Joscha Bach

Implementing Emotion and Motivation in AI Architectures

Tutorial AAAI 2018

From Computation to Emotion

How can a computational system experience emotion?

What is emotion?

What is experience?

Relationship between experiencer and experience?

Emotional states as cognitive configurations

Emotions as cognitive configurations

Emotions as cognitive configurations

Overview

- AGI perspective on minds
- Basic architectural components
- Modeling motivation (MicroPsi model)
- Models of emotion
- Emotion in the Psi theory
- Modeling personality
- Emotion, self and prosociality

Basic perspective on general AI

- Mind as machine

- Machine = computational system
- computation = regular state change

- universal computation: set of computable functions that can compute all computable functions (when given unbounded resources)

Basic perspective on general AI

- Access to universe via discernible differences
  $\rightarrow$ Information
- Meaning of information:
  relationship to changes in other information
- Intelligence: ability to model
- Modeling is function approximation
- Purpose of modeling is regulation,
  to maximize rewards

Basic perspective on general AI

- Classical AI:
  direct modeling of cognitive functionality
  $\rightarrow$ "first order AI"
- Deep Learning:
  systems that model functionality themselves;
  compositional function approximation
  $\rightarrow$ "second order AI"
- Meta Learning:
  systems that learn how to build learning systems
  $\rightarrow$ "third order AI"

Basic perspective on general AI: open questions

- Are humans meta learning systems?
- Is evolution a (slow and ineffective) search for meta learners?

- Is there a class of universal function approximators that can approximate any function that can be approximated by a computer (when given unbounded resources),
  and does it contain itself and us?

