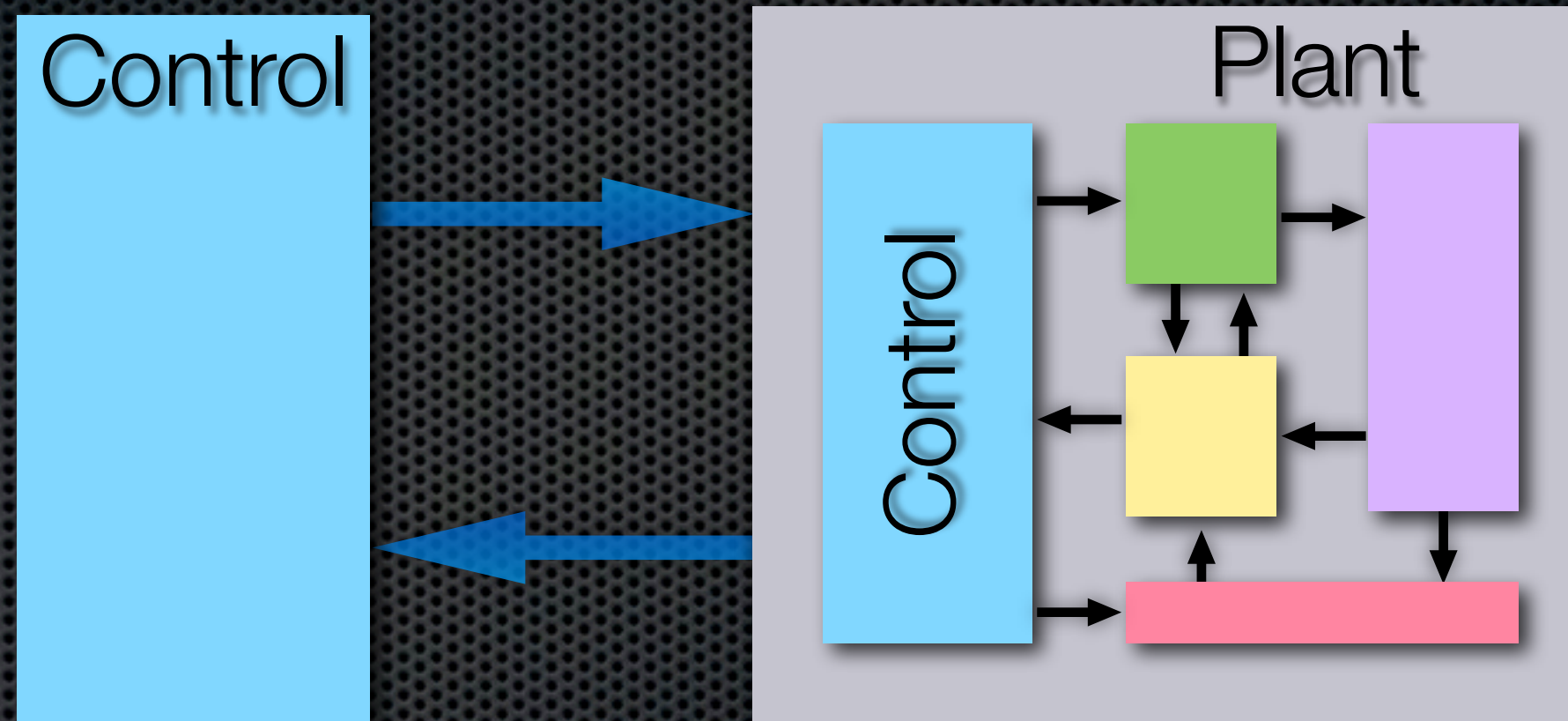


Control a controller ?



Metacognition?
Metacontrol?

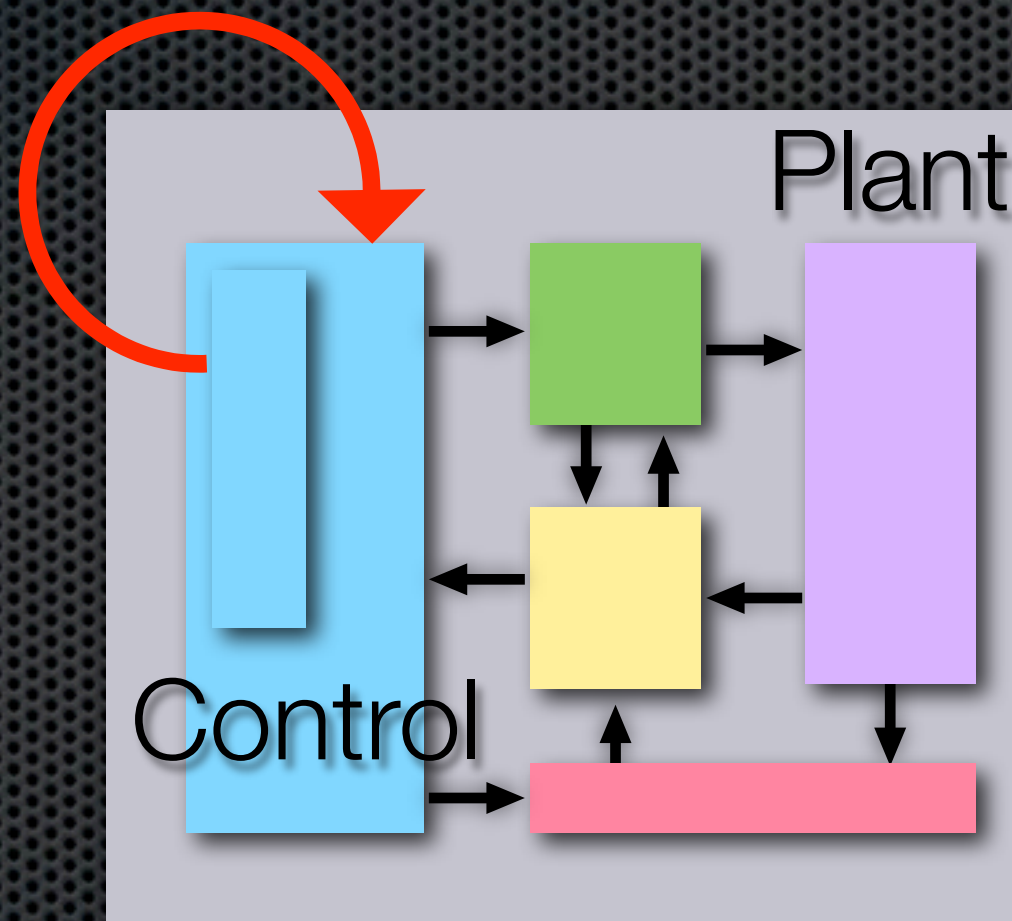
Control a controlled plant



A controlled plant is just **another kind** of plant

The key step

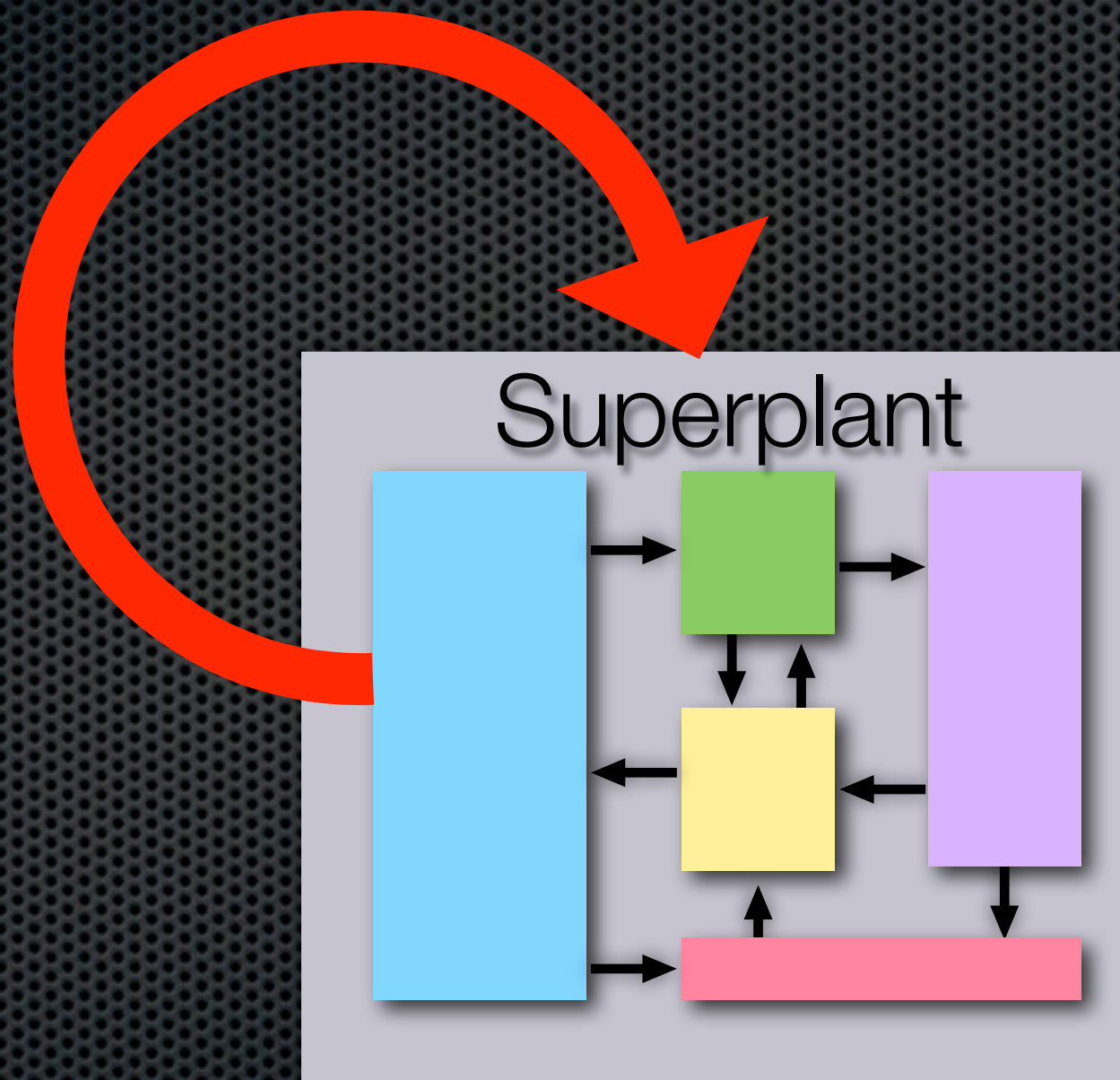
Self-loop



A way of bounding recursion in metacontrol

The key step

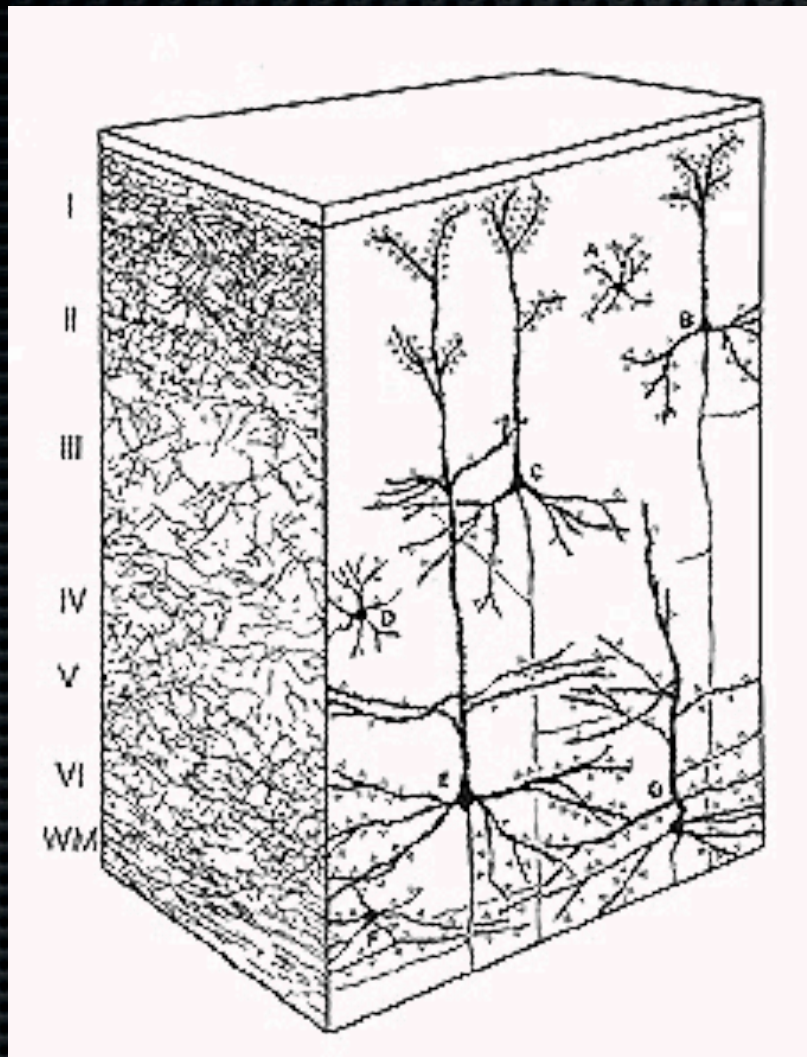
Self-loop



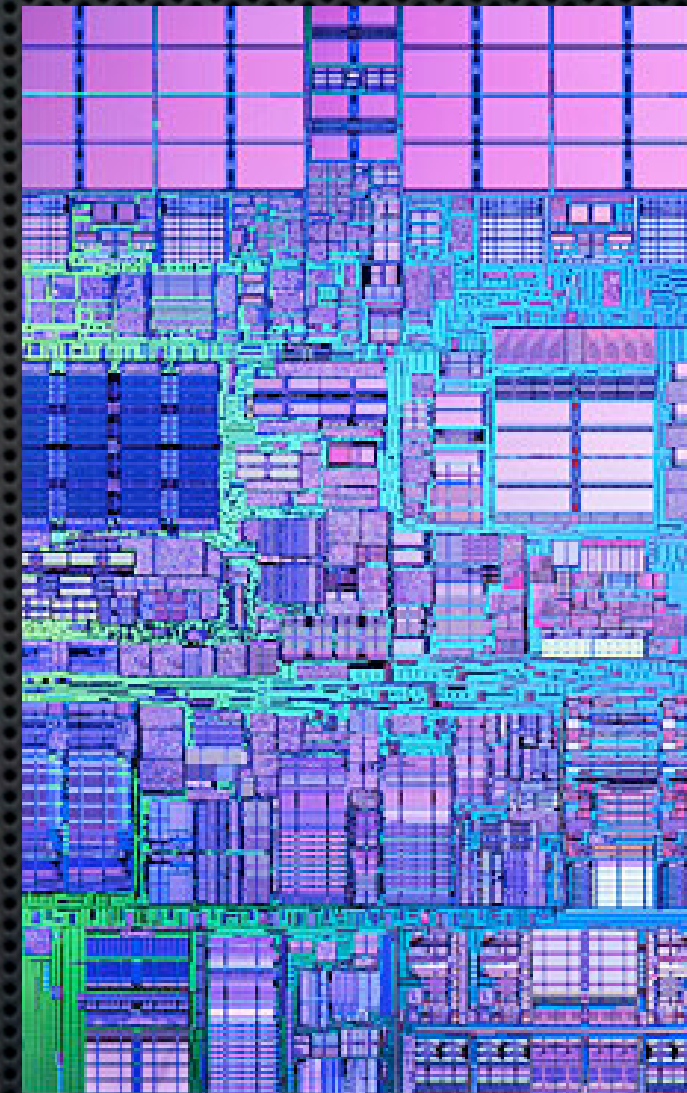
Where to look for solutions?

Cognitive science, neuroscience, etc.

Brain-inspired systems



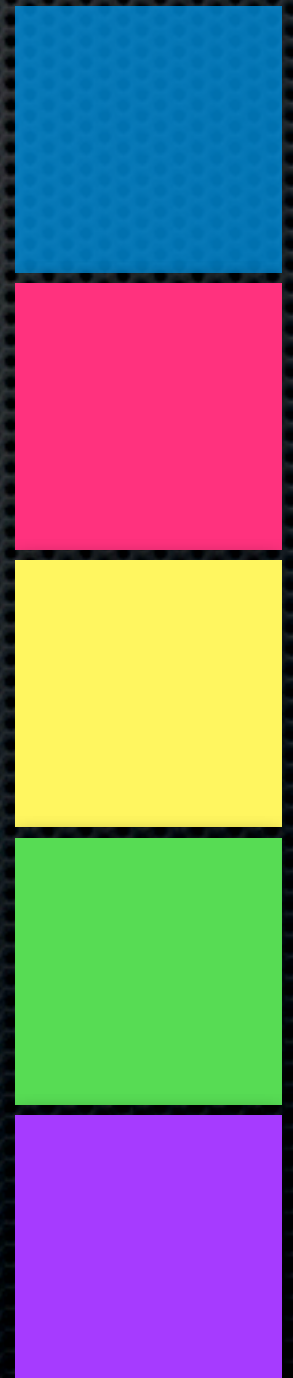
Accidental



Intentional

Consciousness ?

