Cognitive Integration, Spring 2016

Introduction to the MicroPsi Architecture

Joscha Bach
joscha@mit.edu

Overview

In this introduction:

- Requirements for AGI architectures
- Representation and symbol grounding
- Configurable cognition: understanding emotion
- Motivation and autonomy

MicroPsi Architecture

- Origins: Psi theory (Dörner et al. 1999, 2002)

Dietrich Dörner
Universität Bamberg

MicroPsi Architecture

- Origins: Psi theory (Dörner et al. 1999, 2002)
- MicroPsi 1 (2003–2007):
  - Framework for simulating cognitive agents
  - Explorative agents in virtual worlds
  - Evolutionary simulations
  - Learning and classification
  - Robot control
- MicroPsi 2 (2012–current):
  - new simulation framework
  - adaptation for knowledge representation

Basic premises

- Are minds characterized by a single organizing principle, or by a complex set of very particular constraints?

→Human minds are solution to a very specific class of control problems, with generally applicable intelligence as a by-product

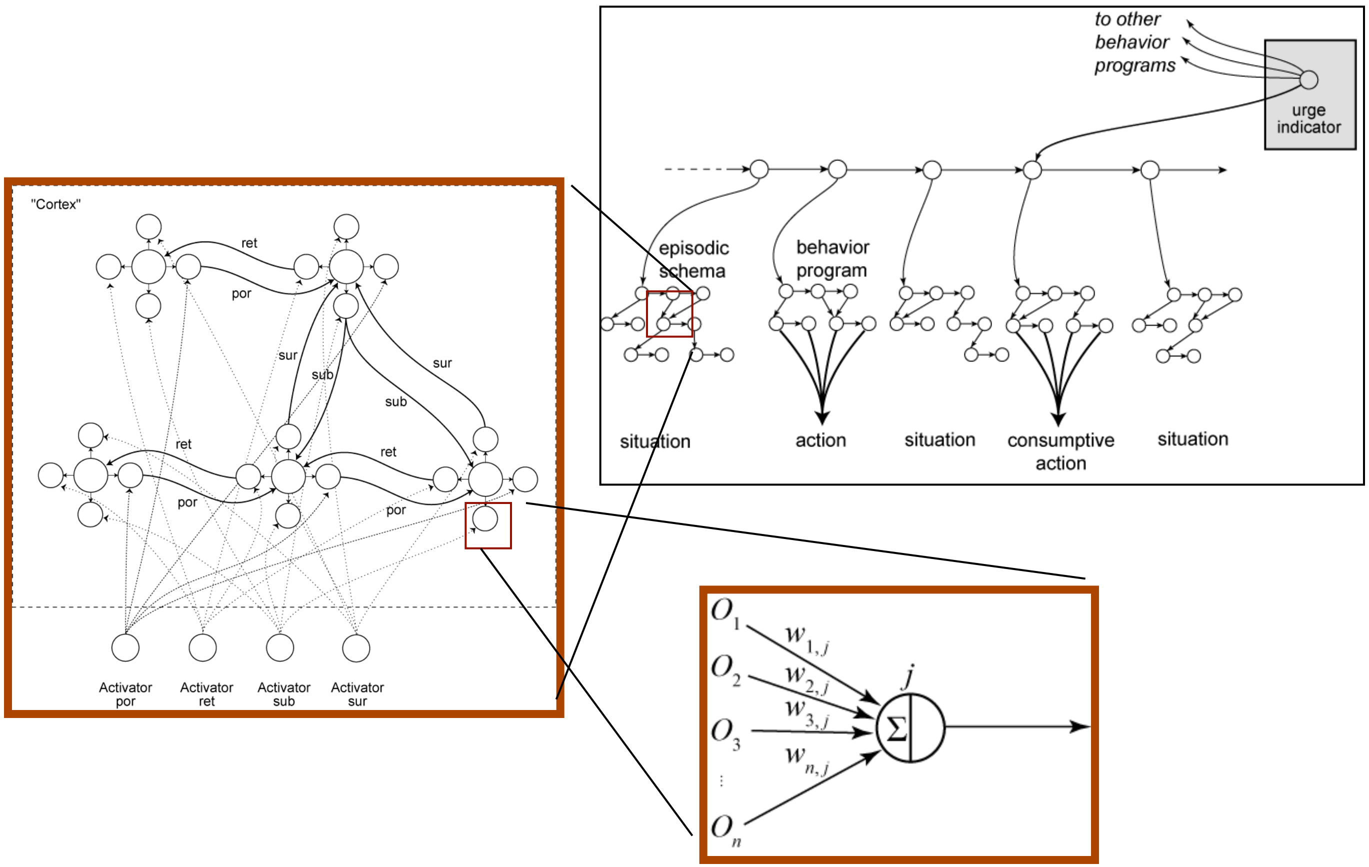
Cognitive Artificial Intelligence

Methods should focus on components and performances necessary for intelligence:

- **Whole, testable architectures**
- **Universal Representations:**
  Grounded neuro-symbolic representations (integrate both symbolic and distributed aspects)
- **(Semi-) Universal Problem Solving:**
  Learning, Planning, Reasoning, Analogies, Action Control, Reflection ...
- **Universal Motivation:**
  Polythematic, adaptive goal identification
- **Emotion and affect**

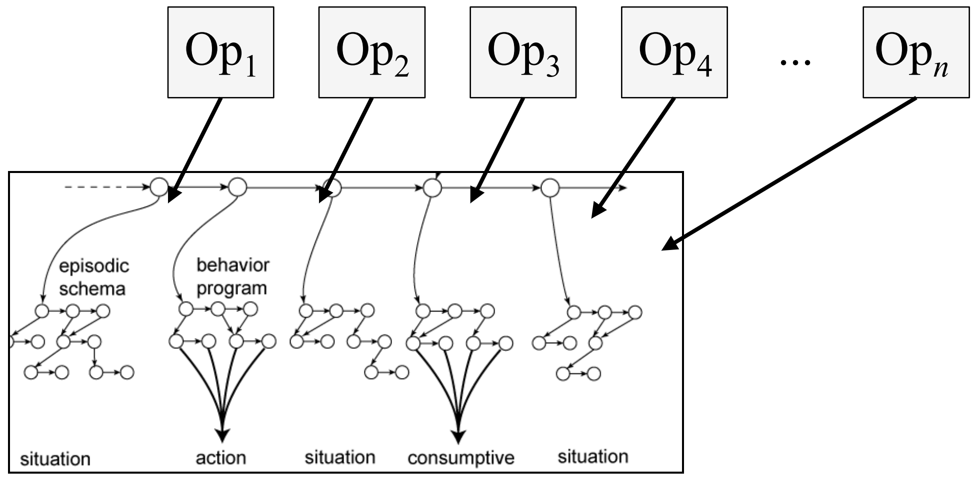
Components for Cognitive AI

- Universal mental representations
  (compositional + distributed → neurosymbolic)



Components for Cognitive AI

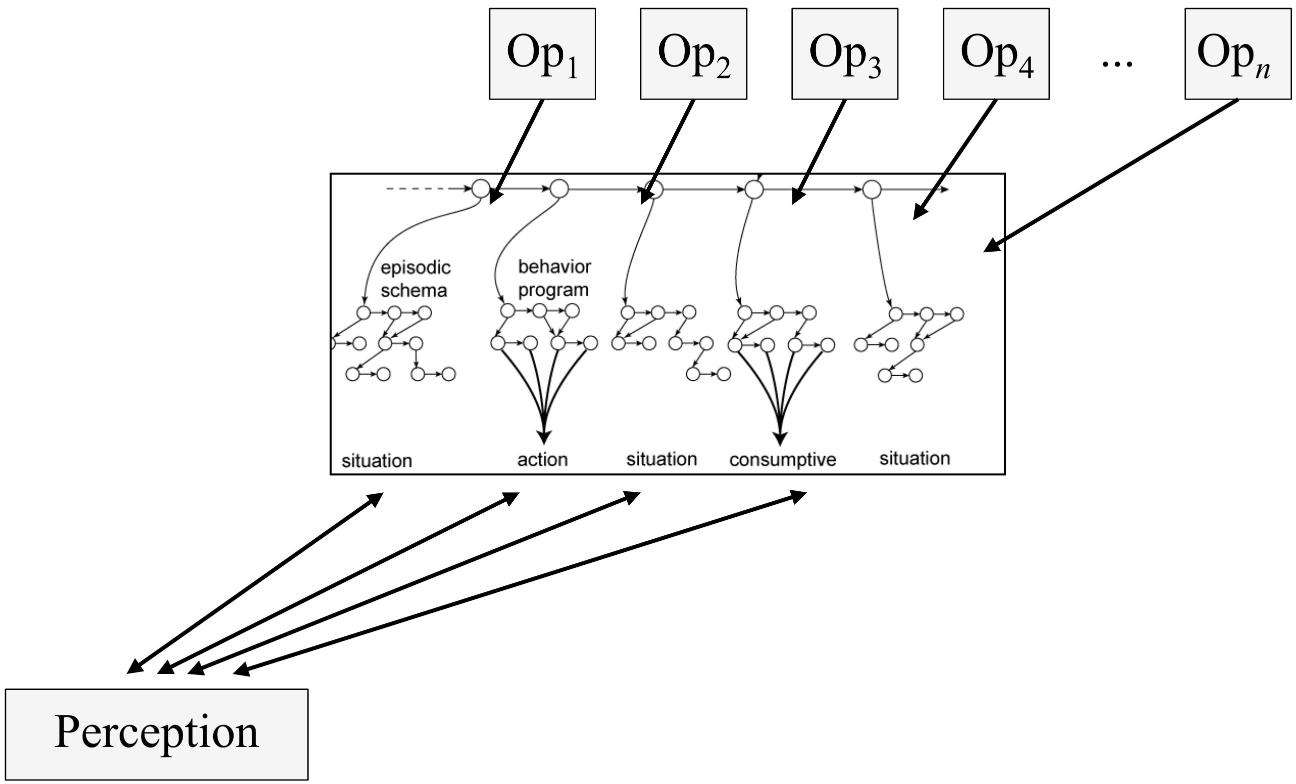
- (Semi-) General problem solving: Operations over these representations



(neural learning, categorization, planning, reflection, consolidation, ...)