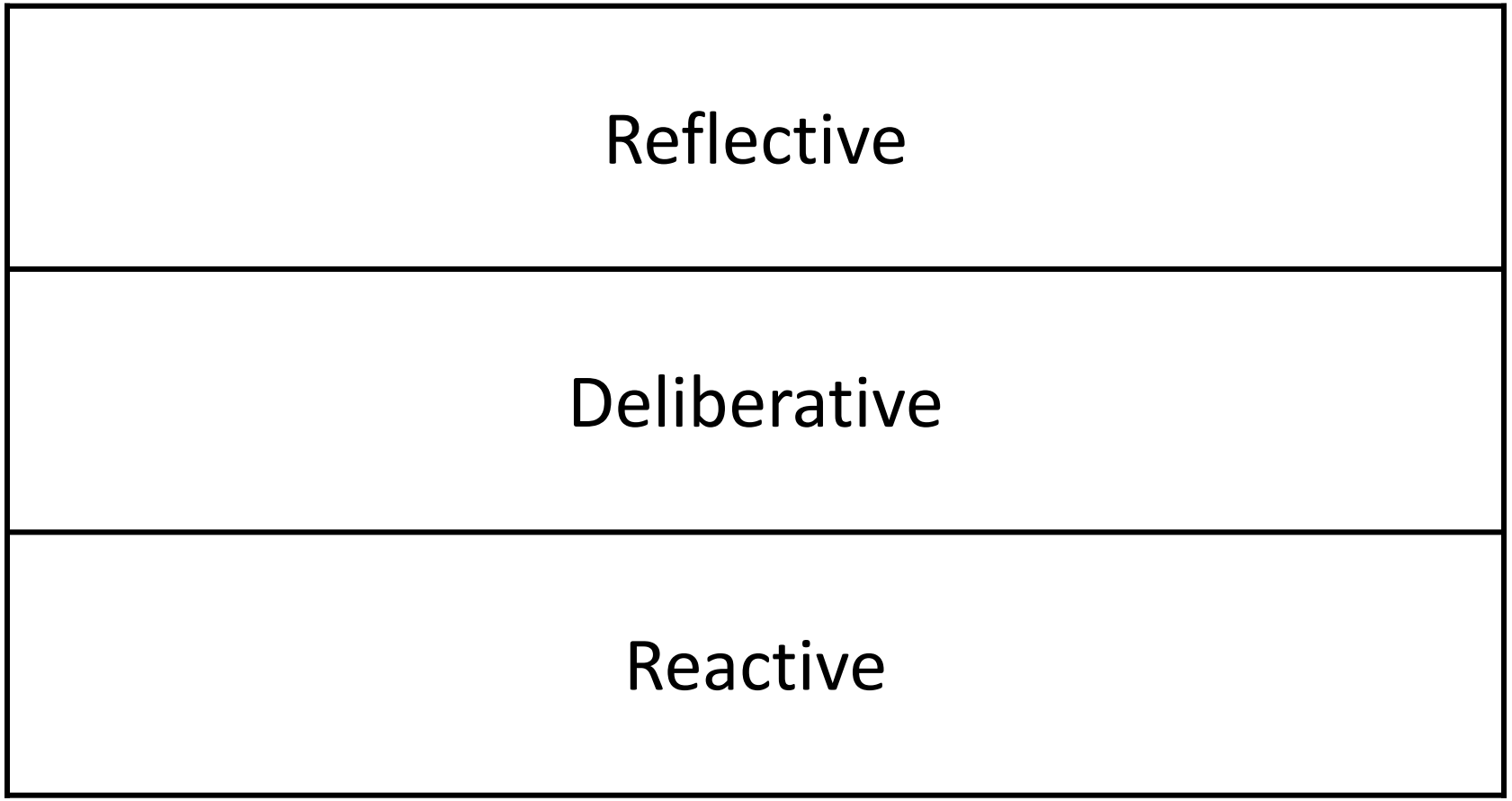
AI perspective on the mind

- Constructed architectures (Minsky, Simon, Newell) — classical cognitive architectures