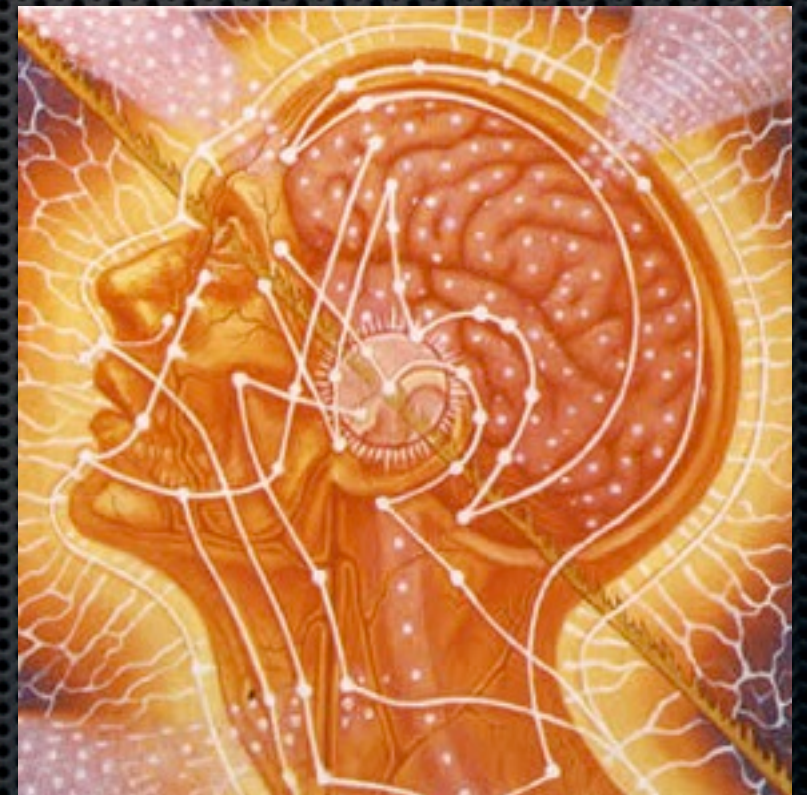
- ✧ Measuring the **world**
 - ✧ Access consciousness
- ✧ Experiencing the **world**
 - ✧ Phenomenology (feeling qualia)
- ✧ Experiencing the **self**
- ✧ Knowing **others'** minds



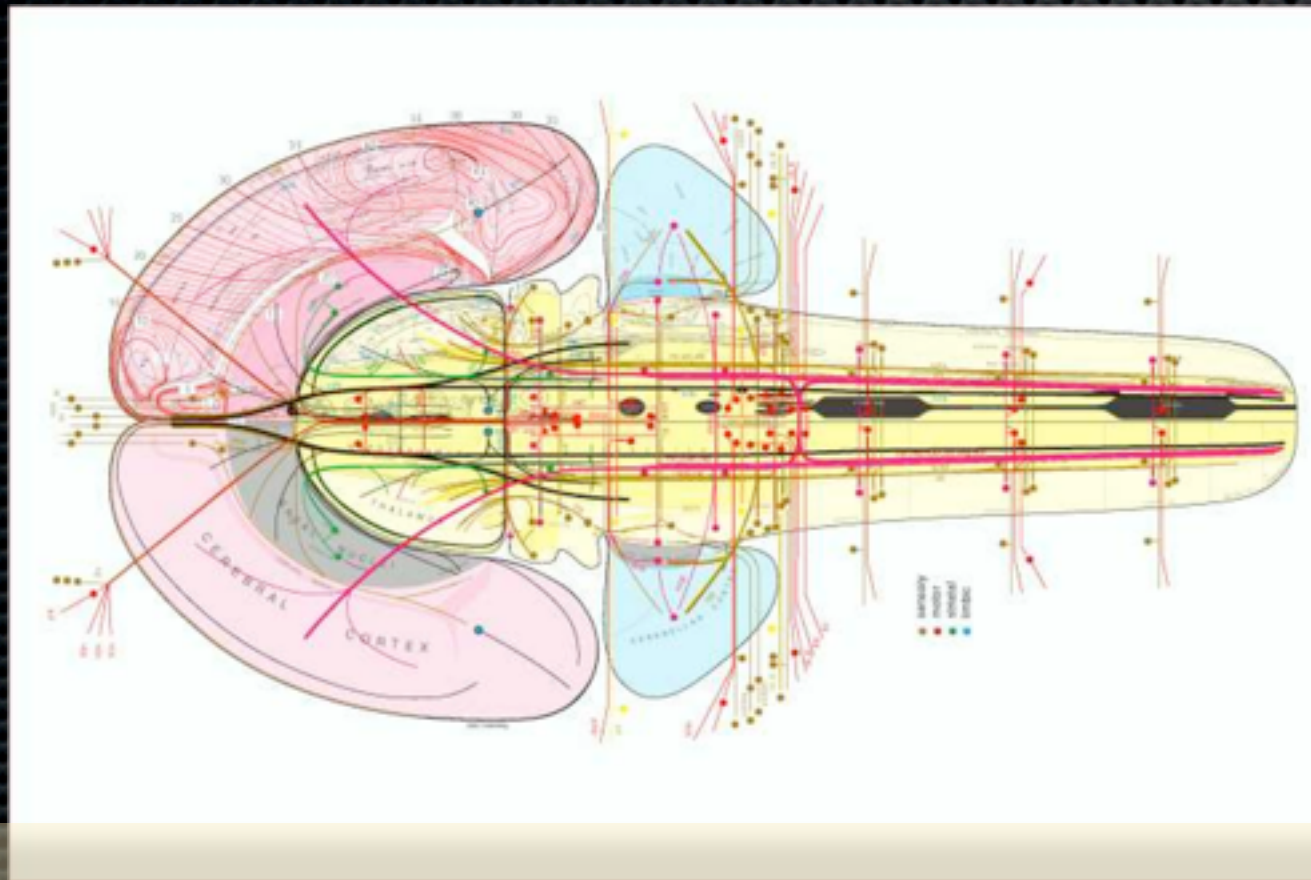
Some BIG problems

- ✦ Too many scattered foci on **biological self-awareness**
 - ✦ Psychology
 - ✦ Neuroscience
 - ✦ Philosophy
 - ✦ Anthropology
 - ✦ ...
- ✦ Cognitive science **does not have a clear grasp** of the issue of consciousness;
- ✦ Less of its brain architecture

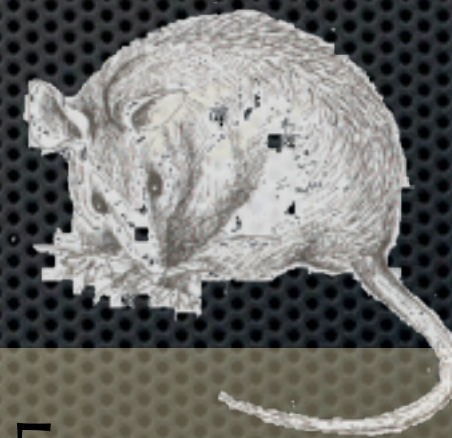
Cognitive
Science



Looking at brains



Rat brain



1

3×10^2

10^5

10^{10}

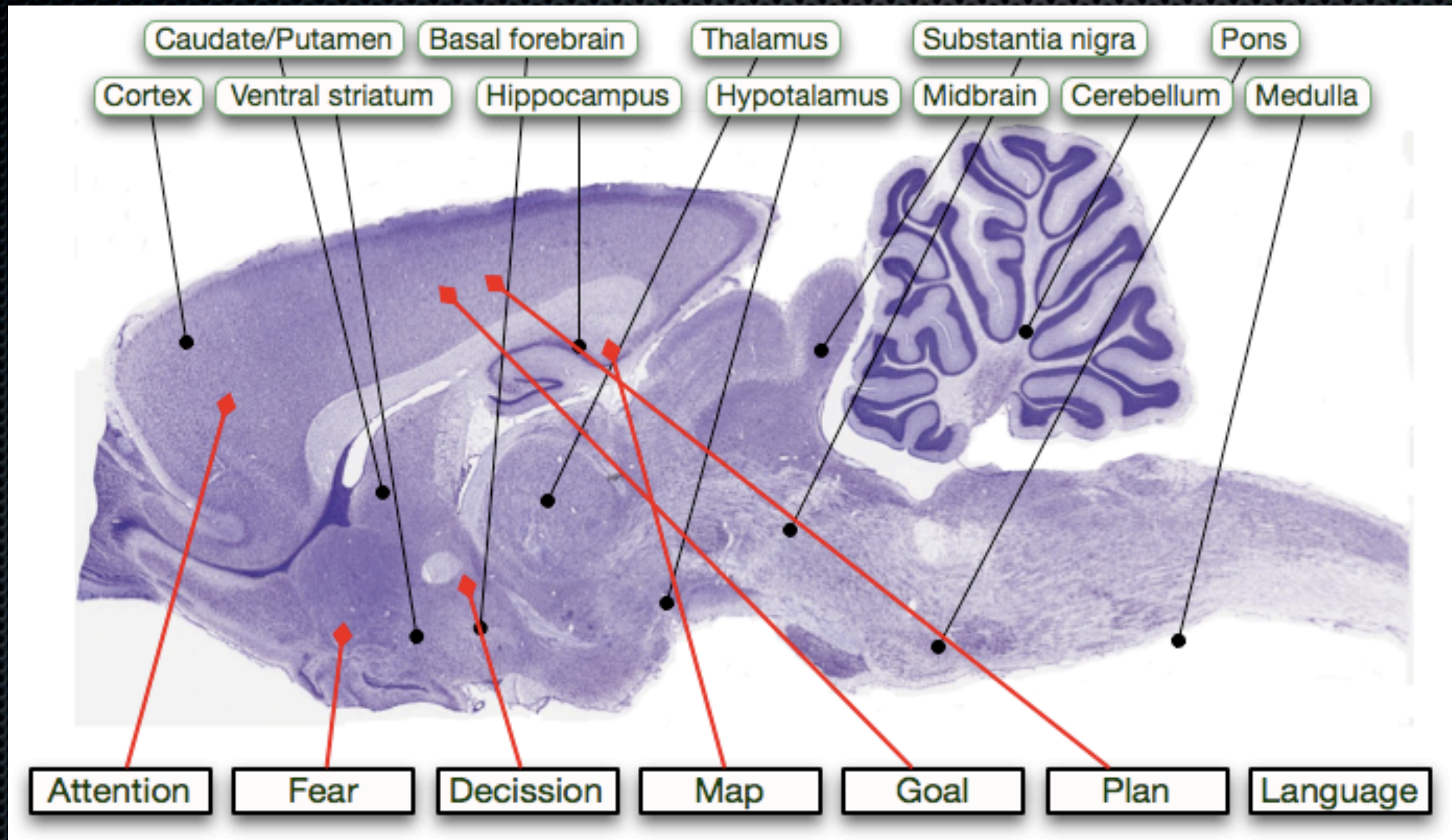
E. Coli

C. Elegans

D. Melanogaster

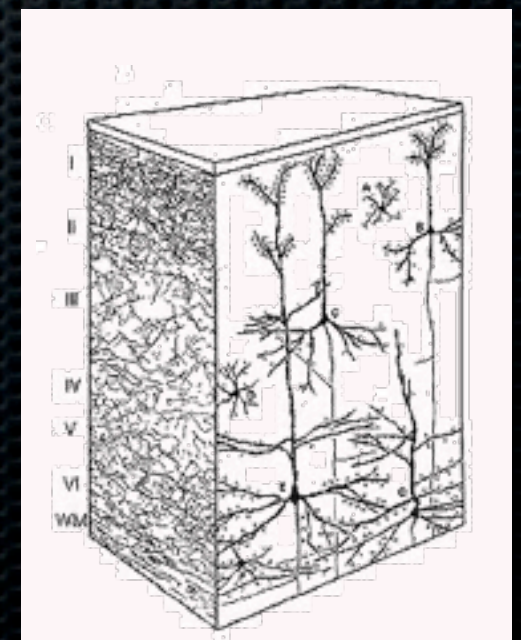
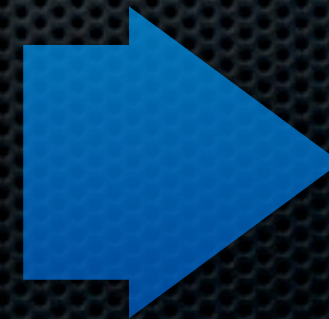
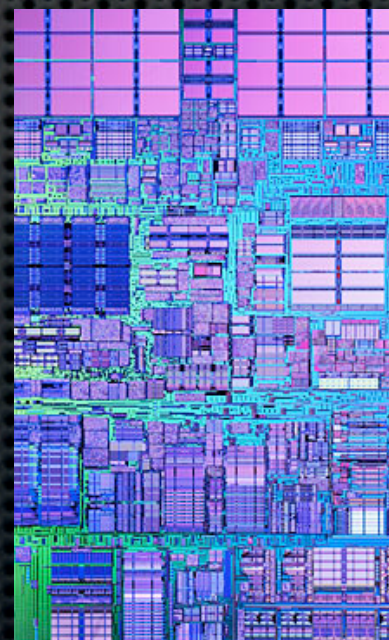
H. Sapiens

Function vs. structure



The other way round

- ✦ It seems that bio-inspiration is yet far from providing solid insights into the **architecture of self**
- ✦ At least to the level of resilience required by technical systems (esp. safe-critical ones)
- ✦ My impression is that what will happen is that self-aware **ICT will illuminate** cognitive science concerning these issues



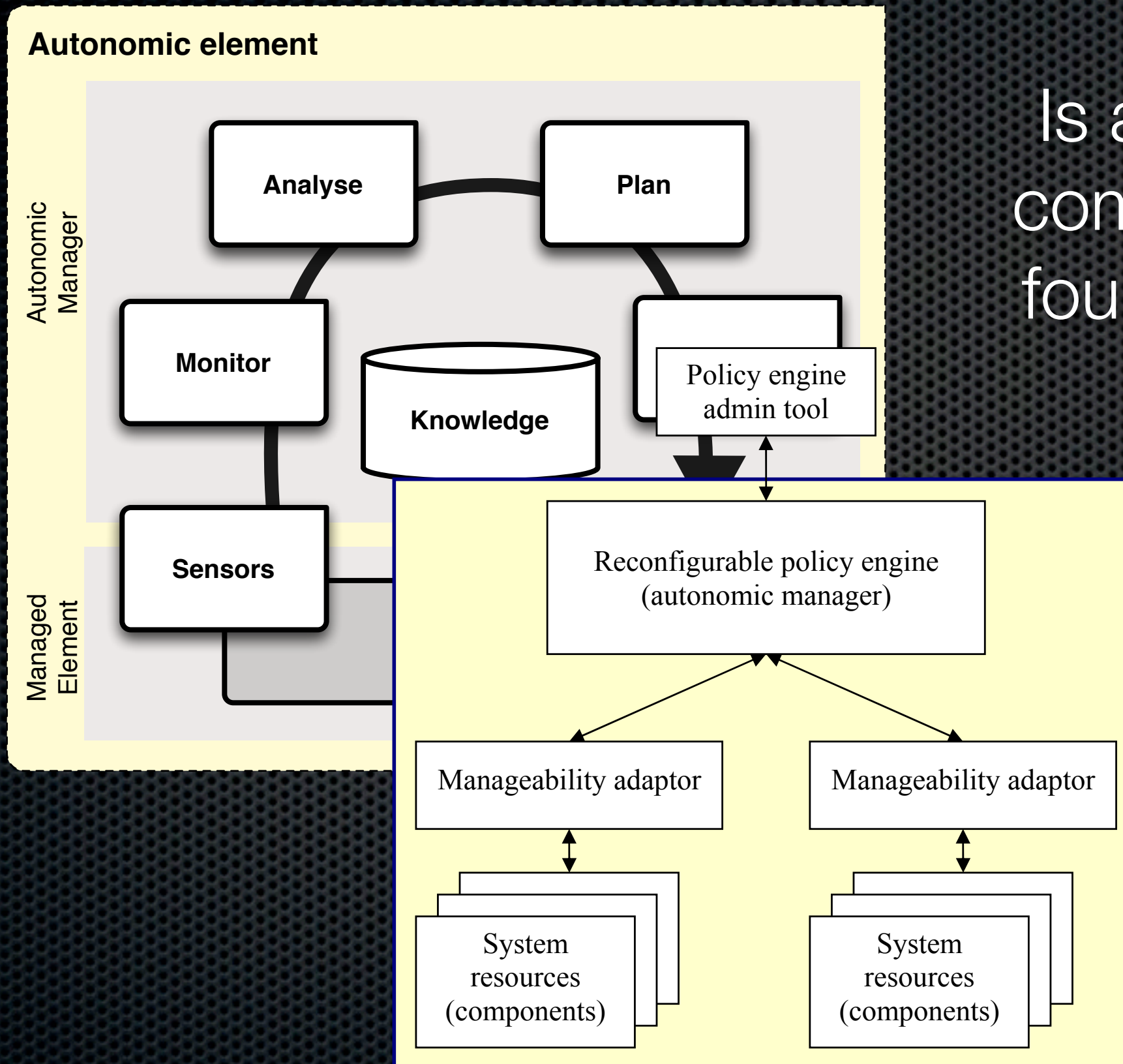
Fundamentals for Self

Basic cognitive patterns for self-awareness

An Approach to “Self”