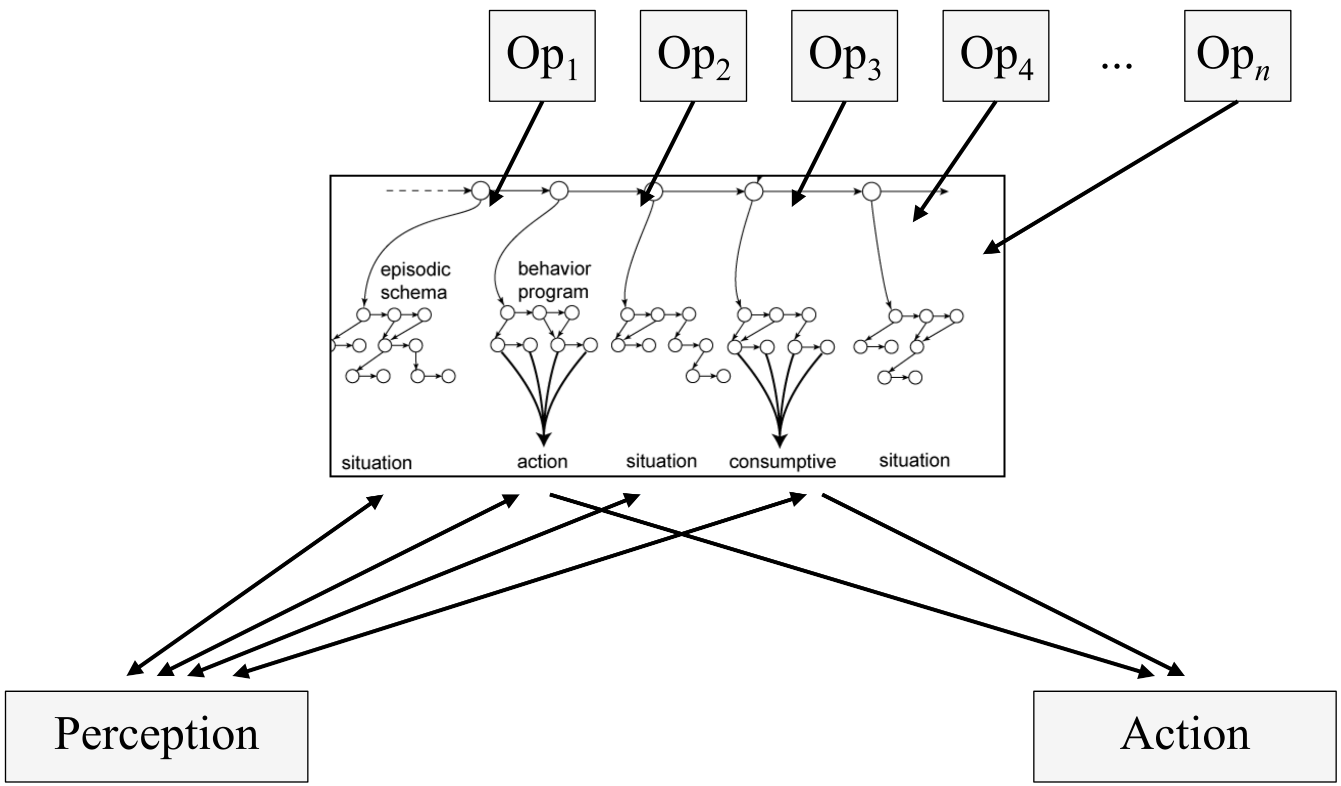
Components for Cognitive AI

- Perceptual grounding



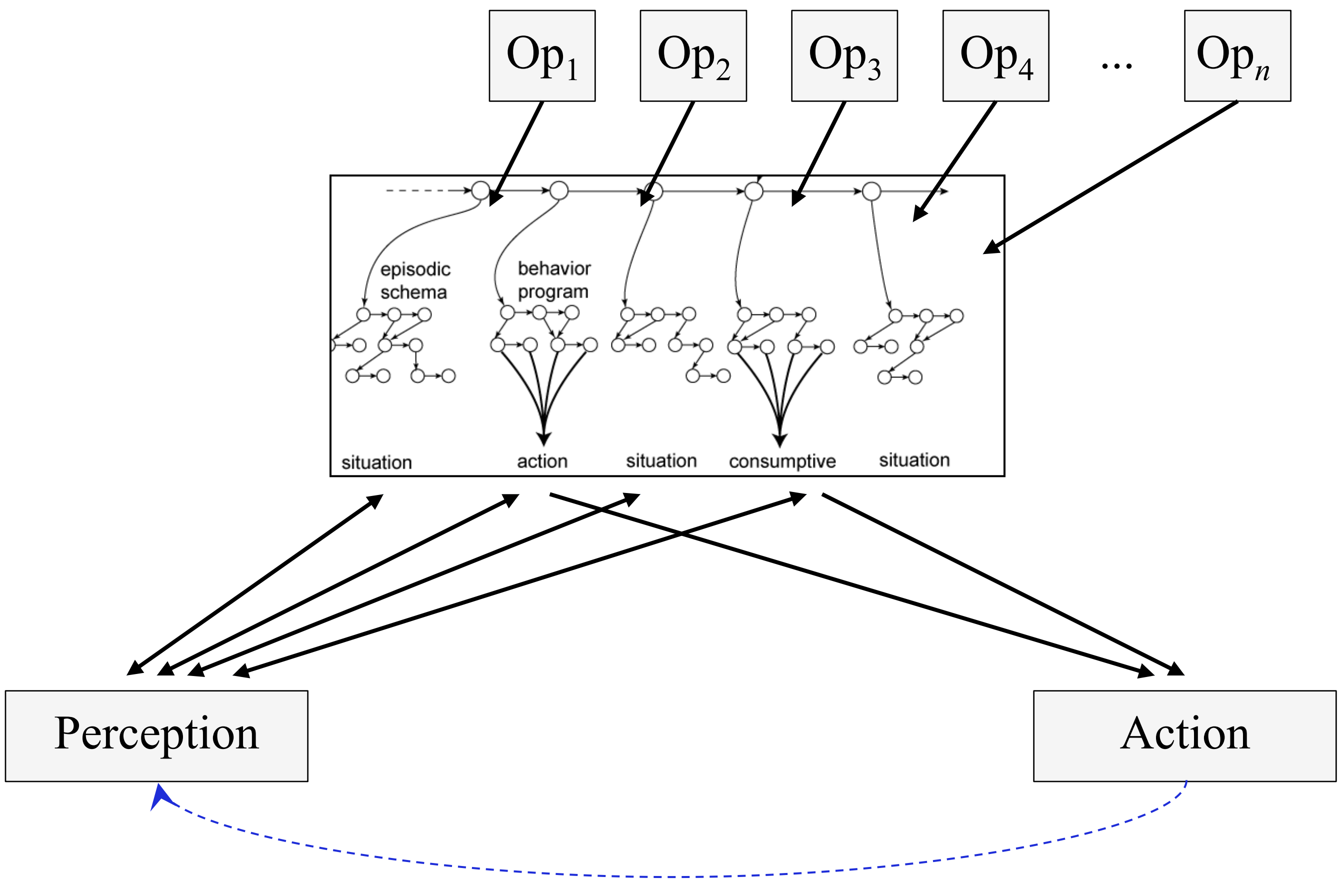
Components for Cognitive AI

- Perceptual grounding and action



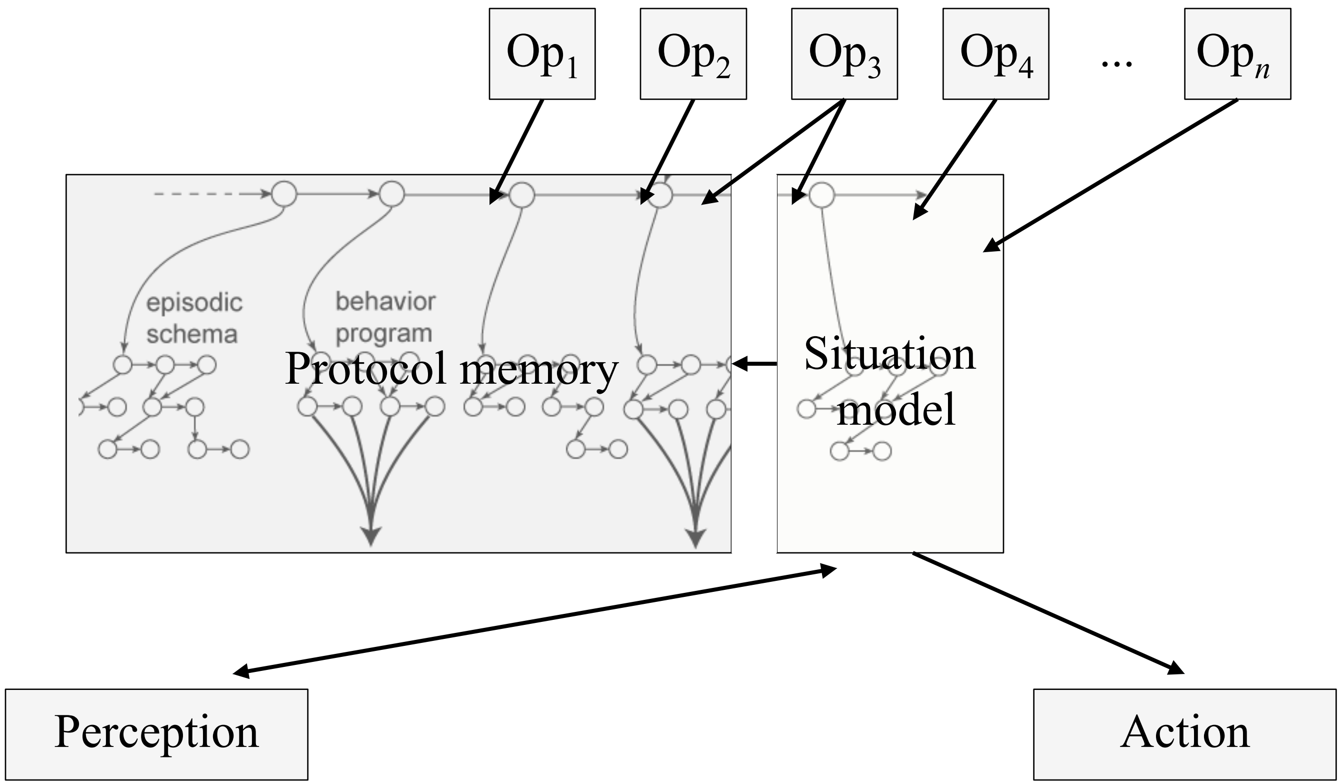
Components for Cognitive AI

- Perceptual grounding and action



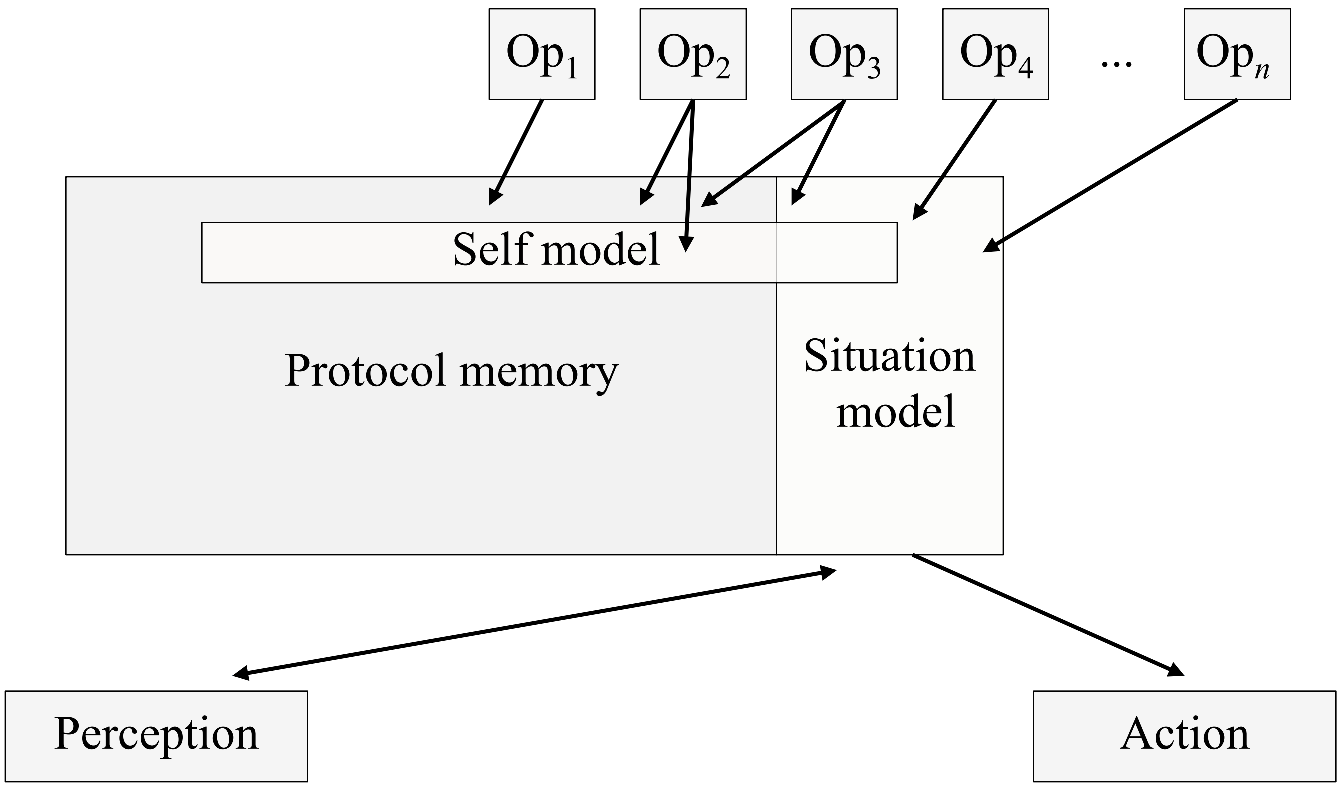
Components for Cognitive AI

- Model of current situation, and protocol of past situations



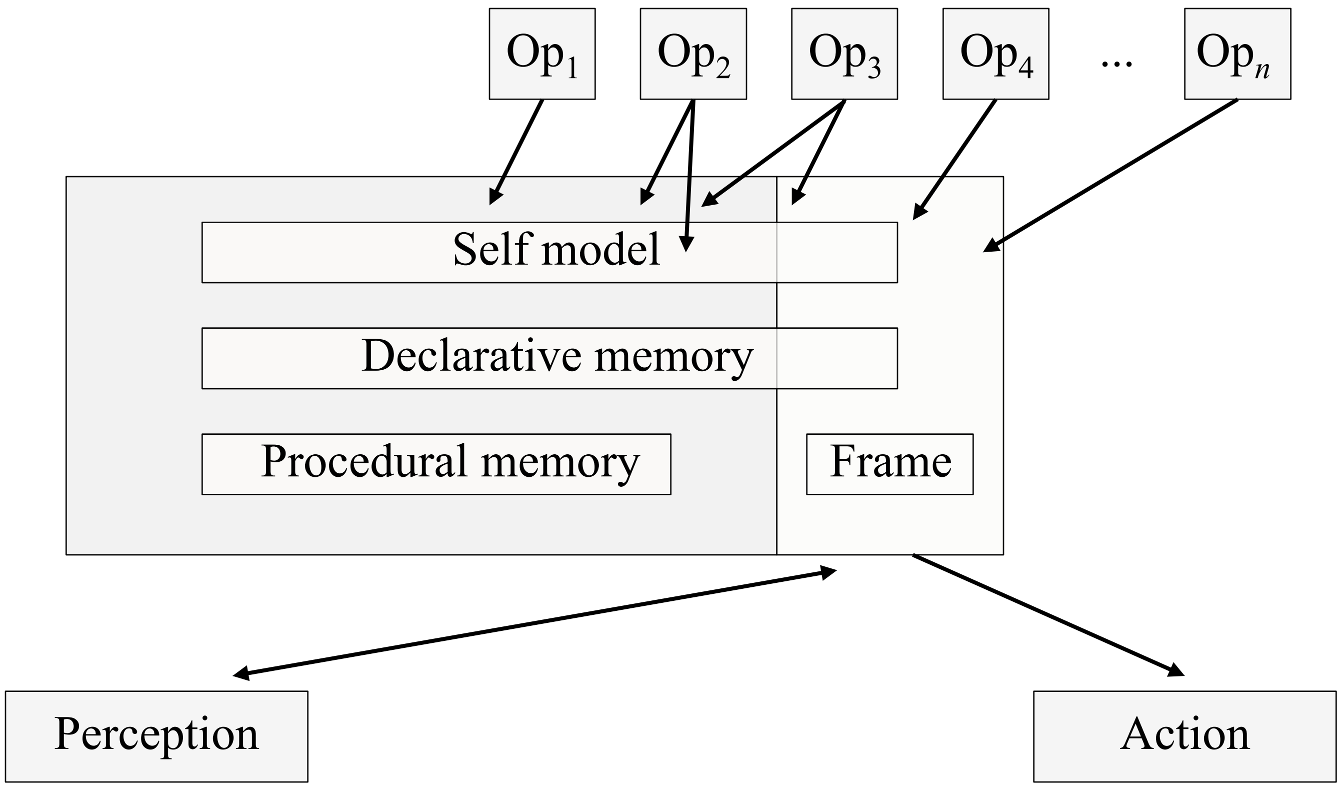
Components for Cognitive AI

- Model of self



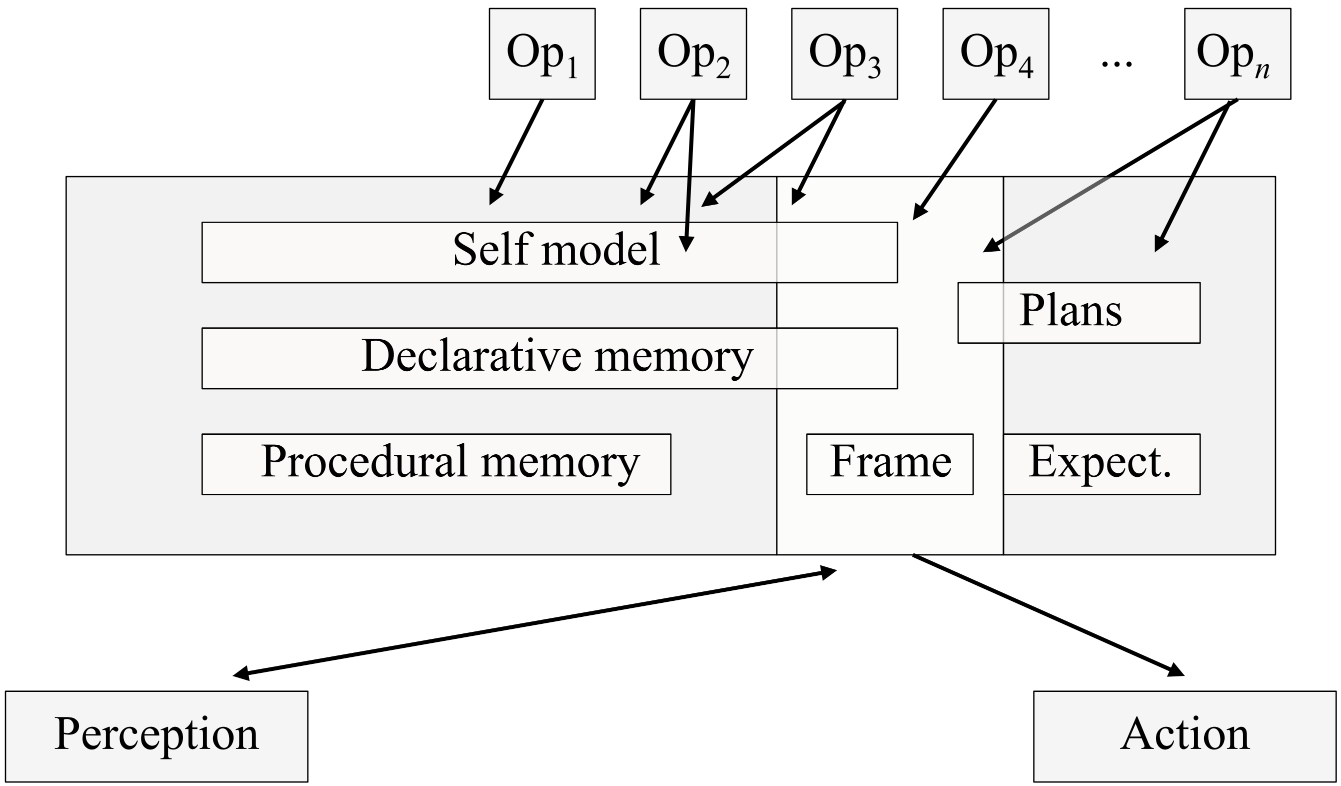
Components for Cognitive AI

- Abstractions of objects, episodes and types



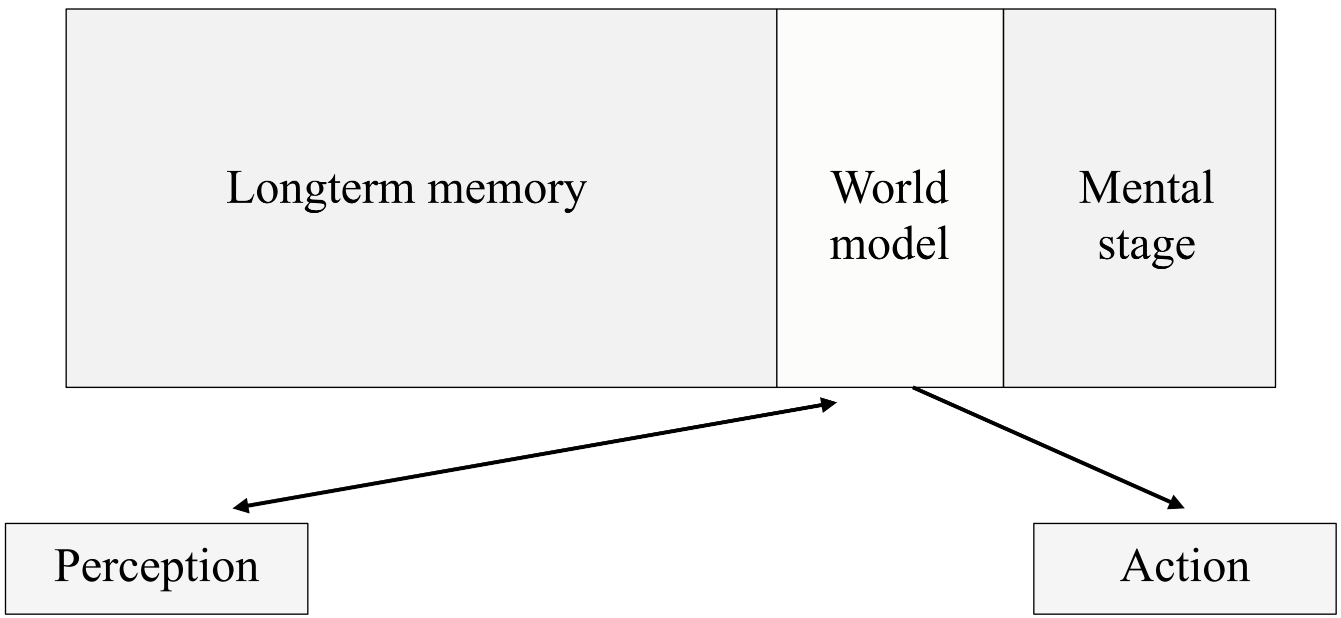
Components for Cognitive AI

- Anticipation of future developments



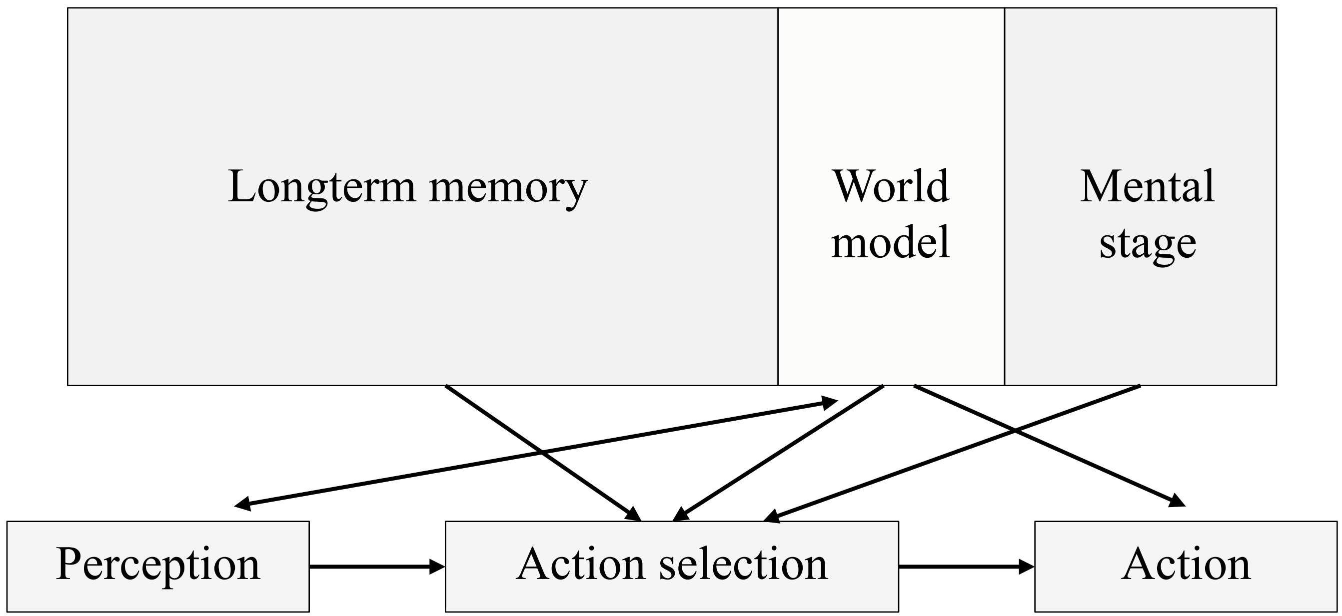
Components for Cognitive AI

- Action selection and executive control



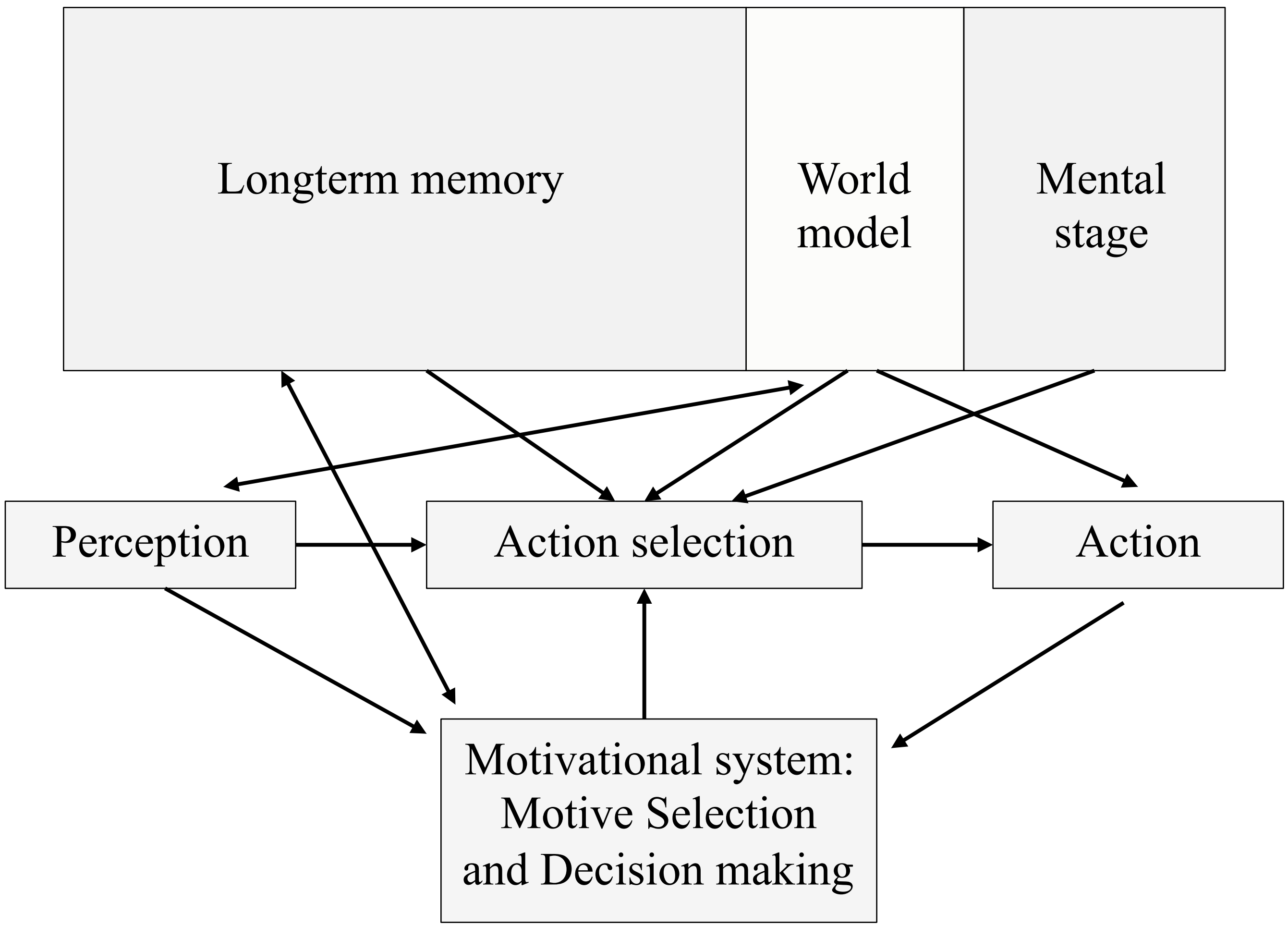
Components for Cognitive AI

- Action selection and executive control



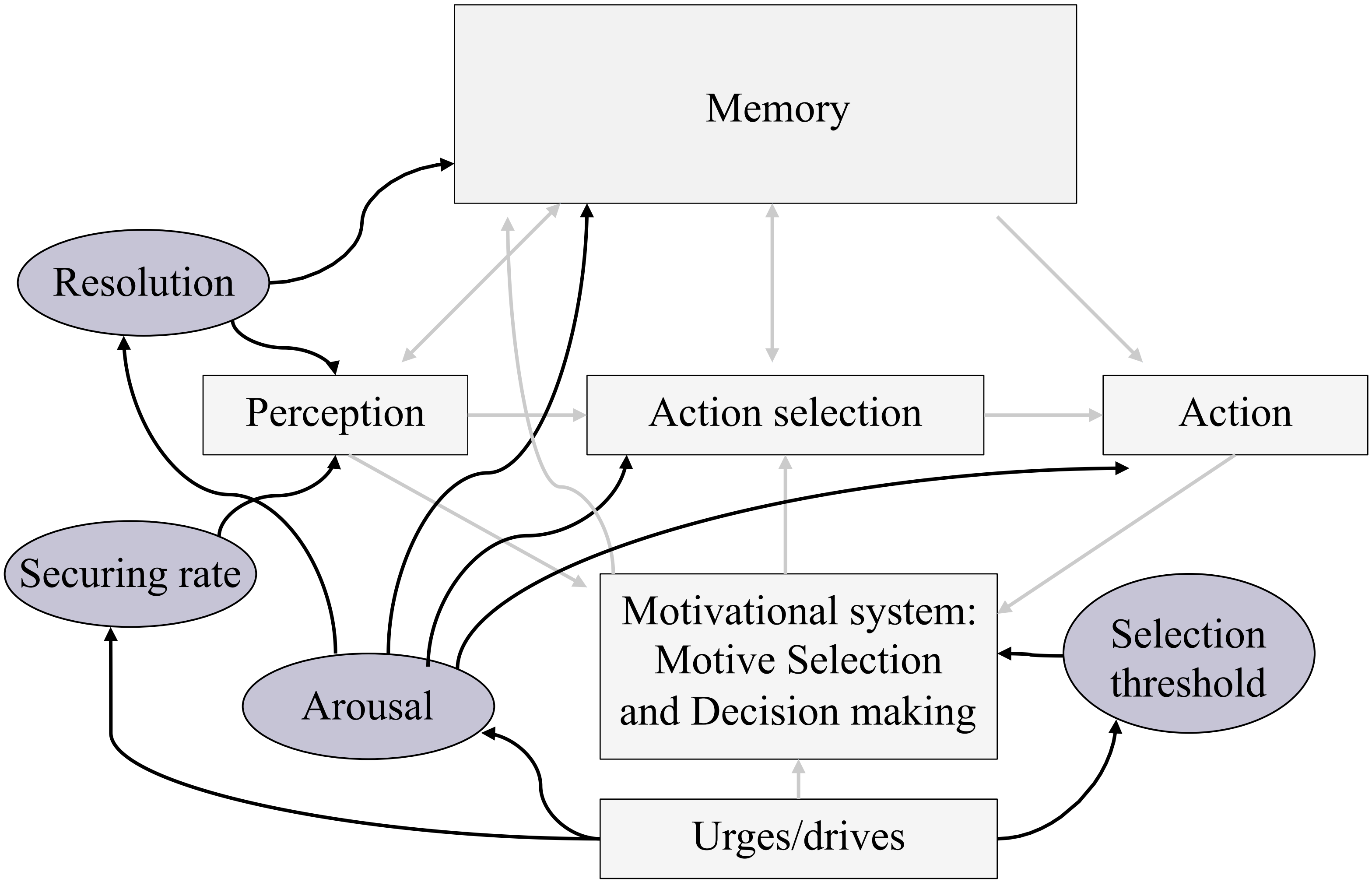
Components for Cognitive AI

- Universal motivation: autonomous identification of goals



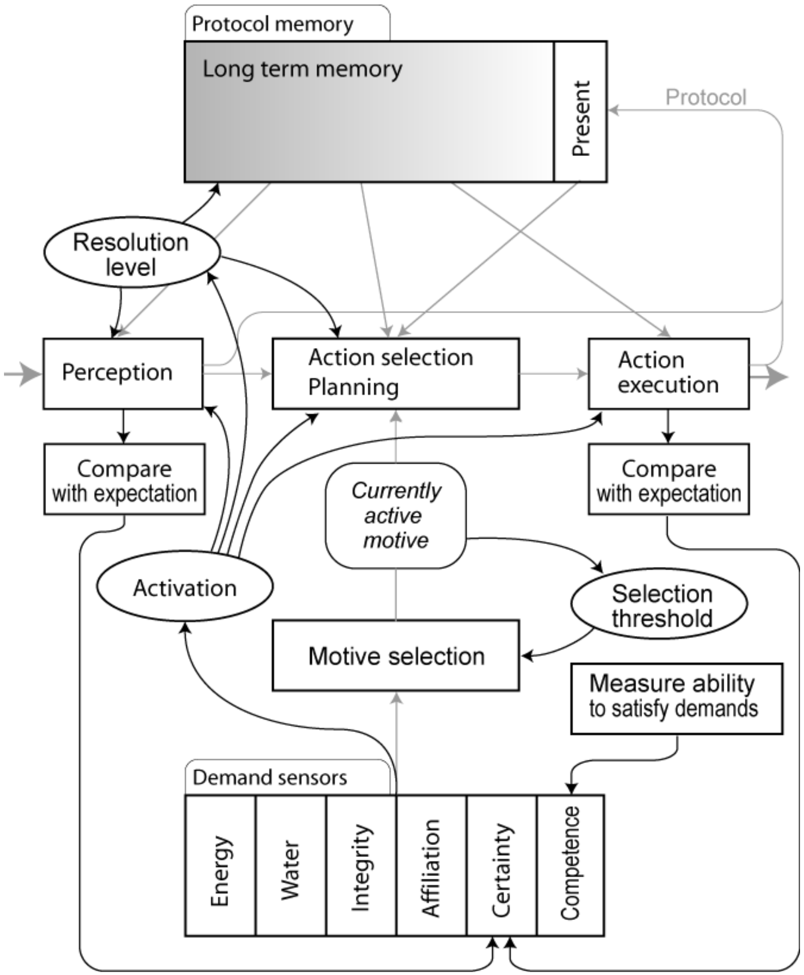
Components for Cognitive AI

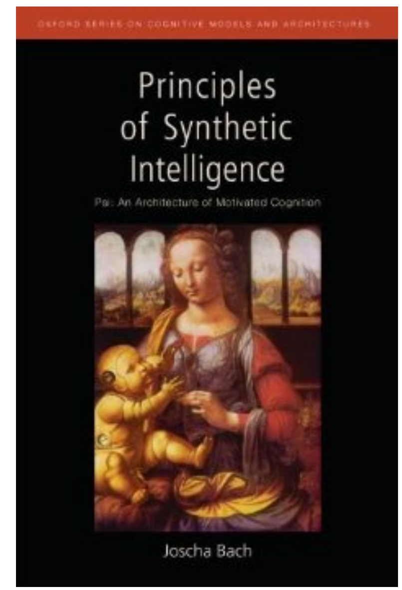
- Emotional modulation and affect



Components for Cognitive AI

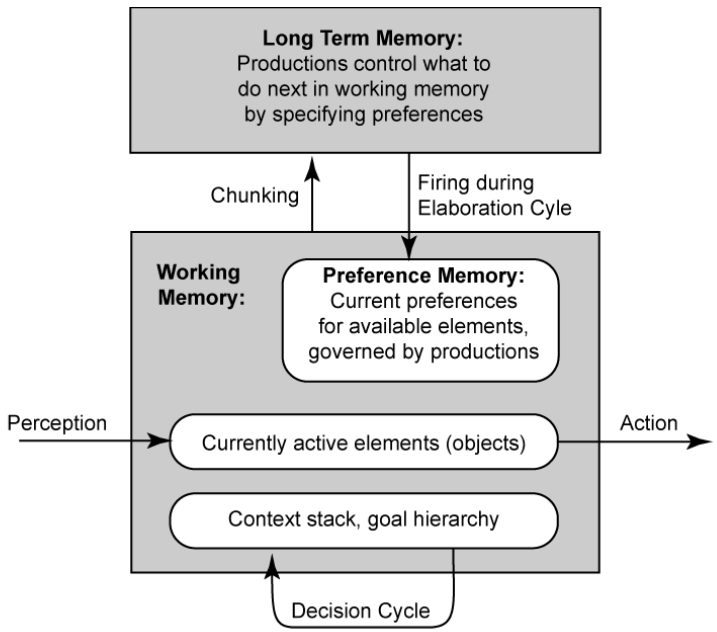
- Whole, testable architectures



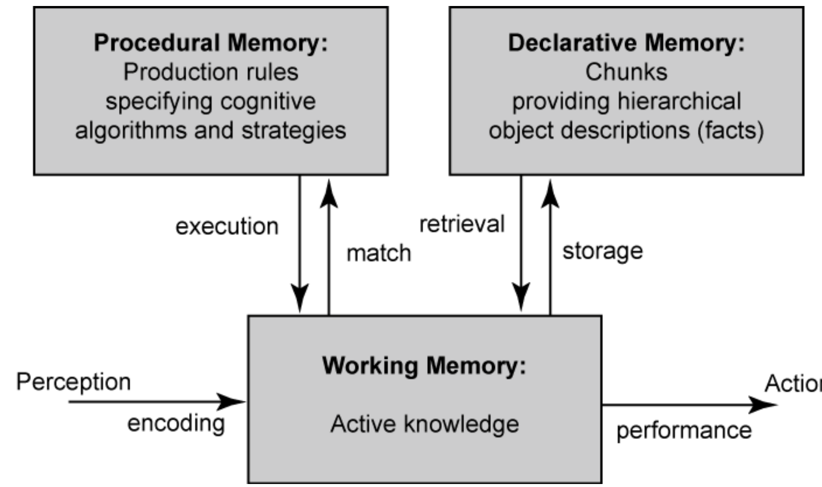


Principles of Synthetic Intelligence
(Bach 2003, 2009)

Cognitive Architectures



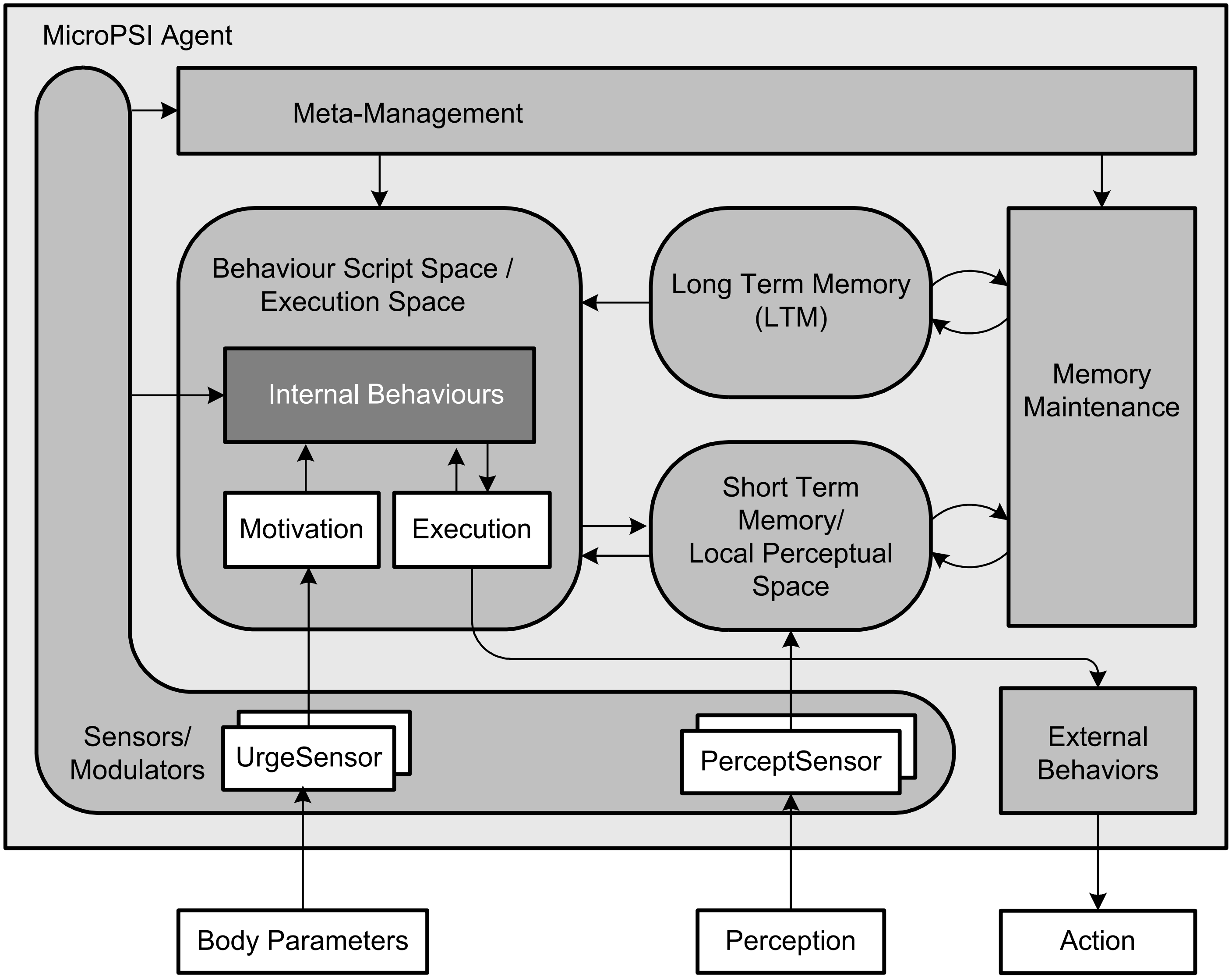
Soar



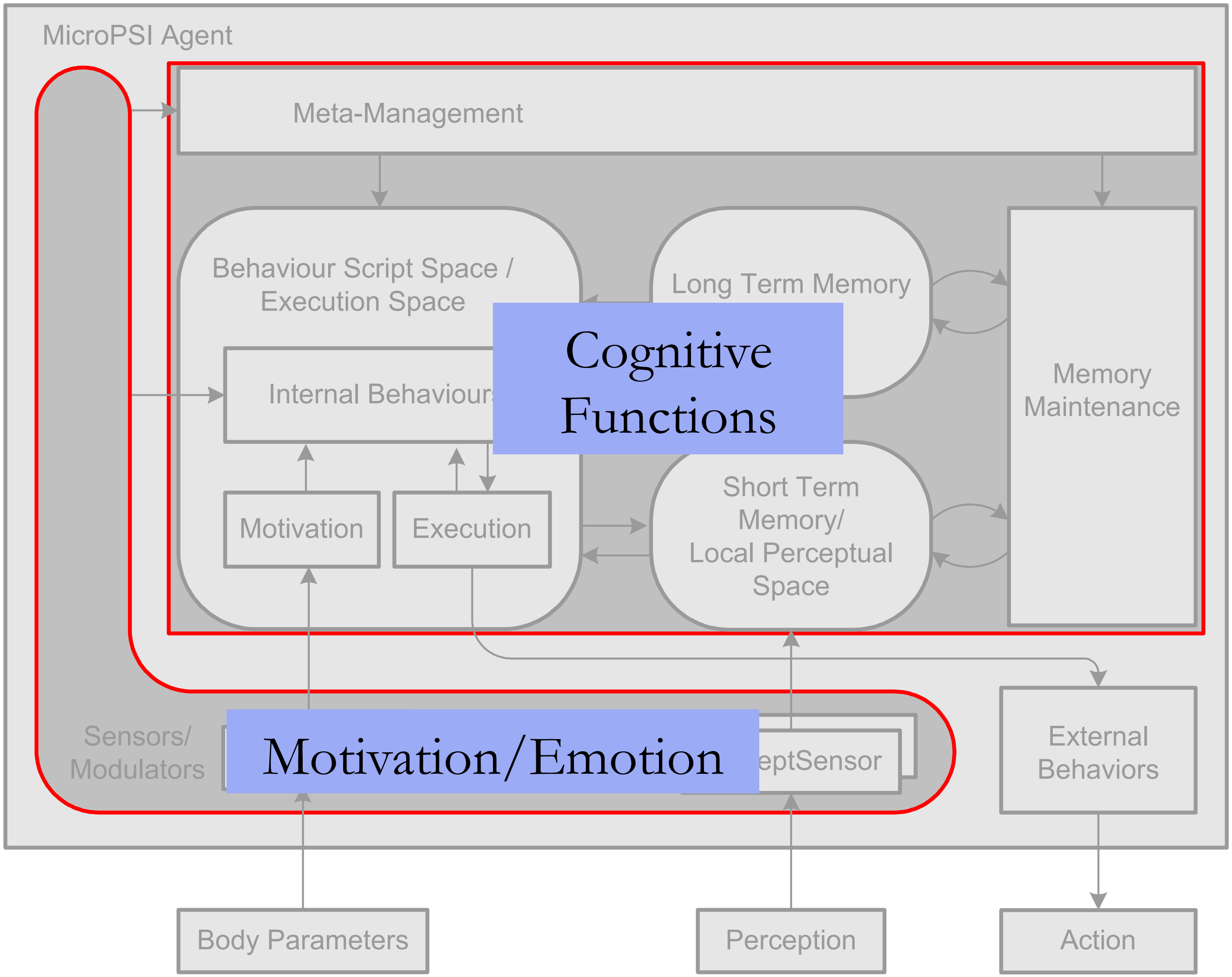
ACT-R

"Classical Cognitive Architectures" tend to focus on cognition as an isolated problem solving capability.