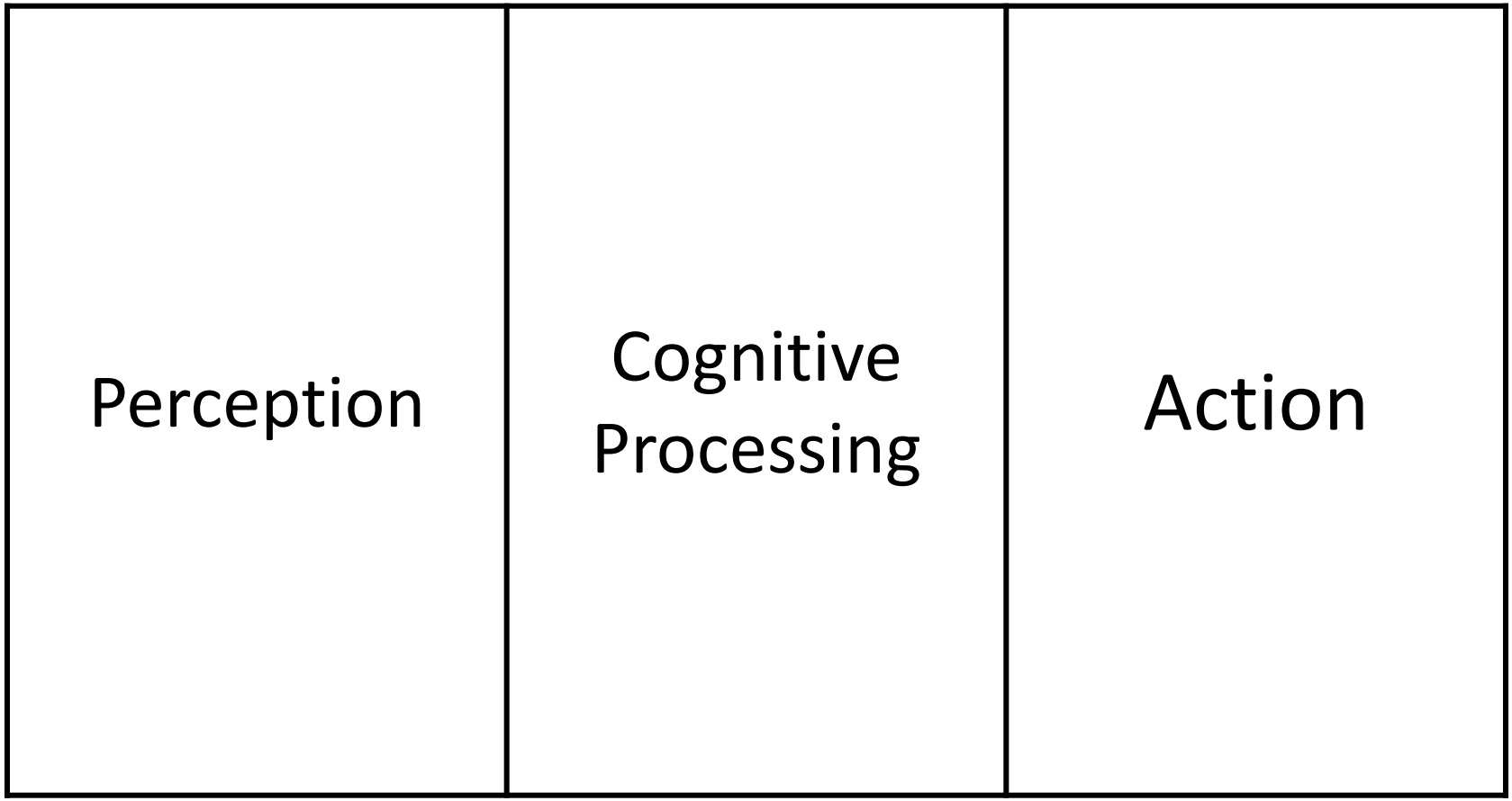
- Generated architectures (Schmidhuber, Hutter, …): general recursive function approximation + reward system