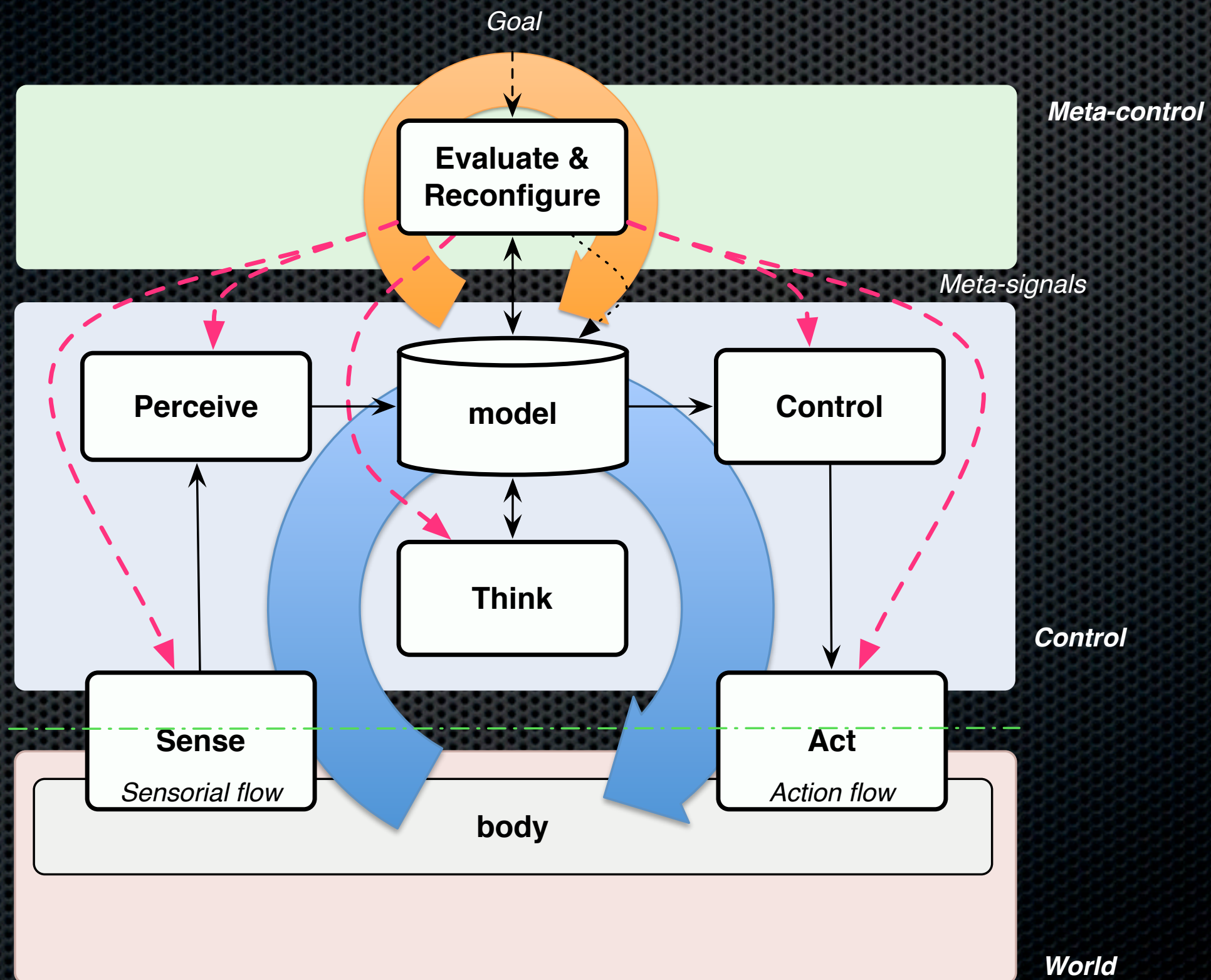
- ✦ Enhance the control capabilities -robustness- of autonomous agents by means of exploitation of self representations (both in perception and action)
- ✦ Sophisticated cognitive competences -self-consciousness- shall emerge from a generalisation and integration of metacontrol mechanisms

Autonomic computing



Is autonomic computing the foundation for self-X?

A general epistemic pattern ?



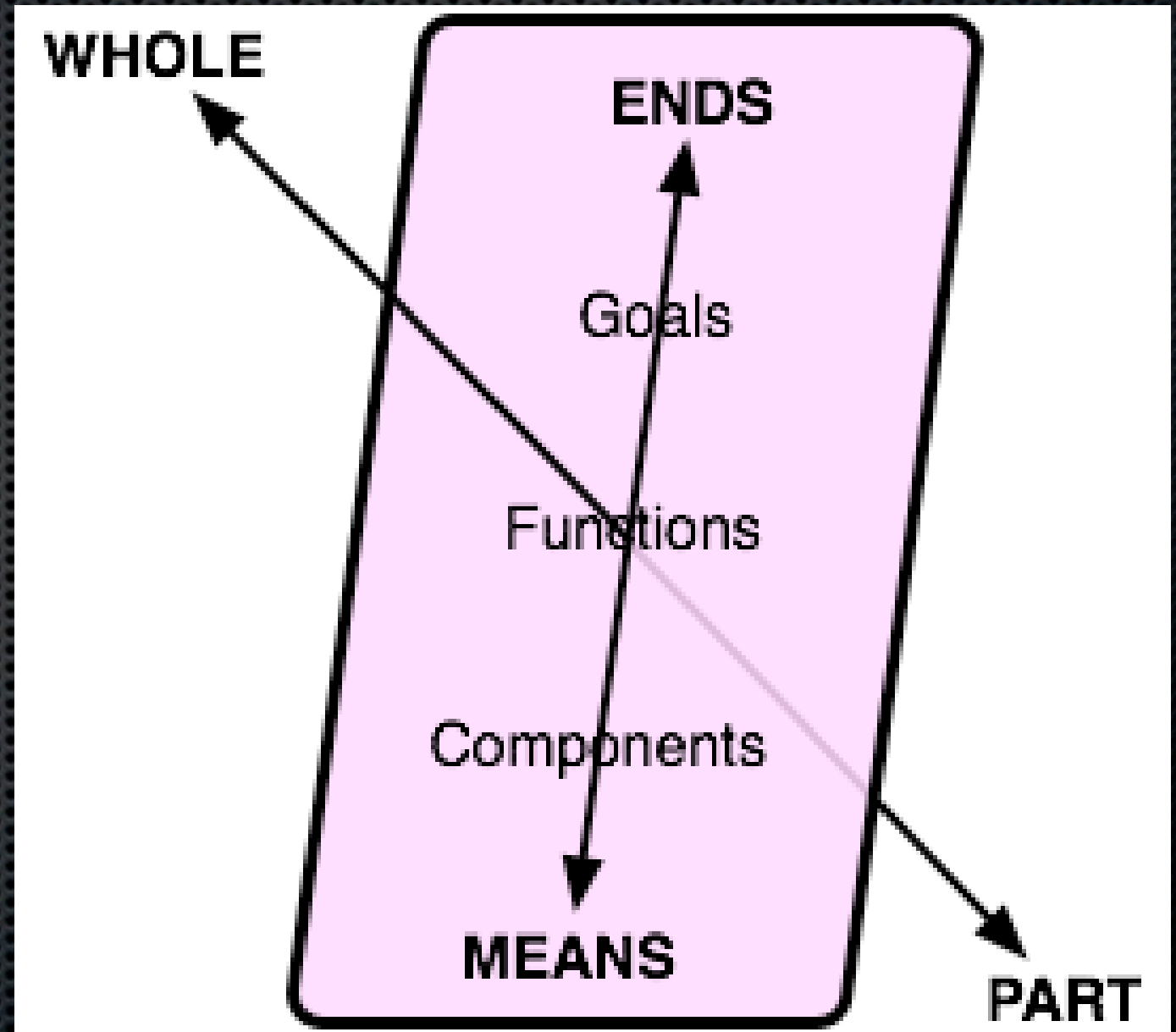
Structure and Function

- ✦ Focus on **function** instead of **structure**
- ✦ *“an object with a set of goals, in a well defined environment, exerts an activity (function) and at the same time experiences how its internal structure evolves through time keeping its identity”.*
[Le Moigne]

The Concept of “Function”

“Some define function to be equivalent to behavior whereas others ... define functions as purely intentional concepts”
[Lind 1994]

Dual **Causal** / **Intentional**
viewpoints on function



The core objective

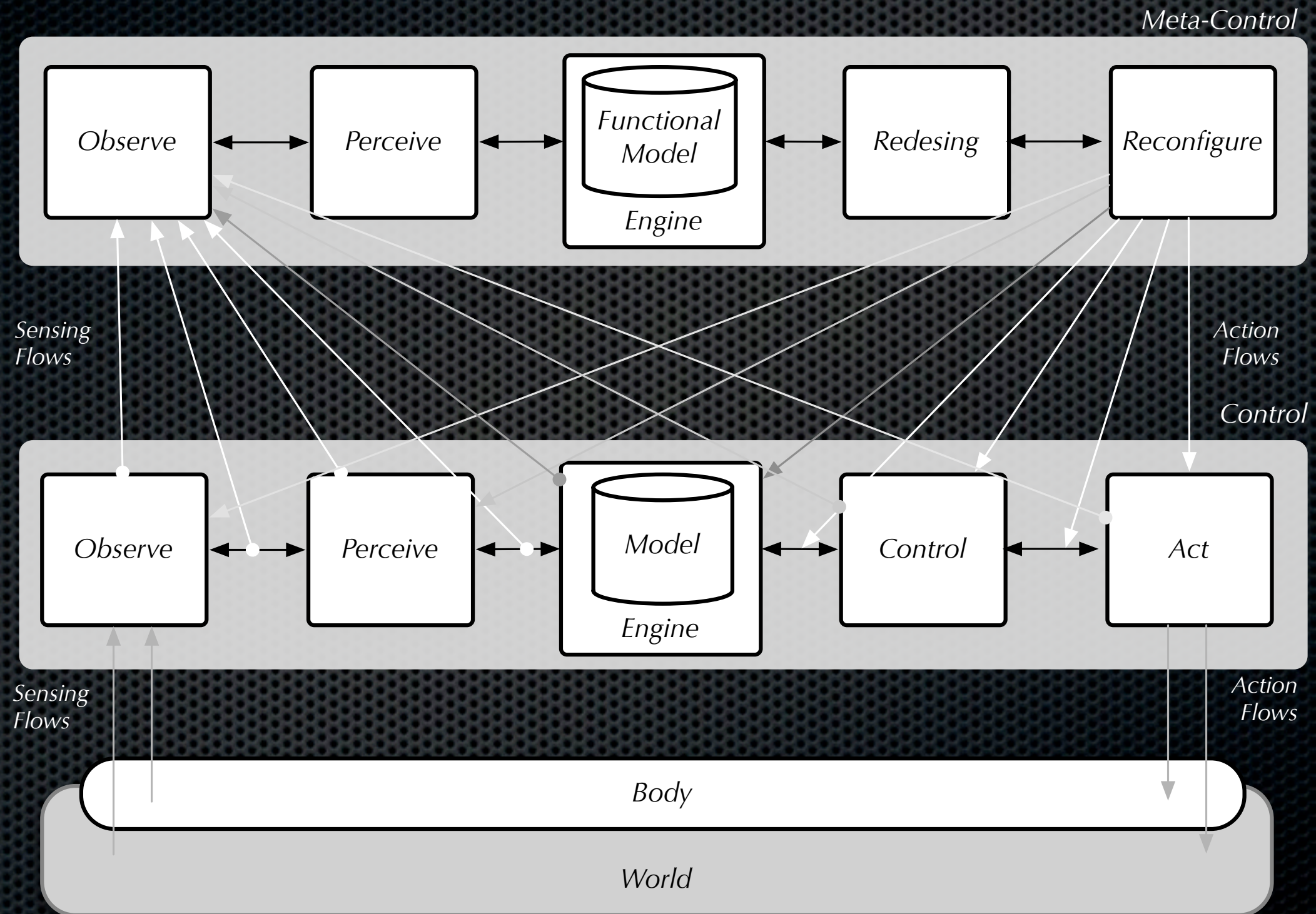
- ✦ The core objective of self-aware systems must be the provision of:

Functional robustness
in the intentional sense

The vision

- ✦ The vision is that:
 - ✦ Model-based perception of functional state + Model-based action on functional state
- ✦ Can render a system that can be
 - **Self-aware** (using a model of self in self- perception)
 - **Robust** (based on world/self model integrated use)
 - **Autopoietic** (using the self-model in a model-driven run-time synthesis)

Cognitive functional metacontrol



Consciousness as function

- ✦ Control mechanisms organise into hierarchies/heterarchies of functional assemblies
- ✦ Layering happens: e.g. Reflexes-Drives-Instincts-Cognitions
- ✦ Metacontrollers (functionally) perceive and (functionally) organise assemblies to maximise value expectancy (the core of emotion understanding)
- ✦ (Emotional) metacontrol mechanisms scale up to consciousness

Principles

• **Model-based cognition:** A cognitive system exploits models of other systems in their interaction with them.

• **Model isomorphism.** An embodied, situated, cognitive system is as good performer as its models are.

• **Anticipatory behavior.** Maximal timely performance is achieved using predictive models.

Principles

Unified cognitive action generation. Generate action based on an integrated, scalable, unified model of task, environment and self in search for global performance maximisation.

Model-driven perception. Perception is realised as the continuous update of the integrated models used by the agent in a model-based cognitive control architecture by means of real-time sensorial information.

Principles

System awareness. An aware system is continuously perceiving and computing meaning - future value- from the continuously updated models.

System self-awareness. A conscious system is continuously generating meanings from continuously updated self-models in a model-based cognitive control architecture.

Principles

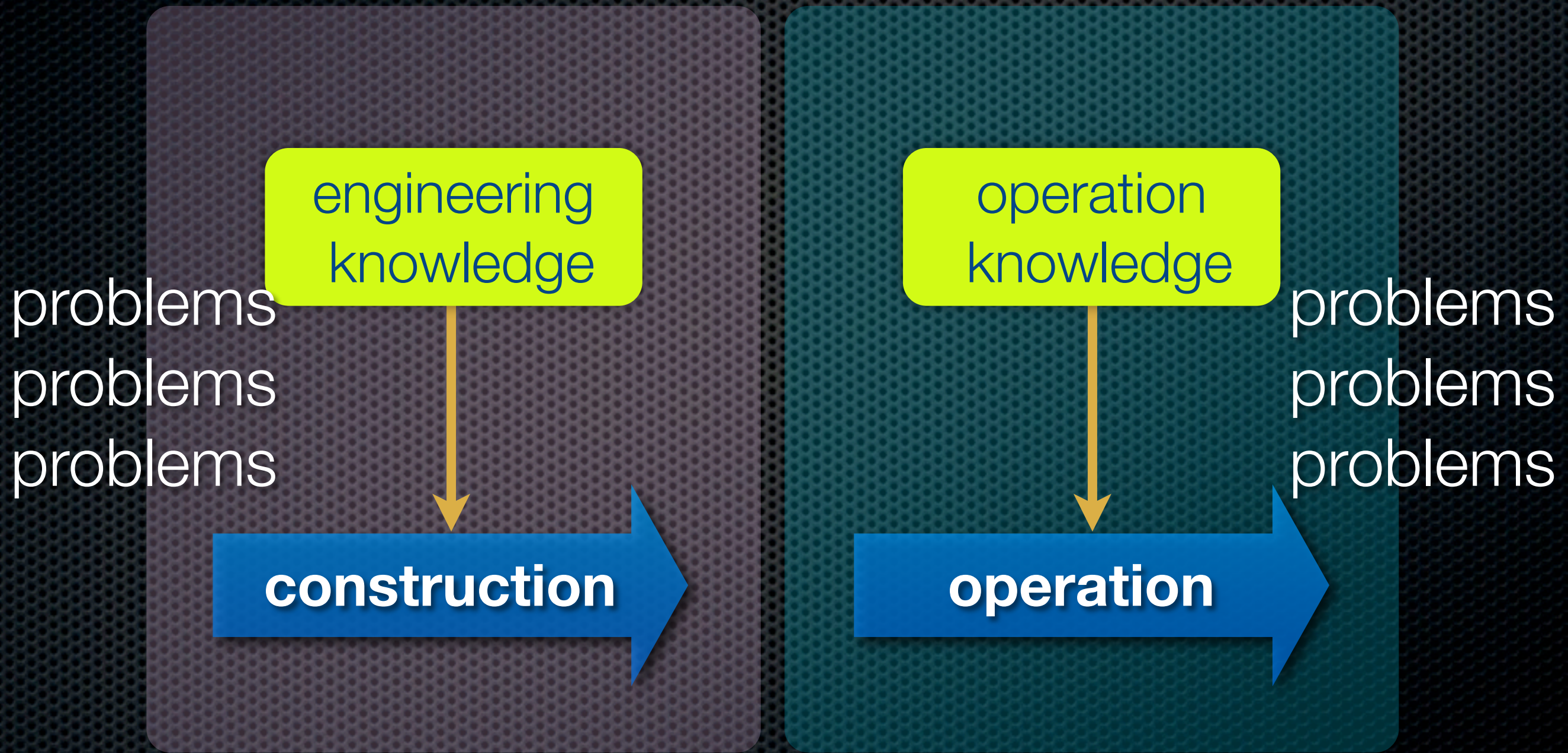
Emotional metacontrol: Emotional systems help reconfigure the system to maximise present/future value using control system patterns.

System attention: Attentional mechanisms allocate both physical and cognitive resources for system perceptive and modelling processes so as to maximise performance.