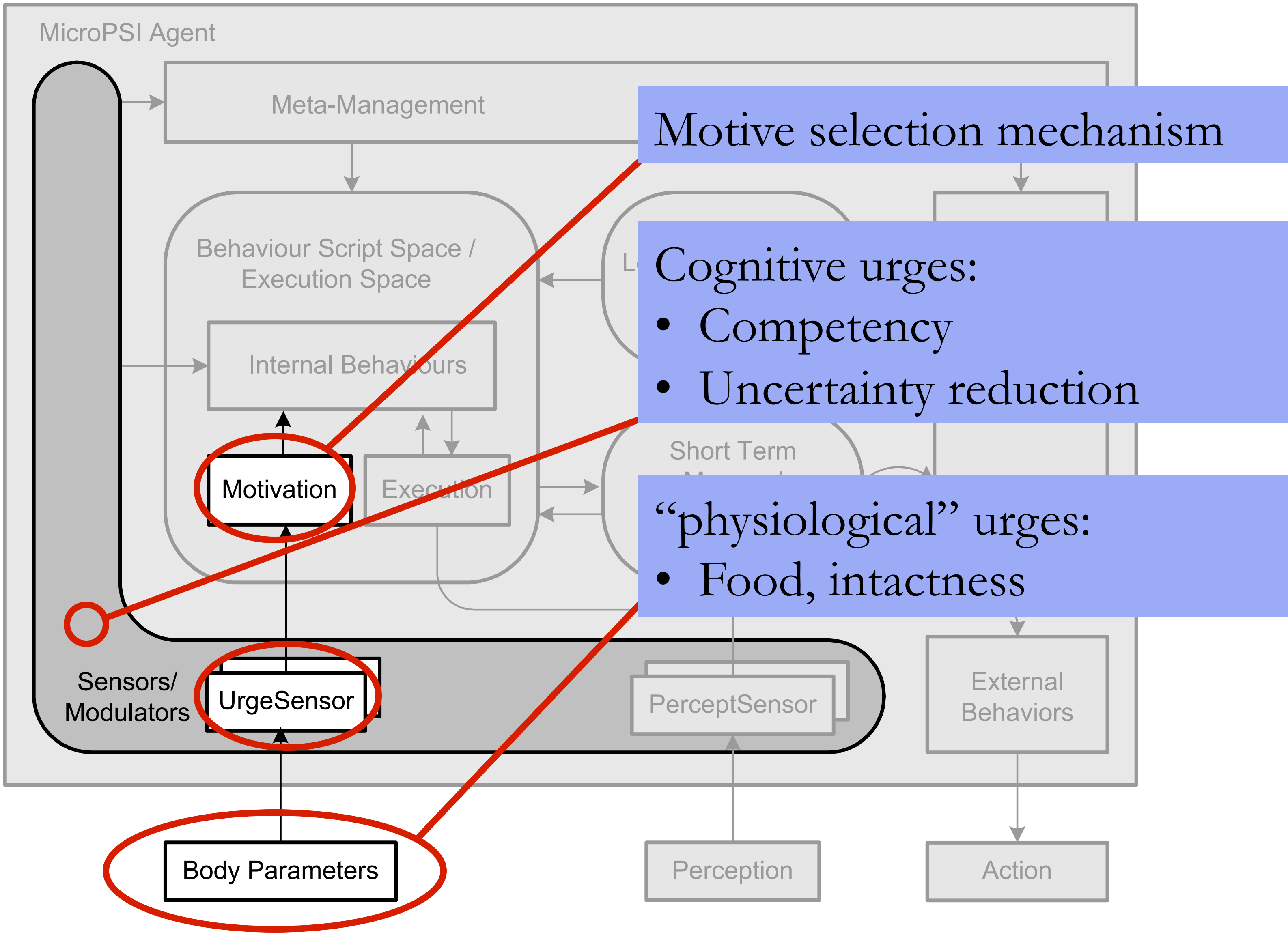
MicroPsi Architecture–simplified



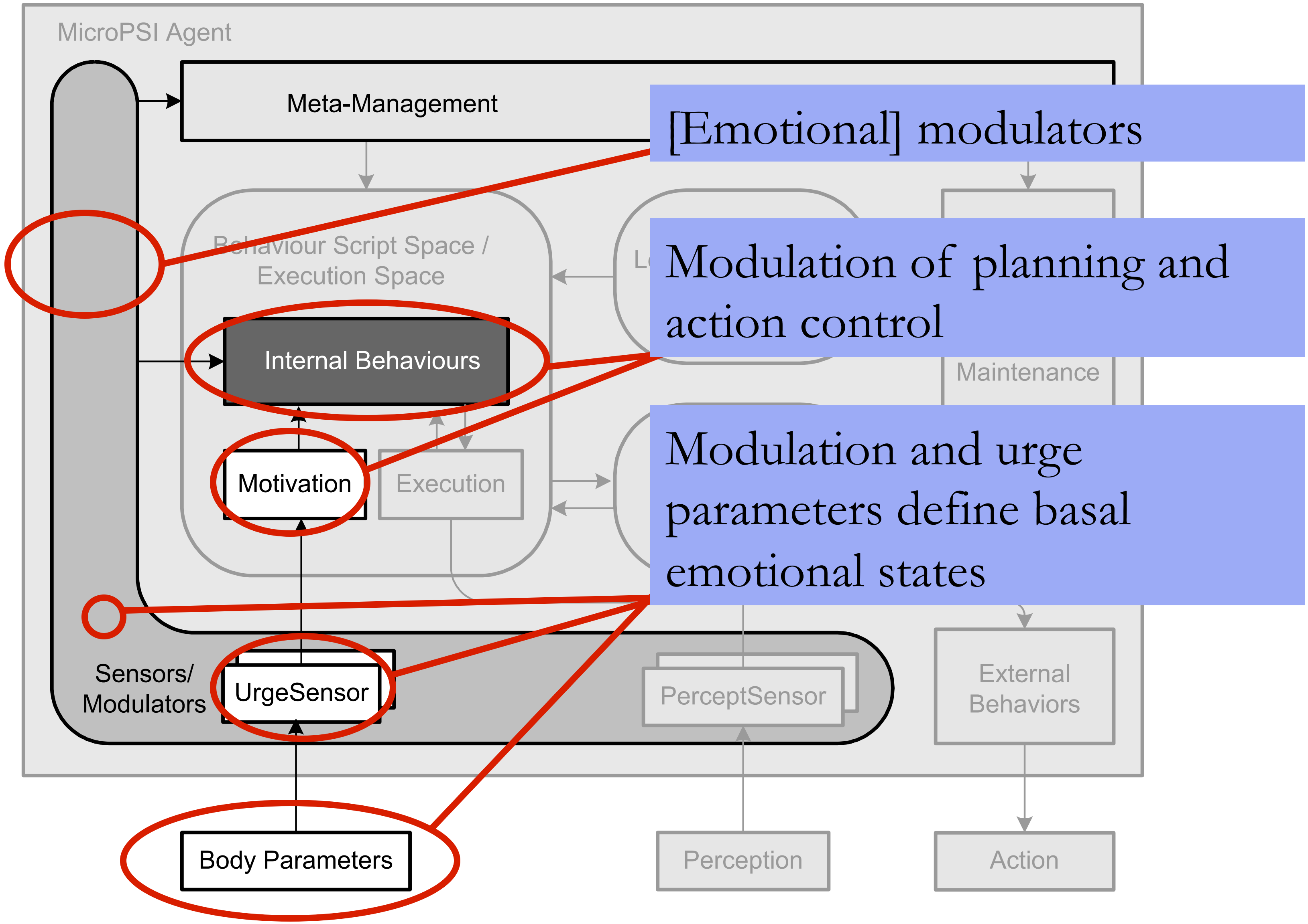
MicroPsi Architecture–simplified



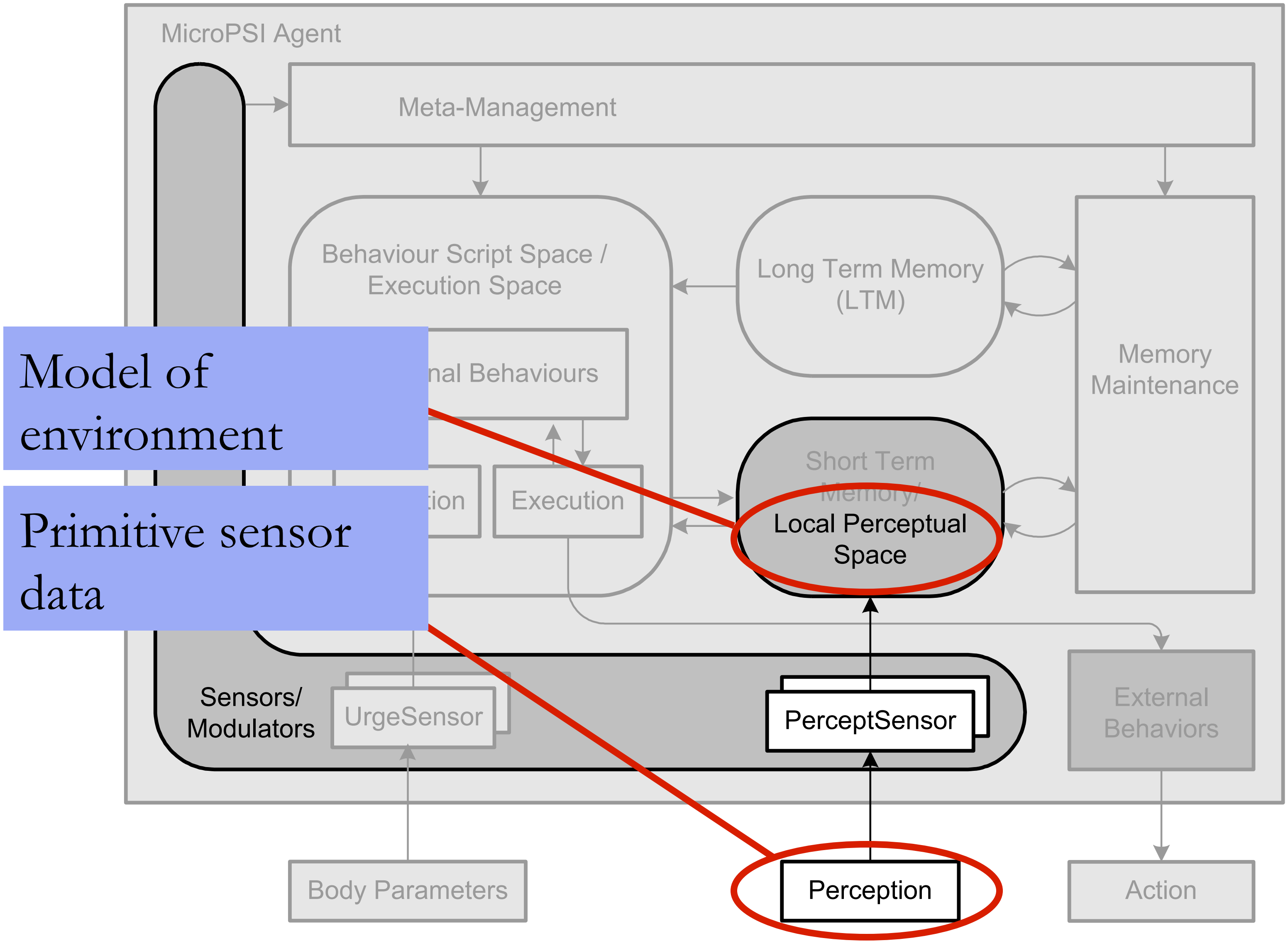
MicroPsi Architecture–simplified



MicroPsi Principles: Emotion



MicroPsi Principles: Perception

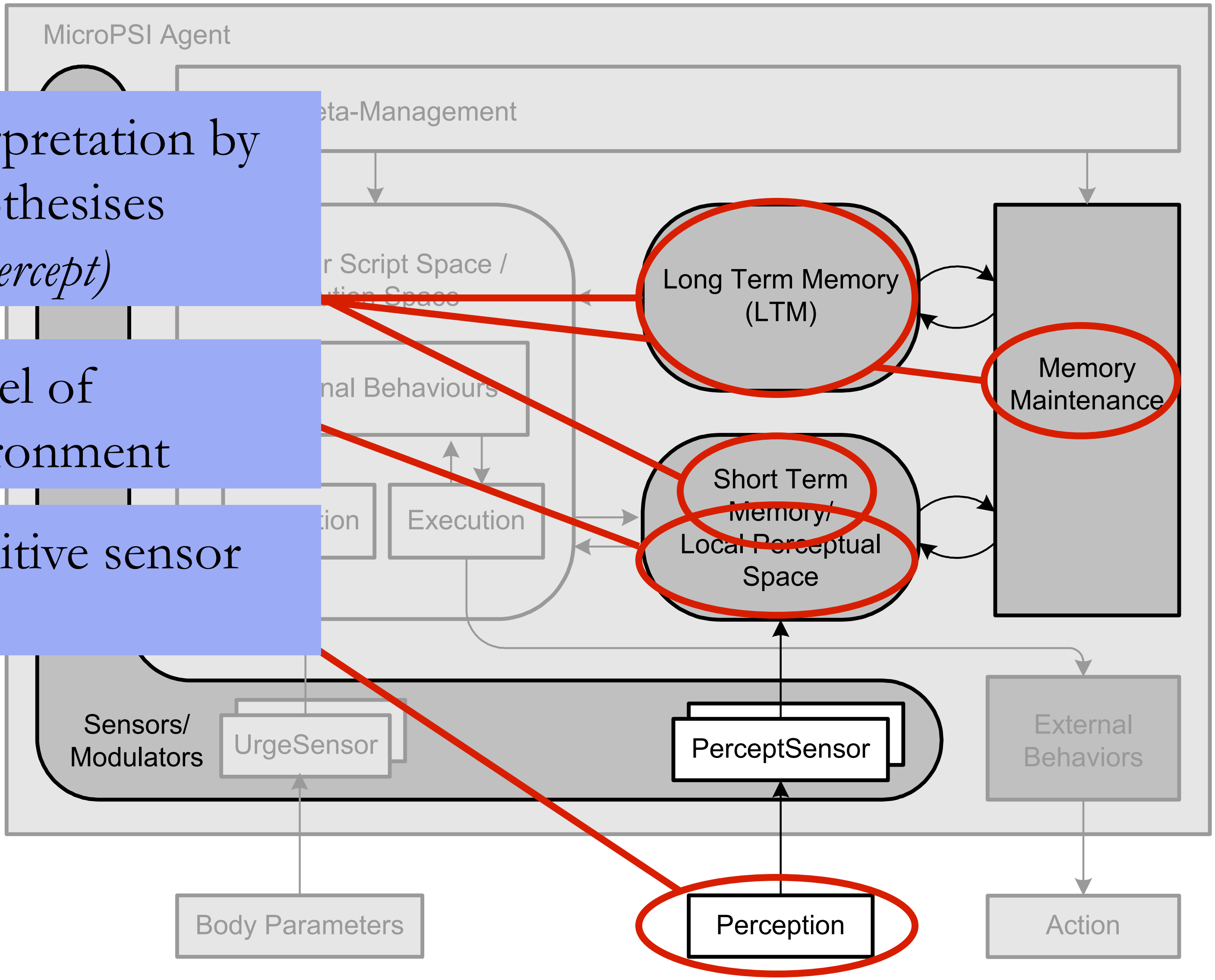


MicroPsi Principles: Perception

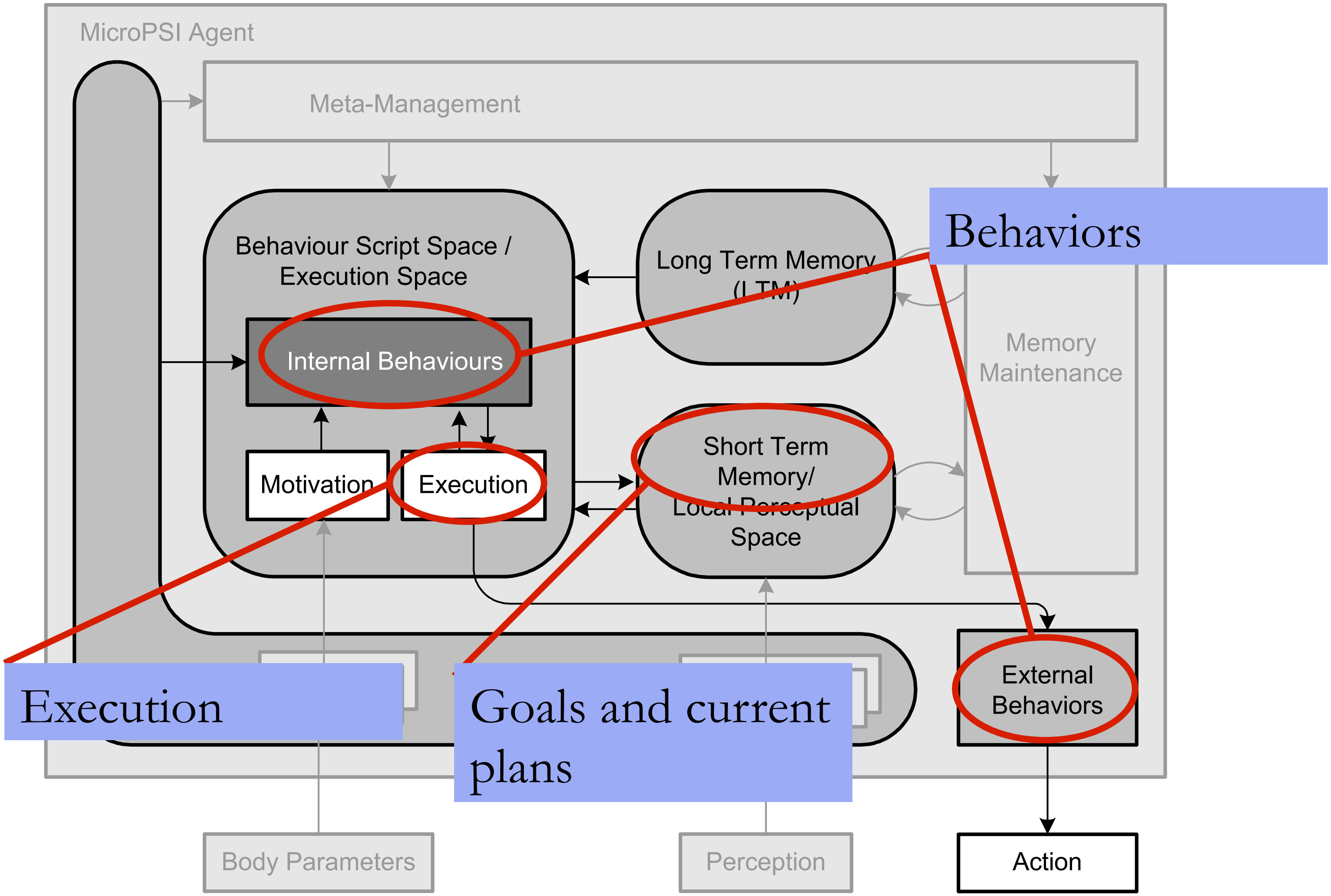
Interpretation by hypothesises *(Hypercept)*

Model of environment

Primitive sensor data



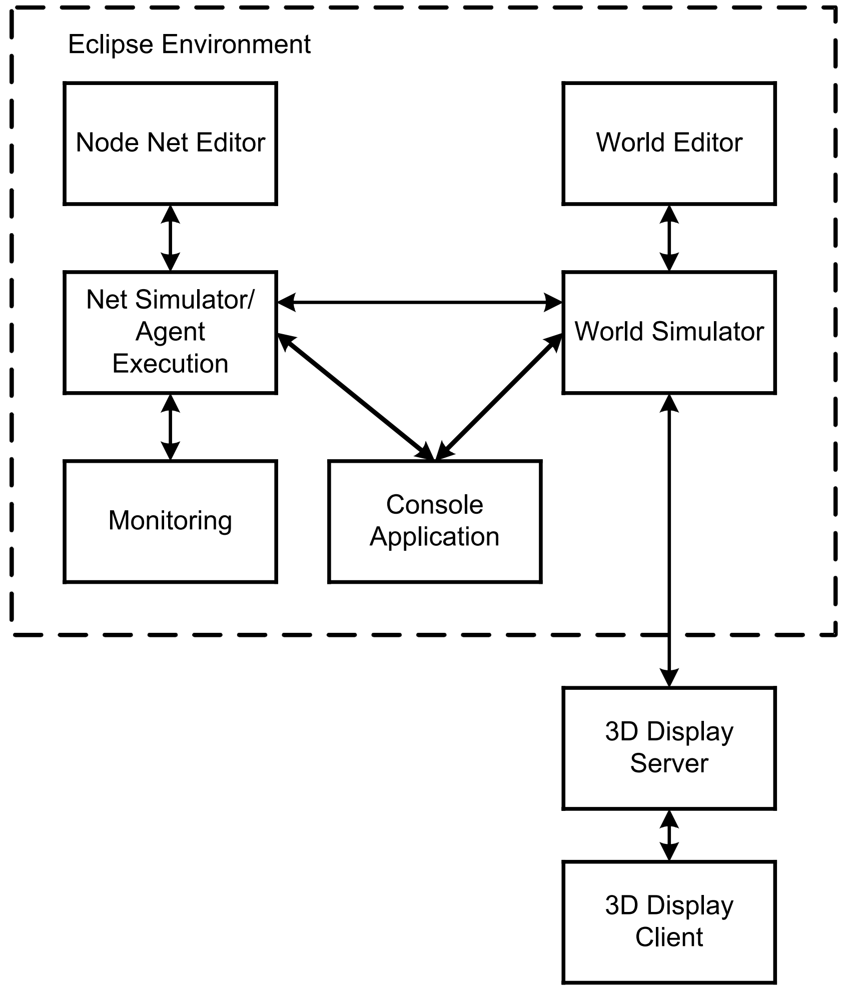
MicroPsi Principles: Action



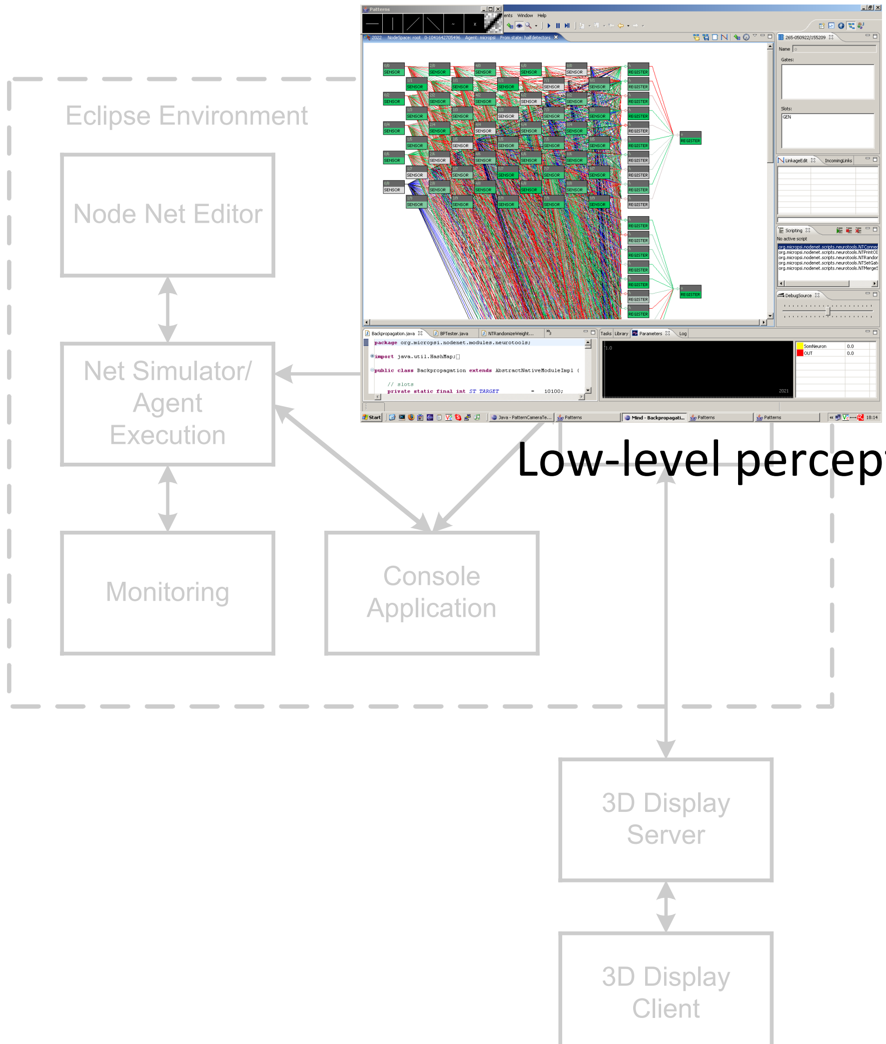
Agent Functionality

- Episodic learning
- Goal-directed behavior, motivational system
- Emotional modulation
- Hypothesis based perception
- Simple planning
- Execution of hierarchical plans

Implementation: MicroPsi (Bach 03, 05, 04, 06)

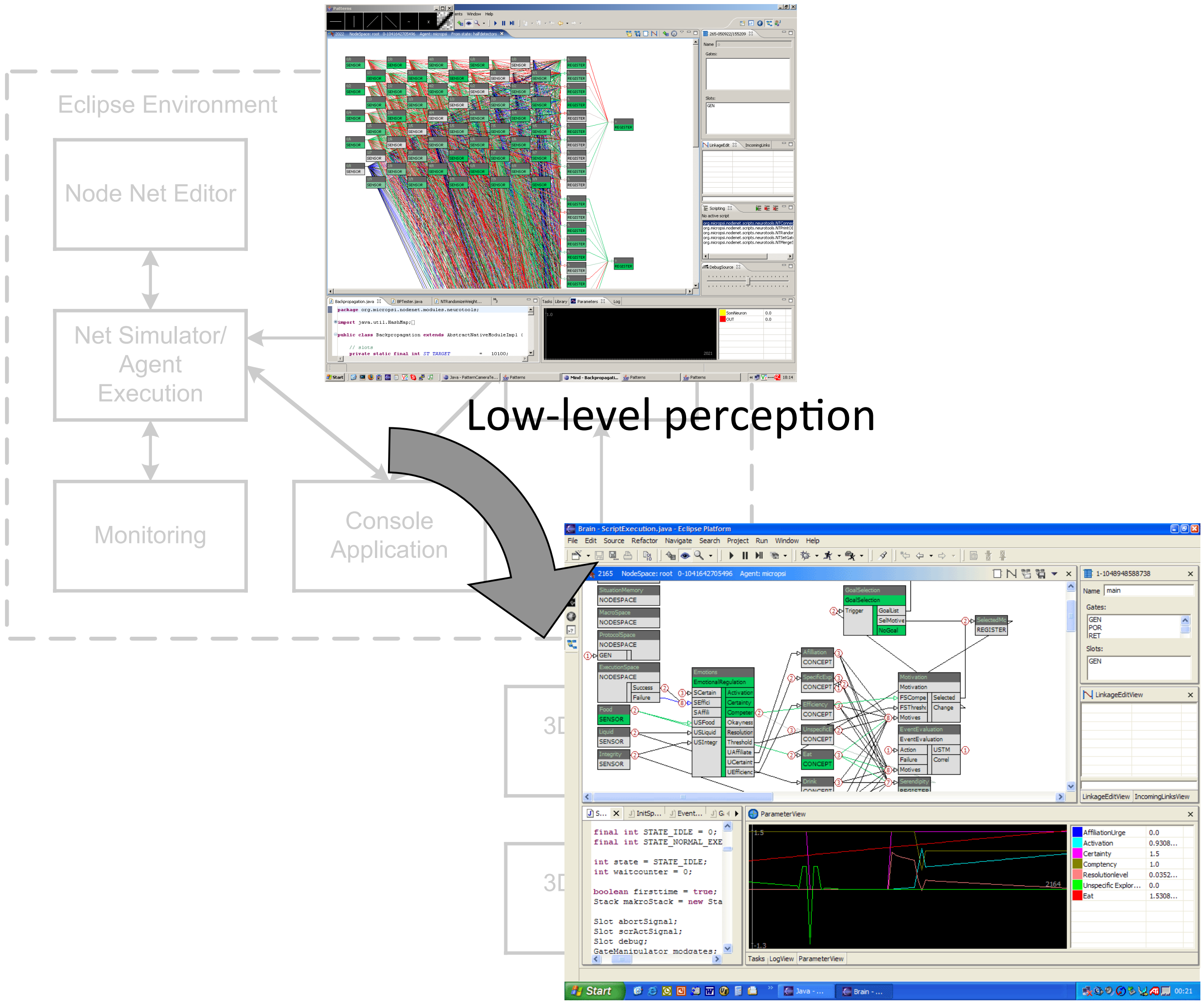


Implementation: MicroPsi (Bach 03, 05, 04, 06)



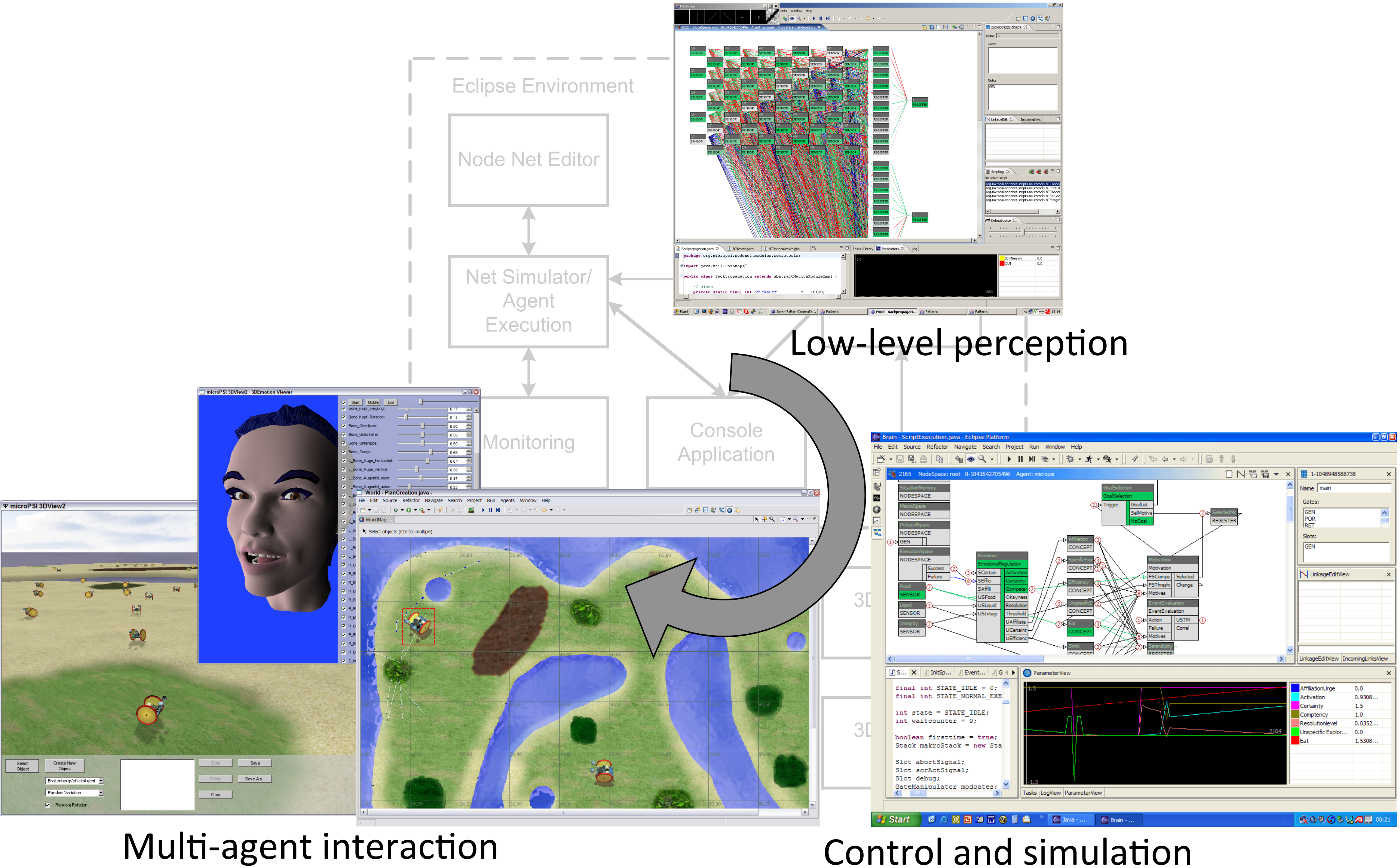
Low-level perception

Implementation: MicroPsi (Bach 03, 05, 04, 06)

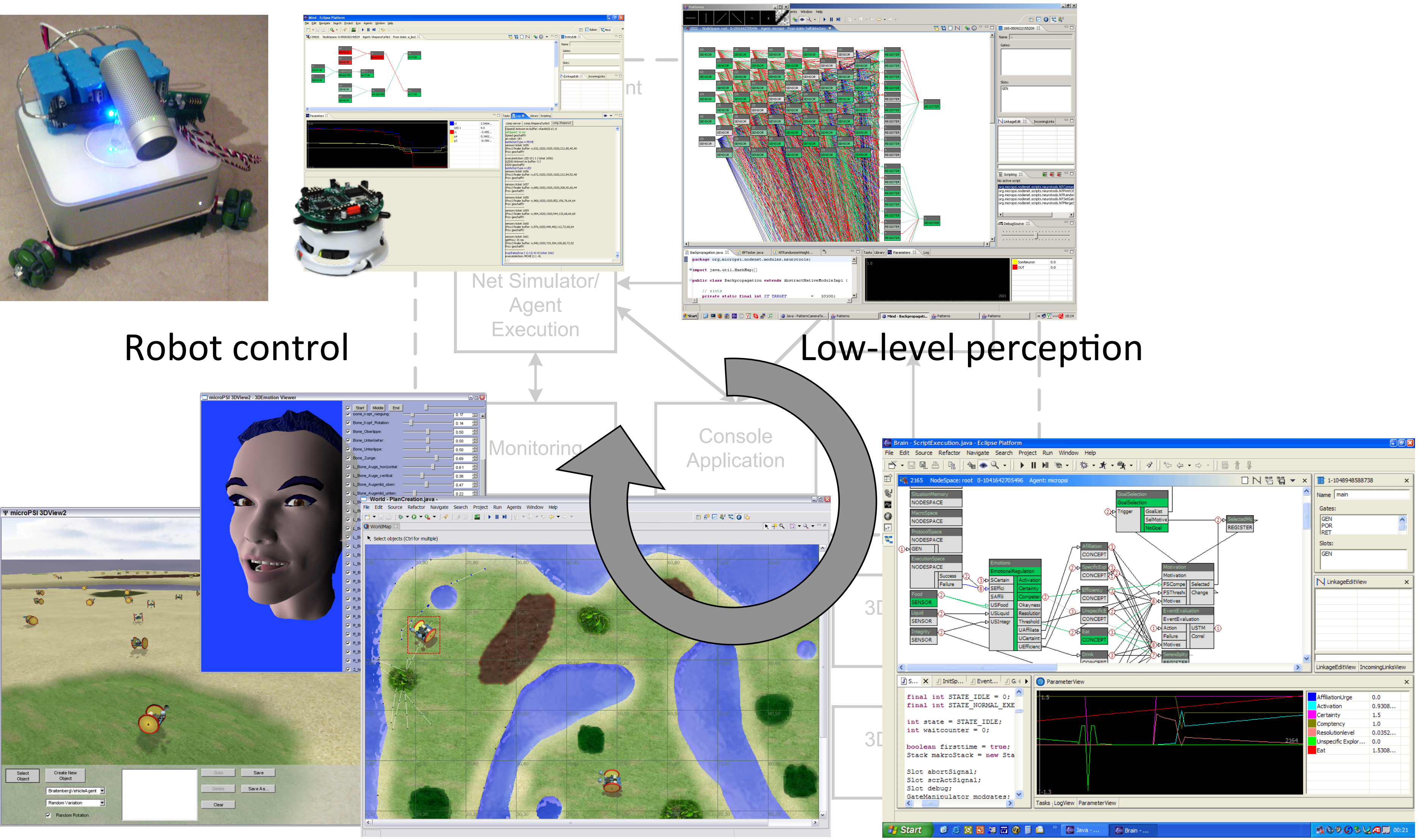


Control and simulation

Implementation: MicroPsi (Bach 03, 05, 04, 06)



Implementation: MicroPsi (Bach 03, 04, 05, 06)



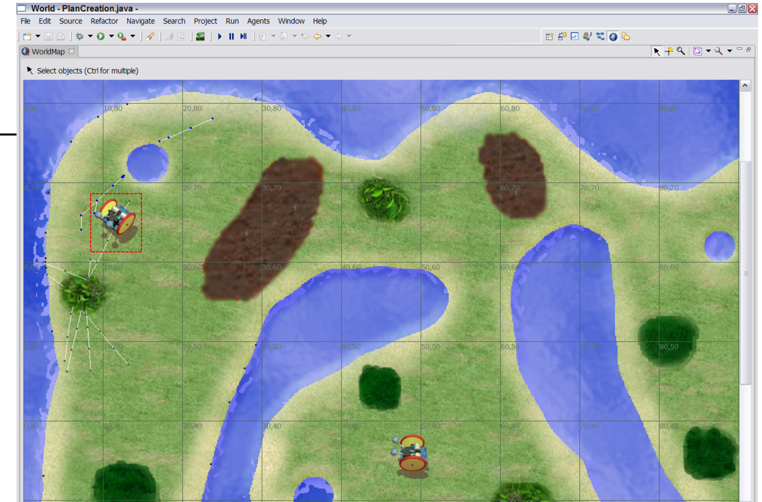
Robot control

Low-level perception

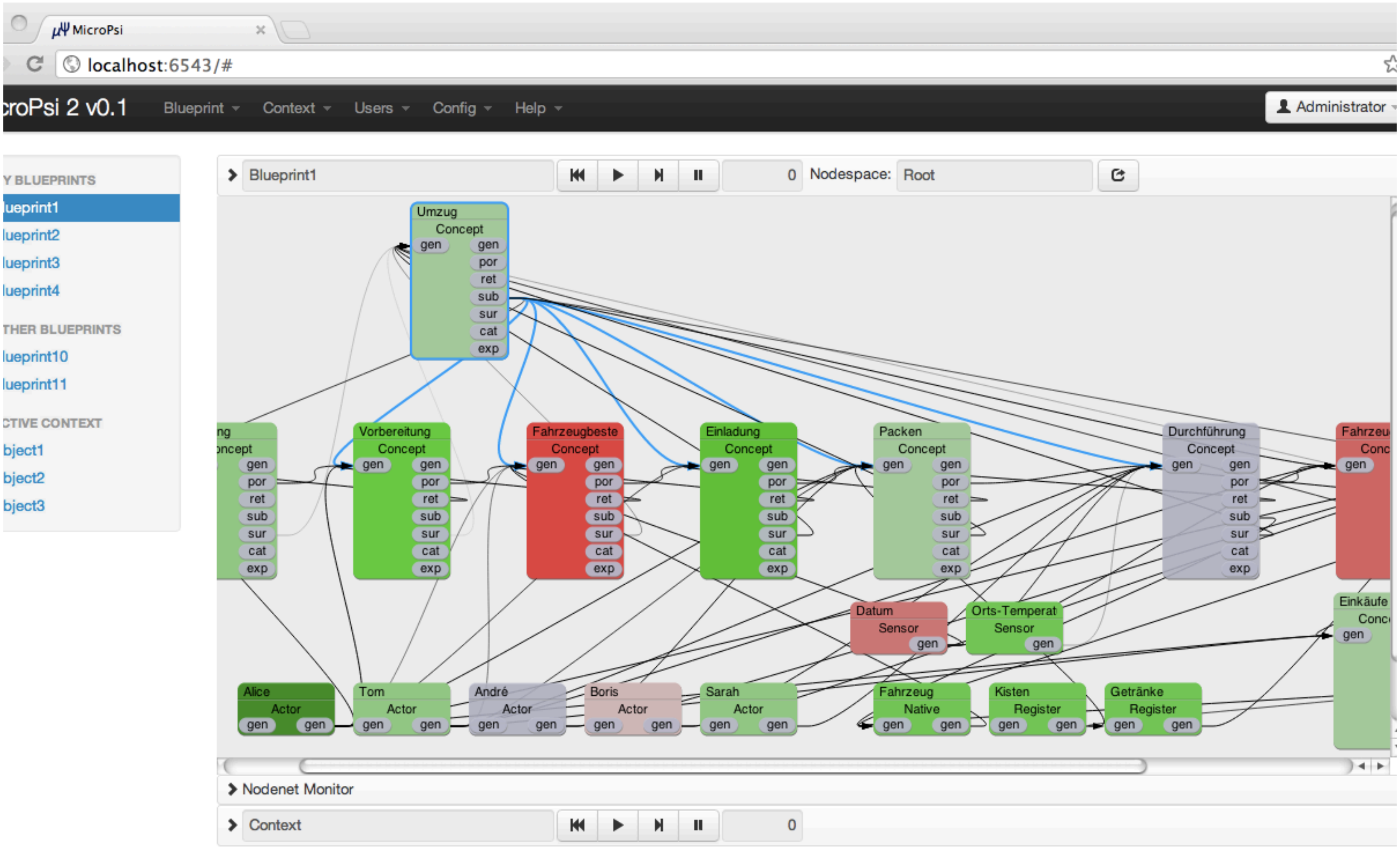
Multi-agent interaction

Control and simulation

Different simulation environments



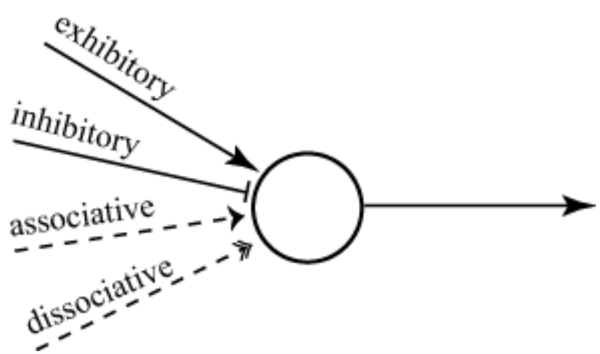
Implementation: MicroPsi 2 (Bach, Welland 12)