- Hybrid perspective:
  mostly generated, but with complex prerequisites and biases

Layers of Cognition



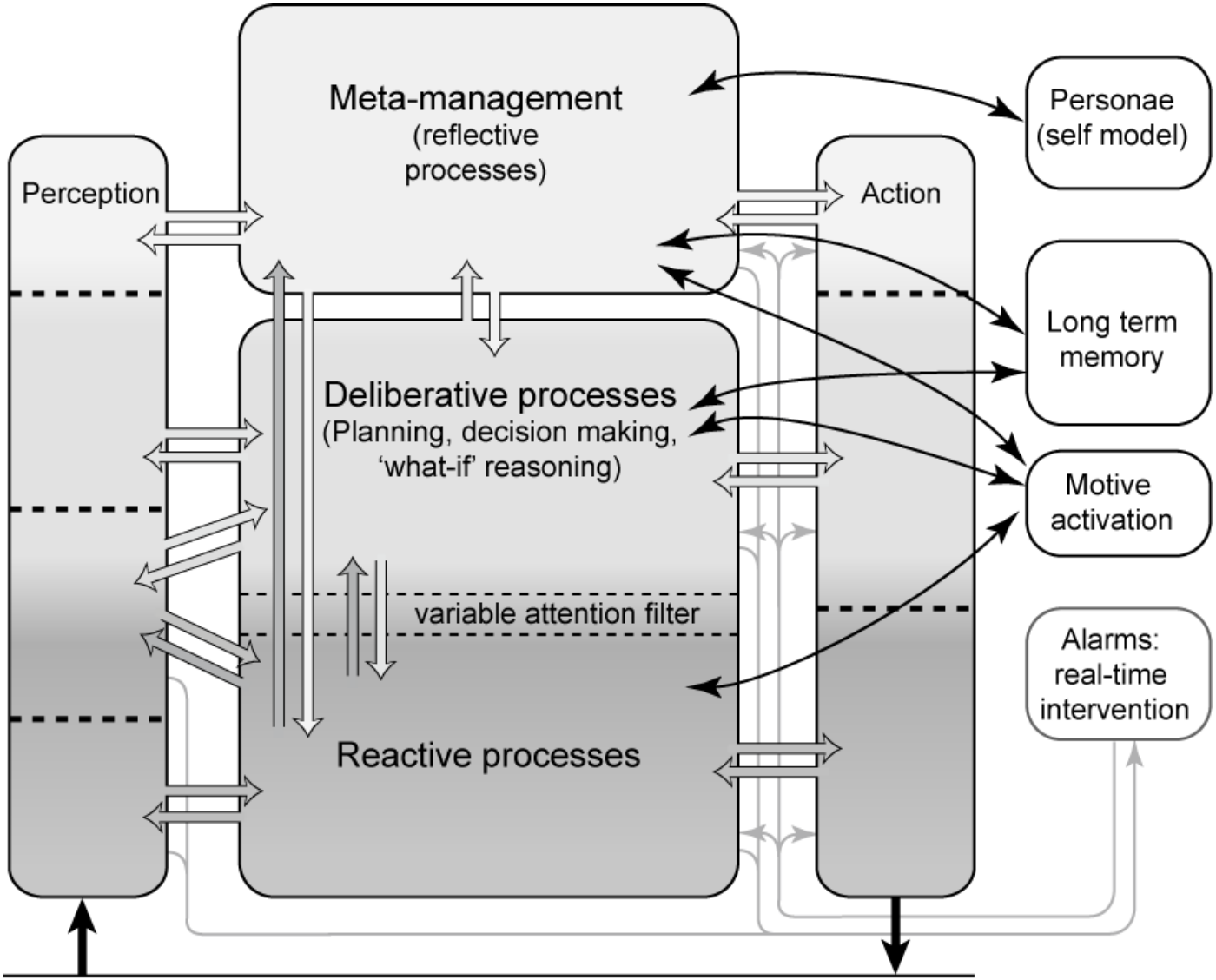
Columns of Cognition



Cognitive Grid

Reflexive Perception	Meta-Management	Management Action
Deliberative Perception	Planning, Reasoning	Deliberative Action
Reactive Perception	Reflexes	Reflexive Action

Conceptual Analysis: HCogAff (Sloman 2001)