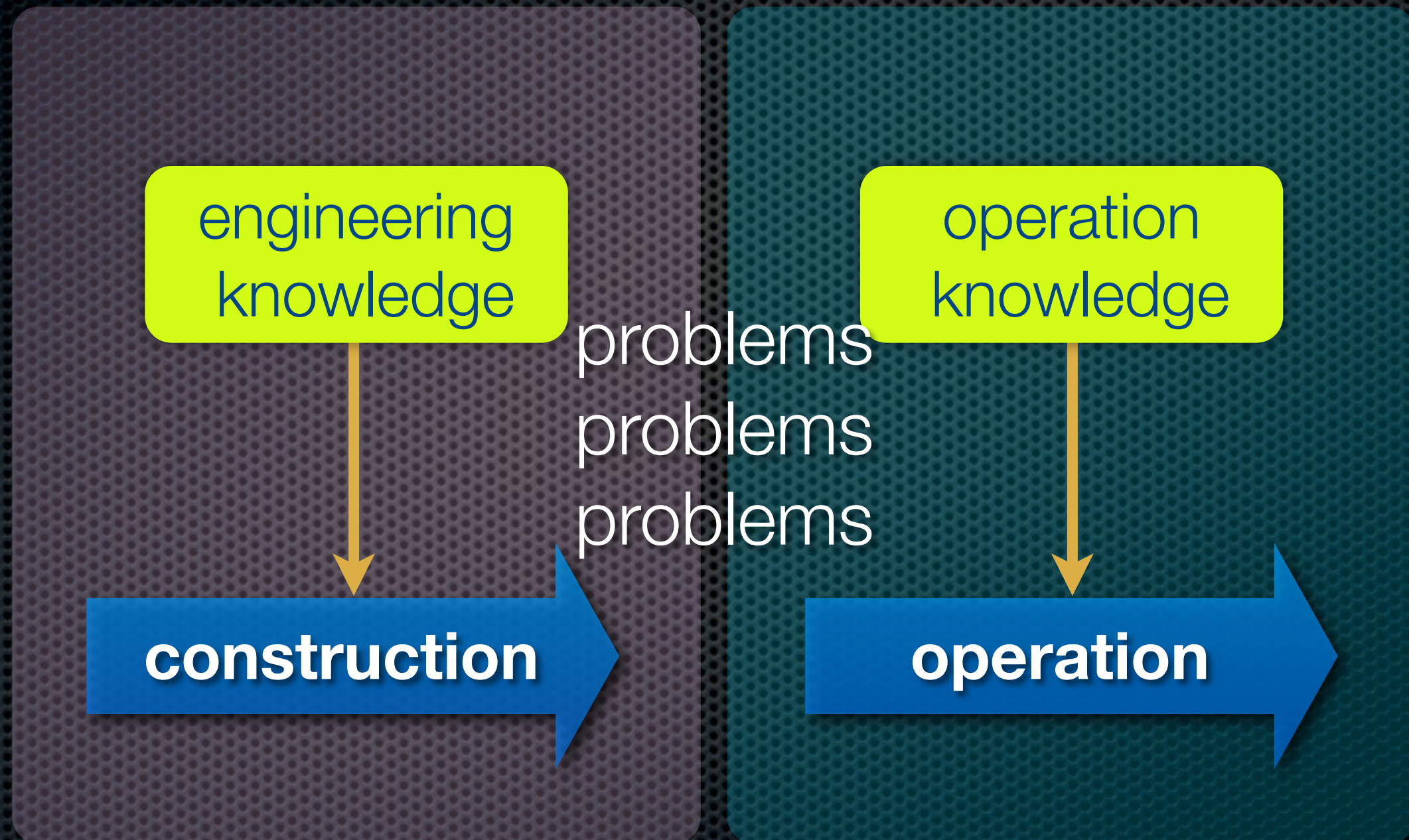
The Concept of “Autonomy”

- ✦ This dual analysis of function [causal / intentional] gives a clue for the elusive understanding of “autonomy”
- ✦ **Autonomy** = Ability to keep intentional function while suffering causal function change (functional robustness)
- ✦ All control levels -autonomic, cognitive, emotional- may contribute to this.

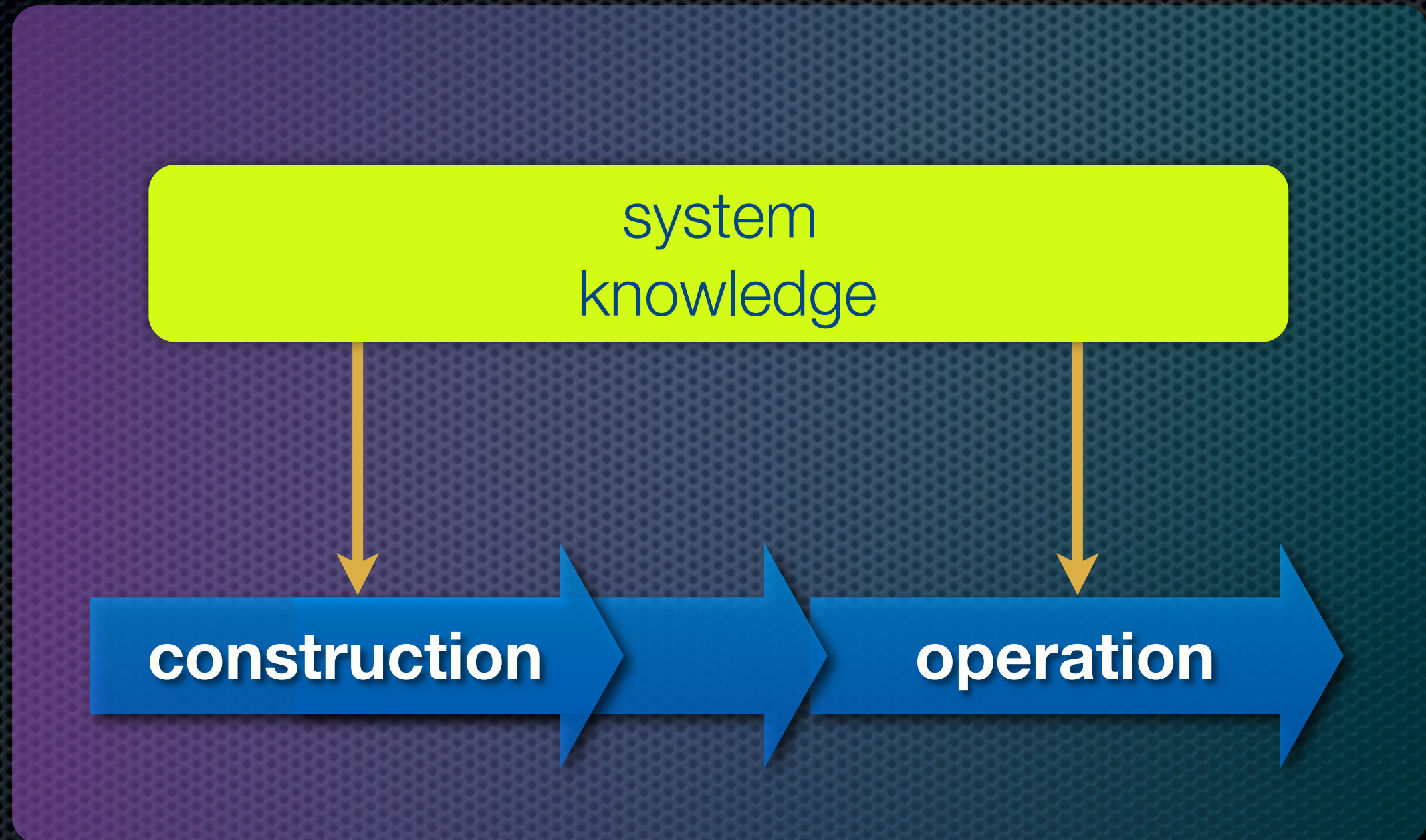
Two system epochs



A single problem set



Removing the gap



Self-engineering systems (predictable autopoietic)

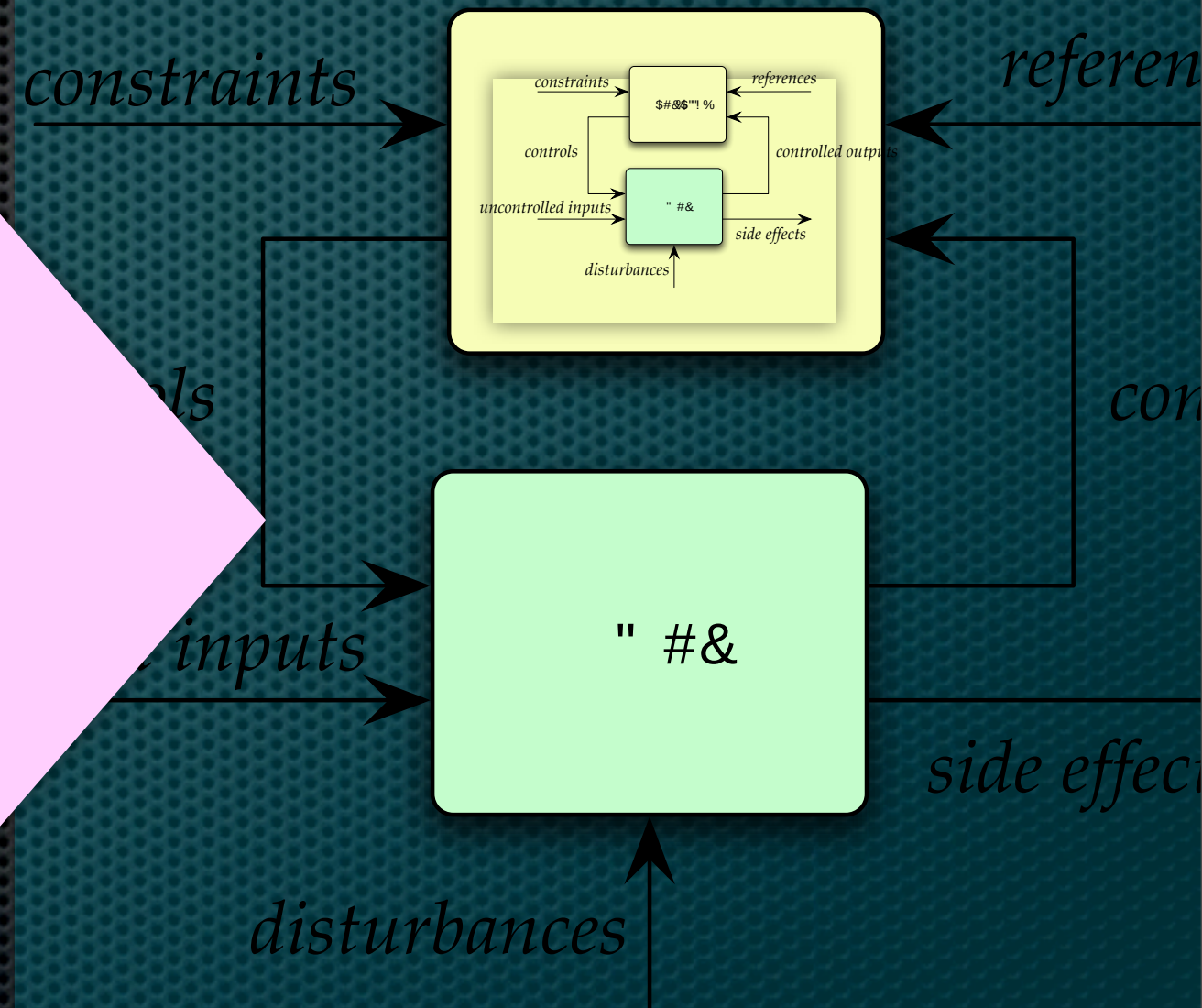
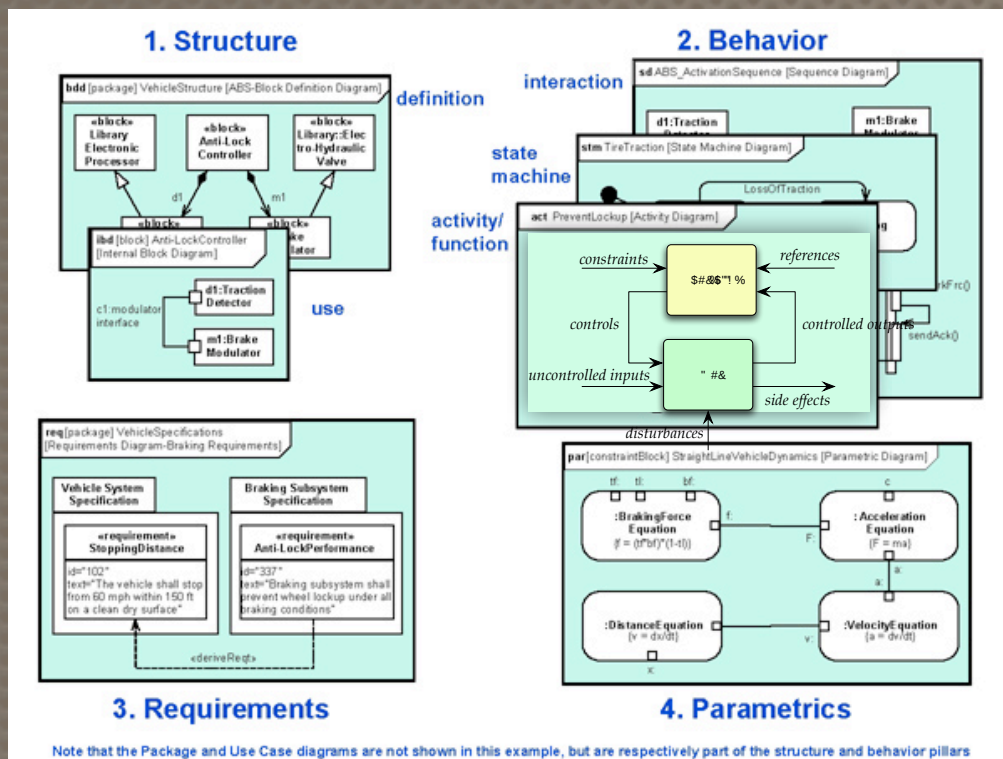
Formalise and reify architectural concepts

- ✦ This vision requires the **formalisation of core architectural design concepts**
- ✦ E.g.: action, agent, algorithm, coupling, environment, component, function, grounding, object, goal, percept, organisation, property, relation, model, resource, structure, structure, system, time-invariant relation, effect, autonomy, pattern, state, role, behaviour, constraint, etc.

Themes to consider

- ✦ Real-time control (sense-perceive-act)
 - ✦ Ontology of control tasks/objects
- ✦ Deep/wide Integration
 - ✦ Sharing ontologies horizontally and vertically
- ✦ Self-management
 - ✦ Explicit self model-based control
- ✦ Model Centric Engineering
 - ✦ Software engineering for Self-awareness

from system models to model-based reflective systems



A final proposal

Robust Autonomy by Model-based Self-Awareness

Establish the grounds for predictable self-
engineering of systems for functional robustness

Self-Awareness: Emerging views and potential impacts

James L. Crowley

Professor, Grenoble INP

Laboratoire Informatique de Grenoble

INRIA Grenoble Rhône-Alpes Centre de Recherche

Self-Awareness: Emerging views and potential impacts

Outline

- Awareness
 - Models from Cognitive Psychology
 - Applications for Informatics
- Self-Awareness
 - Models from NeuroBiology
 - Applications for Informatics
- Why study machine self-awareness?
- Potential impacts on technology and society

But first a warning....

We are all incompetent....

The scientific study of the technology for building Autonomous Self-Aware systems is in pre-paradigm.

We do not have:

- A unified scientific community that share problems and problem solutions
- a working theory for how to build such systems.
- Consensus on the fundamental concepts

We do have

- Potential contributions from many established disciplines
- Communities of interested researchers.
- An historic grand challenge that has intrigued humanity for at least 3000 years.

Introduction

We are all incompetent....

The scientific study of the technology for building Autonomous Self-Aware systems is in pre-paradigm.

We do not have:

- A unified scientific community that share problems and problem solutions
- a working theory for how to build such systems.
- Consensus on the fundamental concepts

We do have

- Potential contributions from many established disciplines
- Communities of interested researchers.
- An historic grand challenge that has intrigued humanity for at least 3000 years.

Introduction

Method:

- Explore (a few) concepts and theories from the human sciences: Human factors, cognitive science and NeuroBiology
- Examine how to use these as paradigms for designing artificial systems.

Concepts to examine:

Awareness, Autonomy, Self-Awareness.

Awareness

- Vigilance against danger or difficulty.
- Having knowledge of something.
- The ability to perceive, to feel, or to be conscious of events, objects or sensory patterns.
- Conscious of stimulation, arising from within or from outside the person

Models of awareness have been studied and applied for human factors in aviation since at least 1914.

Human Factors: Mica R. Endsley

QuickTime™ and a
decompressor
are needed to see this picture.

QuickTime™ and a
decompressor
are needed to see this picture.

Mica Endsley, Ph.D., P.E.

PhD USC 1990

editor-in-chief of the Journal of Cognitive
Engineering and Decision Making

President: SA Technologies

Specialty: Cognitive Engineering

Application Domain: Aviation and critical
systems.

Human Factors: Mica R. Endsley

Method: Task analysis using Situation Awareness Model

Key Publications:

MR Endsley, DJ Garland, (2000), Situation awareness: analysis and measurement, Lawrence Erlaum Associates.

Endsley, M.R., (2000), Theoretical underpinnings of situation awareness: A critical review, Ch1, Situation awareness analysis and measurement, pp3-32.

M.R. Endsley, B. Bolte, D.G. Jones (2003), "Designing for situation awareness: An approach to user-centered design", Taylor & Francis, New York,

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. Human Factors 37(1), pp 32-64.