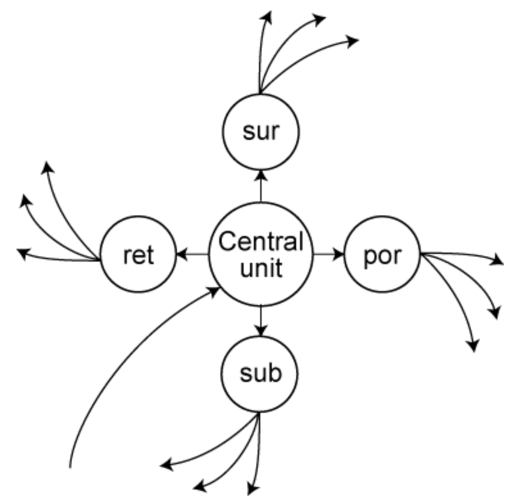


MicroPsi 2

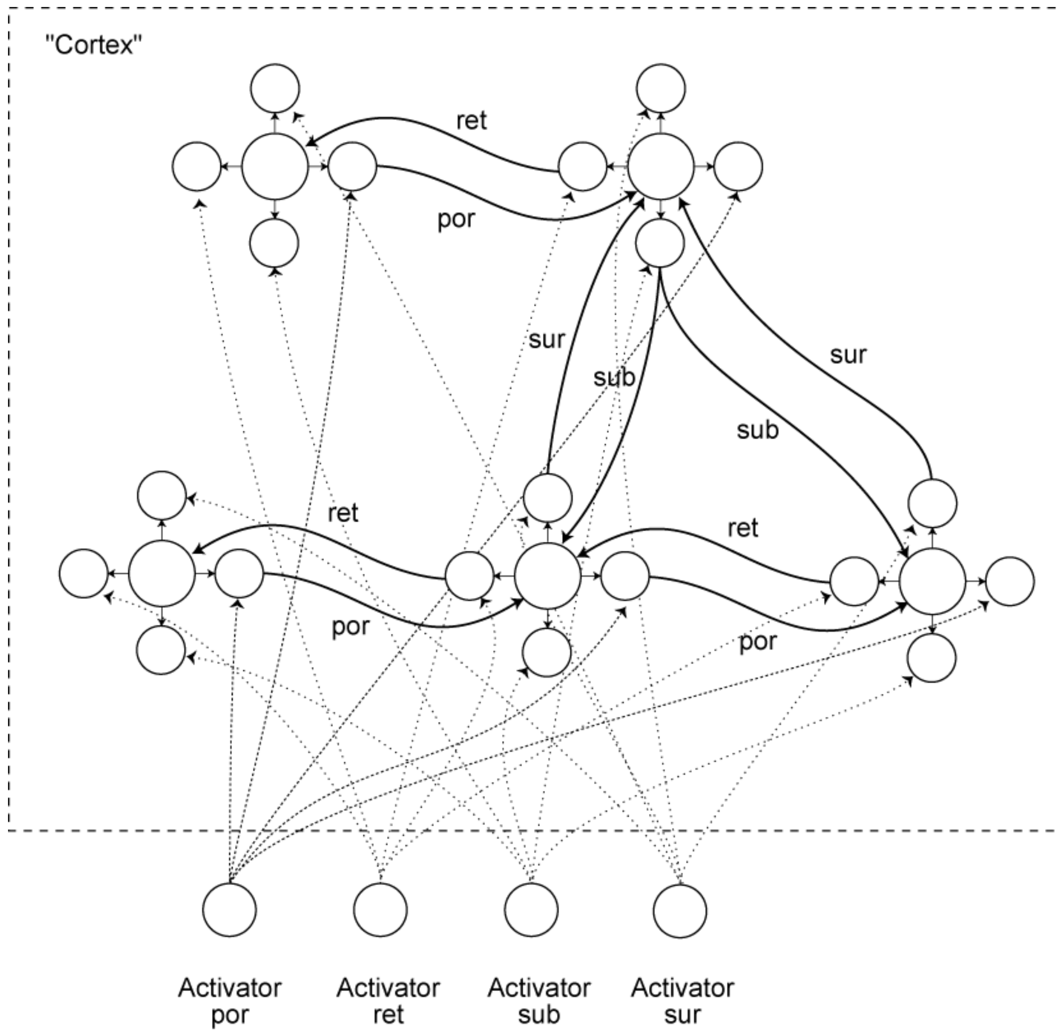
- New software basis (more lightweight and modular)
- Minecraft as (optional) simulation platform (implementation: Kemper 2014)
- Currently:
  design of social games for testing of motivation model