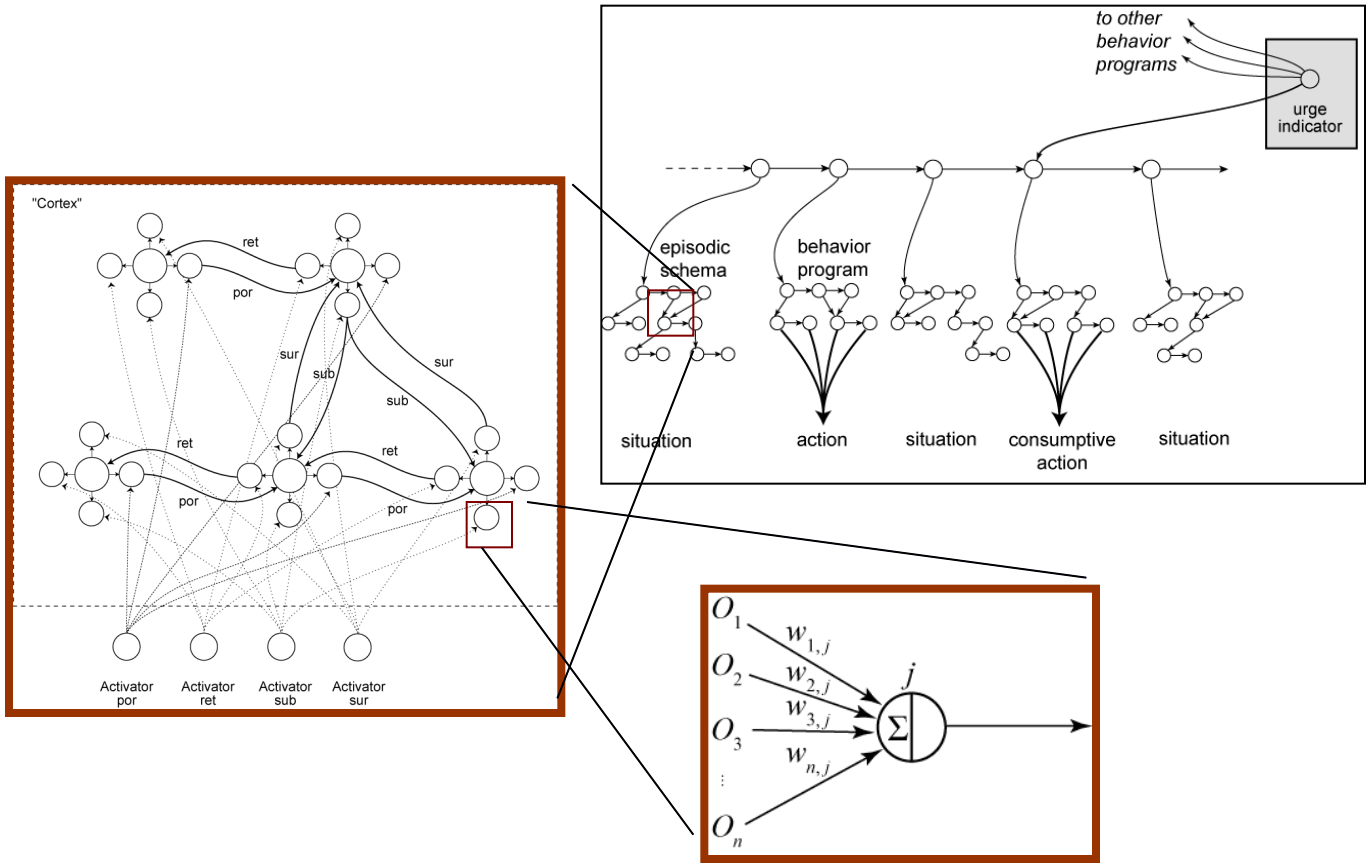
Artificial General Intelligence

Methods should focus on components and performances necessary for intelligence:

- **Whole, testable architectures**
- **Universal Representations:**
  Dynamic model of environment, possible worlds, and agent
- **(Semi-) Universal Problem Solving:**
  Learning, Planning, Reasoning, Analogies, Action Control, Reflection ...
- **Universal Motivation:**
  Polythematic, adaptive goal identification
- **Emotion and affect**

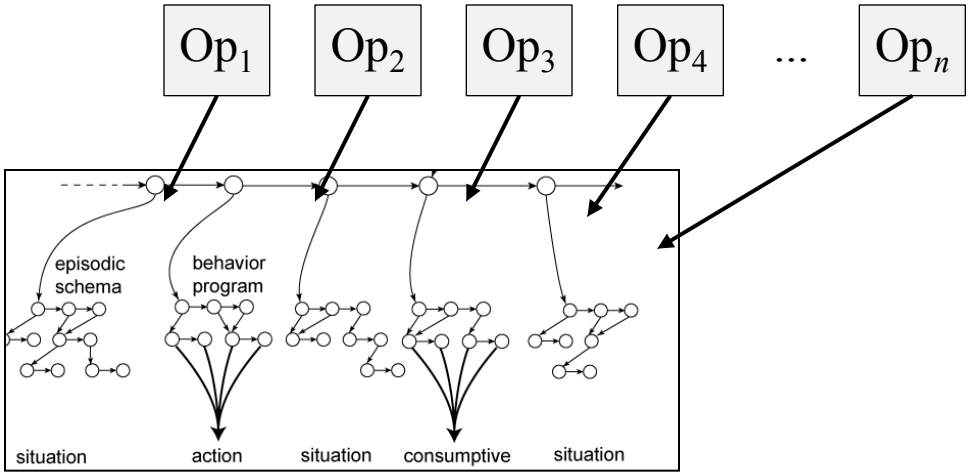
Components for Cognitive AI

- Universal mental representations
  (compositional + distributed → neurosymbolic)