Endsley, M.R., (1995), Measurement of situation awareness in dynamic systems, Human Factors: The Journal of the Human Factors and Ergonomics Society, vol 37, no 1, pp 65-84, 1995.

are needed to see this picture.

are needed to see this picture.

Endsley's Model of the Dynamic Decision Making Process.
Figure 2 from (Endsley 2000)

Levels in Situation Awareness

Situation Awareness (Endsley): The Perception of [relevant] elements of the environment in a volume of space and time, the comprehension of their meaning and the projection of their status in the near future.

Levels in Situation Awareness

- 1: Sensing: Sensing of entities relevant to task
- 2: Assimilation: association of percepts with models that predict and explain.
- 3: Projection: Forecast events and dynamics of entities

To use Endsley's model for artificial systems we need a programmable theory for perception, assimilation and projection.

Cognitive Science: Philip N. Johnson-Laird

QuickTime™ and a
decompressor
are needed to see this picture.

Philip N. Johnson-Laird

QuickTime™ and a
decompressor
are needed to see this picture.

PhD Psychology, 1967, University College London
Stuart Professor of Psychology at Princeton Univ.
1971-1973: Inst. of Advanced Study, Princeton U.
1973-1989: Laboratory of Exp. Psychology, Univ of
Sussex
1989- Applied Psychology Unit, Princeton Univ.

Cognitive Science: Philip N. Johnson-Laird

Research Method:

Situation Models as theory of mental models for natural language and inference.

Publications:

Johnson-Laird, Philip N . How We Reason. Oxford University Press.
2006

Johnson-Laird, Philip N . Computer and the Mind: An Introduction to Cognitive Science. Harvard University Press. 1998

Johnson-Laird, Philip N with Ruth M. J. Byrne. Deduction. Psychology Press, 1991

Johnson-Laird, Philip N. Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness. Harvard University Press, 1983

Situation Models:

A model for human cognitive abilities

P. Johnson-Laird 1983 – Mental Models.

A model to describe human cognitive ability to

- 1) Provide context for story understanding
- 2) Interpret ambiguous or misleading perceptions.
- 3) Reason with default information
- 4) Focus attention in problem solving

Situation Models:

A model for human cognitive abilities

Application Domains in Cognitive Psychology

- Understand spoken narration
- Understanding text and literature
- Spatial reasoning
- Social interaction
- Game playing strategies
- Controlling behavior

Situation Models: Concepts

P. Johnson-Laird 1983 – Mental Models

Situation: Relations between entities

Entities: People and things. Correlated observations.

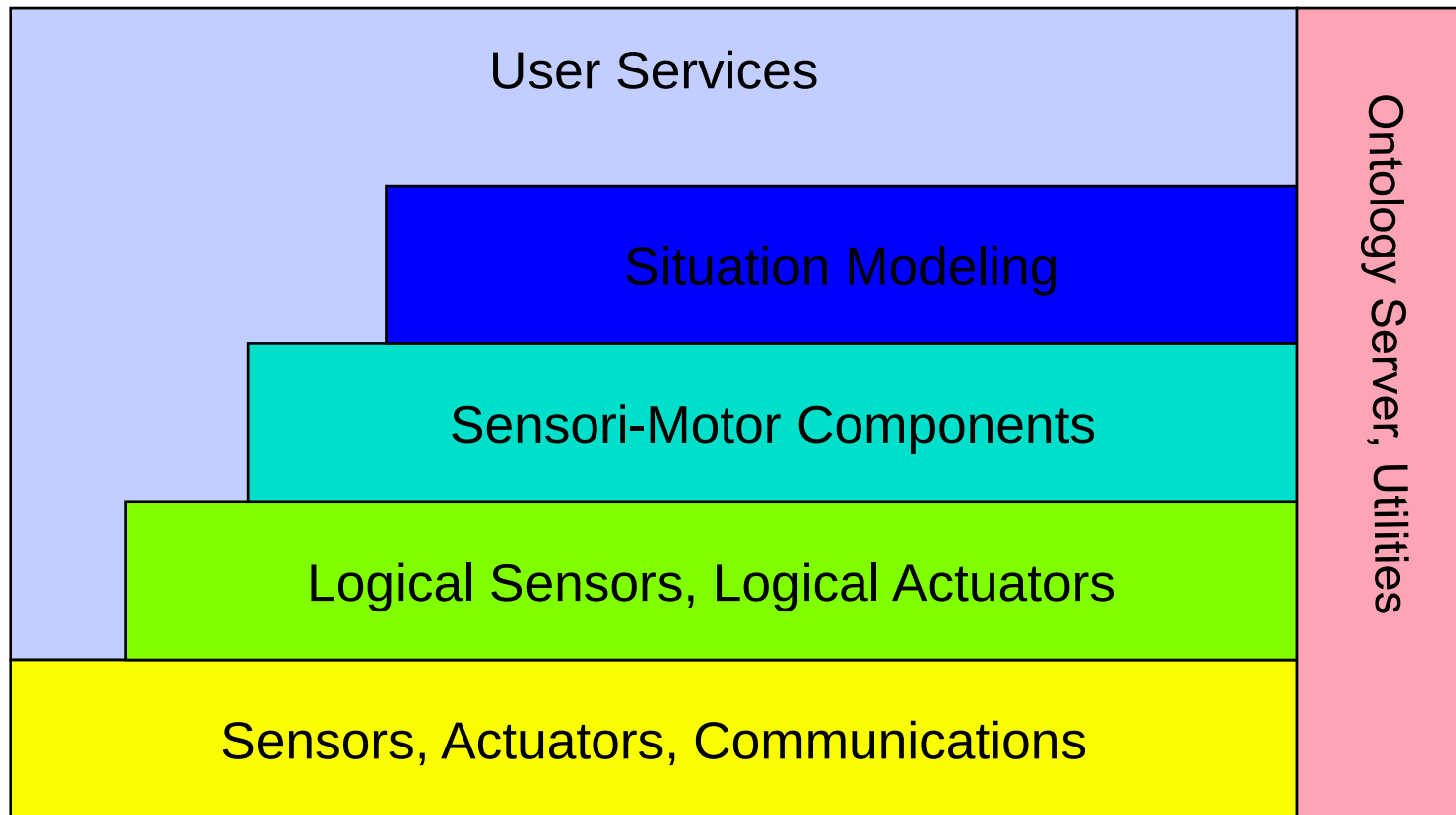
Relations: An N-ary predicate ($N=0,1,2,\dots$)

Examples: John points-at board, John looks-at audience

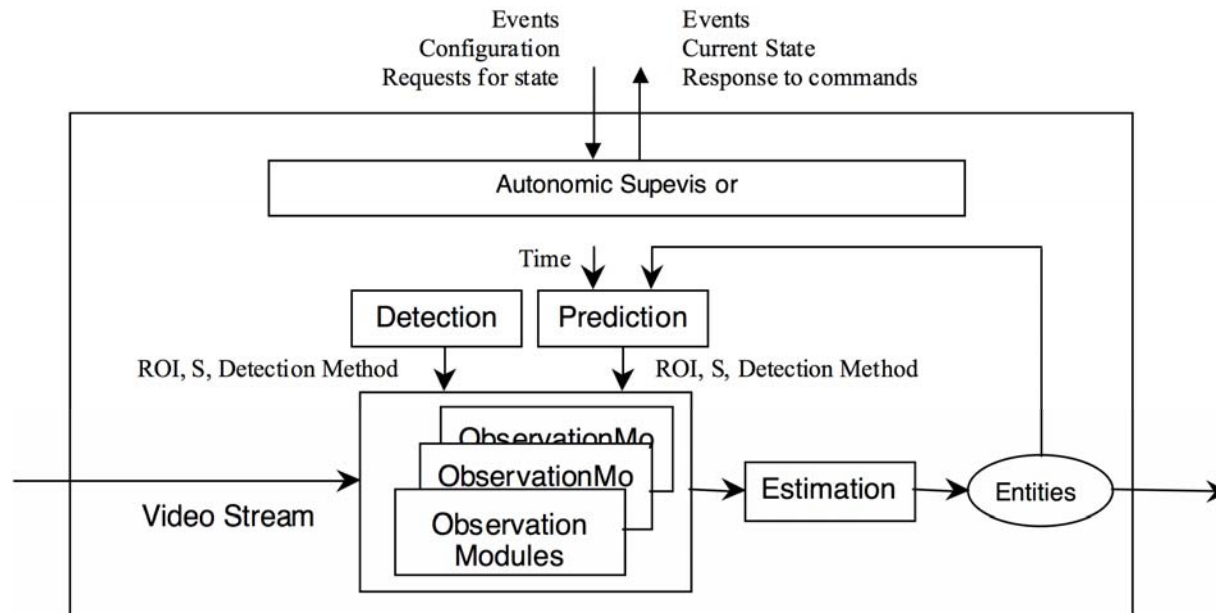
Situation Models can be used as a programming theory.

Software Architectural Reference Model

(ICT CHIL - Computers in the Human Interaction Loop)



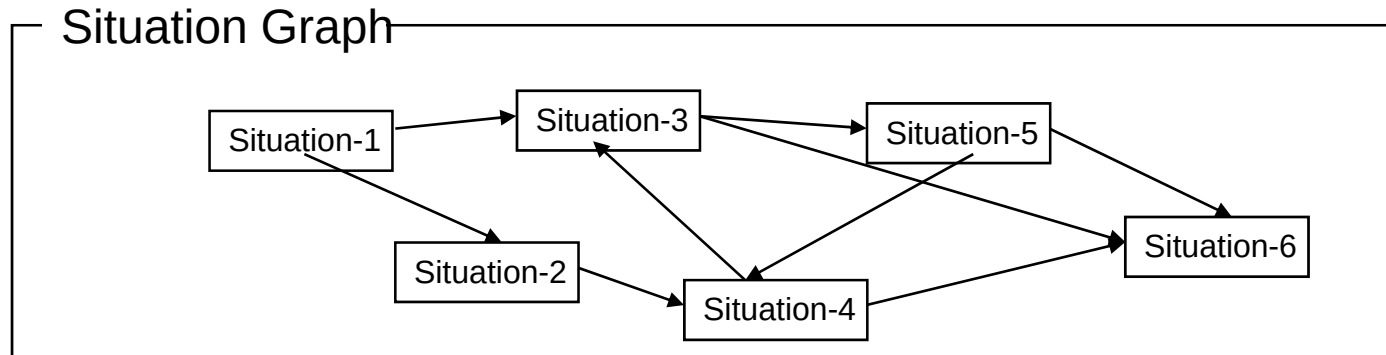
Assimilation: Perceptual Component



Autonomic Supervisor Provides:

- Module execution
- Interpretation for external communication
- Parameter Regulation
- Description of State and Capabilities

Projection: Situation Graph



A situation graph describes a state space of situations

Each Situation provides information for:

- Focus of Attention: entities and relations for the system to observe
- Default information (Context)
- Predictions about possible next situation
- System actions (prescribed, allowed or forbidden)

Situation Models: A Tool for Observing and Understanding Activity

Examples of applications of situation models

- Lecture Recording system
- Meeting event recording system
- Automatic Privacy filter (for mediaspace)
- Monitoring activity for assisted living
- Teaching polite interaction to robots
- Automated Video surveillance
- Customer observation systems.

An example: Observing a card Game

Actors:

- Jerome, Sonia and Stan

Objects:

- Card table, cards

Roles:

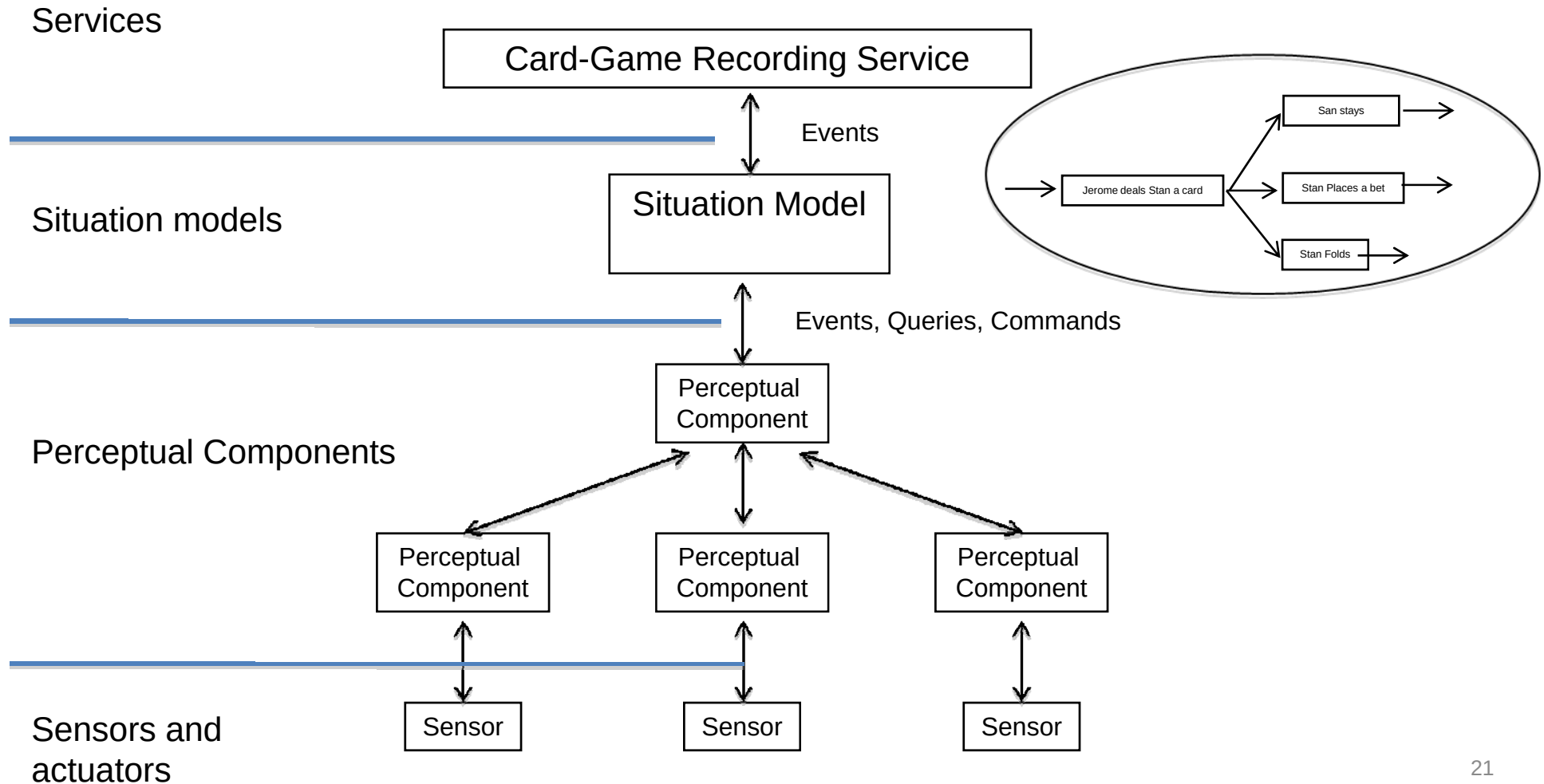
- Dealer, player

Relations:

- <Dealer> deals to <player>
- <player> talks to <player>
- <player> Folds
- <player> makes a bet
- <player> talks to <x>



An example: Recording events in a card Game



Awareness

Machine awareness (in the sense of Endsley or Johnson-Laird) is within the current state of the art in machine perception

Awareness of self requires more: A sense of self.

Awareness of Self

Self-awareness: Aware of oneself as an individual.

Self: Consciousness of one's own identity

Consciousness:

- an alert cognitive state in which a subject is aware of himself and his situation
- knowing and perceiving; having awareness of surroundings and sensations and thoughts;

In medicine, consciousness is assessed by observing a patient's alertness and responsiveness to stimuli

Autonomy and Self

Assertion: a sense of self requires autonomy

Autonomy:

- From auto "self" + nomos, "law": one who gives oneself his/her own law
- Existing as an independent entity
- Self-governing, Self-protecting.
- Able to self-maintain functional integrity.

Autonomous System: An independent system that controls itself and maintains its own integrity

Autonomic Computing provides an enabling technology for Autonomy.

Autonomic Computing

Autonomic computing : a metaphor inspired by natural self-governing systems, and the autonomic nervous system found in mammals.

March 2001 Keynote address to the National Academy of Engineers by Paul Horn (IBM vice president)

Autonomic computing systems are systems that manage themselves given high-level objectives from administrators.

A scientific community for autonomic computing for computer operating systems has emerged.

Concepts from autonomic computing are also having impact on Computer Vision, Robotics, and Computer Networks.

Autonomic Nervous System (ANS)

The ANS regulates the homeostasis of physiological functions

The ANS is not consciously controlled.

Commonly divided into three subsystems:

- Sympathetic nervous systems (SNS)) (fight or flight)

- Parasympathetic nervous system (PNS) (rest and digest)

- Enteric nervous systems (ENS) (the second brain)

The ANS maintains internal homeostasis.

Homeostasis

Homeostasis :

- the ability of a system to regulate its internal environment and to maintain a stable, constant condition. Typically used to refer to a living organism.

Defined by Walter Bradford Cannon (1932) from ancient Greek (homos, Similar), (histēmi, Standing still).

Inspired by the milieu interieur described by Claude Bernard in 1865.