Representation in Psi: Neurons and Quads

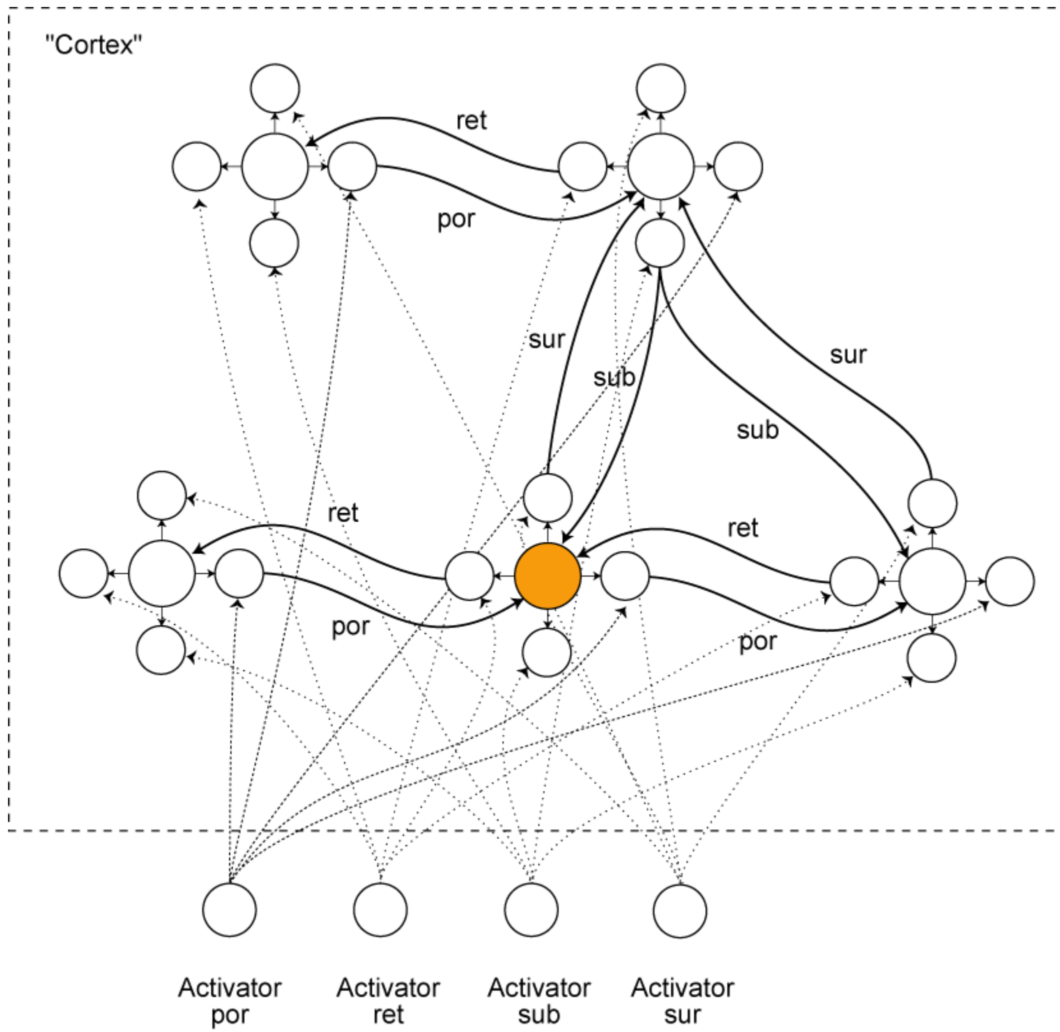




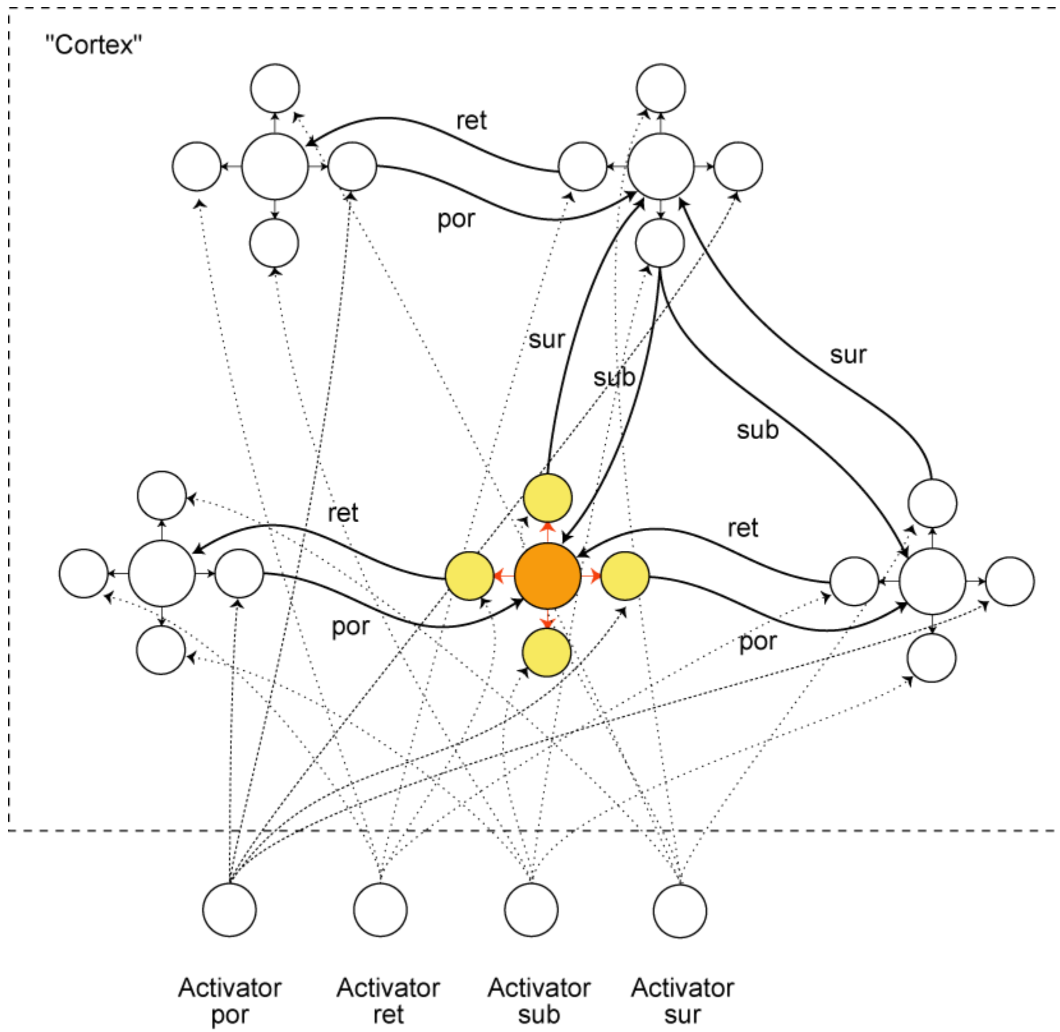
“Cortex” and Activators



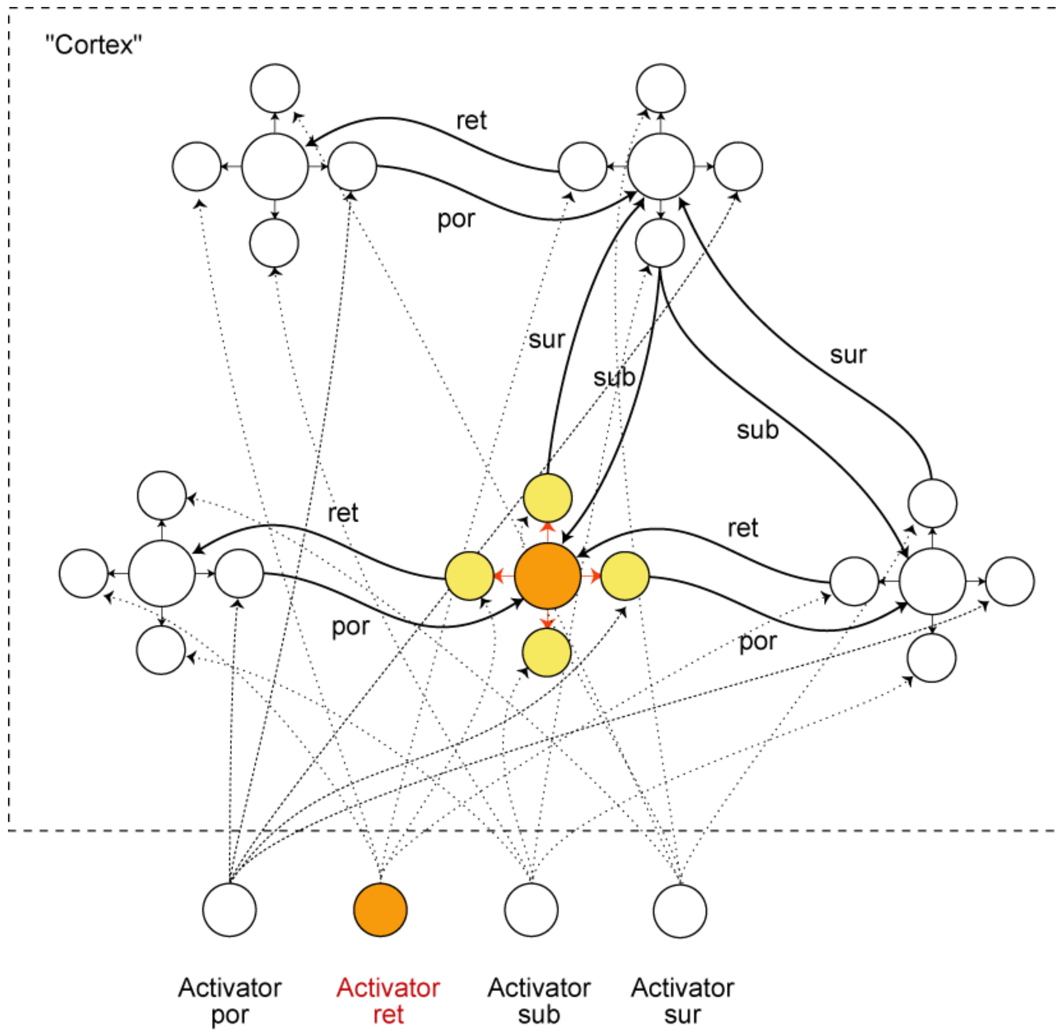
Spreading Activation in Quads



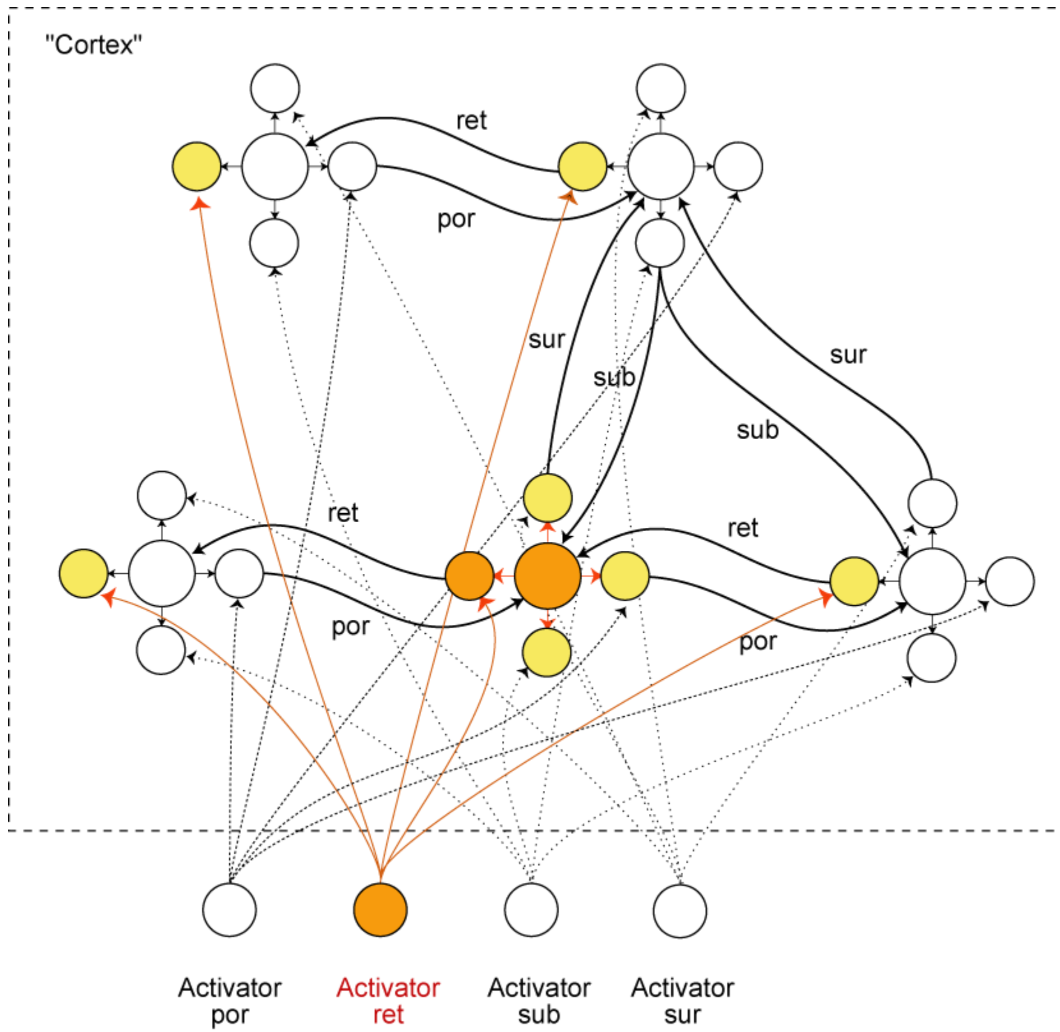
Spreading Activation in Quads



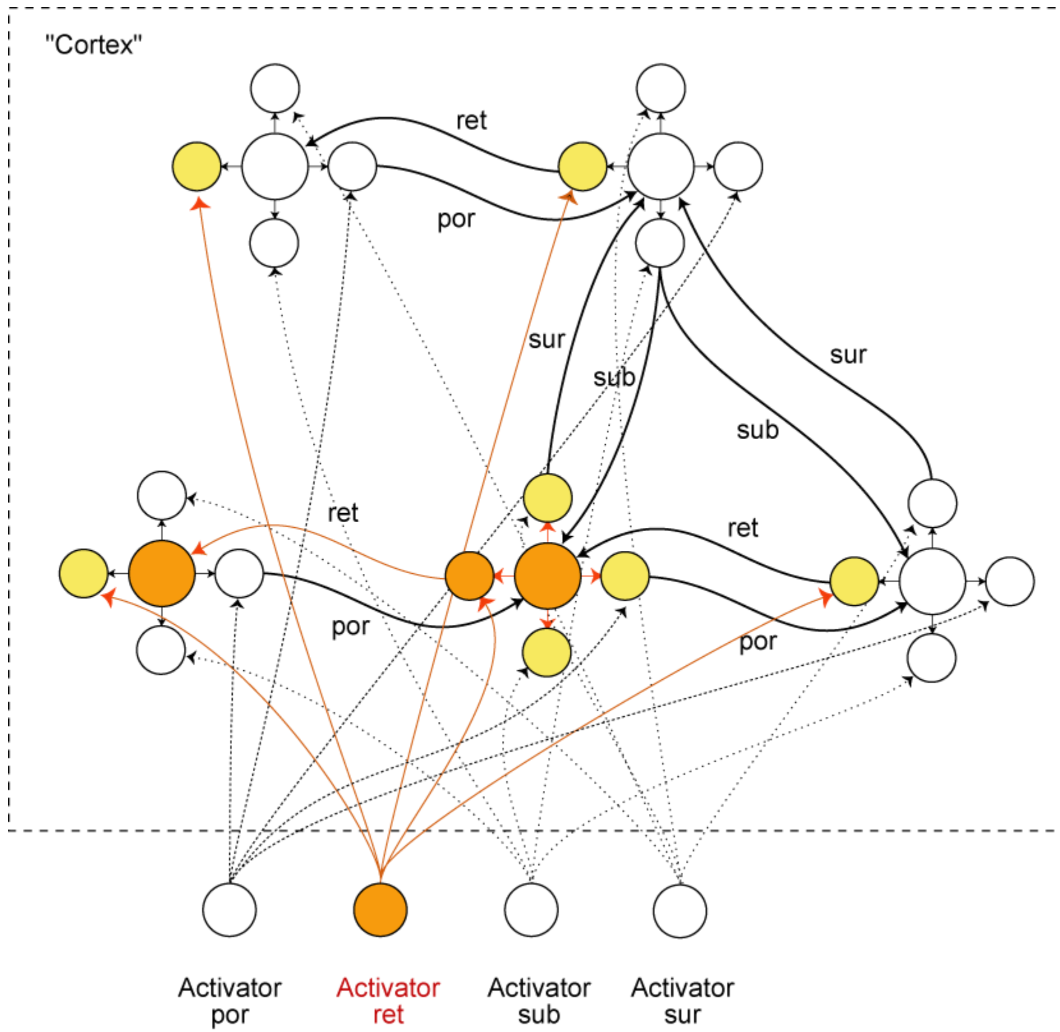
Spreading Activation in Quads



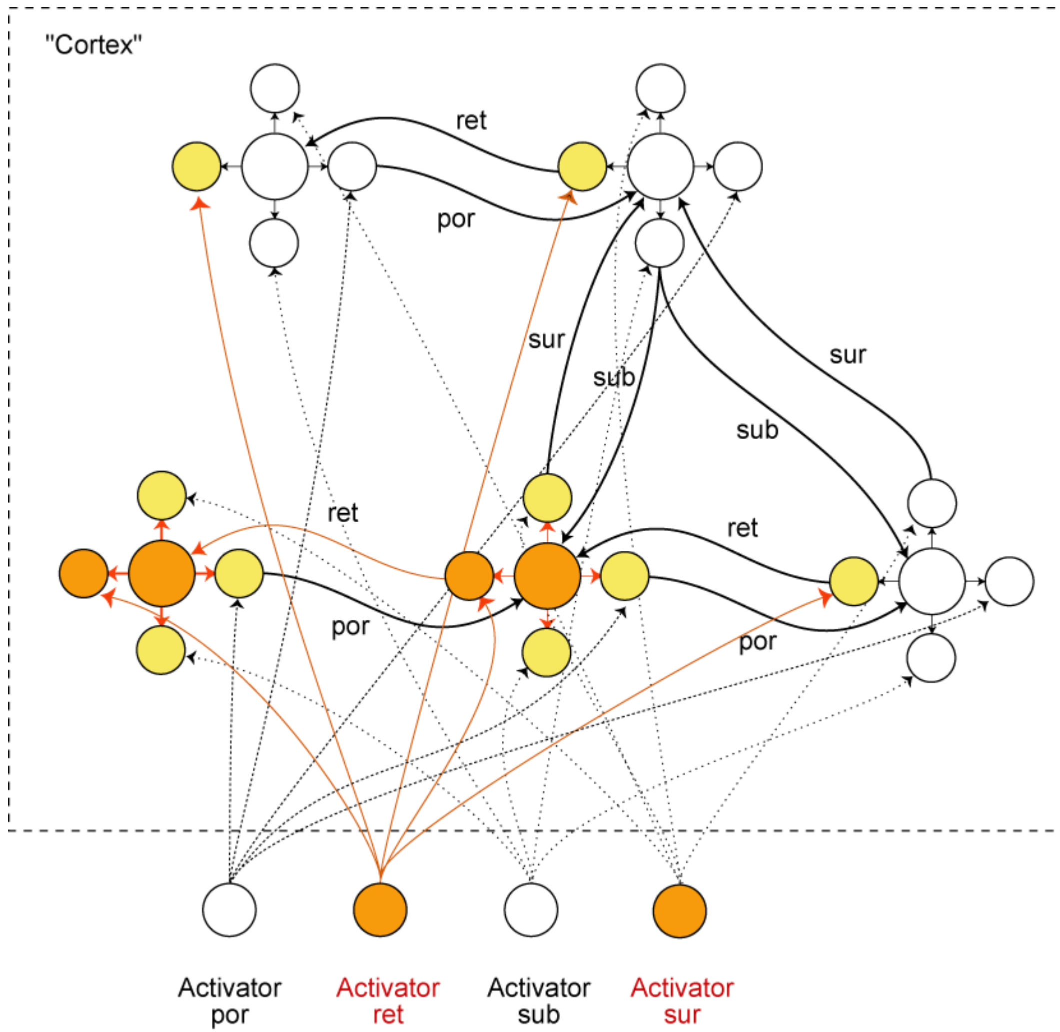
Spreading Activation in Quads



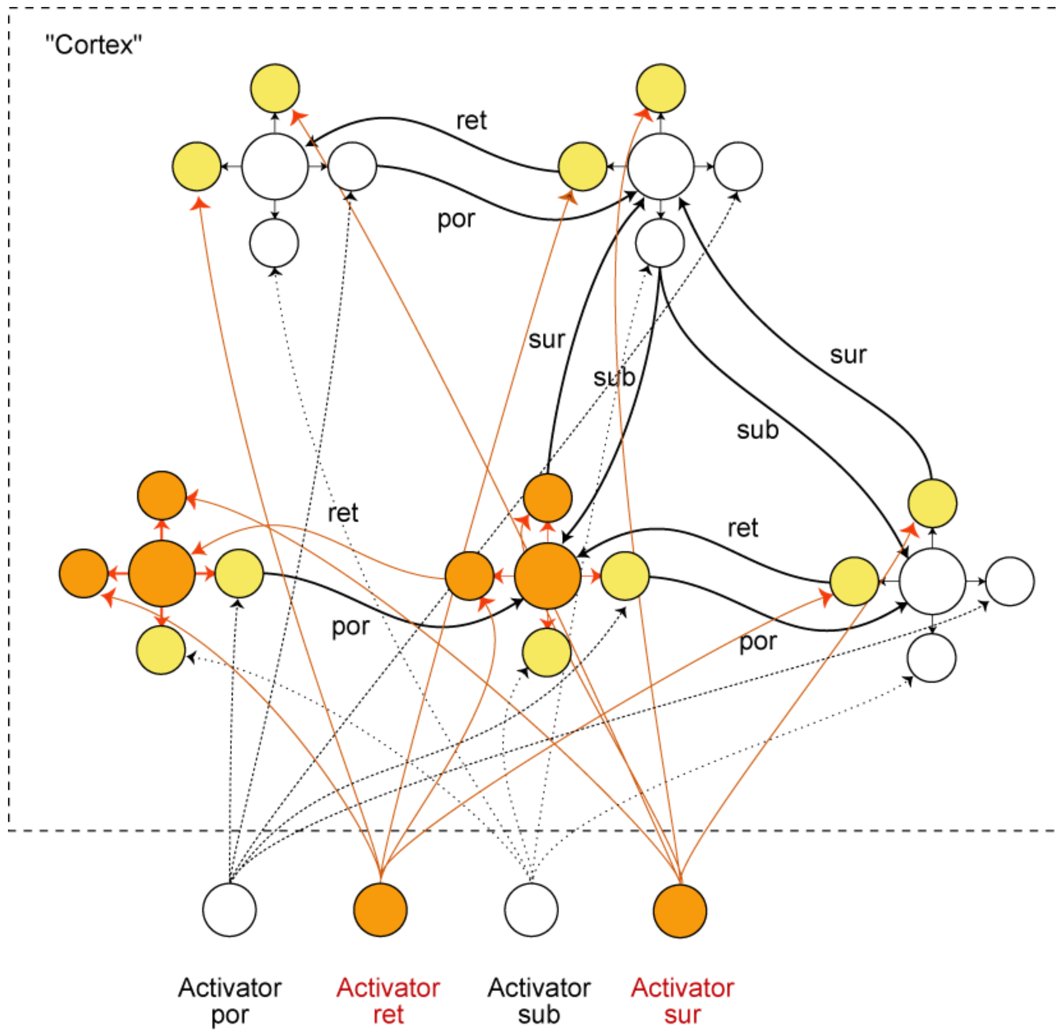
Spreading Activation in Quads



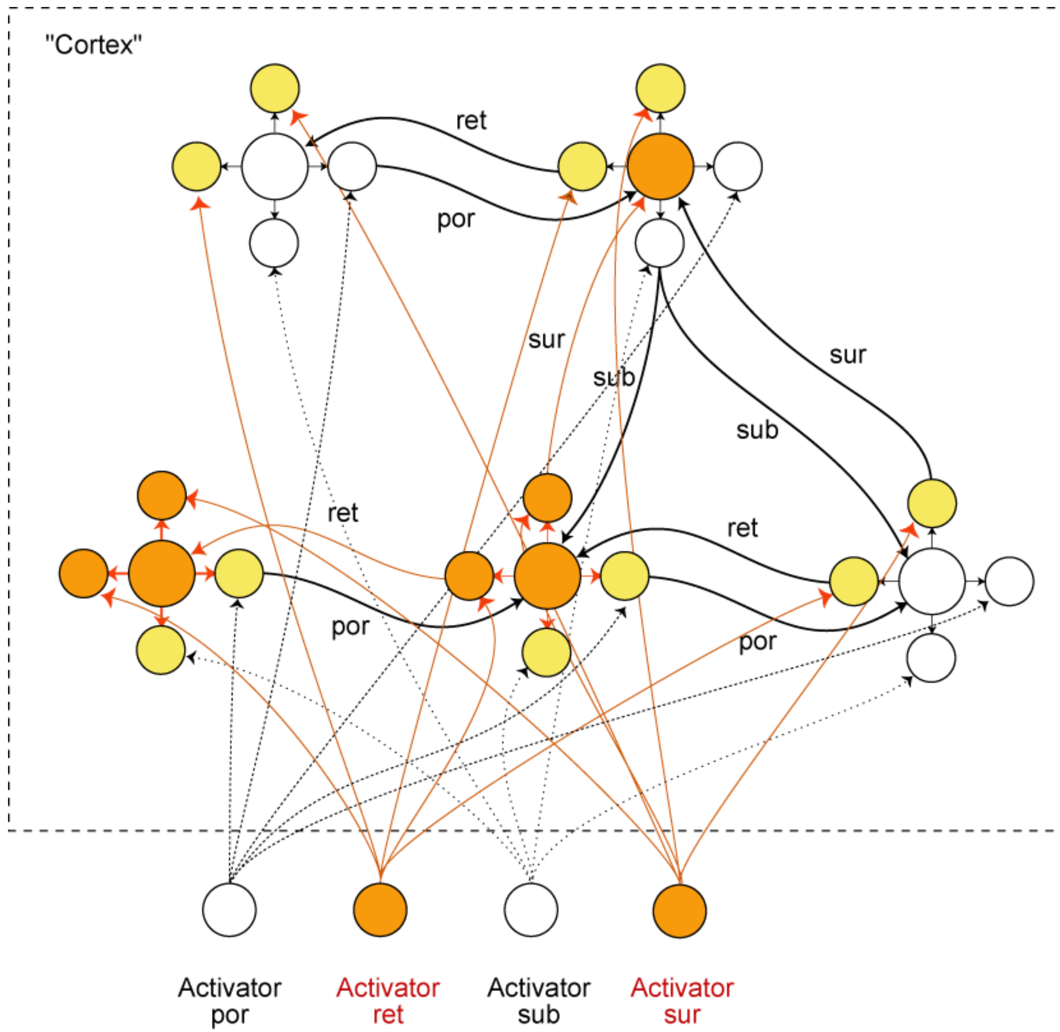
Spreading Activation in Quads



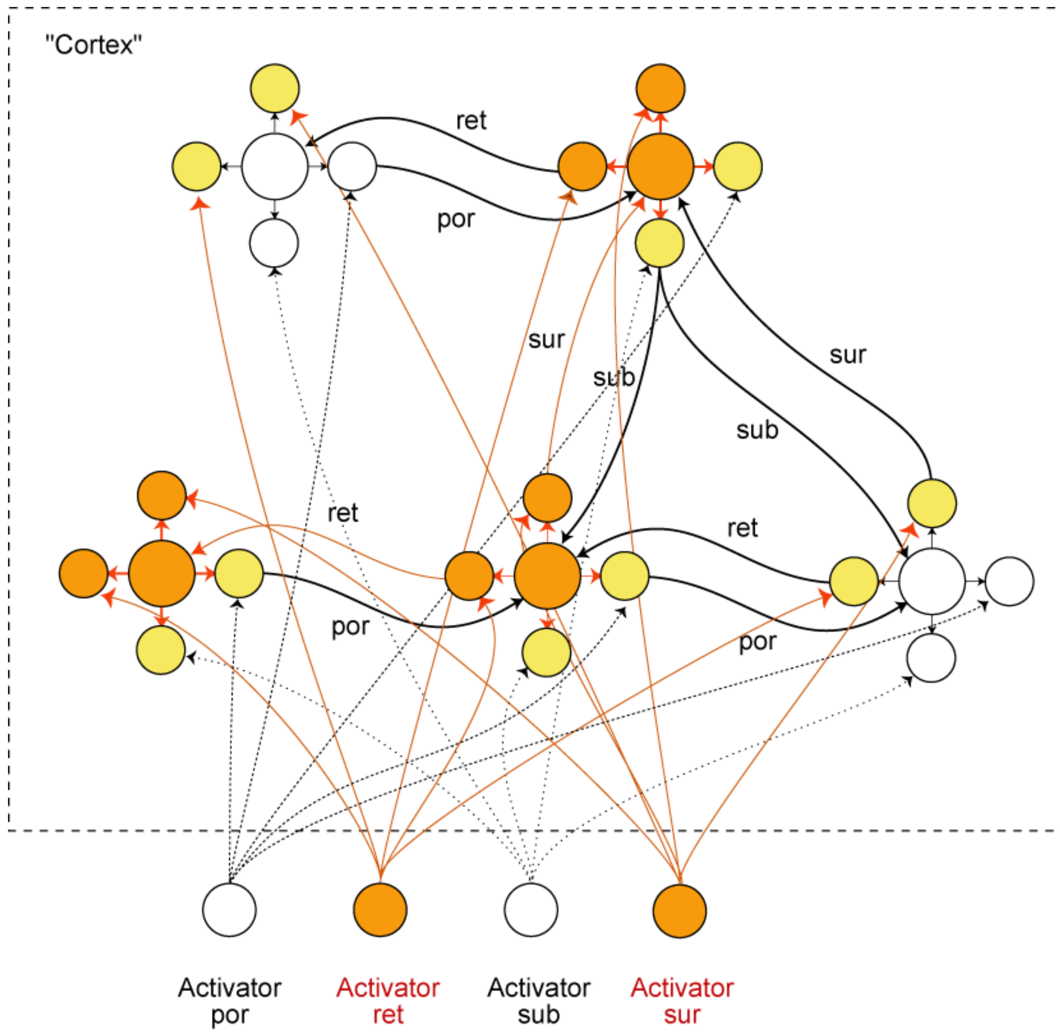
Spreading Activation in Quads



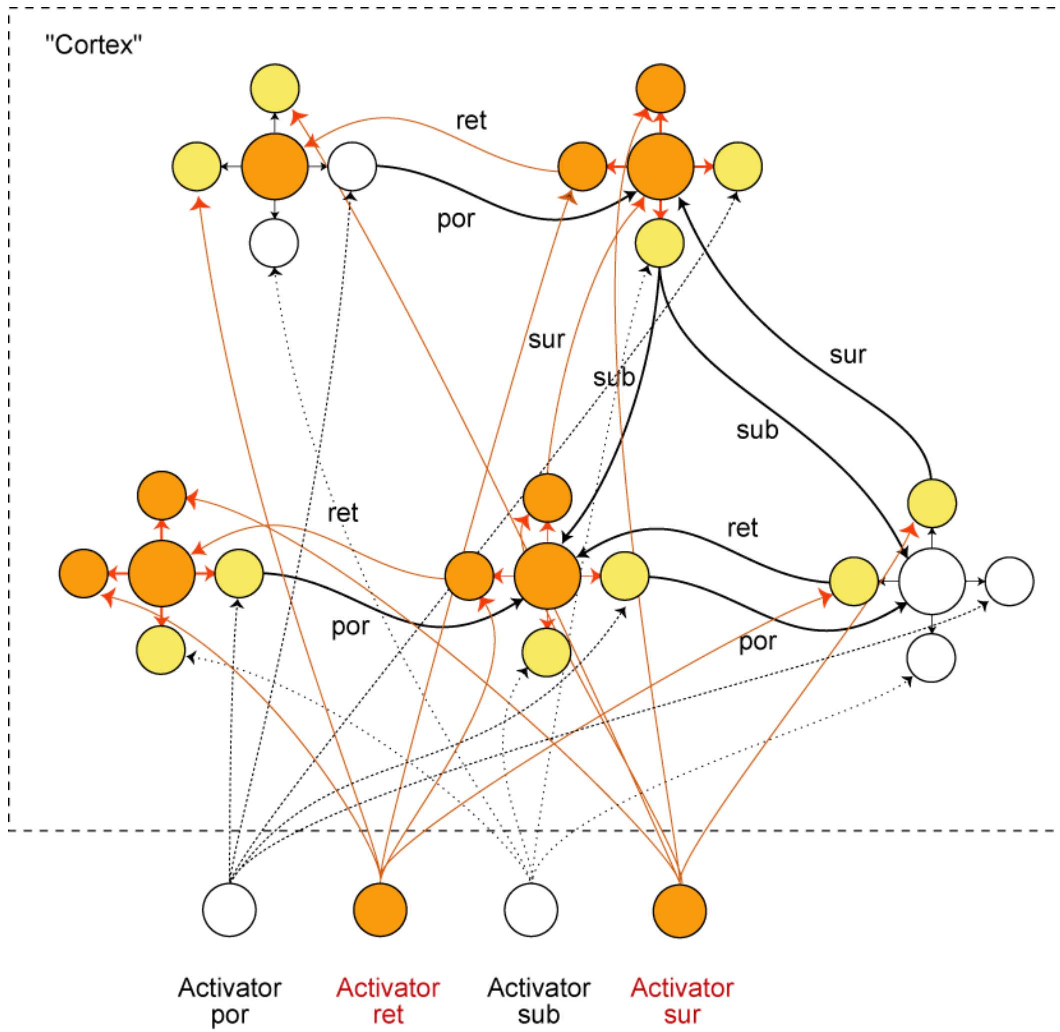
Spreading Activation in Quads



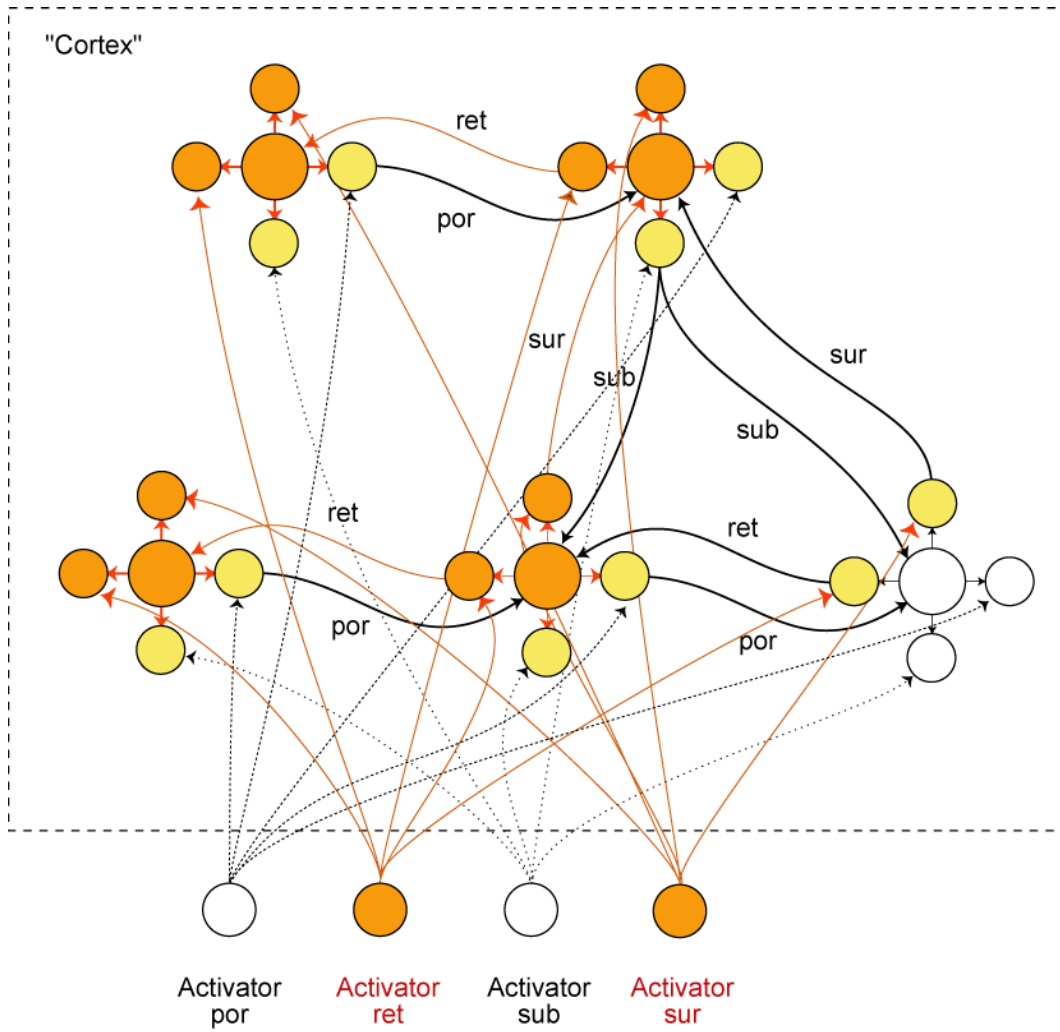
Spreading Activation in Quads



Spreading Activation in Quads

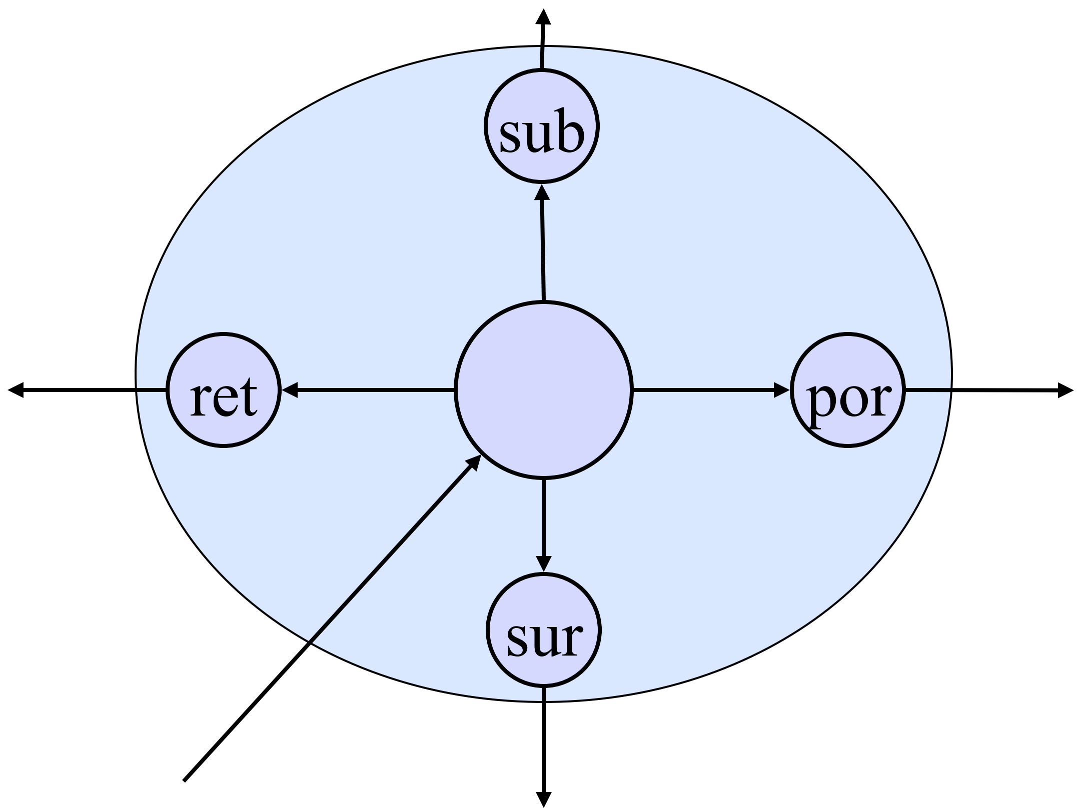


Spreading Activation in Quads



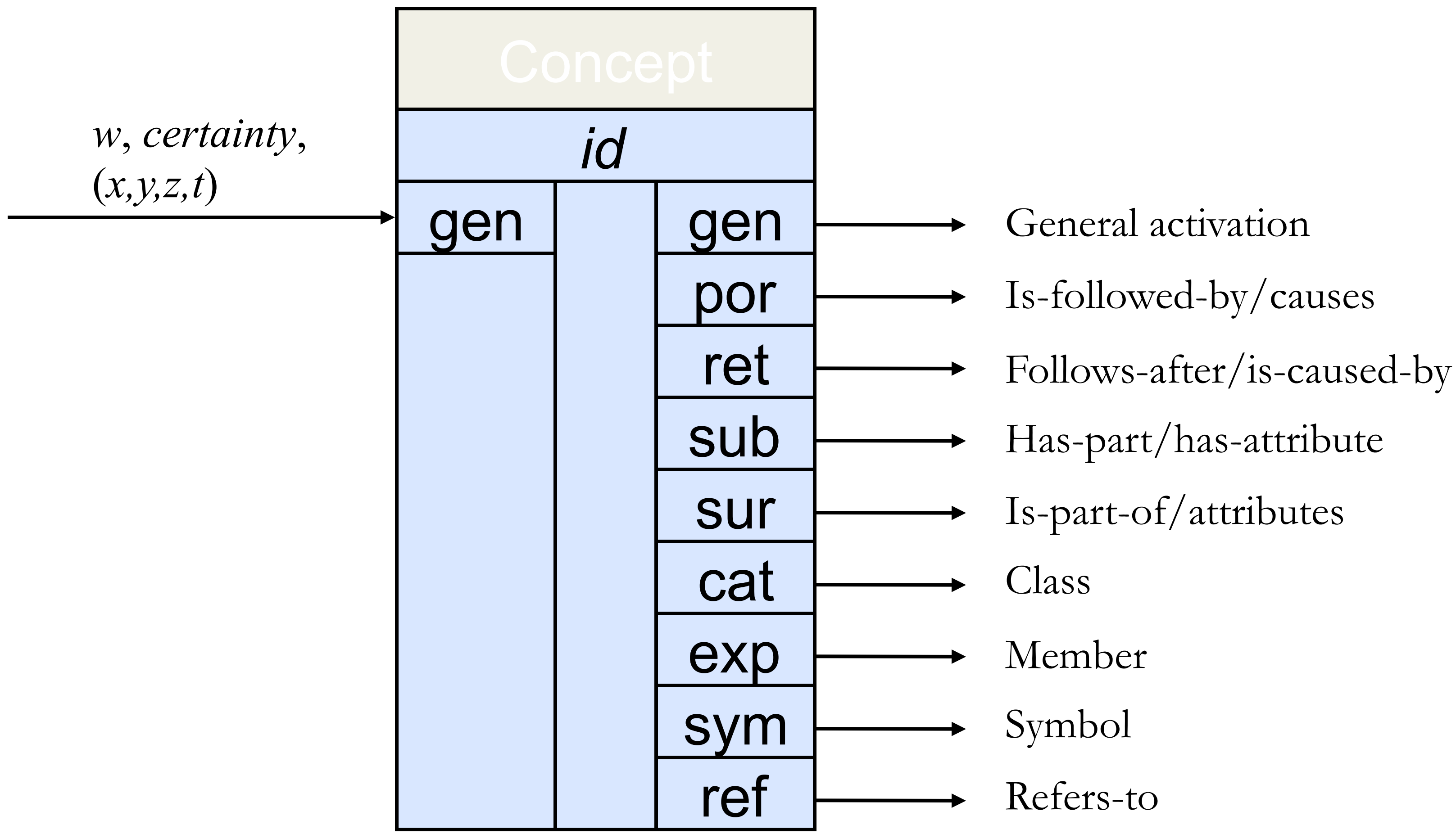
From Quads to Node Nets

→ "Quad"

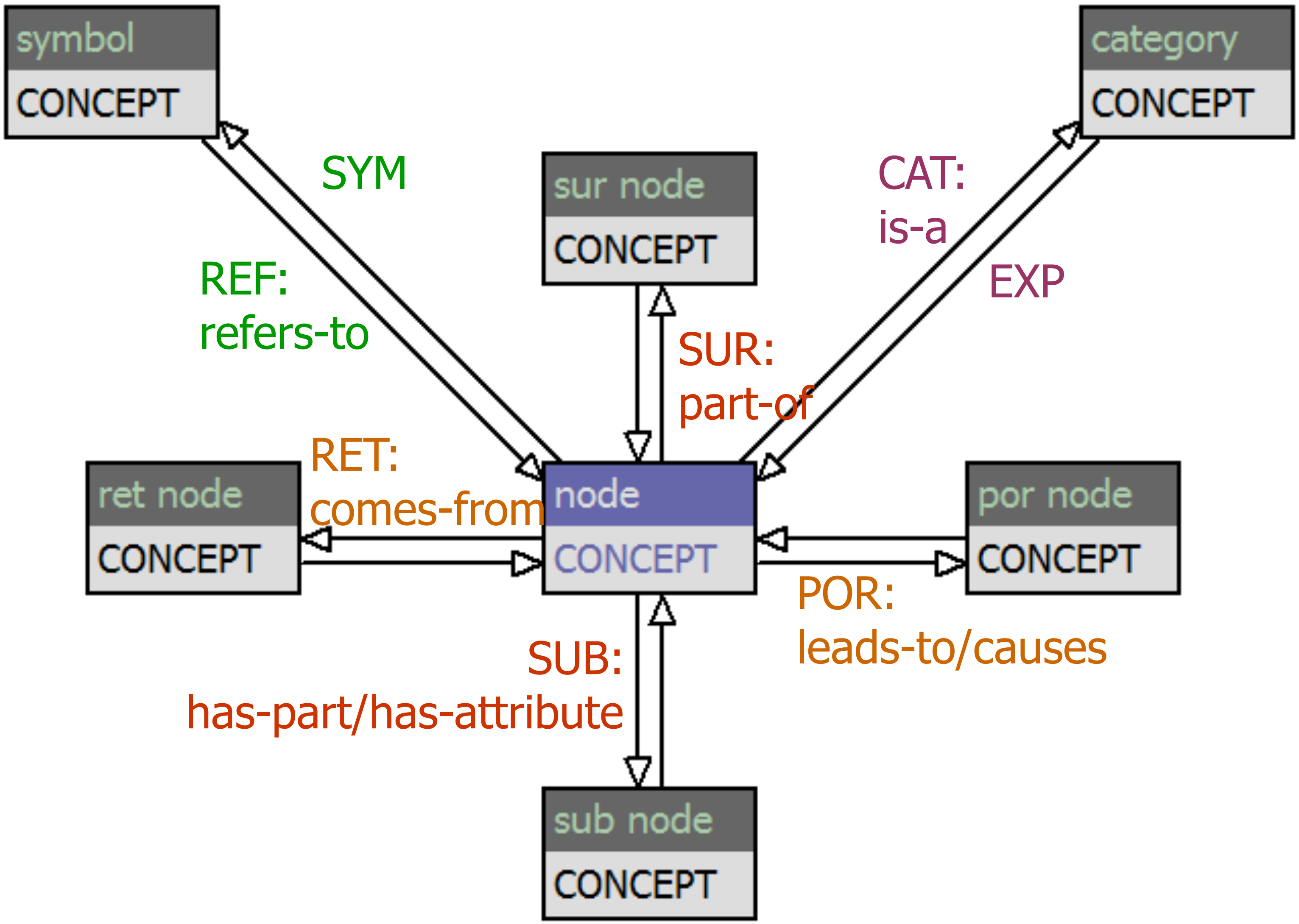


Node Nets: Concept Nodes

Basic Building Block: **Concept Node**

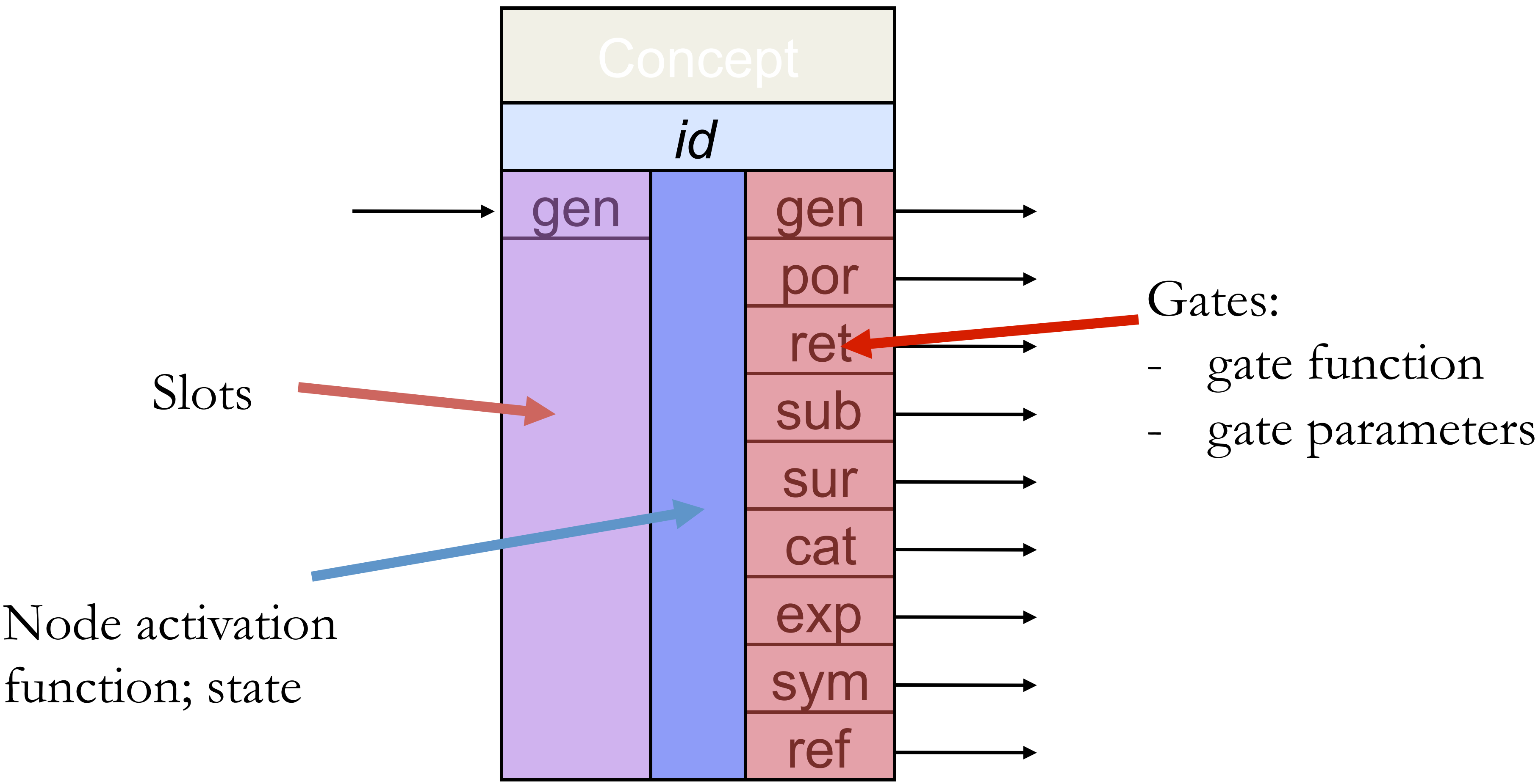


Node Nets: Link Types



Node Nets: Concept Nodes

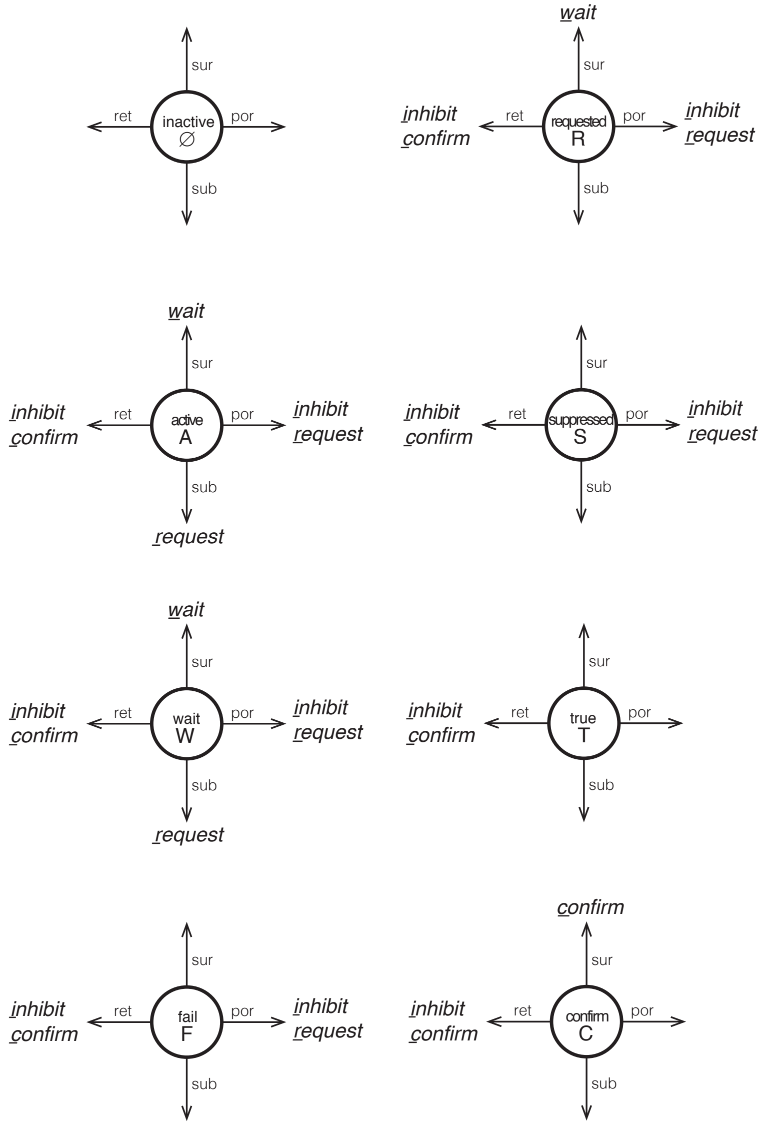
Basic Building Block: **Concept Node**

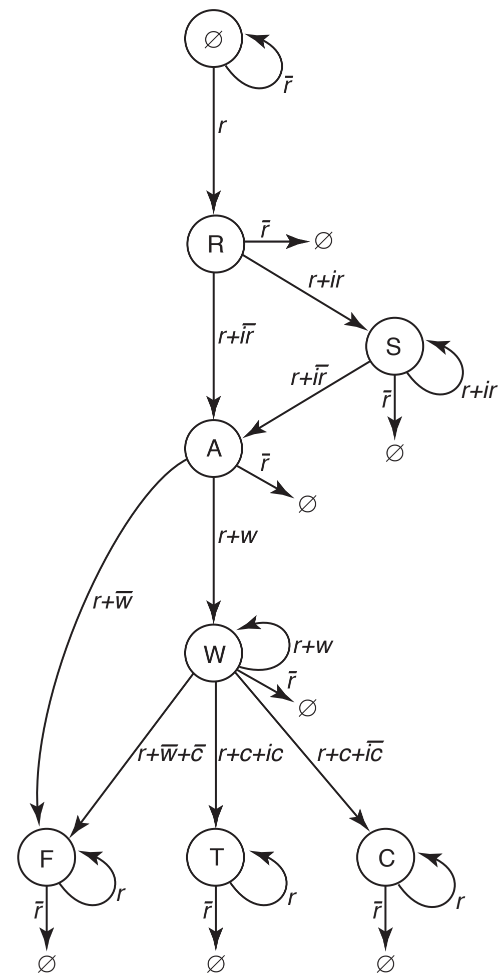


Building agents with node nets

- Graphical representation + python shell
- Node types:
  - individual NN elements (sigmoidal neurons etc.)
  - concept nodes
  - script nodes (state machines for distributed execution)
  - sensors and actuators
  - node spaces
  - control nodes
  - native modules
  - new: multi-state nodes

Request Confirmation Networks



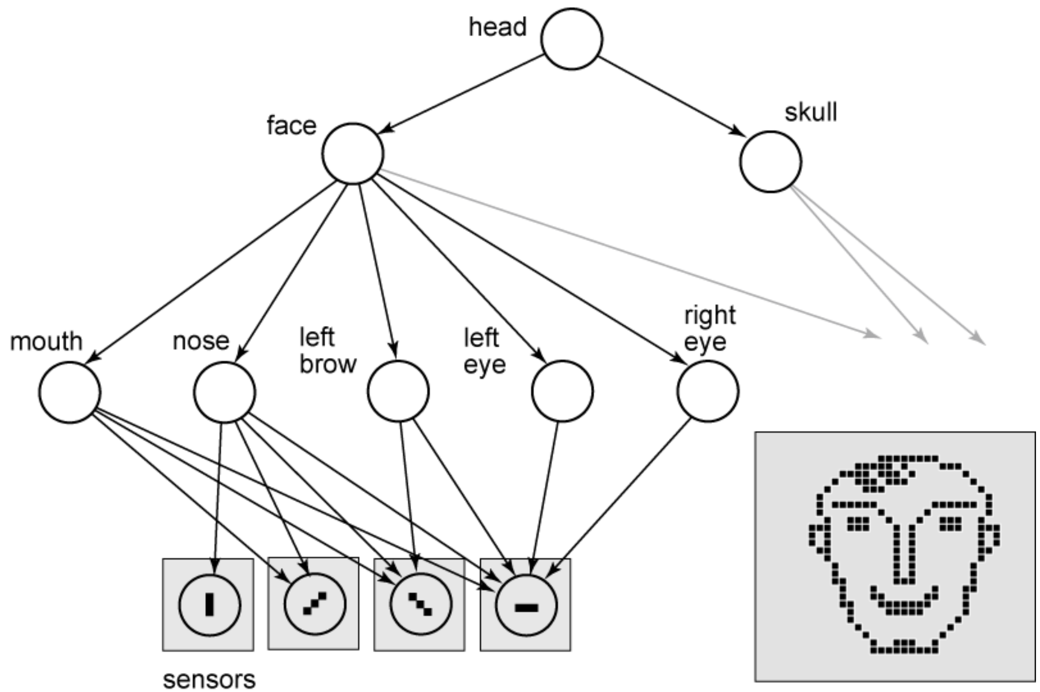


Using hierarchies for perception

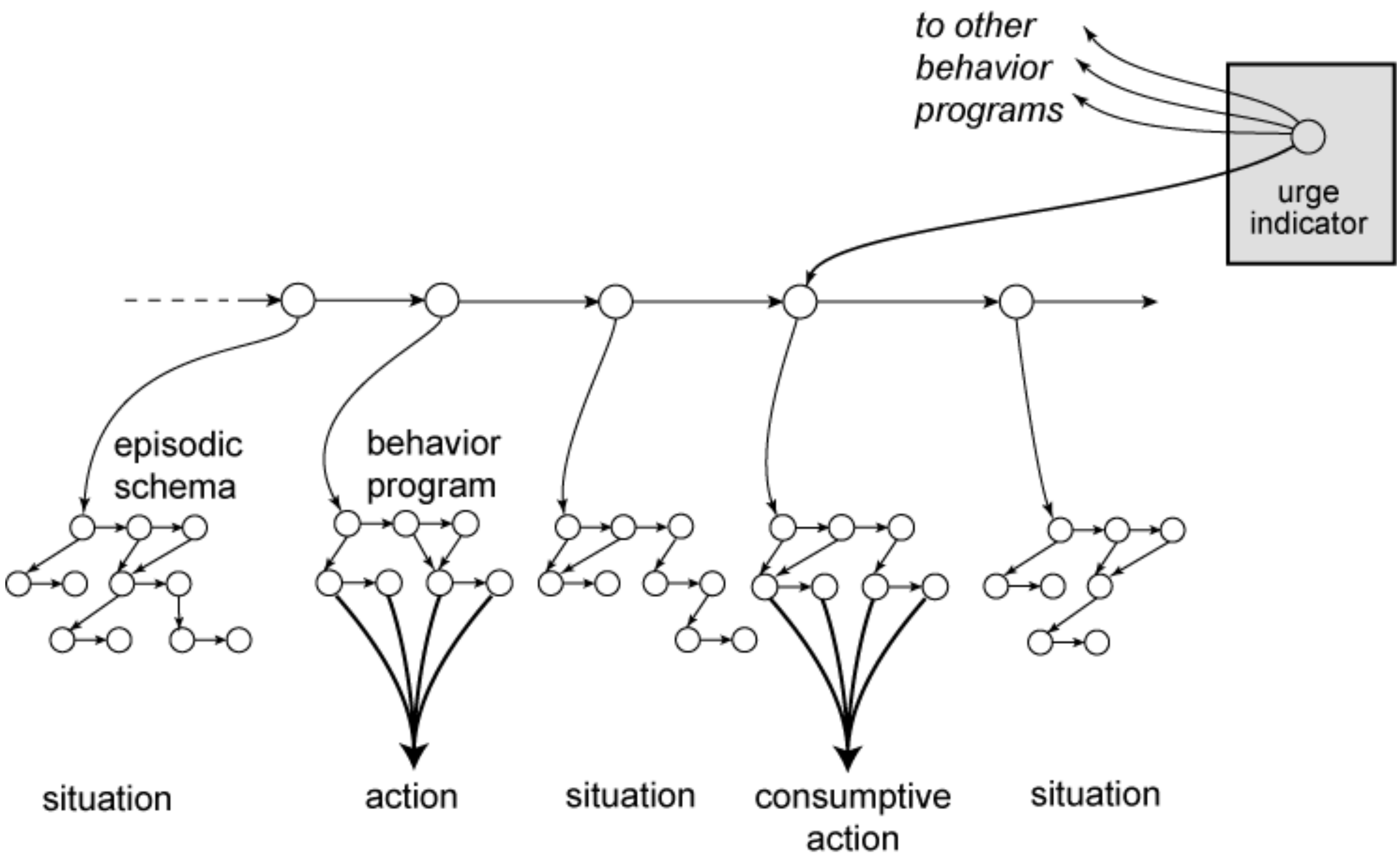
HyPercept (Hypothesis based Perception, Dörner 1999)

- Features from environment trigger sensors
- Sensors activate object hypotheses (activation spreads upwards)
- Object hypotheses are successively tested (activation spreads downwards)
- Sensors confirm or invalidate hypotheses
- Finally:
  - If only one hypothesis left: successful recognition
  - If multiple hypotheses left: further exploration
  - If no hypothesis left: new object

Hierarchical Representation



Protocol Memory



Emotion and Motivation in MicroPsi

- Goal: Computational Model of Emotion

Why is emotion so interesting?

- Obviously: applications
- *Phenomenal aspect* of emotion – *feeling*:

  Emotion may be a critical juncture when it comes to understanding the mind

→ What is emotion?

Emotion

- Having an emotion is different from behaving as if having an emotion

- What is it like to have an emotion?
- Can emotion only be simulated, or can an artificial system be in an emotional state?

→ is having an emotion a way or an aspect of information processing?