Components for Cognitive AI

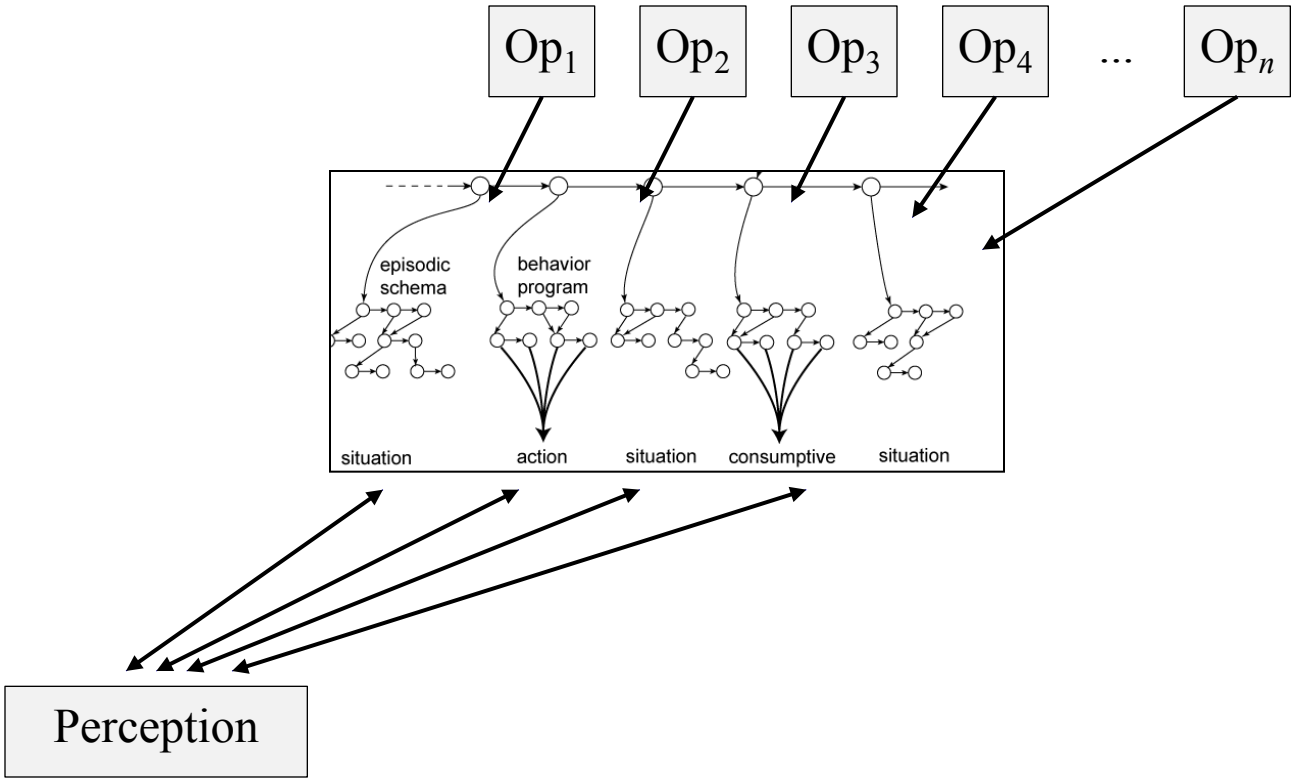
- (Semi-) General problem solving: Operations over these representations