- Homeostasis is fundamental to life.
- Humans extend homeostasis to their external environment.
- Awareness (Vigilance against danger or difficulty) is part of the extension of homeostasis.

Autonomic Properties for Software Systems

The essence of Autonomic Computing is Self Management.
(Klephart and Chess, 2002).

Self-Configuration: Automatic configuration of components;

Self-Healing: Automatic discovery, and correction of faults;

Self-Optimization: Automatic monitoring and control of resources to ensure the optimal functioning with respect to the defined requirements;

Self-Protection: Proactive identification and protection from arbitrary attacks.

Autonomic Properties for Computer Vision

As used of Perceptual Components in the ICT CHIL project.

(JL Crowley, Autonomic Computer Vision, Int. Conf on Vision Systems, 2007.)

- Auto-configuration: The component can configure its own operating parameters
- Auto-regulation: The component adapts parameters to maintain a desired process state.
- Self-description: The component provides descriptions of the capabilities (to component registry) and the current state of the process (on request).
- Self-monitoring: The component supervisor estimates state and quality of service for each processing cycle.
- Self-repair: The component can detect and correct conditions by reconfiguring modules.

The use of Self in this context is abusive! (pet-peeve).

Systems and services do not (currently) have a sense of self.

Neuroscience: Antonio R. Damasio

QuickTime™ and a
decompressor
are needed to see this picture.

QuickTime™ and a
decompressor
are needed to see this picture.

Dr. Antonio Damasio

Professor of Neuroscience USC

USC Brain and Creativity Institute Studies

Research: neurobiology of the mind, especially
neural systems for memory, language,
emotion, and decision-making.

Neuroscience: Antonio R. Damasio

Research Method:

Correlates loss of cognitive abilities with damage to neural structures

Books :

- Descartes' Error: Emotion, Reason, and the Human Brain 1994
- The Somatic marker hypothesis and the possible functions of the prefrontal cortex, 1996
- **The Feeling of What Happens: Body and Emotion in the Making of Consciousness, 1999**
- Looking for Spinoza: Joy, Sorrow, and the Feeling Brain, 2003

Emotions and Appetites

Somatic marker mechanism (Damasio): Theory of how perceptions interact with emotions.

Emotions regulate homeostasis in reaction to stimuli from the external world.

Homeostatic emotions are feelings evoked by internal body states, which modulate behavior.

Examples: Thirst, hunger, feeling hot or cold (core temperature), feeling sleep deprived, salt hunger and air.

Appetite: An instinctive physical desire, especially one for food or drink. Appetites motivate much of human behaviours

A. Damasio : Some vocabulary

Images: Neural patterns with structure representing sensory stimuli:
Visual, auditory, olfactory, gustatory, somatosensory (body).

Images Depict physical properties of entities
May be driven by senses or by internal stimuli (dispositions)
Are dynamic (spatio-temporal).
Are not always conscious. (example: Blind-sense).

Sensori-motor maps: express interaction between an organism and an entity.

Second order maps: represent changes in Sensori-motor maps.

Categories of Self (A. Damasio)

Core-Self: A transient entity; recreated for each object with which the brain interacts

Proto-Self: A representation of core self in a second order neural map (image).

Autobiographic Self: Permanent dispositional records of core-self experiences. Systemized memories of situations in which core consciousness was involved with invariant characteristics of life.

Location of some proto-self structures

QuickTime™ and a
decompressor
are needed to see this picture.

QuickTime™ and a
decompressor
are needed to see this picture.

QuickTime™ and a
decompressor
are needed to see this picture.

From Damasio, The Feeling of What Happens: Body and Emotion in the Making of Consciousness, 1999, figure 5.1 pp 155

Proto-Self and Second Order Maps

QuickTime™ and a
decompressor
are needed to see this picture.

From Damasio, The Feeling of What Happens: Body and Emotion in the Making of Consciousness, 1999, figure 6.3 pp 193

Self Awareness (A. Damasio)

Consciousness

Core Consciousness: an imaged, non-verbal account of the organisms own state.

Extended Consciousness: Awareness of entities and events in relation to autobiographical self.

QuickTime™ and a
decompressor
are needed to see this picture.

Neurobiological locus of self-awareness

OutTime™ and
decompressor
are needed to see the picture.

Damage to structure front brain stem(A) causes locked-in-syndrom: Consciousness with inability to move (except vertical eye movement).

Damage to structure back brain stem (B) causes coma.

A technology for self-awareness?

Why create a technology for self-awareness?

- 1) A Grand Challenge for humanity (on the order of human flight, travel to the moon or mapping the genome).
- 2) Pure scientific (curiosity driven) research.
- 3) (personal view) A potentially powerful enabling technology for machine learning and development.

For 60 years the quest for self-Aware machines has been side-tracked in a quest for machine “intelligence”.

What do we mean by Intelligent?

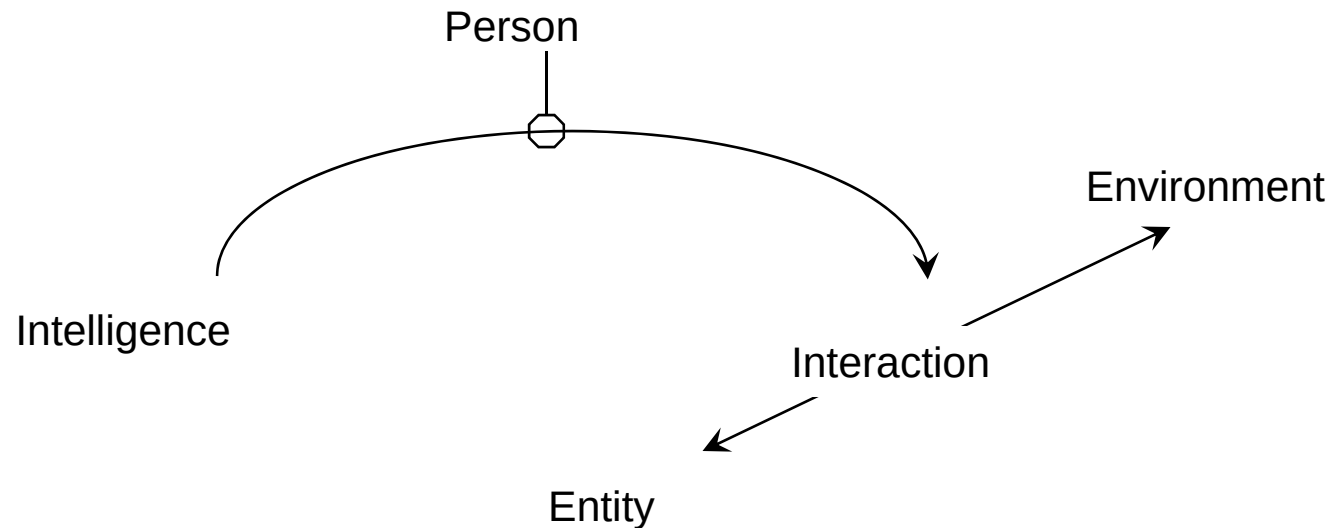
Intelligence describes the interaction of an entity with its environment.*

Intelligence is a description (an ascribed property)

Intelligence describes an entity that interacts.

*Cognitive Systems Research Roadmap (2002), European Commission, ECVision Network (David Vernon, Editor).

What do we mean by Intelligent?

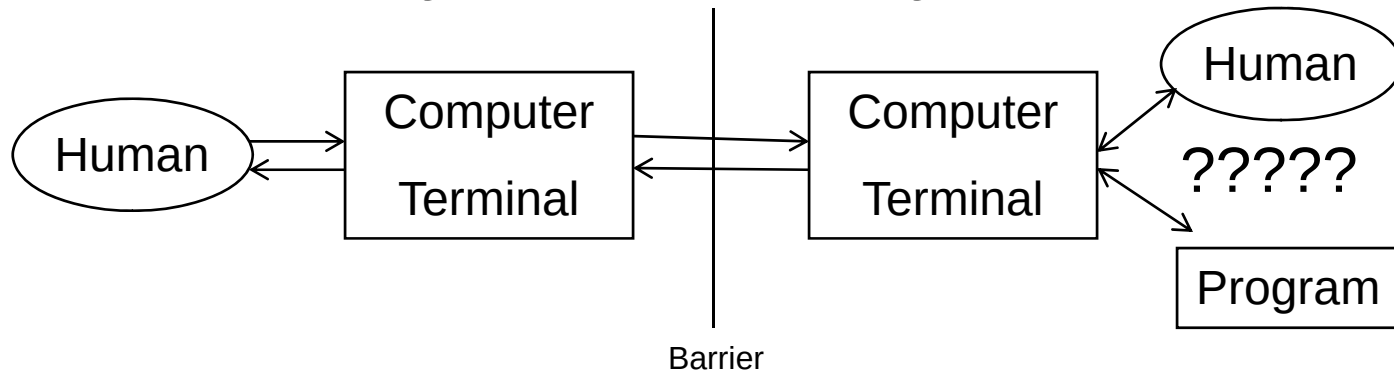


Intelligence describes the interaction of an entity with its environment.

Human level intelligence requires an ability for social interaction. (the Turing Test).

Early Computing: Turing's View of Intelligence

The Turing Test: The imitation game



Alan Turing changed the question from "Can machines think?" to "Can a machine behave as a human?"

Turing claimed that a machine would exhibit intelligence if it exhibited behaviour that could not be distinguished from a person.

The Turing test measures abilities for social interaction.

Social Interaction for Personal Robots



QuickTime™ and a
decompressor
are needed to see this picture.



There is no technology for personal robots to learn polite interaction.

- My Roomba cleaning robot is aware but not self aware.
- My Aibo robot interacts but does not have a second order model of his own sensori-motor maps.
- There is no technology to endow a robot with a "sense" of "self"

Computer Science: Marvin Minsky

QuickTime™ and a
decompressor
are needed to see this picture.

Marvin Minsky

Professor MIT

PhD in mathematics Princeton (1954)

Fundamental contributions to AI, cognitive psychology, mathematics, computational linguistics, robotics, and optics.

QuickTime™ and a
decompressor
are needed to see this picture.

Social Common Sense

Minsky distinguishes

Common sense: The collection of shared concepts and ideas that are accepted as correct by a community of people.

Social Common Sense: shared rules for polite, social interaction that govern behavior within a group

Situated Social Common Sense: Social common sense appropriate to the current situation.

Situated Social Common Sense

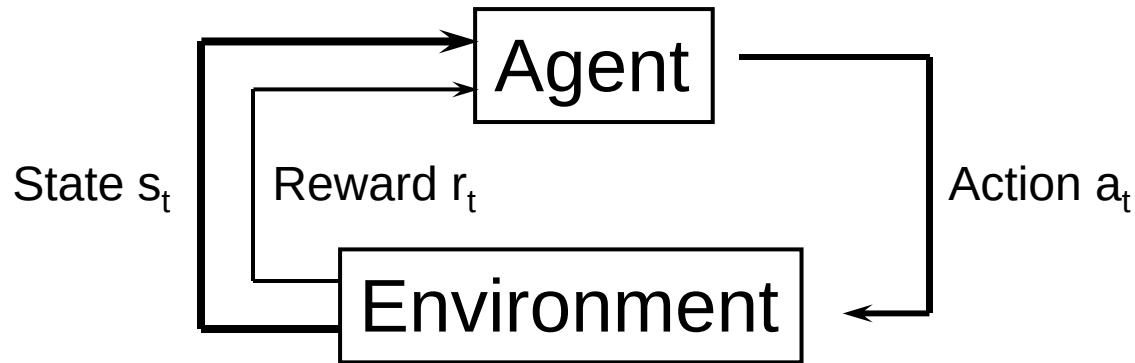
Assertions:

- Politeness is a problem of Situated Social Common Sense
- Politeness requires understanding social situation

Social Common Sense varies over individuals and groups.

⇒ Social Common Sense must be learned through interaction

Reinforcement Learning



Algorithm to learn a policy

$\pi: \textit{situation} \rightarrow \textit{action}$

Q-Learning [Watkins 89] :

- Learns policy and Q-Value at the same time
- Based on temporal different and learning rate.

Fundamental Hard Problem: Credit Assignment; Assigning rewards to actions.

Q-Learning

Estimates the value of state-action pairs

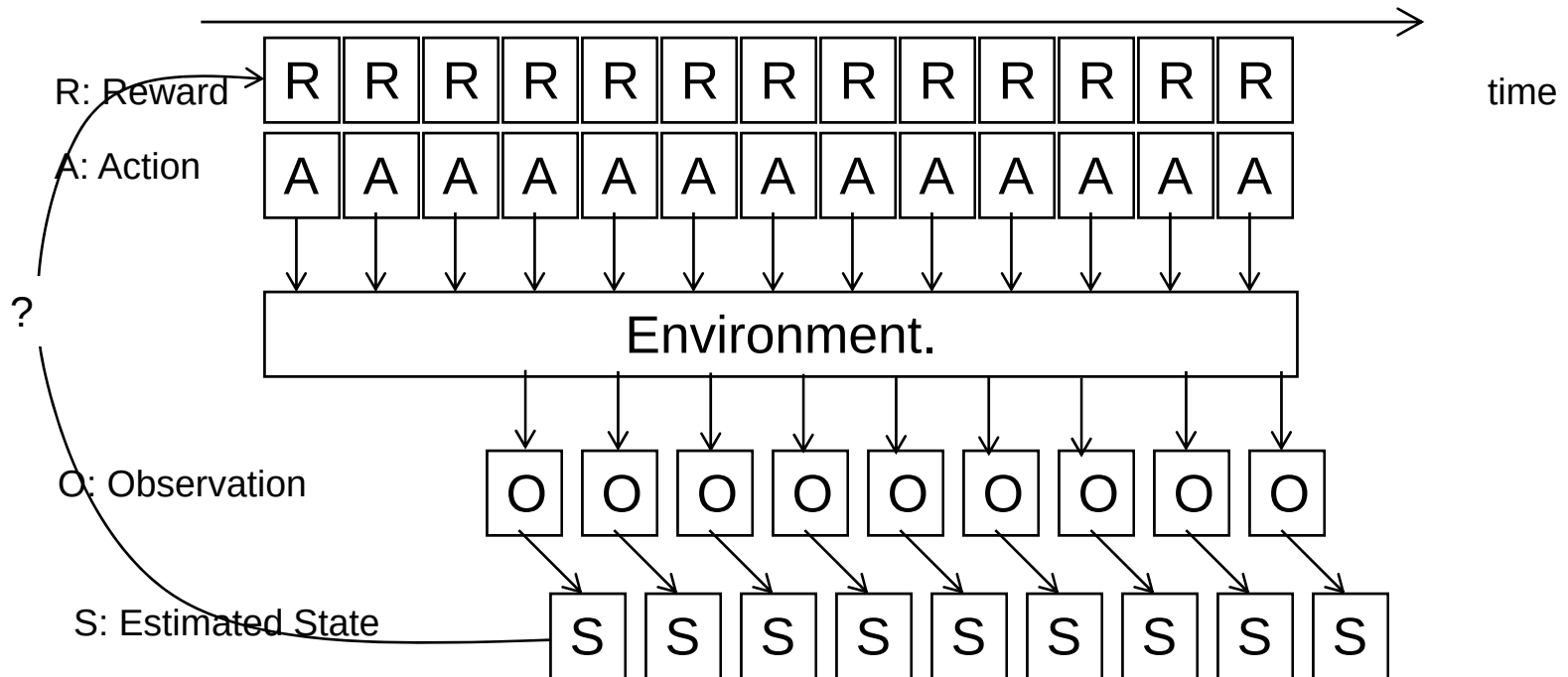
$$Q_{t+1}(s_t, a_t) = (1 - \alpha_t) \cdot Q_t(s_t, a_t) + \alpha_t \left(R(s_t, a_t) + \gamma \max_{a'} Q_t(s_{t+1}, a') \right)$$

Asynchronous Q-learning: Variable learning rate with (s, t)

$$\alpha_t = \frac{1}{(1 + t_{s,a})^w} \text{ and } w \in \left(\frac{1}{2}, 1 \right)$$

Immediate reward fails completely in complex interactions.

The Credit Assignment Problem



Reinforcement learning works fine for simple actions with fixed delay effects.
Reinforcement learning is impractical for social interaction because of the complexity of the actions and the results.

Why is Social Learning hard?

Learning rate should depend on many social factors, including

- Social Context
- Time of Day
- Emotional stimulation
- Motivation, Attention
- Nature of reward
- Surprise

Hypothesis (personal): The autobiographical "Self" may facilitate learning social common sense by providing abstraction.

Combined with case-based learning, self-awareness may permit learning for social interaction.

Why is Social Learning hard?

Fundamental Problem

The Knowledge Barrier:

The extreme complexity of human activity and individual preferences makes credit assignment for individual actions impossible

Proposed Solution

Use autobiographical self is a method for organising abstract models of behaviour for learning from social interaction.

Possible Impacts on Technology

Self awareness can be used to provide:

- Enhanced system reliability
- Machine Learning for social interaction
- Learning for complex behaviours.

Possible Application Domains

- Personal Robots (Self-aware appliances)
- Teaching machines
- Self-aware Smart Homes
- Self-aware Intelligent Vehicles
- Self-aware Smart Cities

Impacts on Society

- Artificial systems with beyond-human intelligence.
- Social interaction with self-aware robots and services as part of Human Development.
(teaching machines training children?)
- Extension of “human rights” to Machines.
- Legal responsibilities for autonomous self-aware machines?