Modeling emotion

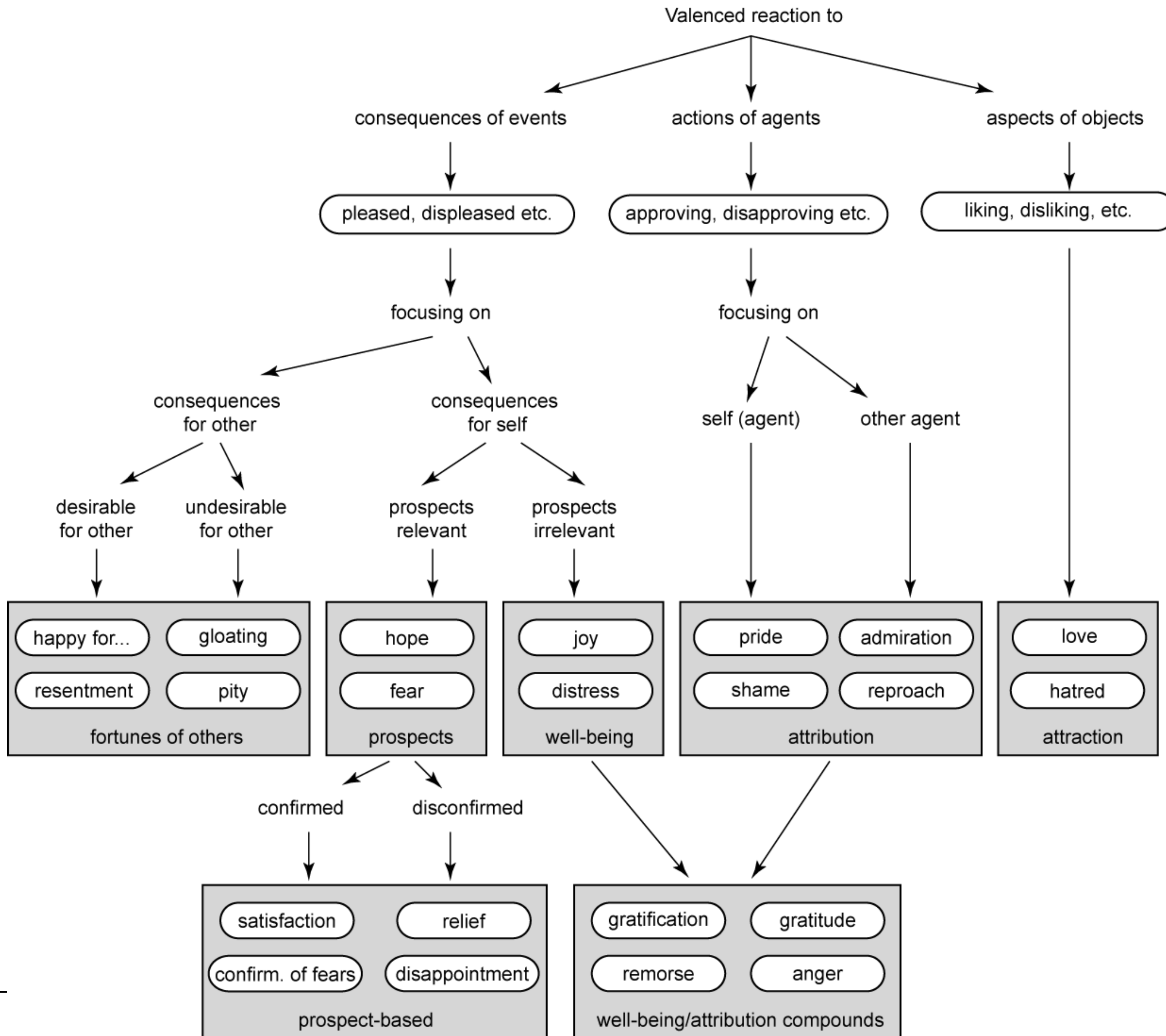
Most models of emotion account for **description**:

- What emotional states can a system have?
- Which events (in the environment) trigger emotions?
- How can emotional states be expressed?
- How can emotional states be recognized?

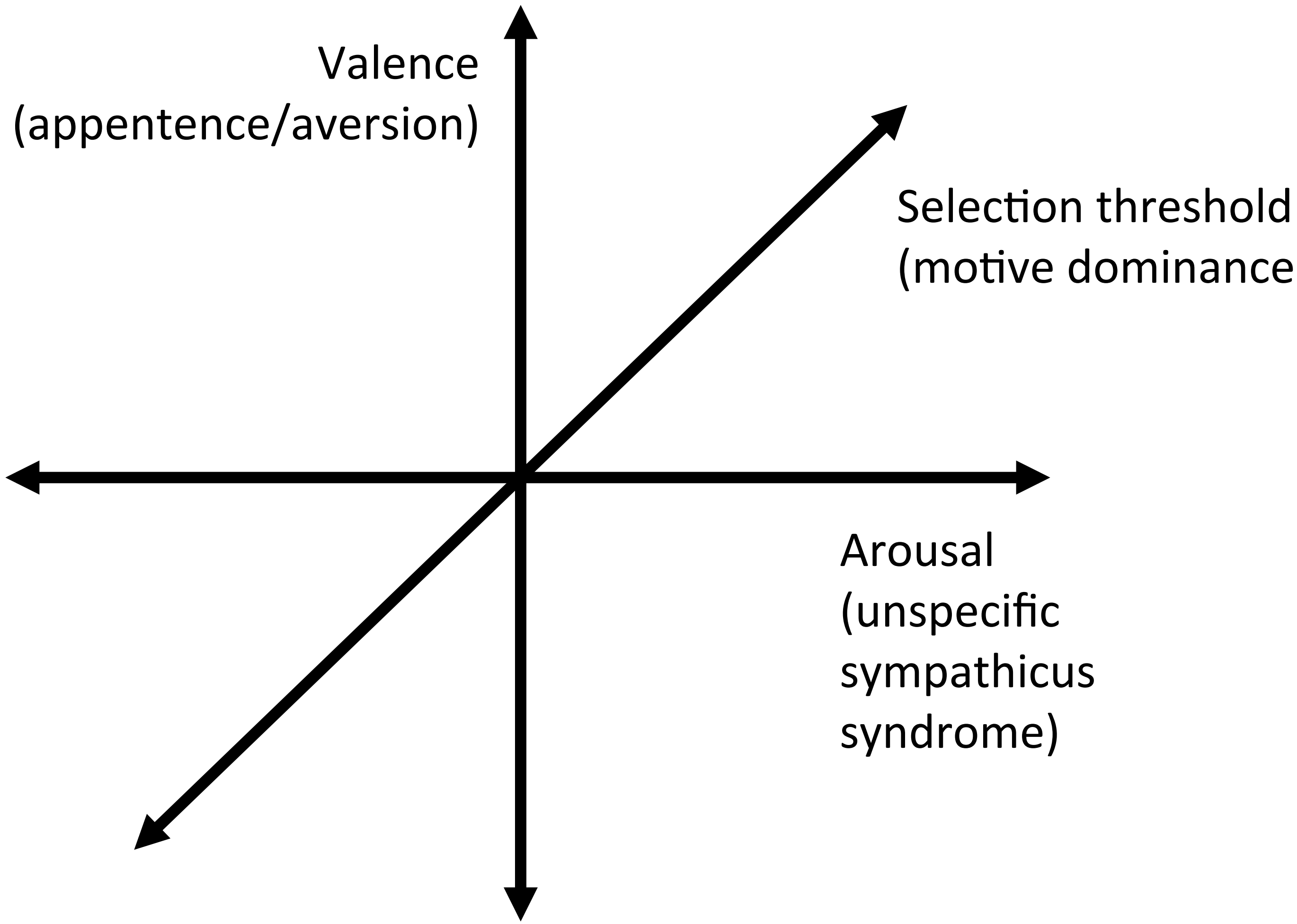
(Plutchik 80; Ortony, Clore & Collins 88; Scherer 93; Gratch & Marsella 04)

vs. "How can a system have an emotion?" (functional)

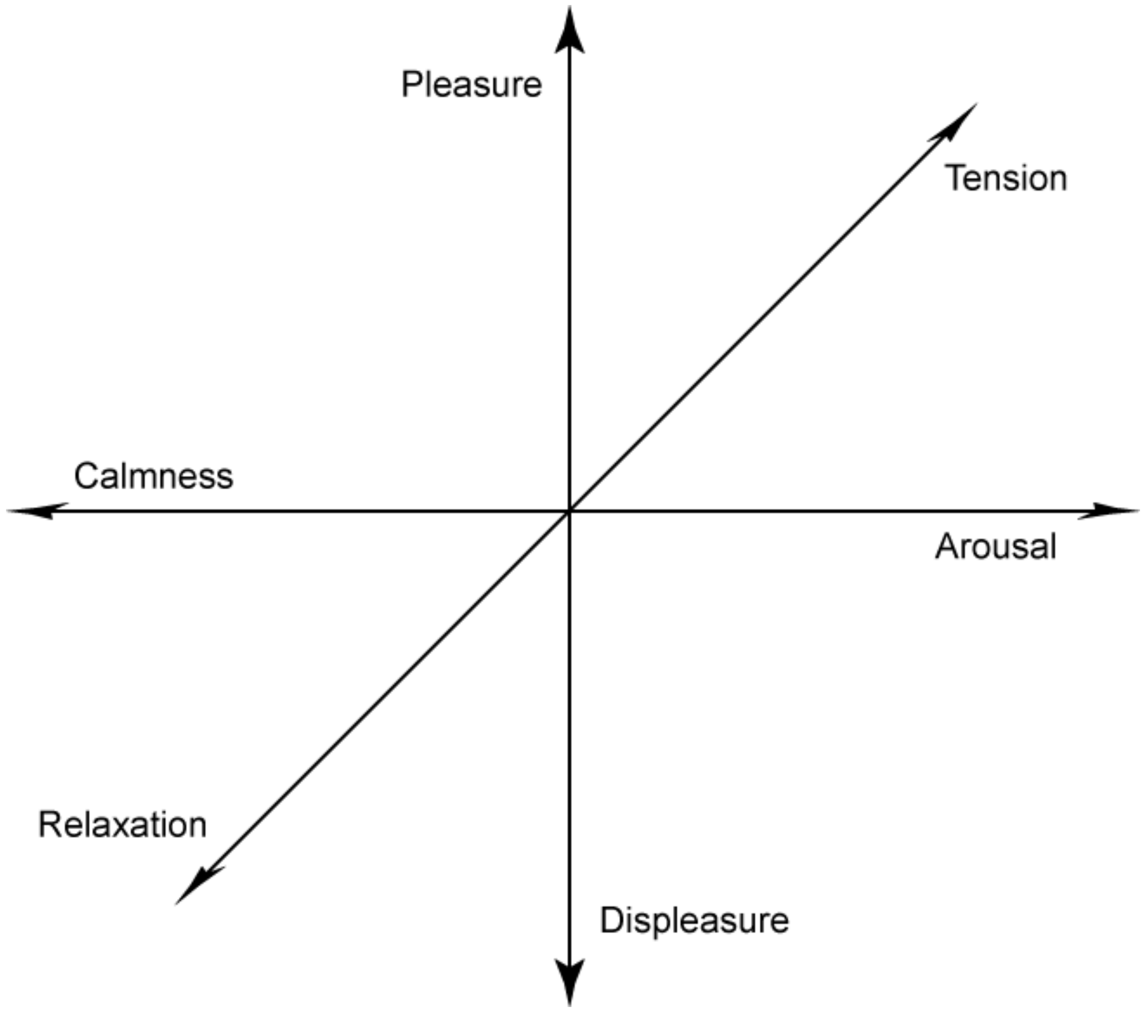
Conceptual analysis: Ortony, Clore, Collins 1988:



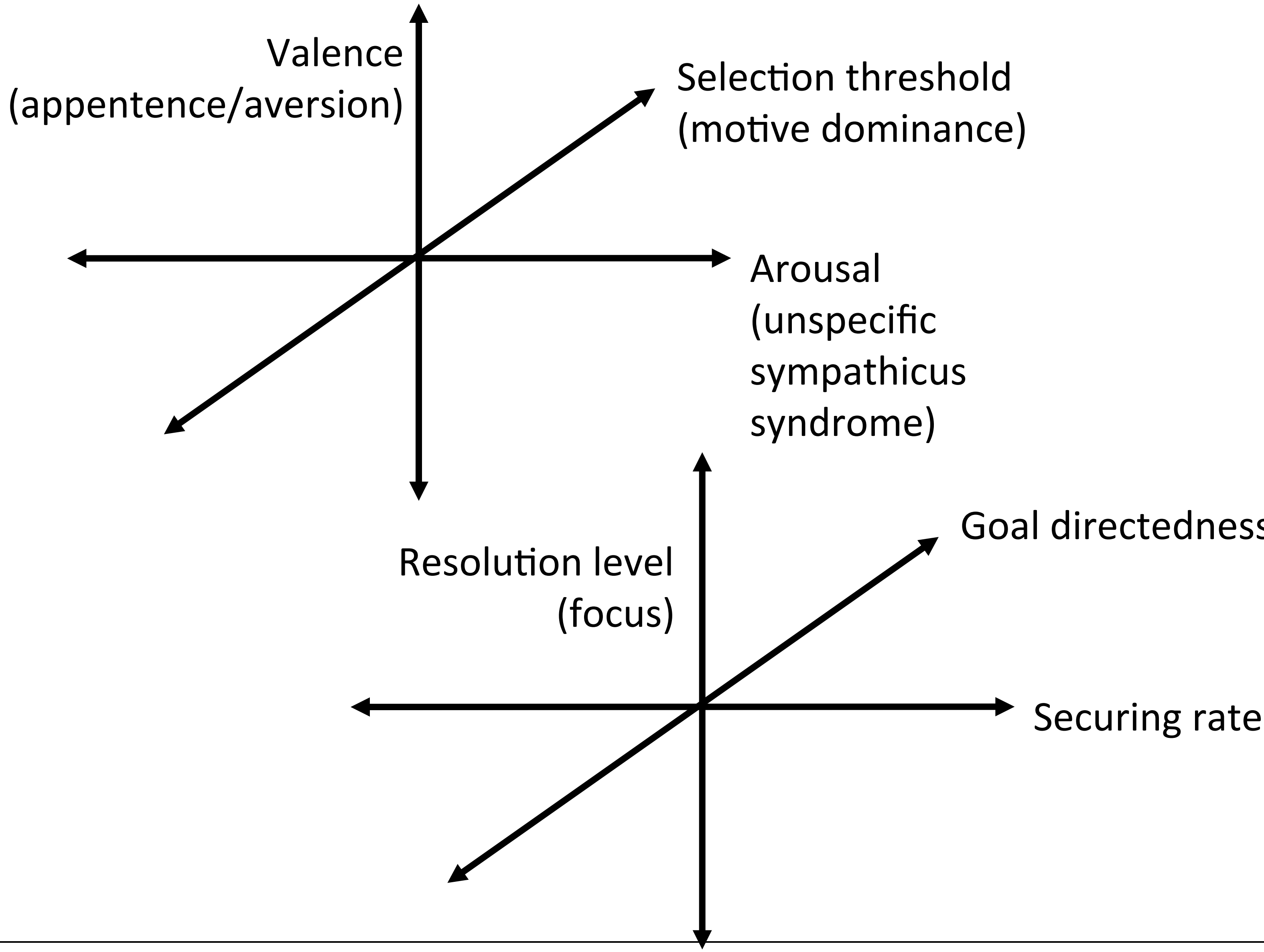
Affective dimensions in the PSI theory (Dörner 1999)



Compare: Affective dimensions (Wundt 1910)



Affective dimensions in the PSI theory



The Psi theory about emotion

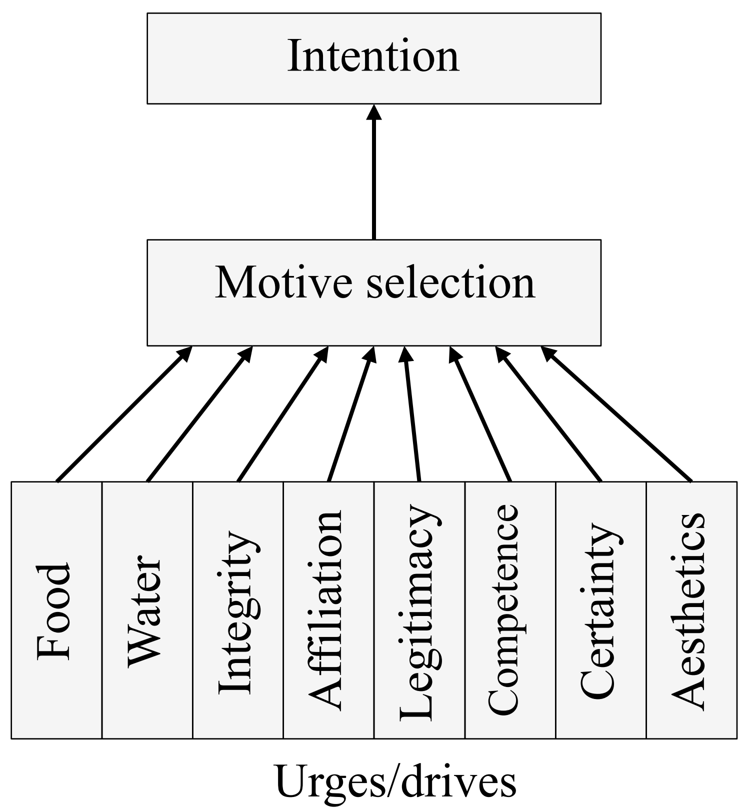
- Affect is seen as a *configuration* of a cognitive system
- Modulators of cognition:
  - arousal, selection threshold, securing threshold, resolution level
  - estimate of competence and certainty
  - pleasure/distress signals → valence
- Affective state is emergent property of modulation
- Directed affects (higher-level emotions) emerge by association of demand with appetive or aversive objects/situations

Purpose of emotional modulation

- Control width, depth and bias of operations on mental representations of the agent
  → modify perception, memory, planning and action selection
- Reduce complexity of cognitive processes

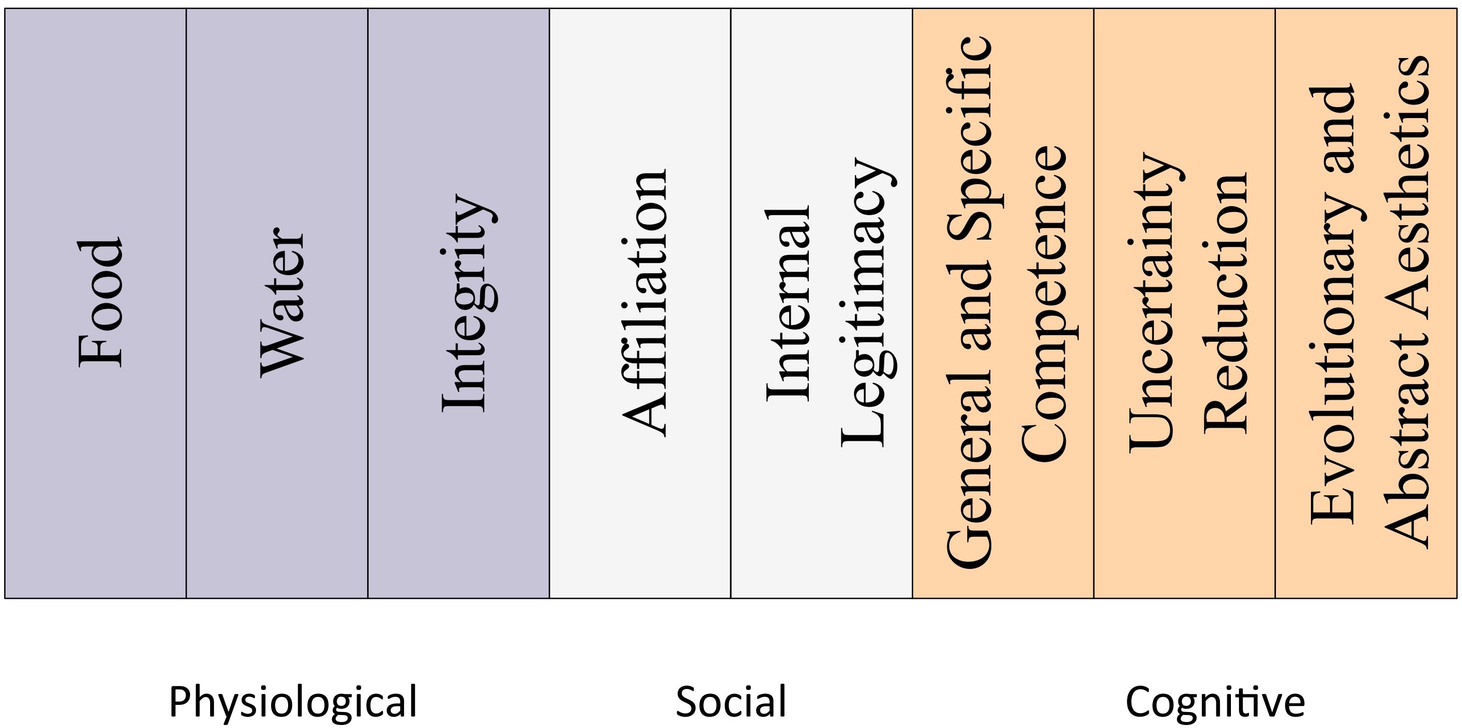
Motivational System

- All goals attempt to satisfy a (hard-wired) demand
  → flexible goals, but (evolutionary) suitable behavior



Motivational System

- Drives correspond to set of demands of the agent



Physiological Drives

- if autonomous regulation of body processes fails
  → actively manage physiology (seek food, water, healing, shelter, rest, warmth, ...)
  → escape perilious situations
  → **implicitly** seek physical survival

Social Drives

- *Affiliation*: structure social interaction beyond rational utility
- increased by 'legitimacy signals', decreased by 'anti legitimacy signals' (and adaptively over time); allows for non-material reward and punishment

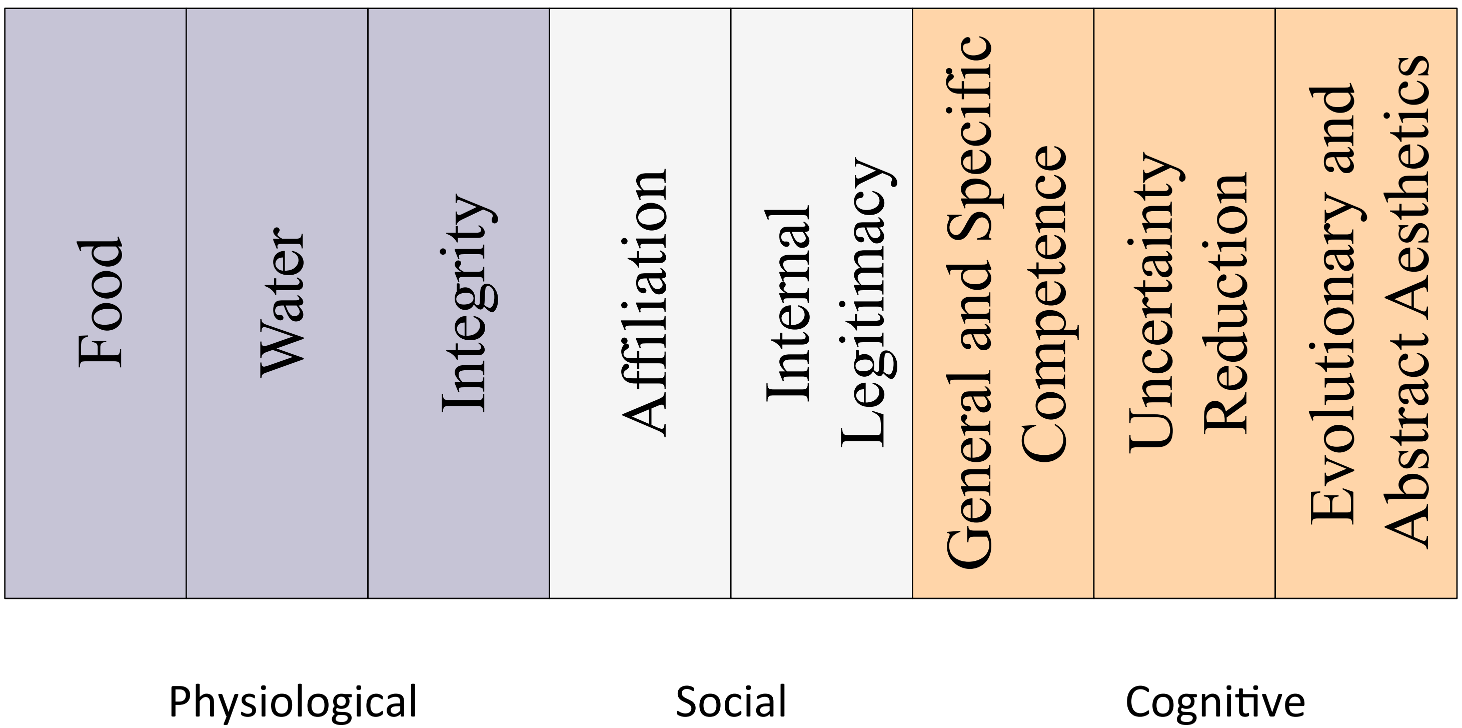
- *external* legitimacy: group acceptance
- *internal* legitimacy: "honor", conformance to internalized social norms

Cognitive Drives

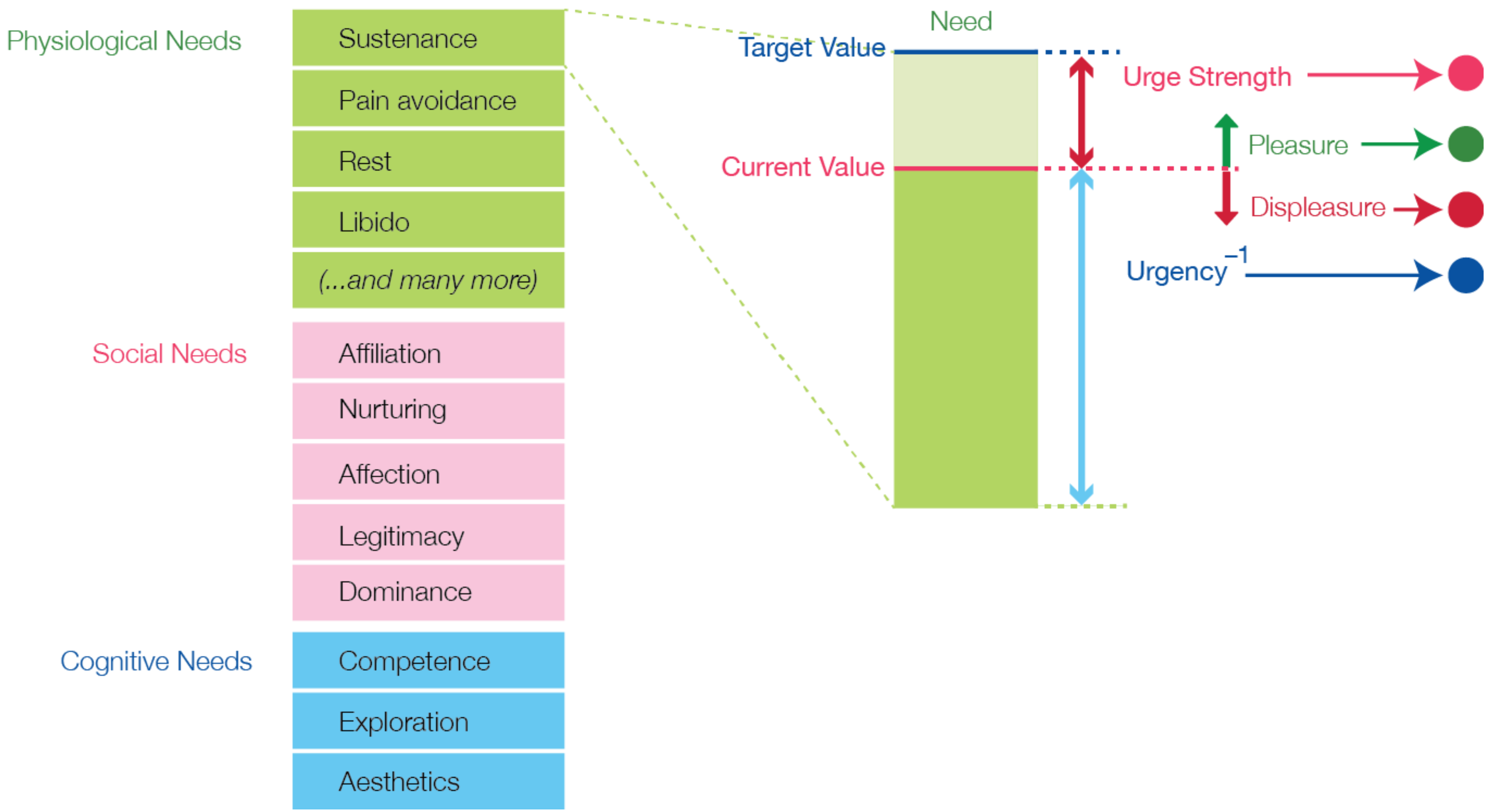
- Competence
  - epistemic (problem specific)
  - general (ability to satisfy demands)
  - effect oriented
- Uncertainty reduction
  - novelty seeking
- Aesthetics
  - evolutionary preferences (stimulus oriented)
  - abstract (representation oriented)