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

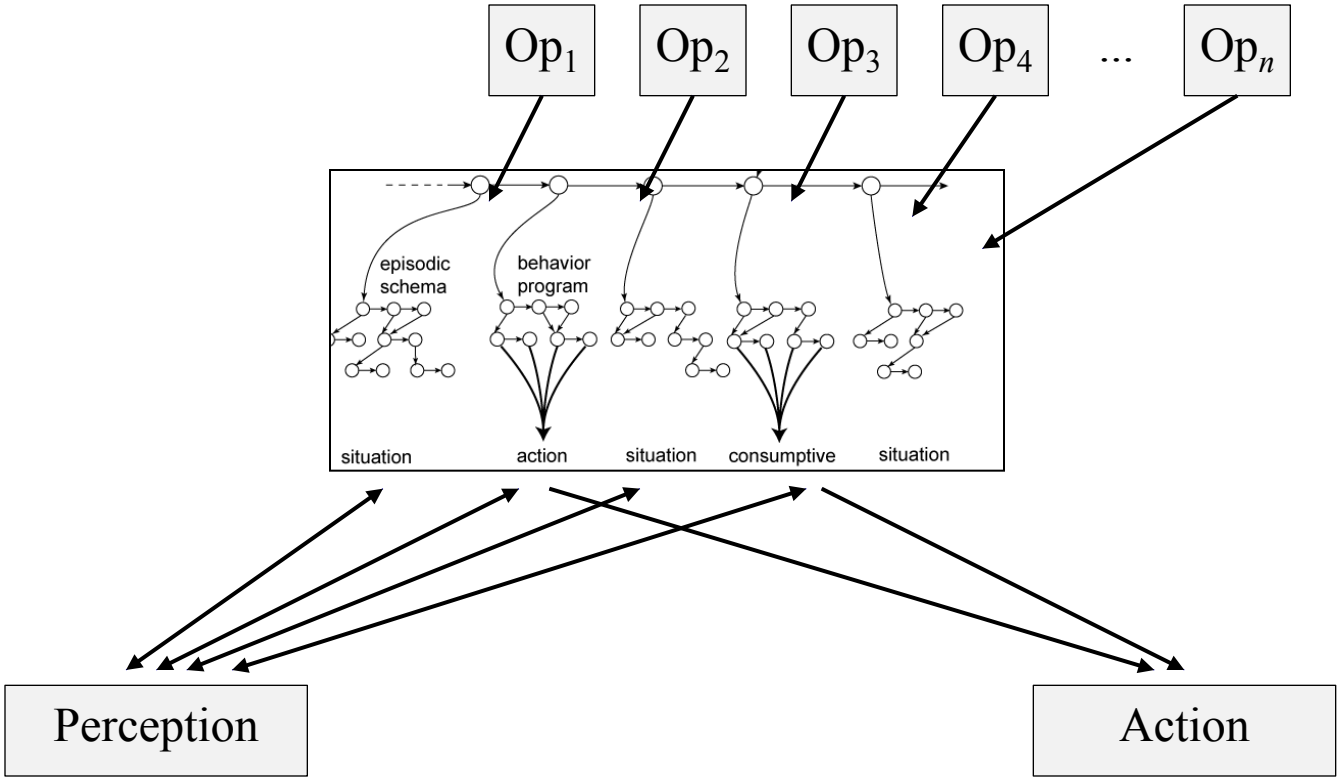
Components for Cognitive AI

- Perceptual grounding



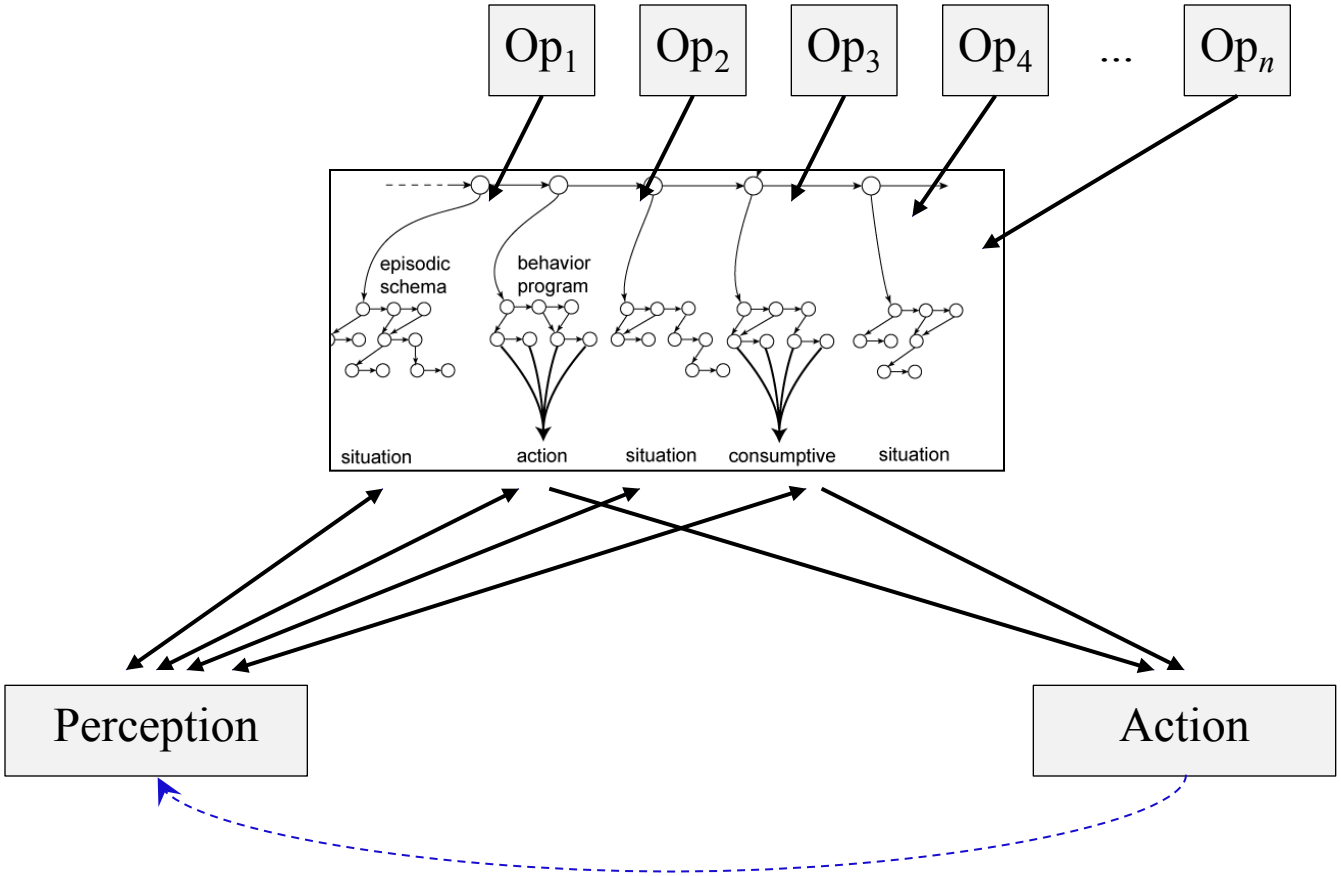