Conclusions

1. The invention of a technology for autonomous self-aware machines is a long grand challenge for humanity
2. Scientific investigation in this area is “pre-paradigm”
3. Emergence of a working theory is (finally) within our reach.
4. A working theory requires convergences of ideas and methods from a variety of disciplines.
5. Success may have long term impact on humanity and the human condition.



**Hasso
Plattner
Institut**

IT Systems Engineering | Universität Potsdam



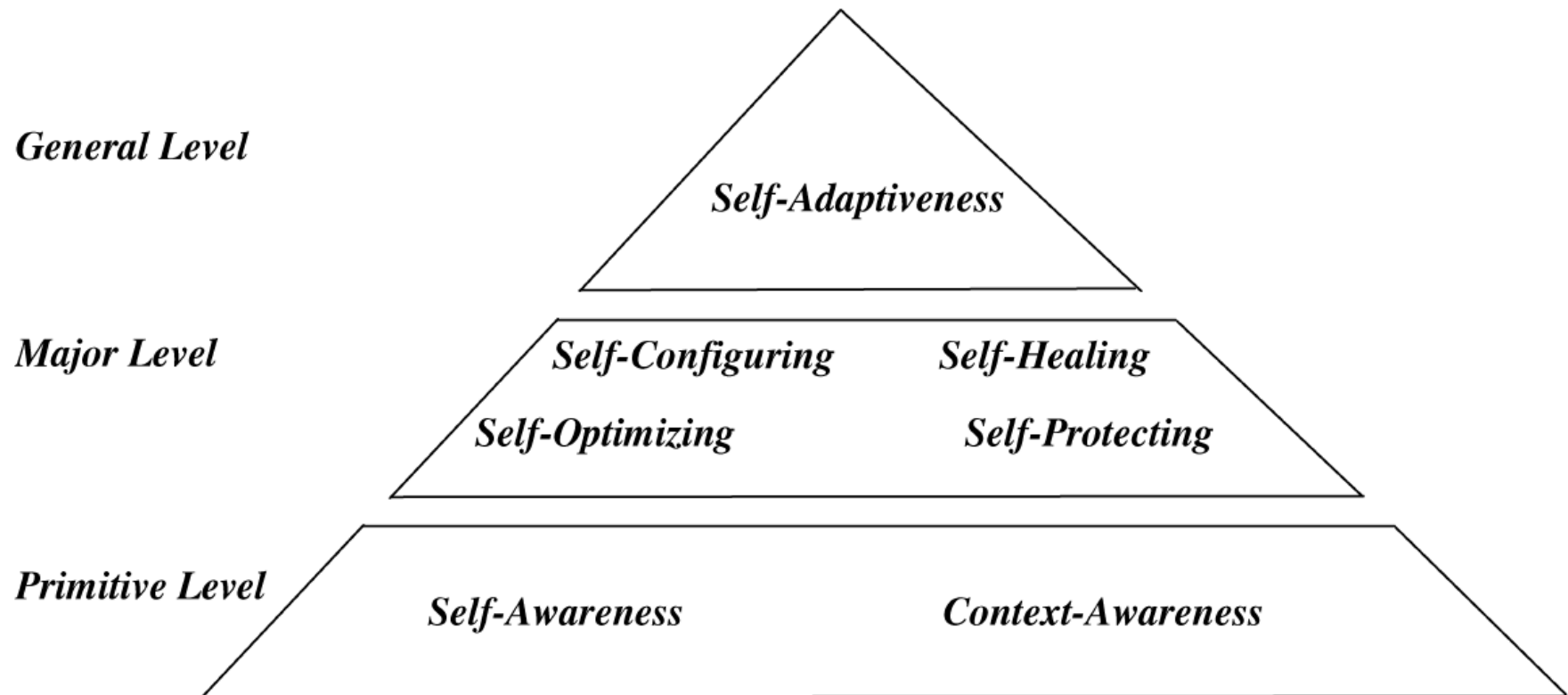
Self-Awareness and Self-Consciousness via Software: An Engineering Perspective

Expert Talk, First CHIST-ERA conference
EUR - Rome, 28 May 2010

Holger Giese
System Analysis & Modeling Group, Hasso Plattner
Institute for Software Systems Engineering at the
University of Potsdam, Germany
holger.giese@hpi.uni-potsdam.de

My Terminology

2



Mazeiar Salehie and Ladan Tahvildari. *Self-adaptive software: Landscape and research challenges*. In ACM Trans. Auton. Adapt. Syst., Vol. 4(2):1--42, ACM, New York, NY, USA, 2009.

Self-awareness and self-consciousness via software will lead to a new generation of adaptive and evolving systems of systems.

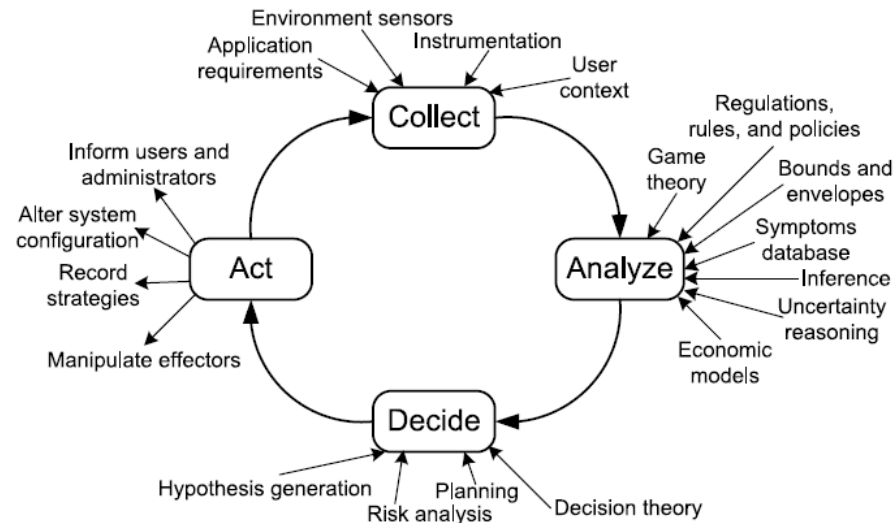
Disclaimer: *Self and consciousness* are considered only in a weak partial sense.

BUT we do know yet

- How to **cost effectively** engineer this capabilities?
- How to engineer **meaningful** and **trustworthy** systems of systems where the subsystem have this capability?

Adaptation Loop & Self-Consciousness

4



Roadmap:

- Requirements
- Modeling
- Feedback loops
- Assurance

Self-Awareness:

- Includes collect
- Enables analyze
- Basis for act

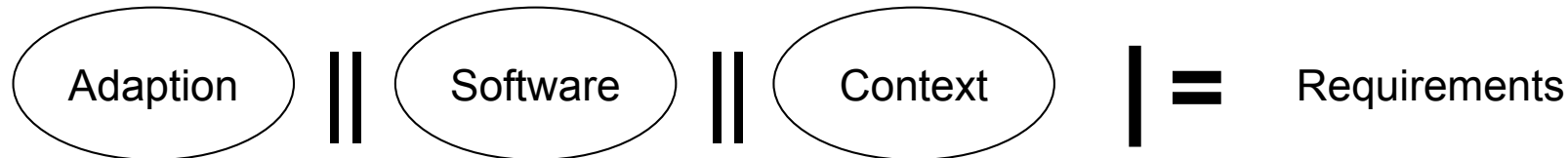
Self-Consciousness:

- Includes collect
- Enables analyze
- Basis for decide (alternatives as well as comparison)
- Basis for act

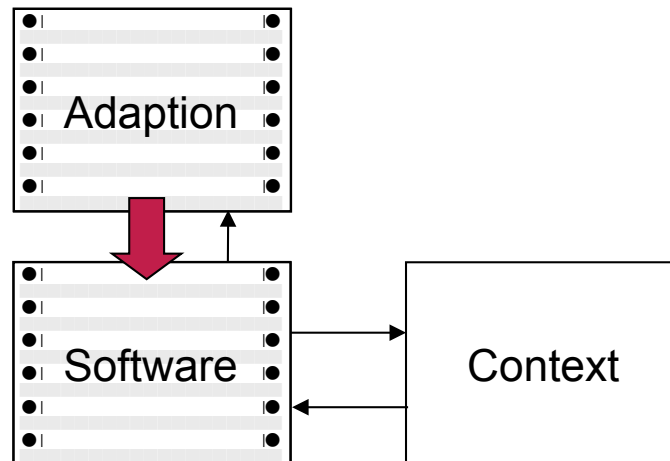
Engineer Monolithic Self-Awareness

5

Development Time:



Runtime:



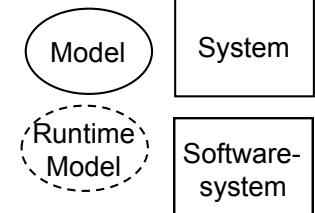
Benefit:

Adaption based on self-awareness (and context-awareness) enables to fulfill the requirements despite dynamic changes of the context (reflected in the software) and software.

Challenge: engineer adaptation loop

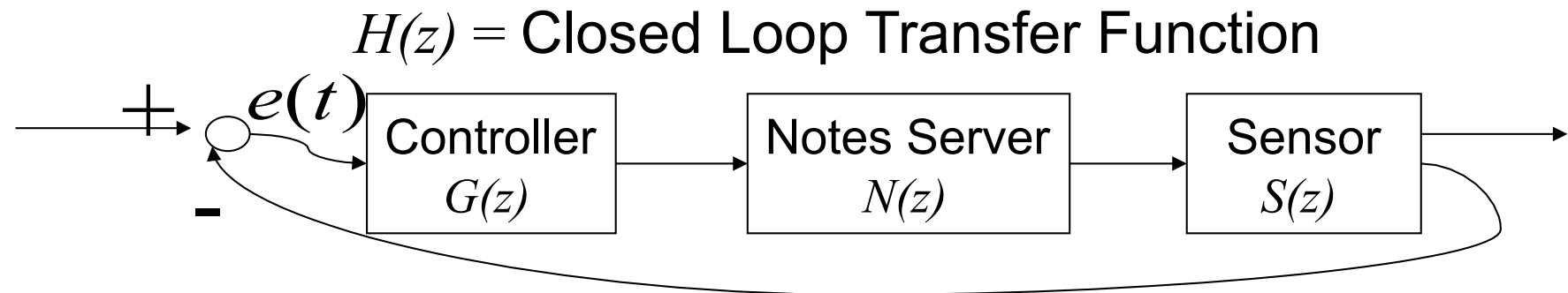
Assumption: known context and software where only parameters change

Legend:



Autonomic Computing: Software & Control Theory

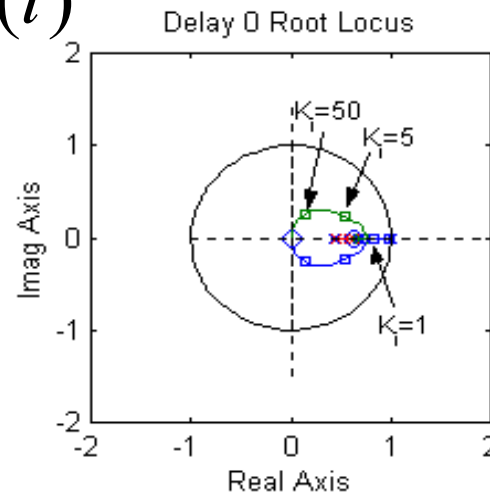
6



Integral Control Law

$$u(t) = u(t-1) + K e(t)$$

**Poles of
 $H(z)$**



Transfer Function

$$N(z) = \frac{b_1}{z - a_1}$$

**Choose a
good K**

Beyond Parameter Changes ...

7

Example: "Ambient Intelligence (AmI): ...
By adding intelligent user interfaces and integrating sensing devices, it should be possible to **identify and model** user activities, preferences and behaviours, and create individualised profiles."

Gasson, Mark; Warwick, Kevin (2007), "D12.2: Study on Emerging AmI Technologies", FIDIS Deliverables 12 (2)

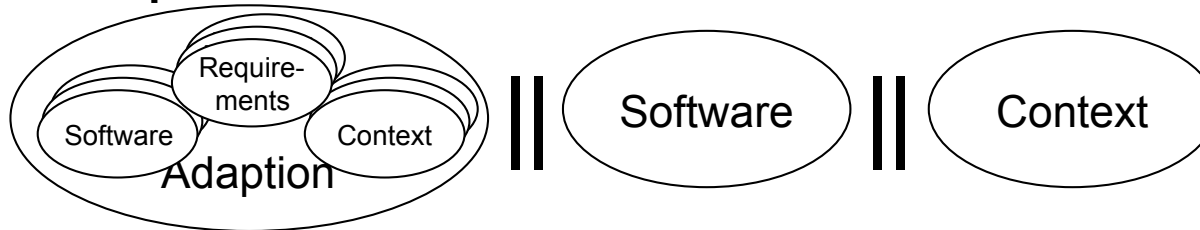


Source: Fraunhofer Verbund Mikroelektronik

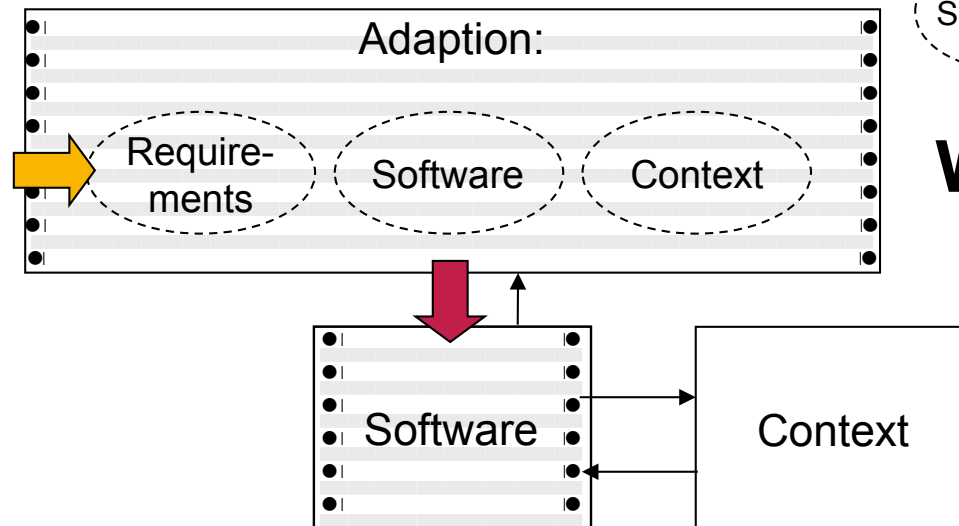
Engineer Monolithic Self-Consciousness

8

Development Time:



Runtime:



$$\text{Software}' = f(\text{Context}, \text{Software}, \text{Requirements})$$

with $\text{Software}' || \text{Context} = \text{Requirements}$

Benefits:

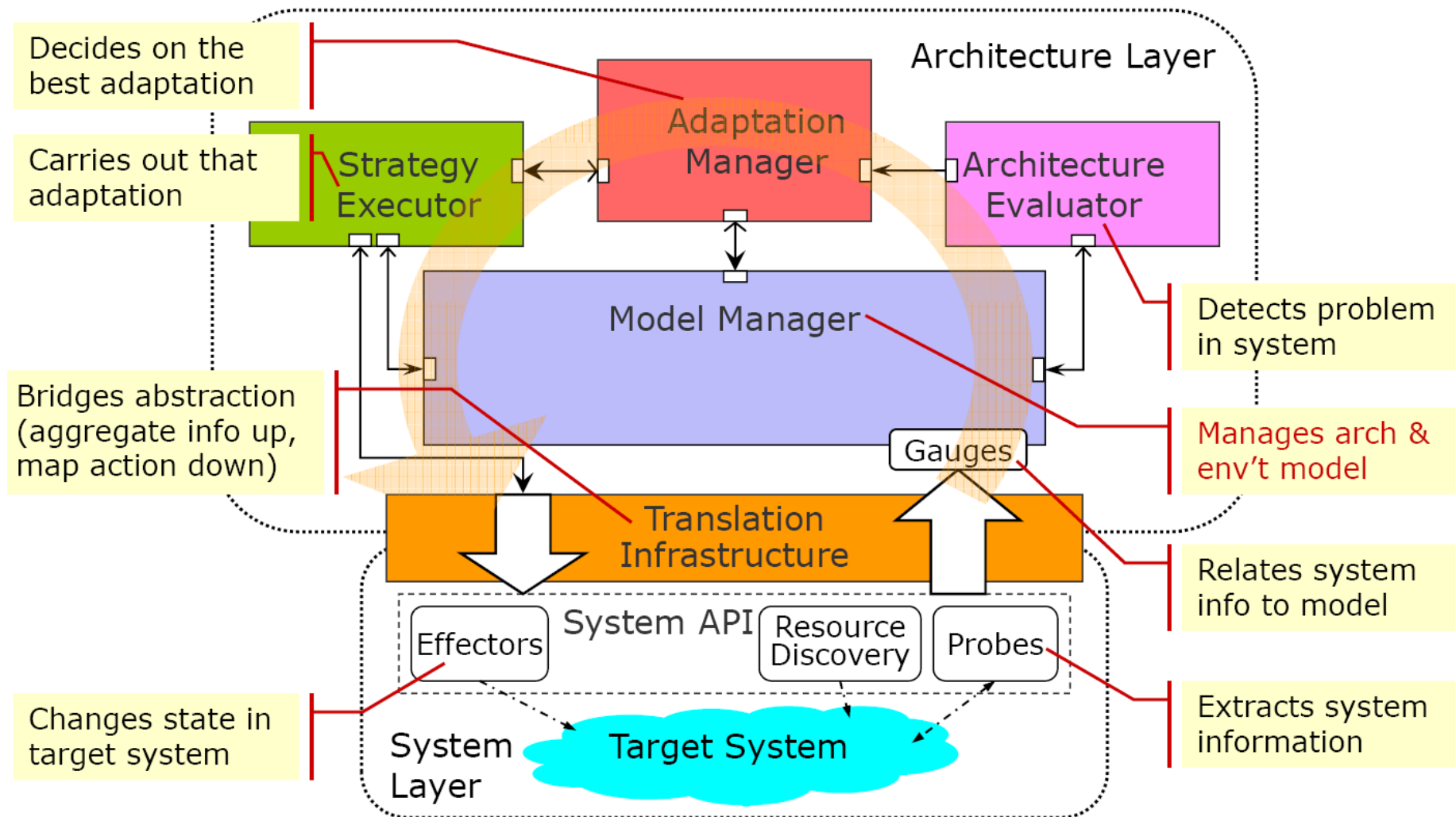
- Covers more possible changes of the software and context
- Possible changes of the software and context can be analyzed at runtime (by f)
- Changes of the software could also be triggered by changes of the requirements