Motivational System

All possible goals correspond to (at least one) demand

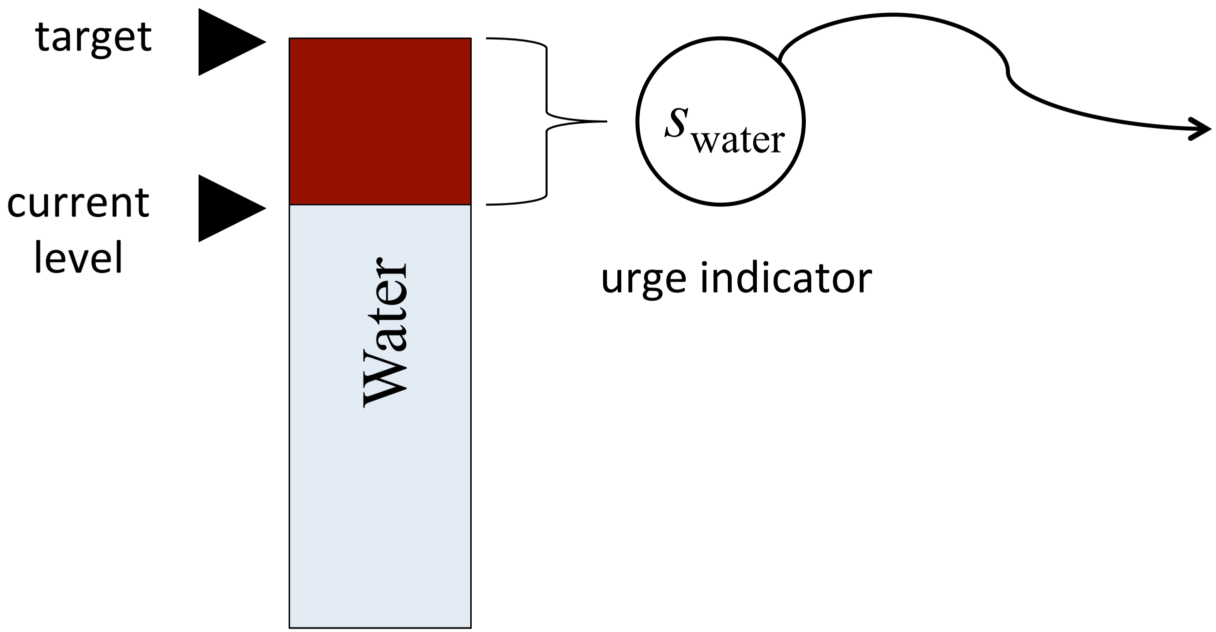


Needs and urges



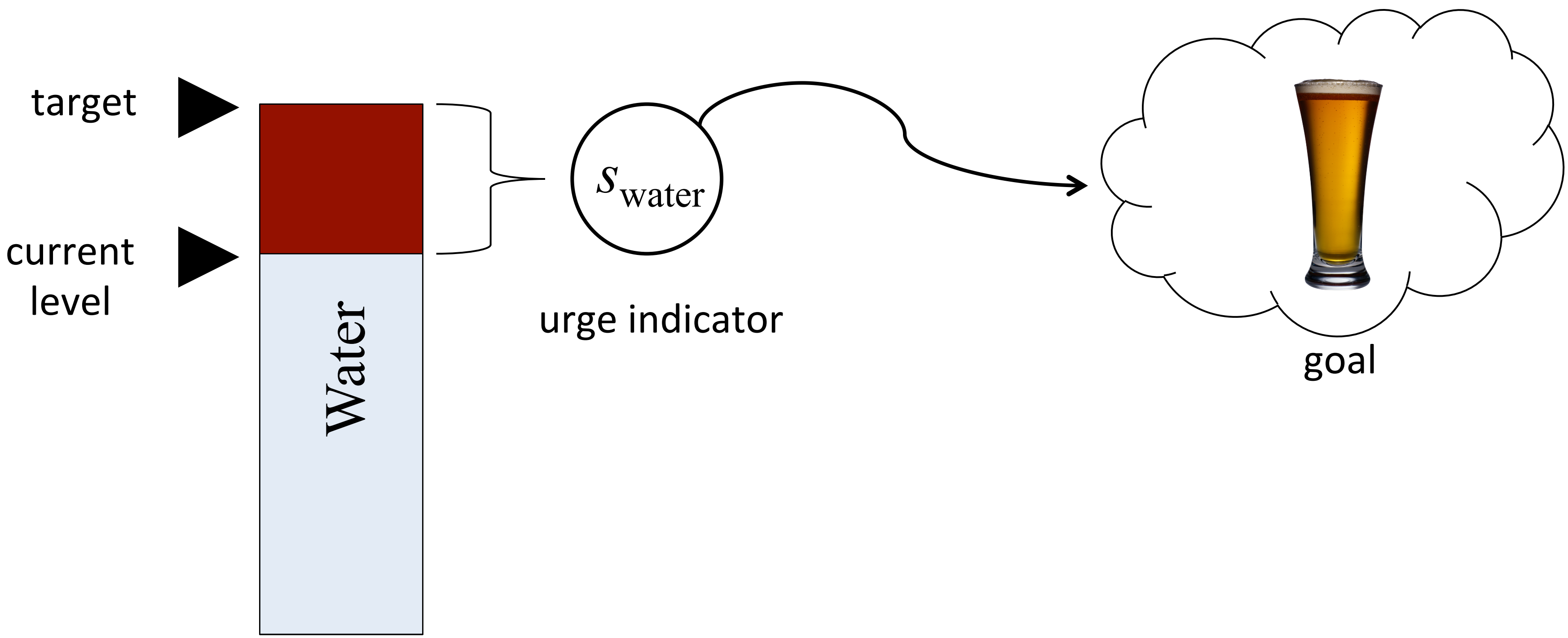
Motivational System

- drive = demand + urge indicator



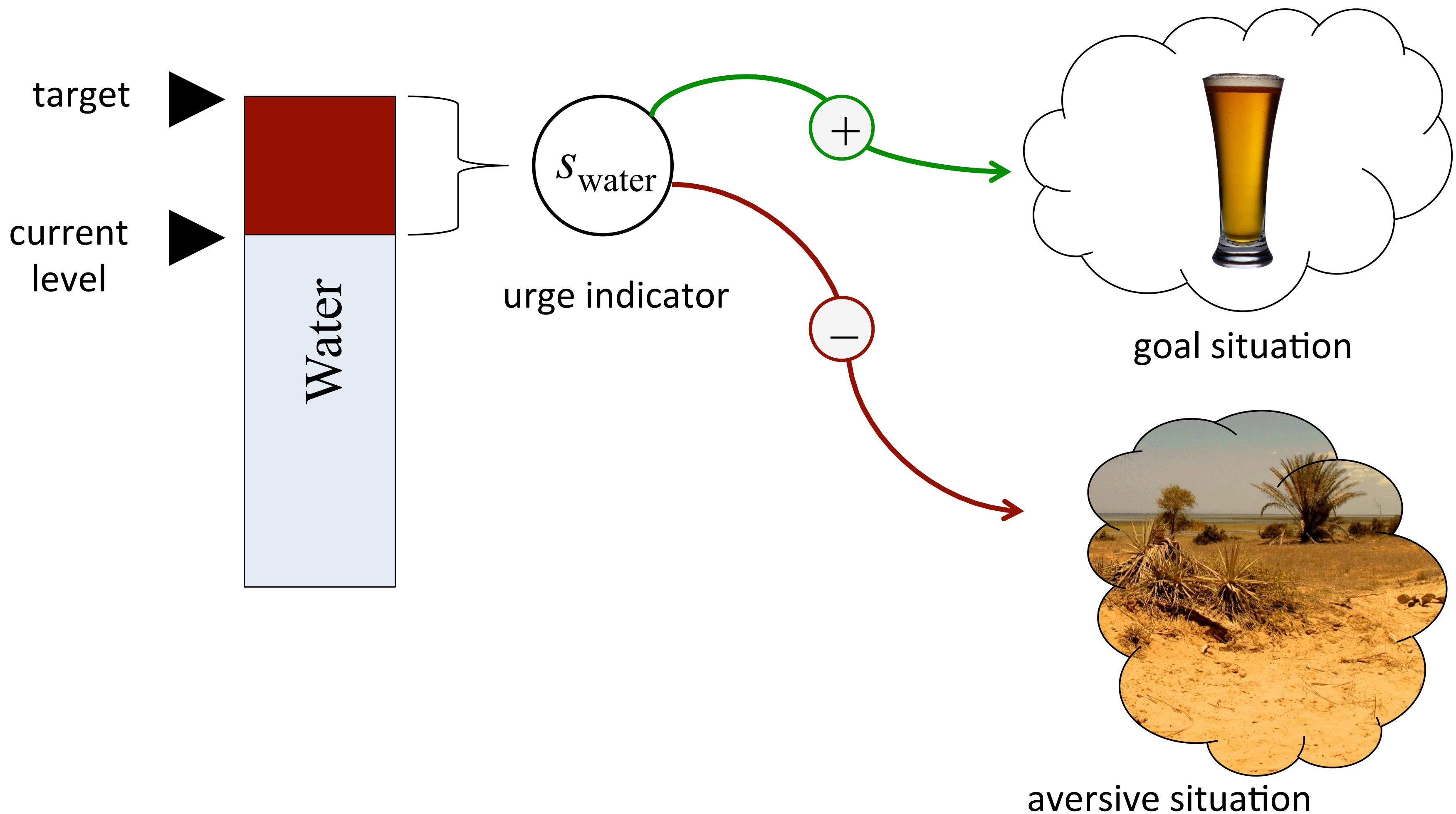
Motivational Learning

- motive = urge + goal situation



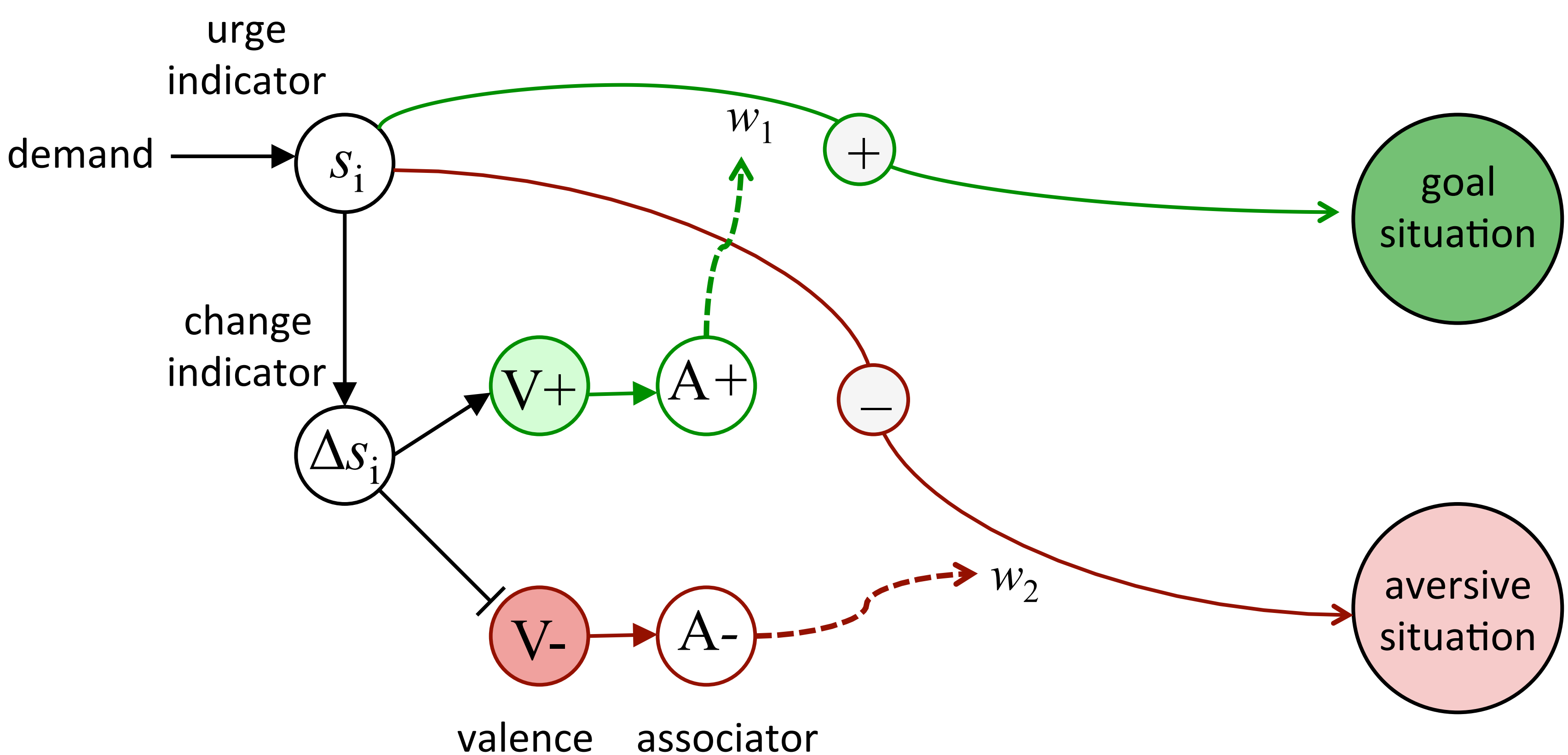
Motivational Learning

- motive = urge + goal situation



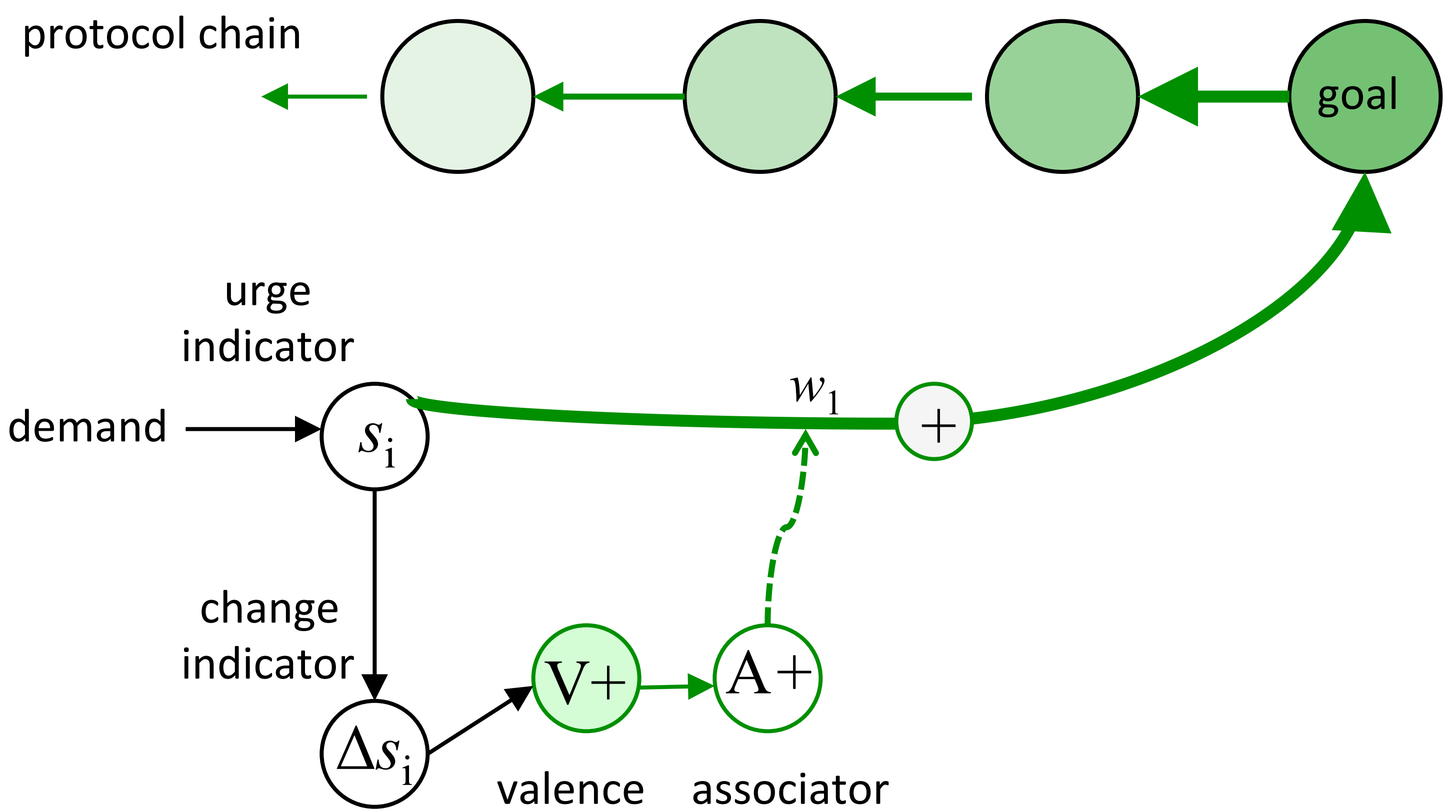
Motivational Learning

- association by learning:



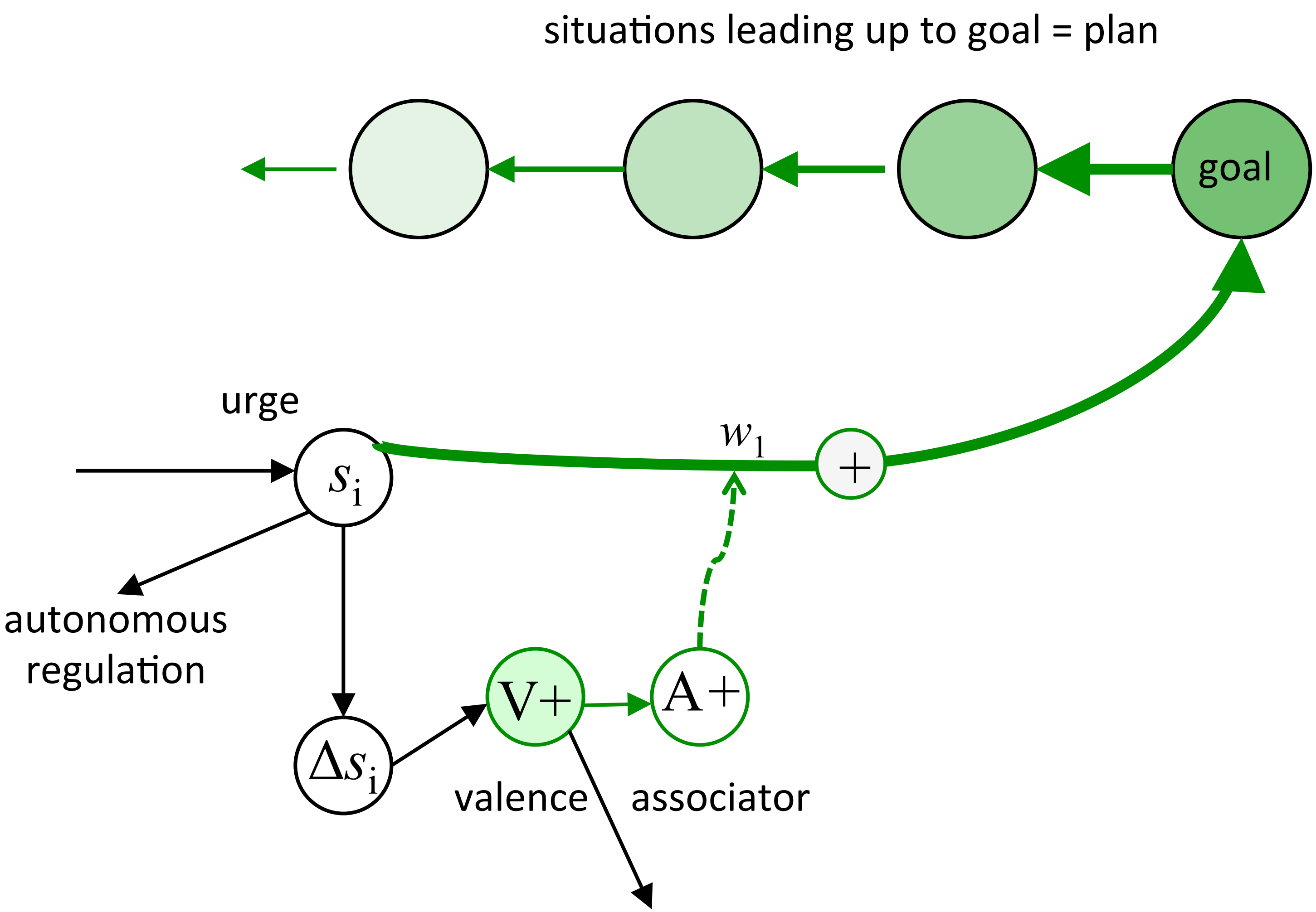
Motivational Learning

- retrogradient reinforcement



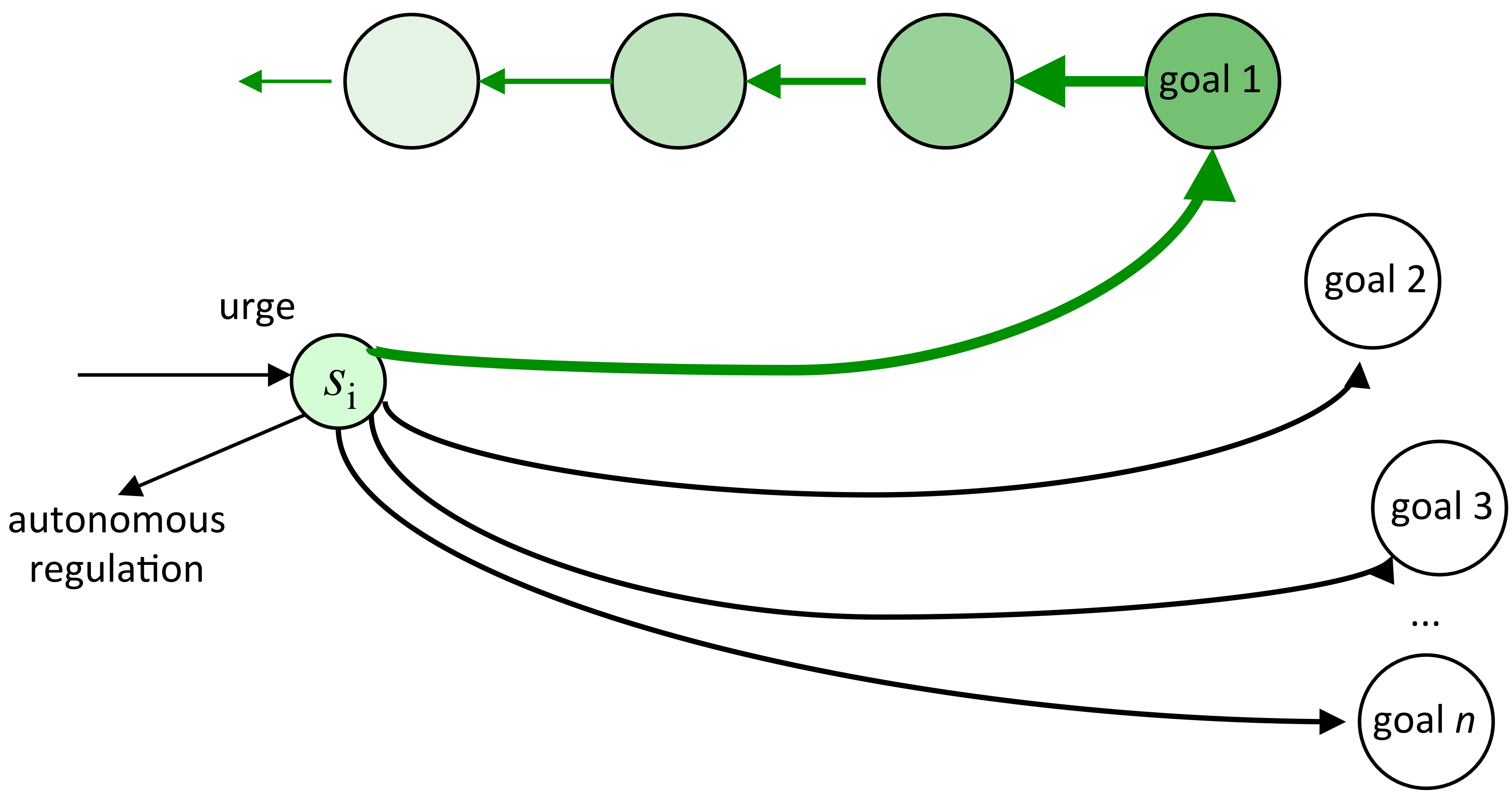
Motivational Learning

Motivator:



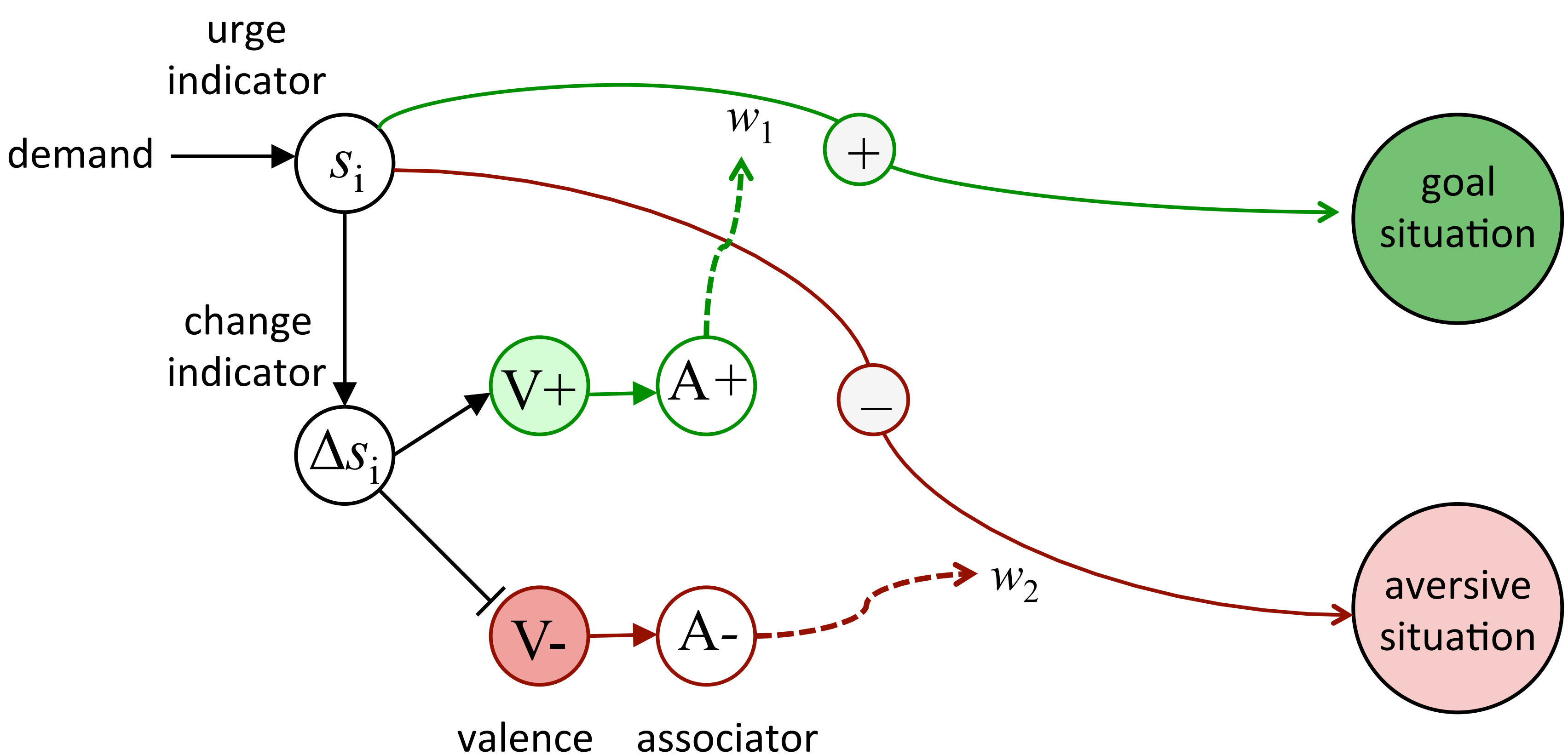
Motivational Learning

Intention:



Motivational Learning

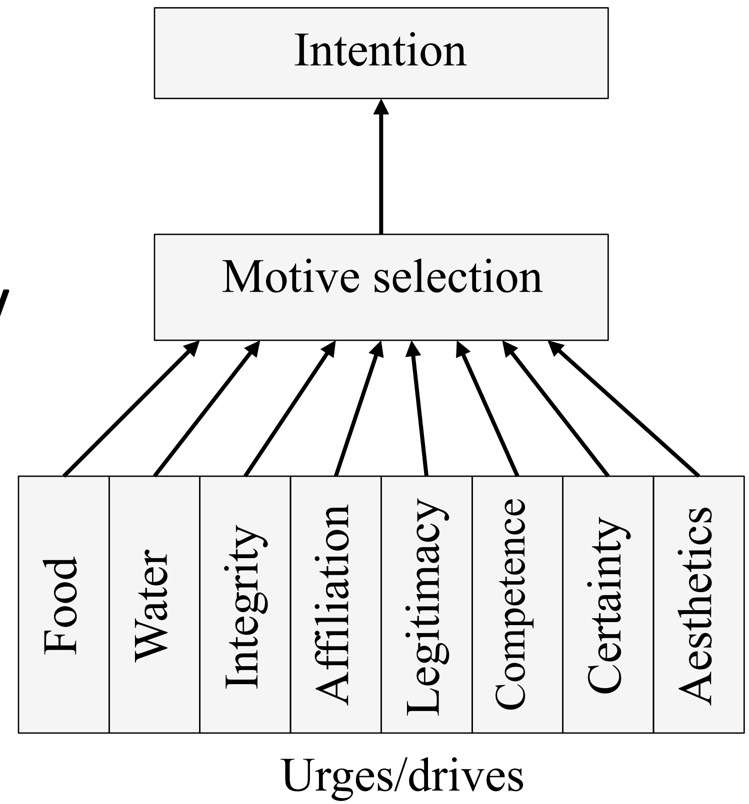
- association by learning:



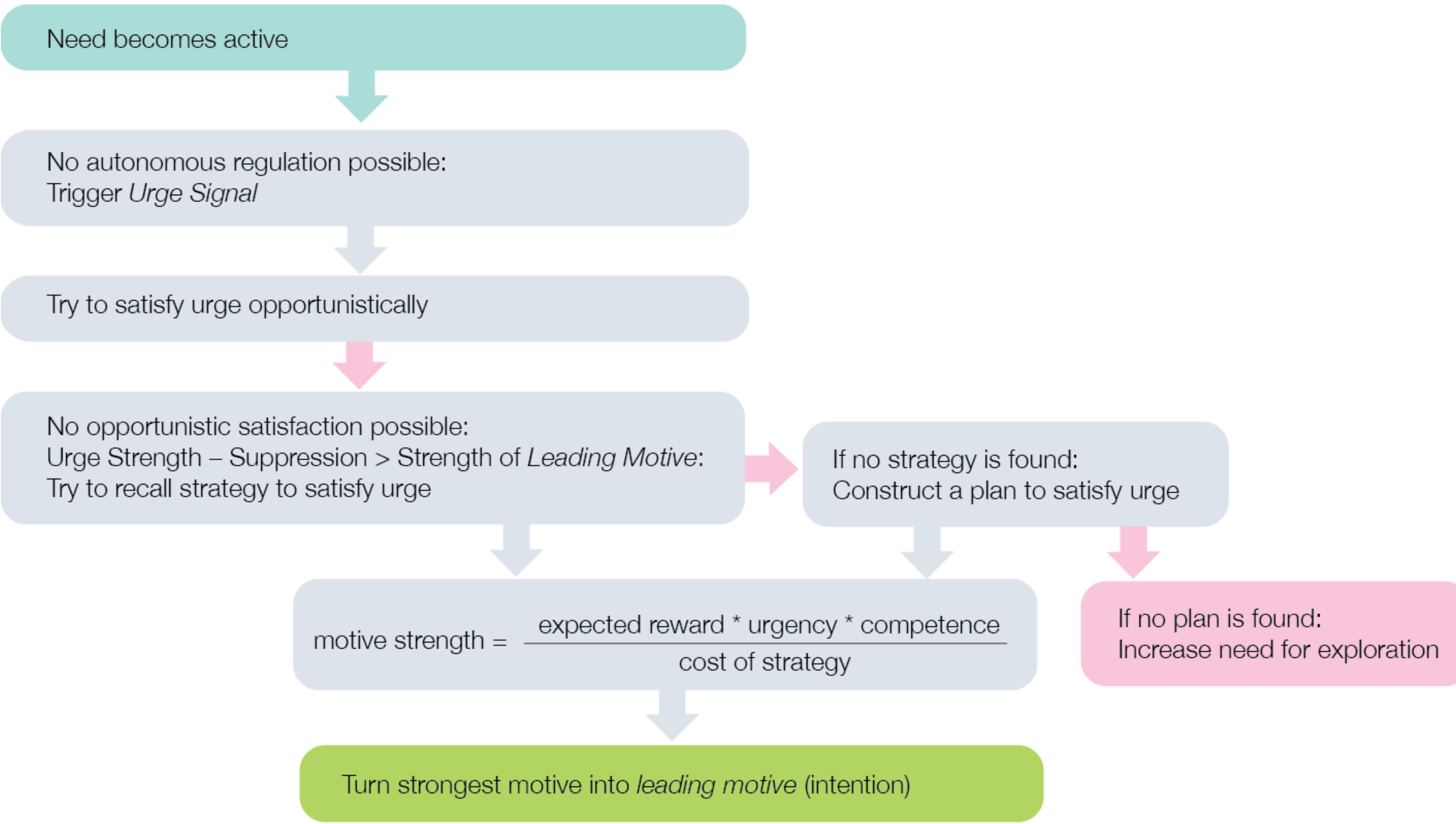
Motive Selection

A motive is raised to an intention based on:

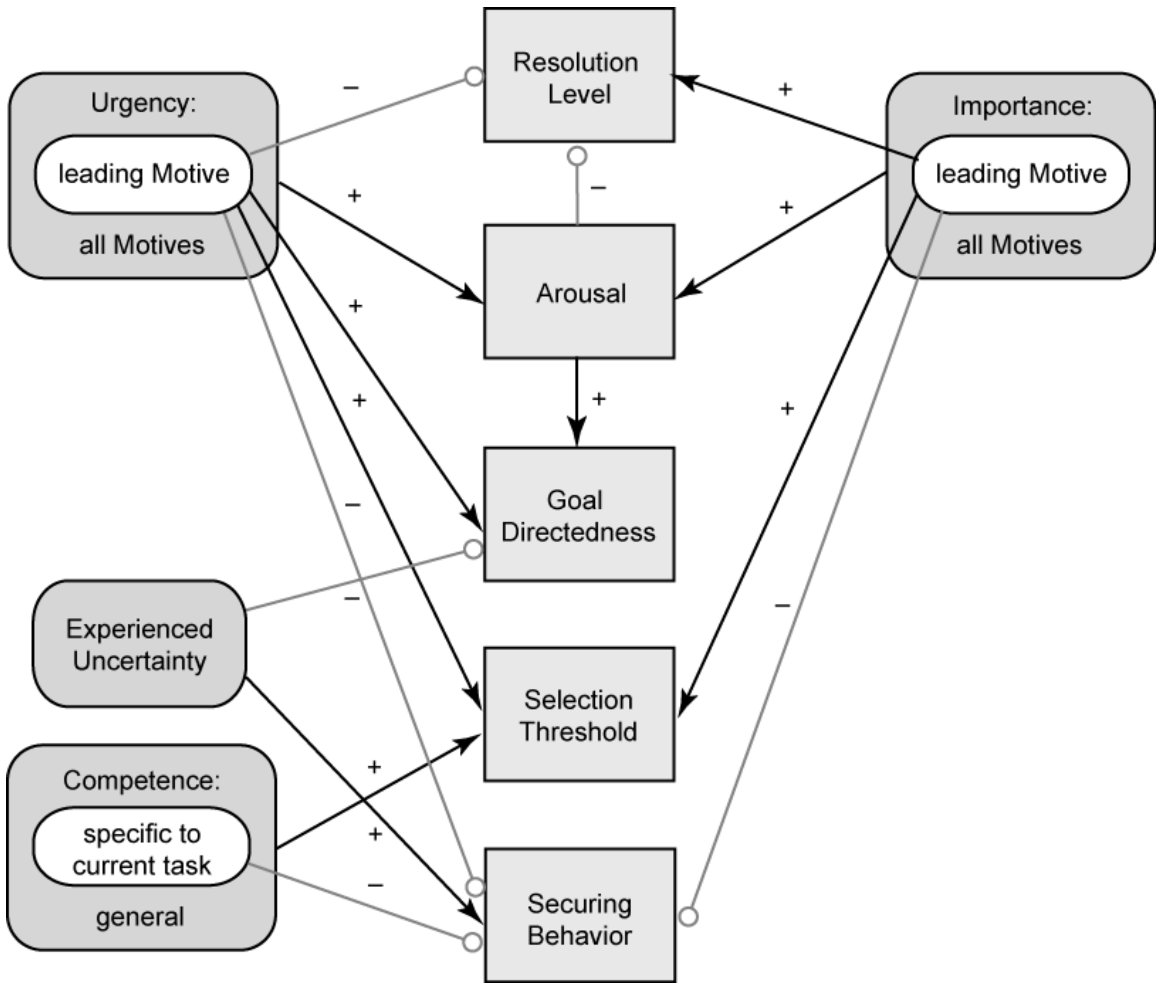
- relative strength of urge
  $\max(s_1, s_2, s_3, \ldots s_n)$
- opportunity
- expected success probability (heuristics) → value/risk
- selection threshold (adaptive; urgency)



Motive selection



Motivation and Modulator Dynamics



Emotions as directed affect + Modulation

Examples:

Fear: anticipation of aversive events (→ neg. valence) + arousal

Anxiety: uncertainty (→ neg. valence) + low competence + arousal, high securing behavior (frequent background checks)

Emotions as directed affect + Modulation

Examples:

Anger: Perceived obstacle (usually agent) manifestly prevented reaching of an active, motivationally relevant goal (→neg. valence), sanctioning behavior tendency (→ goal relevance is re-directed to sanctioning of obstacle), arousal, low resolution level, high action readyness, high selection threshold

Sadness: Manifest prevention from *all* conceived ways of reaching active, relevant goal, without relevant obstacle (→ neg. valence), support-seeking behavior (by increased demand for affiliation), low arousal, inhibition of active goal → decreased action readyness

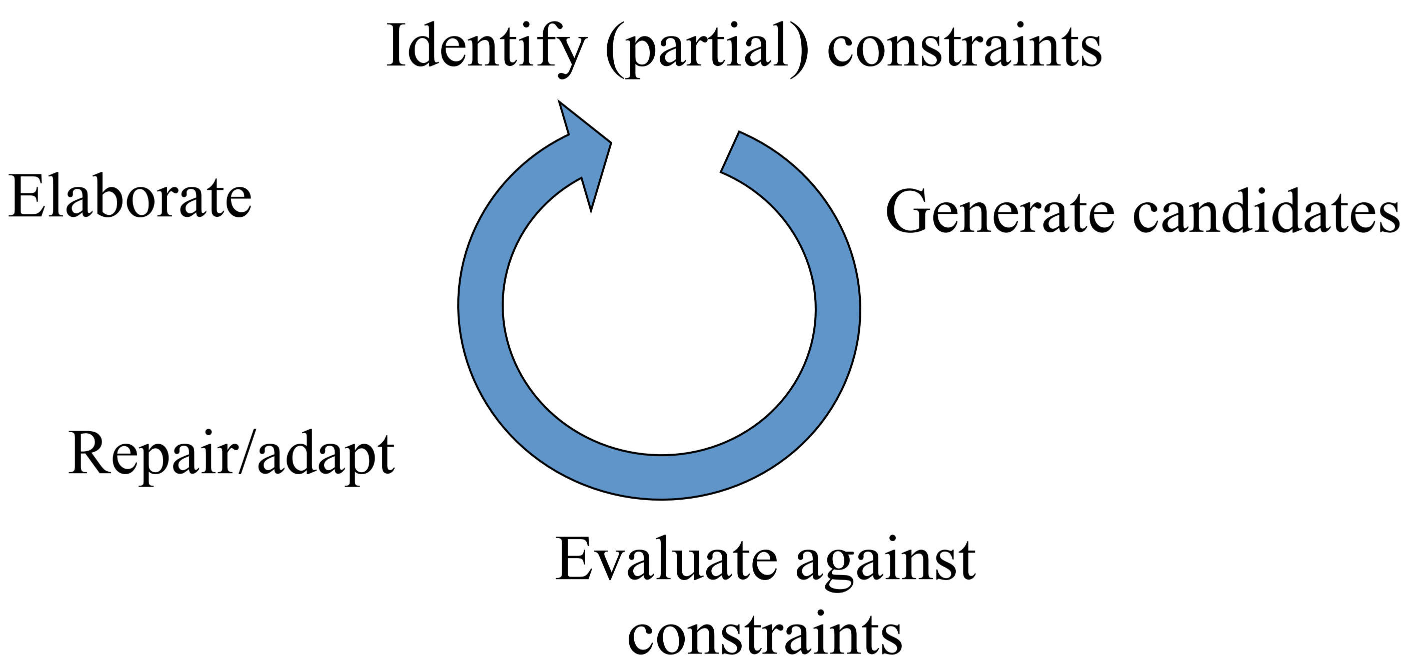
Emotions as directed affect + Modulation

Examples:

- **Pride**: high competence (→ low securing rate), high internal legitimacy, likely coincidence with high external legitimacy
- **Joy**: high arousal + high perceived reward signal from satisfying a demand
- **Bliss**: low arousal + high perceived reward signal from satisfying a demand (since physiological demands often involve high arousal, mostly related to cognitive demands, such as aesthetics)

Example: computational creativity

- Creativity involves exploration of space of possible solutions to an incrementally defined problem

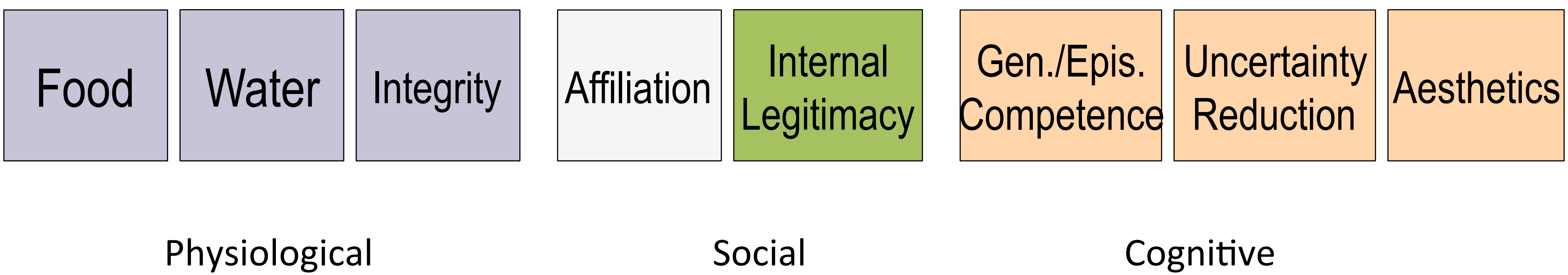


Example: computational creativity

- Creativity as a learning problem:
  - optimize problem representation
  - optimize constraint identification
  - optimize solution traversal
- Source of reward signal:
  - abstract aesthetics (better representations)
  - uncertainty reduction (problem space exploration)
  - competence (specific problem solving skills)
- “Eureka” moment: large “positive” delta during constraint match → reward signal

Example: Mapping to FFM (Big Five)

Demand dynamics:



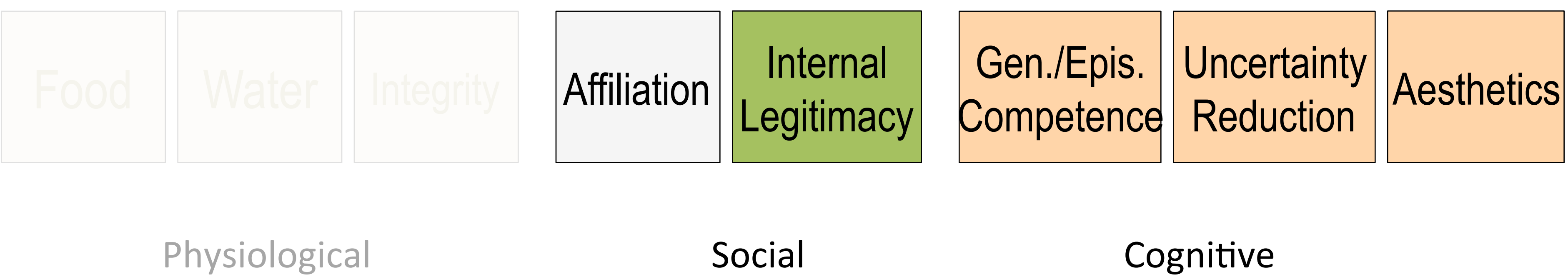
Individual Variations by Parameterizing

Possible grounding of personality properties (FFM):

- **Openness**: appreciation of art and new ideas, curiousity
- **Conscientiousness**: rulefollowing vs. chaotic
- **Extraversion**: tendency to seek stimulation by environment and others
- **Agreeableness**: tendency for cooperativeness and compassion
- **Neuroticism**: emotional stability, effect of failure to self-confidence

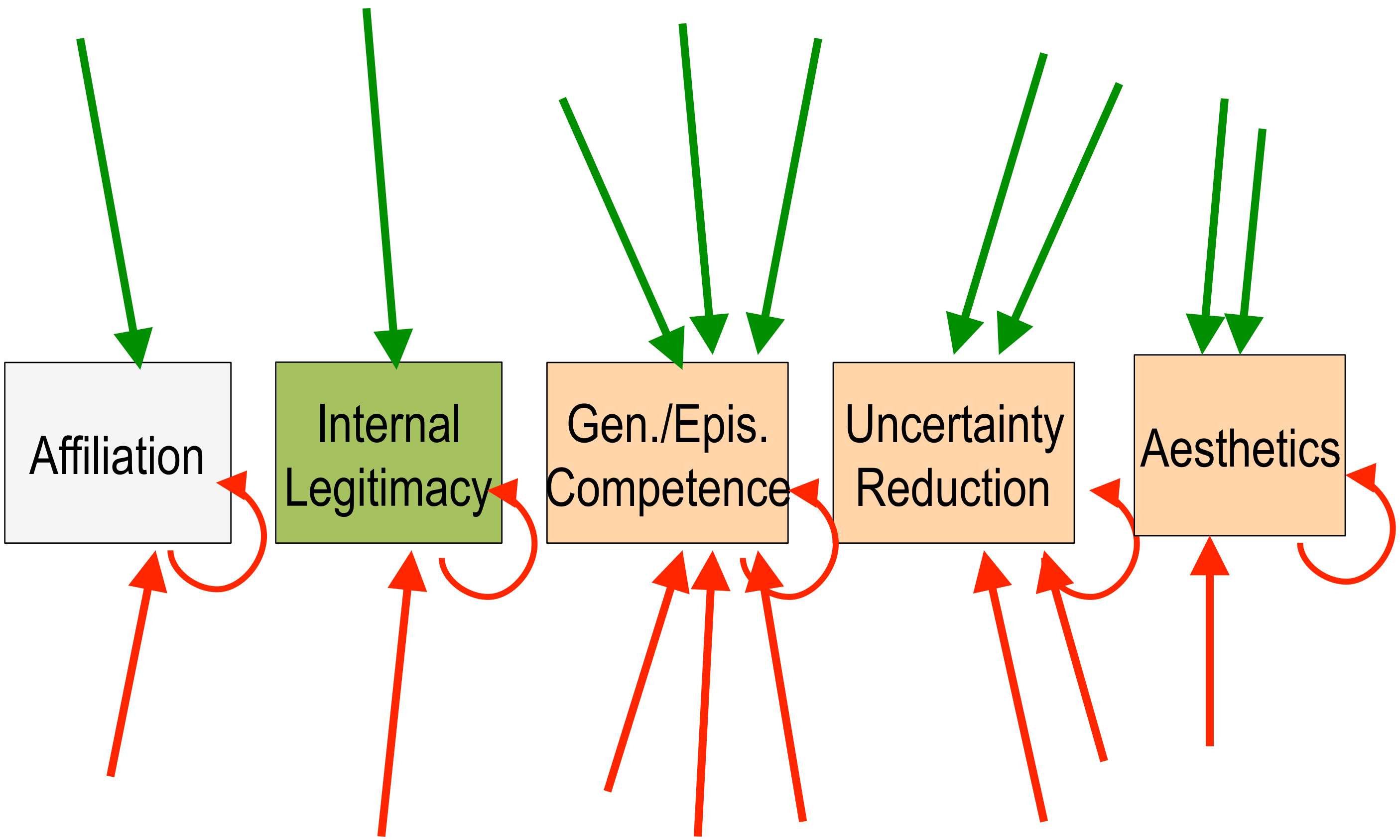
Example: Mapping to FFM (Big Five)

Demand dynamics:



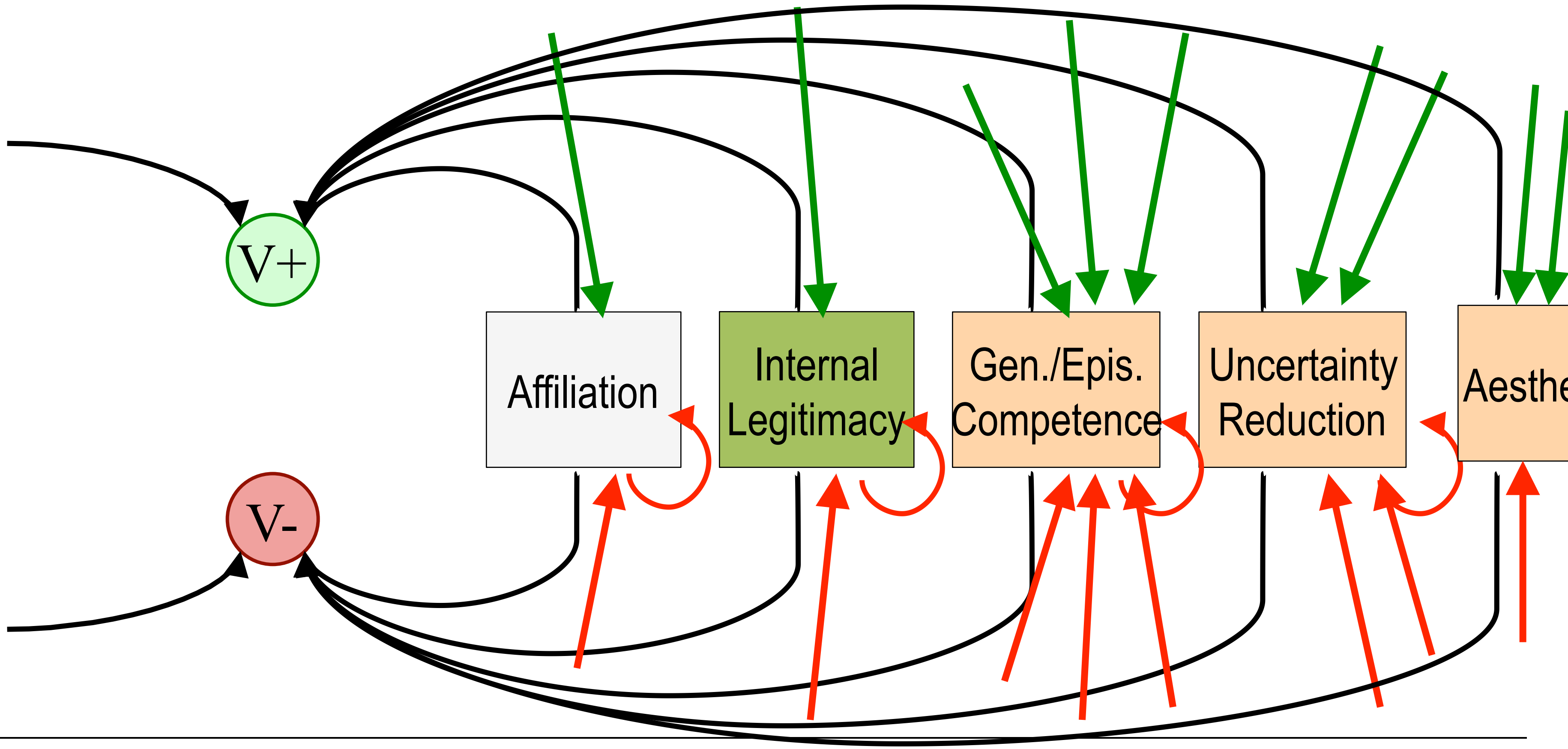
Example: Mapping to FFM (Big Five)

Demand dynamics:



Example: Mapping to FFM (Big Five)

Valence: Pleasure/Pain signals



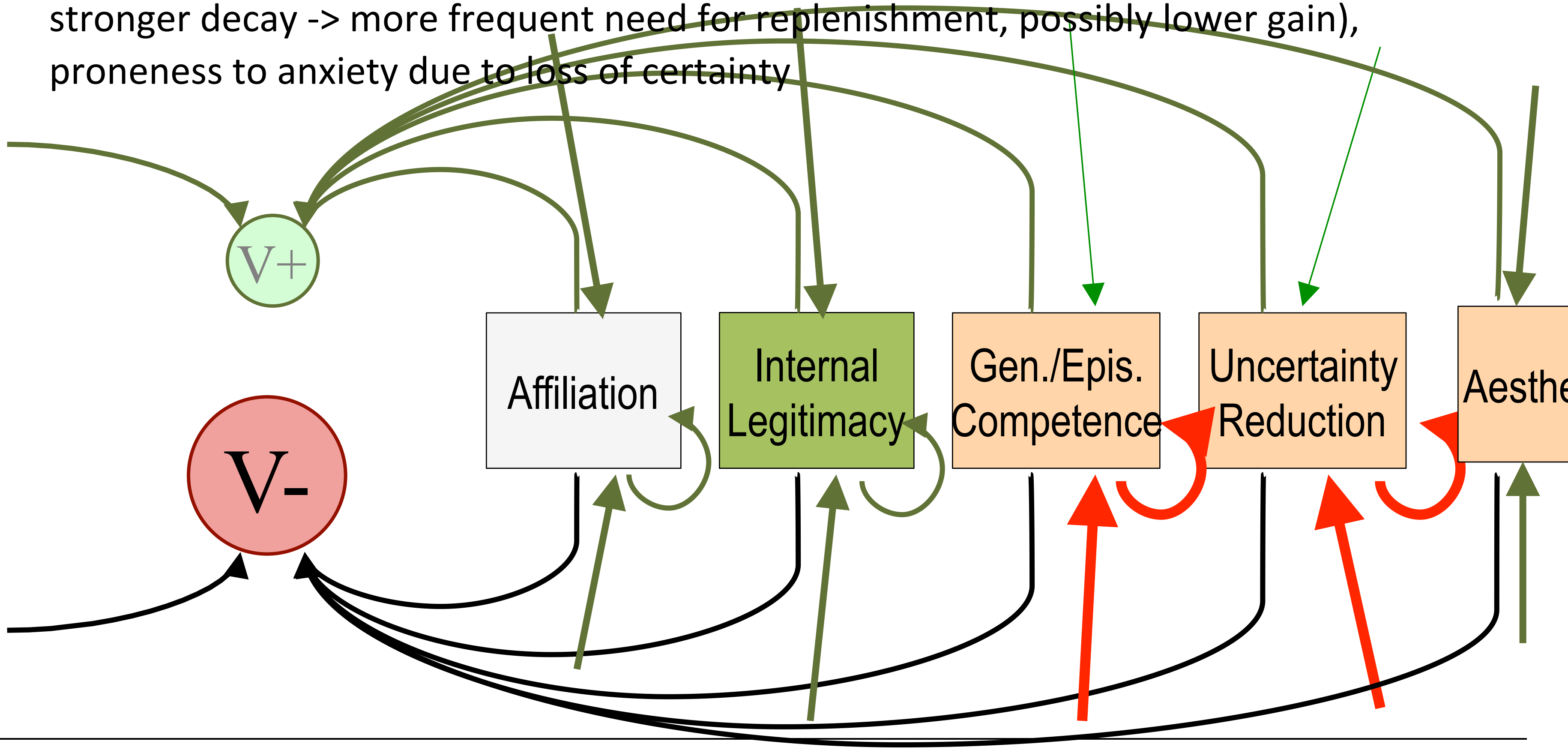
Example: Mapping to FFM (Big Five)

Neuroticism: stronger experience of negative emotions, lower emotional stability

(strong negative reward/stronger loss of competence, certainty;

stronger decay -> more frequent need for replenishment, possibly lower gain),

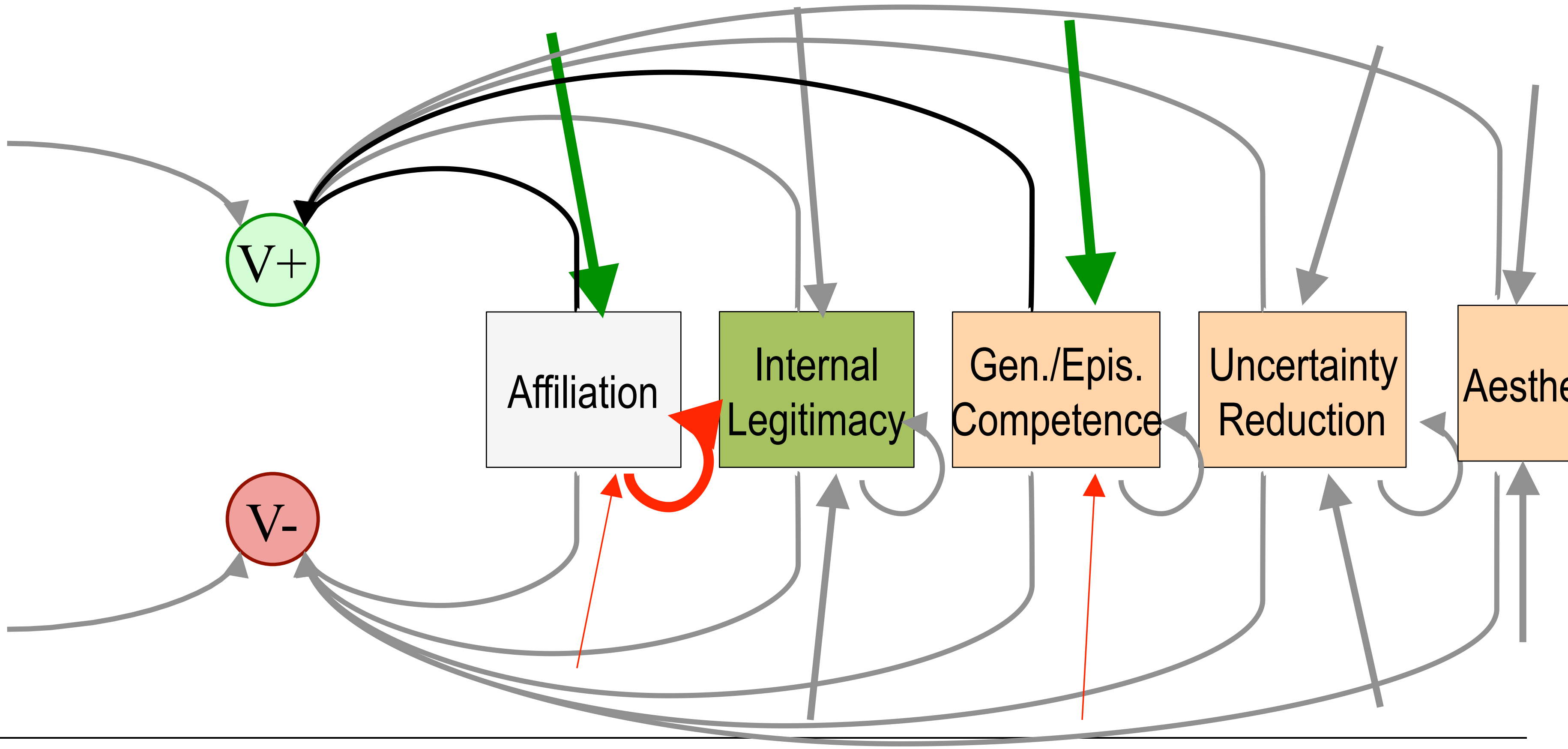
proneness to anxiety due to loss of certainty



Example: Mapping to FFM (Big Five)

Extraversion: surgency, activity in social relations, expressivity

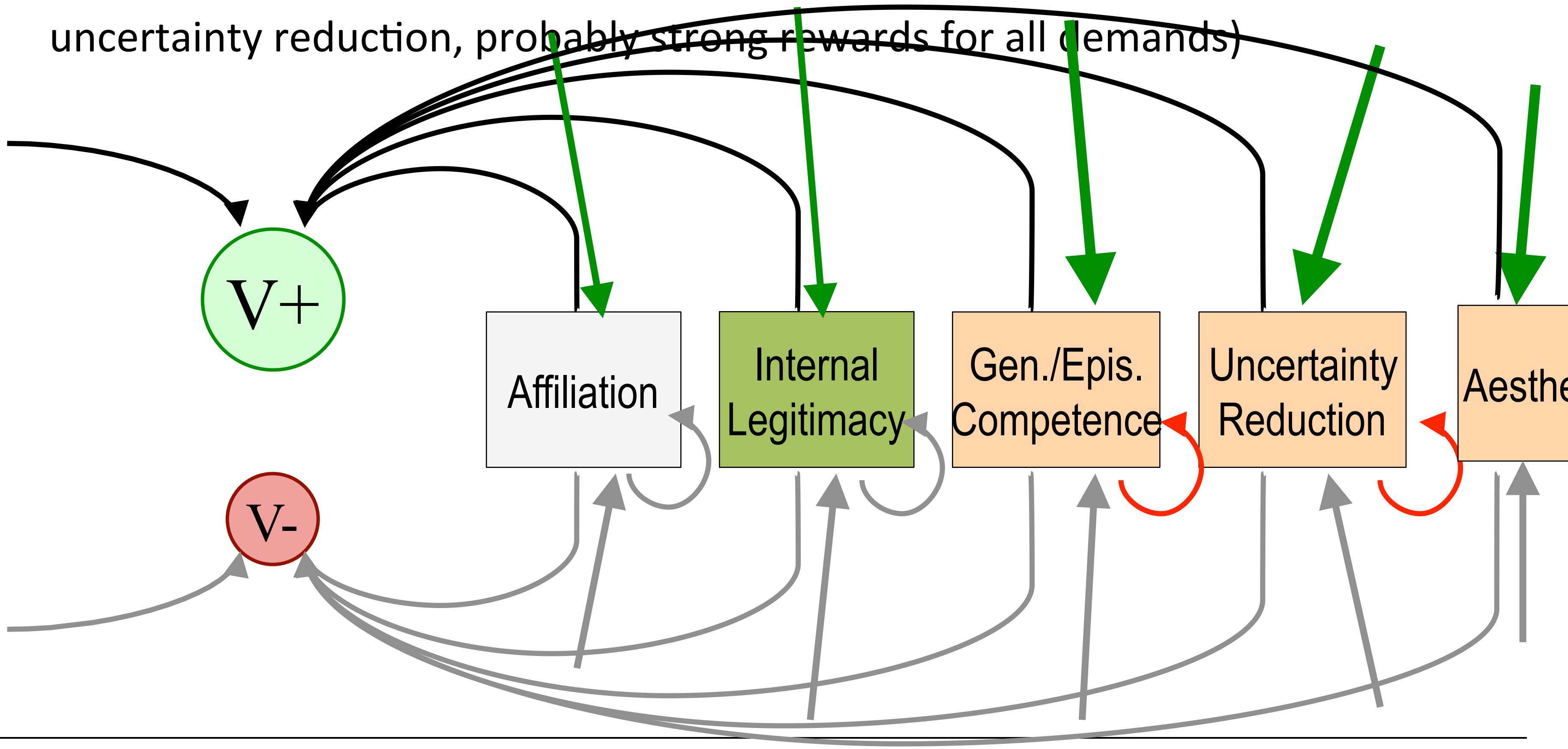
(strong gain for affiliation and competence, high decay of affiliation)



Example: Mapping to FFM (Big Five)

Openness: desire for novelty, intellectual independence, non-conservatism, appreciation for art and new ideas

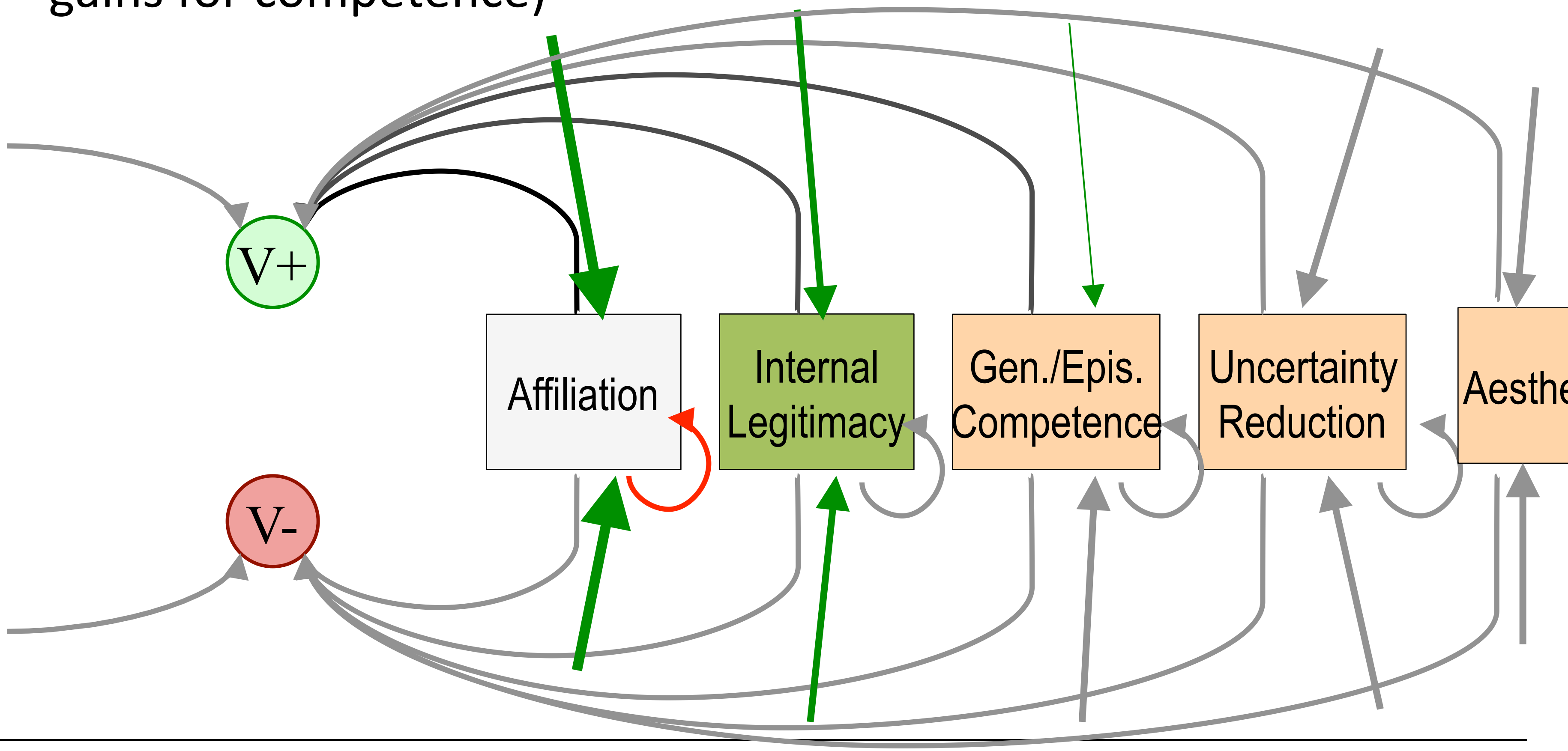
(strong gain for uncertainty reduction, high epistemic competence for uncertainty reduction, probably strong rewards for all demands)



Example: Mapping to FFM (Big Five)

Agreeableness:

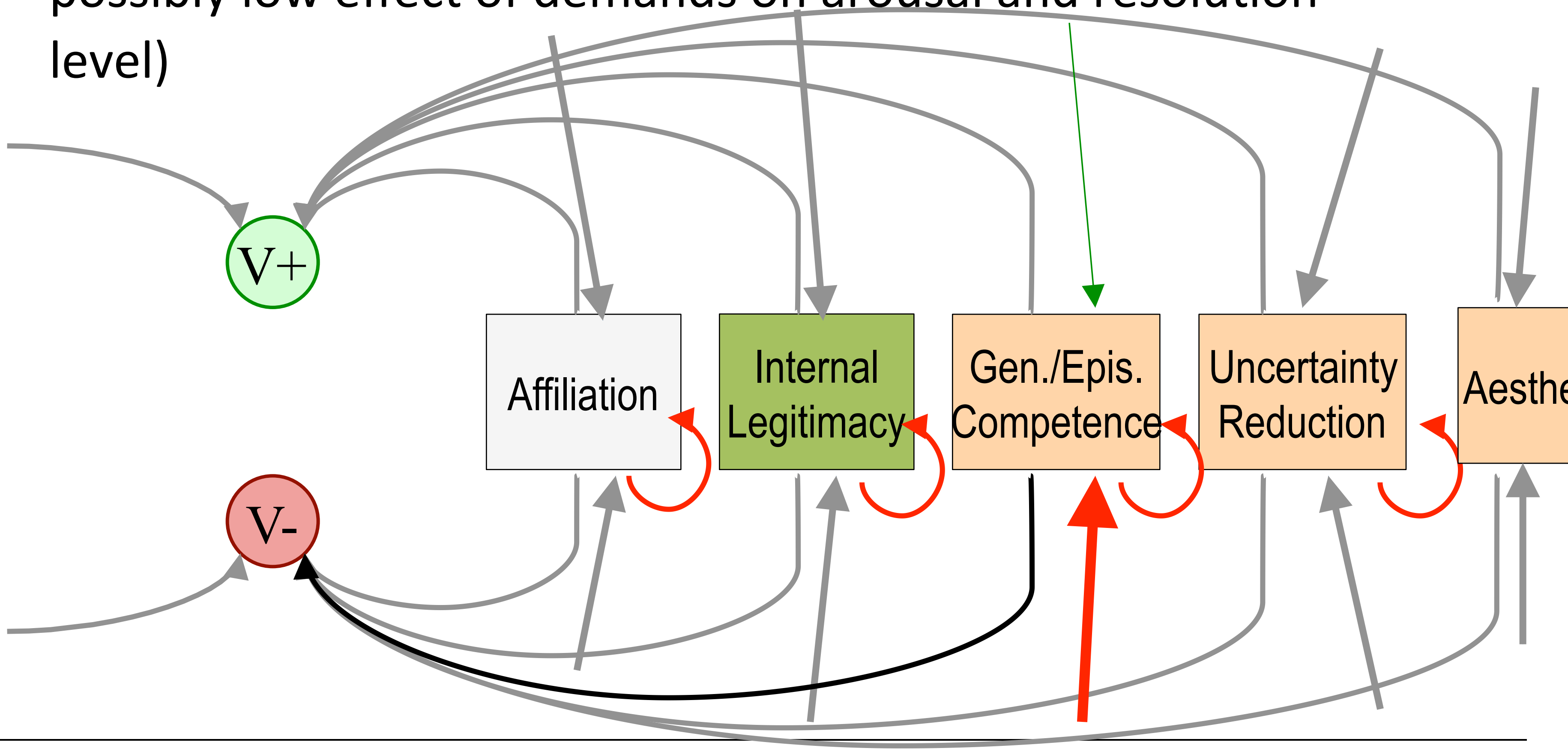
(strong positive and negative reward for affiliation, lower gains for competence)



Example: Mapping to FFM (Big Five)

Conscientousness, Rigidity:

(high loss in competence, high selection threshold, possibly low effect of demands on arousal and resolution level)



Thank you!

- Project home: micropsi.com
- Questions, ideas: joscha@mit.edu