Challenges: engineer + find reasonable f + assurance

Assumption: stable context and requirements

Framework: Reuse Adaptation Infrastructure

9



Rainbow

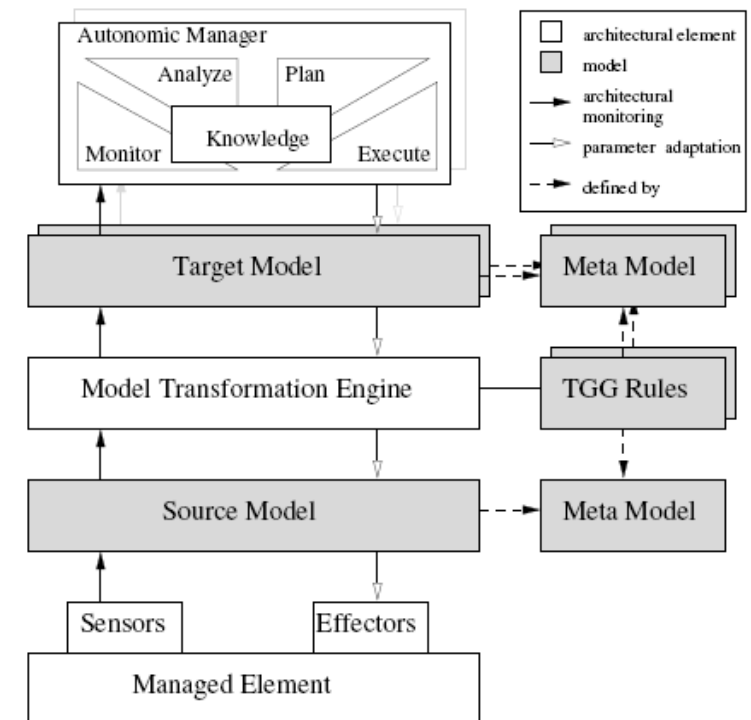
Self-Awareness and Self-Consciousness via Software: An Engineering Perspective

David Garlan, Bradley Schmerl, Shang-Wen Cheng. Architecture-based Self-adaptation. IBM 3rd Proactive Problem Prediction Conference. April 27, 2005

Requirements and Models at Runtime

10

- Requirements are monitored at runtime
- Software is represented by a runtime model (synchronization)
- Generate system specific parts for collect, analyze, decide and act



Vogel, T., Neumann, S., Hildebrandt, S., Giese, H., Becker, B.: Model-Driven Architectural Monitoring and Adaptation for Autonomic Systems. In: Proc. of the 6th International Conference on Autonomic Computing and Communications (ICAC'09), Barcelona, Spain, ACM (15-19 June 2009) accepted paper.

Beyond Monolithic Systems ...

11

Prognoses:

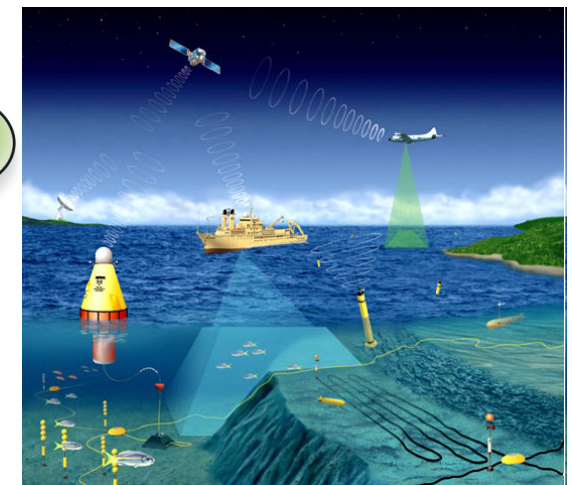
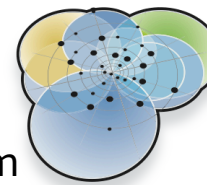
"In the near future, **software-intensive systems** will exhibit **adaptive** and **anticipatory behavior**; they will process knowledge and not only data, and **change their structure dynamically**. Software-intensive systems will act as global computers **in highly dynamic environments** and will be based on and **integrated** with service-oriented and pervasive computing."

M. Wirsing, ed., Report on the EU/NSF Strategic Workshop on Engineering Software-Intensive Systems "Challenges, Visions and Research Issues for Software-Intensive Systems", at ICSE 2004. Edinburgh, UK, May 2004.

"The sheer scale of **ultra large scale systems** will change everything. ULS systems will necessarily be **decentralized** in a variety of ways, developed and used by a wide variety of stakeholders with conflicting needs, **evolving continuously**, and constructed from heterogeneous parts.

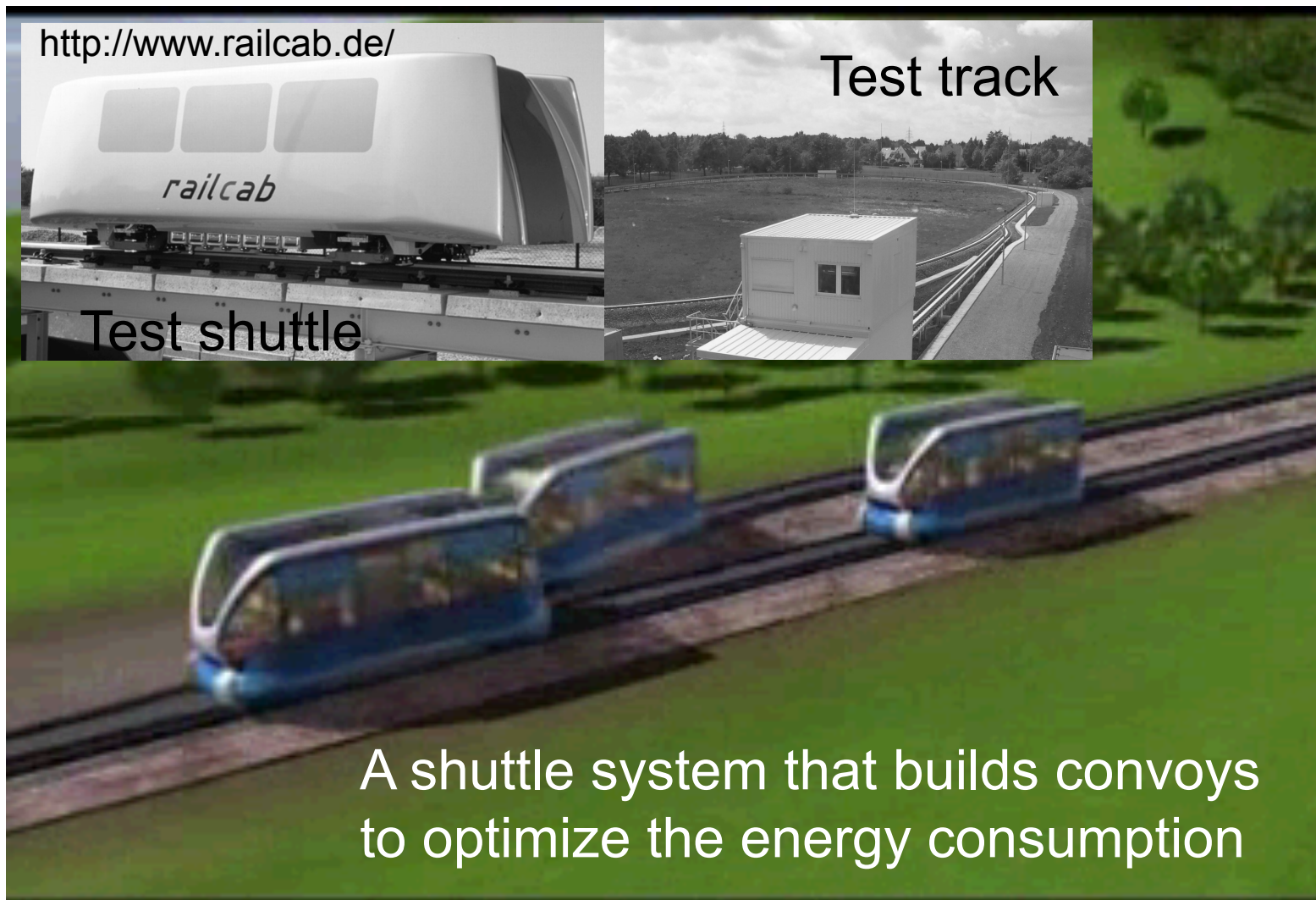
Adaptation is needed to compensate for changes in the mission requirements (...) and operating environments (..)

Northrop, Linda, et al. Ultra-Large-Scale Systems: The Software Challenge of the Future. Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, 2006.



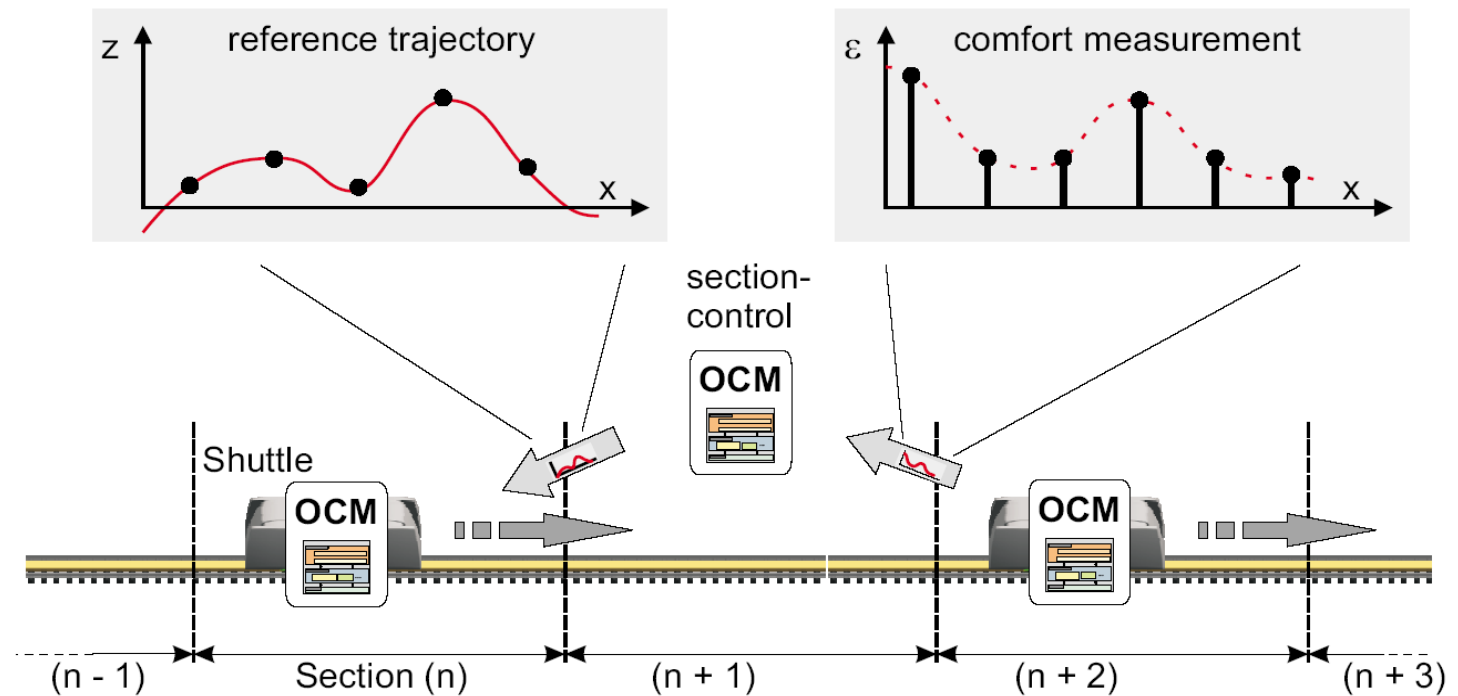
Example: Shuttle System

12



Example for Distributed Self-Consciousness/Adaptation

13



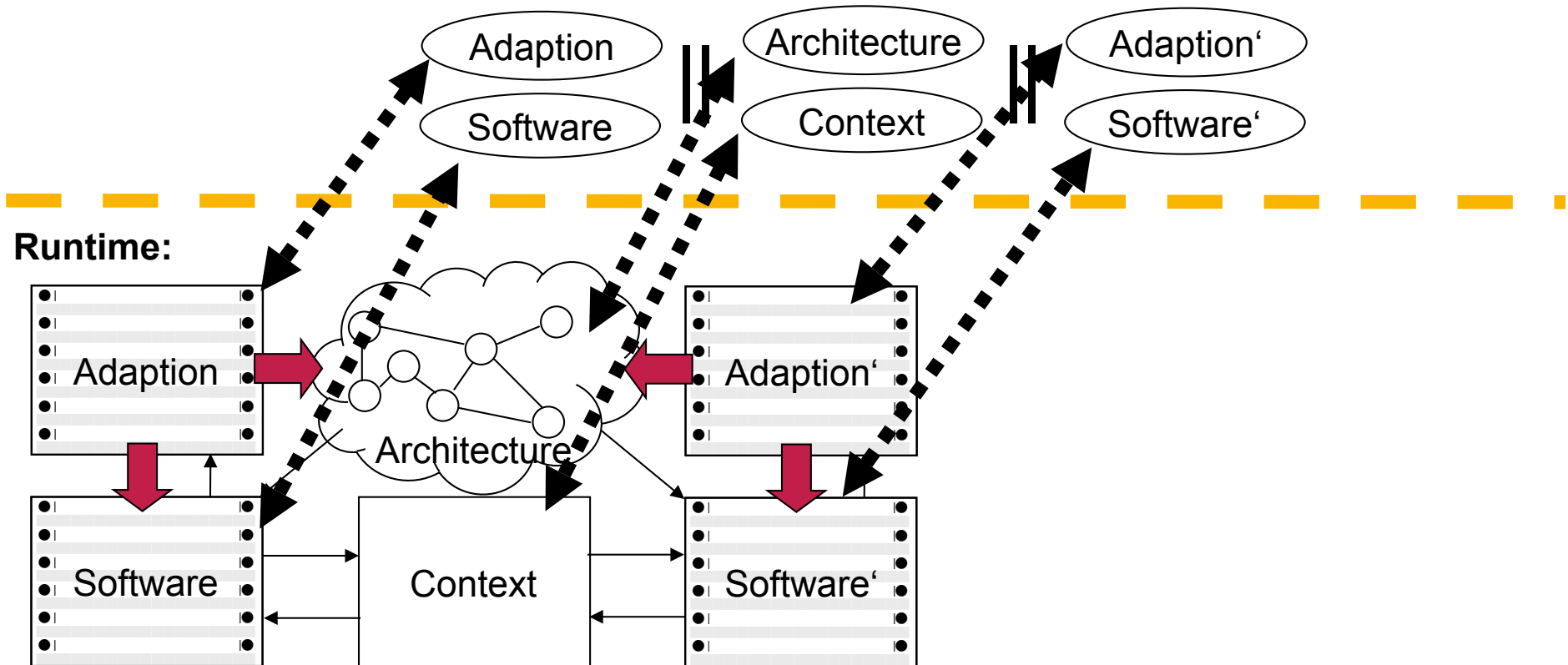
- Distributed learning of a model of the track (context)
- Local learning of a model of the shuttle (self!)
- Planning an adaptation in form of an optimal trajectory
- Trajectory synthesis establishes required assurance

Sven Burmester and Holger Giese and Eckehard Münch and Oliver Oberschelp and Florian Klein and Peter Scheideler, „Tool Support for the Design of Self-Optimizing Mechatronic Multi-Agent Systems, International Journal on Software Tools for Technology Transfer (STTT) 10 (3), 207-222, 2008.

Engineer Distributed Self-Consciousness

14

Development time:



Open Problems:

- (1) Multiple possibly conflicting adaptation loops
- (2) Decentralized organization required => self-organization

Proposed Priorities for the Call

15

- **Cost-Effective Engineering of Systems with Self-Awareness & Self-Consciousness**
 - Forward Engineering via frameworks and models at runtime
 - Reengineering with black-box and white-box models
 - Co-existence of different levels (online application management → self-awareness → self-consciousness)
 - Assurance
- **Engineering of Systems of Systems with Self-Awareness & Self-Consciousness**
 - Decentralized self-consciousness & adaptation
 - Overcoming the heterogeneity
 - Decentralized assurance