Components for Cognitive AI

- Perceptual grounding and action



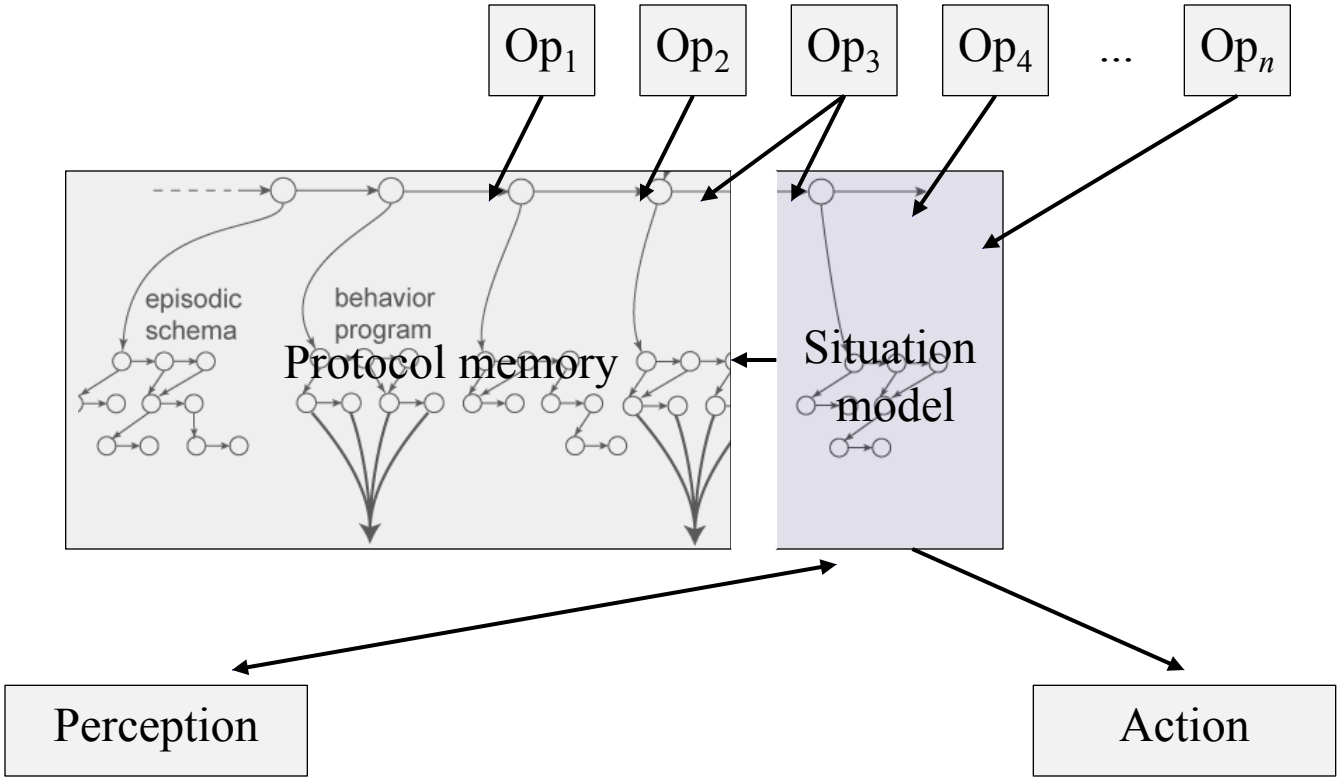
Components for Cognitive AI

- Perceptual grounding and action



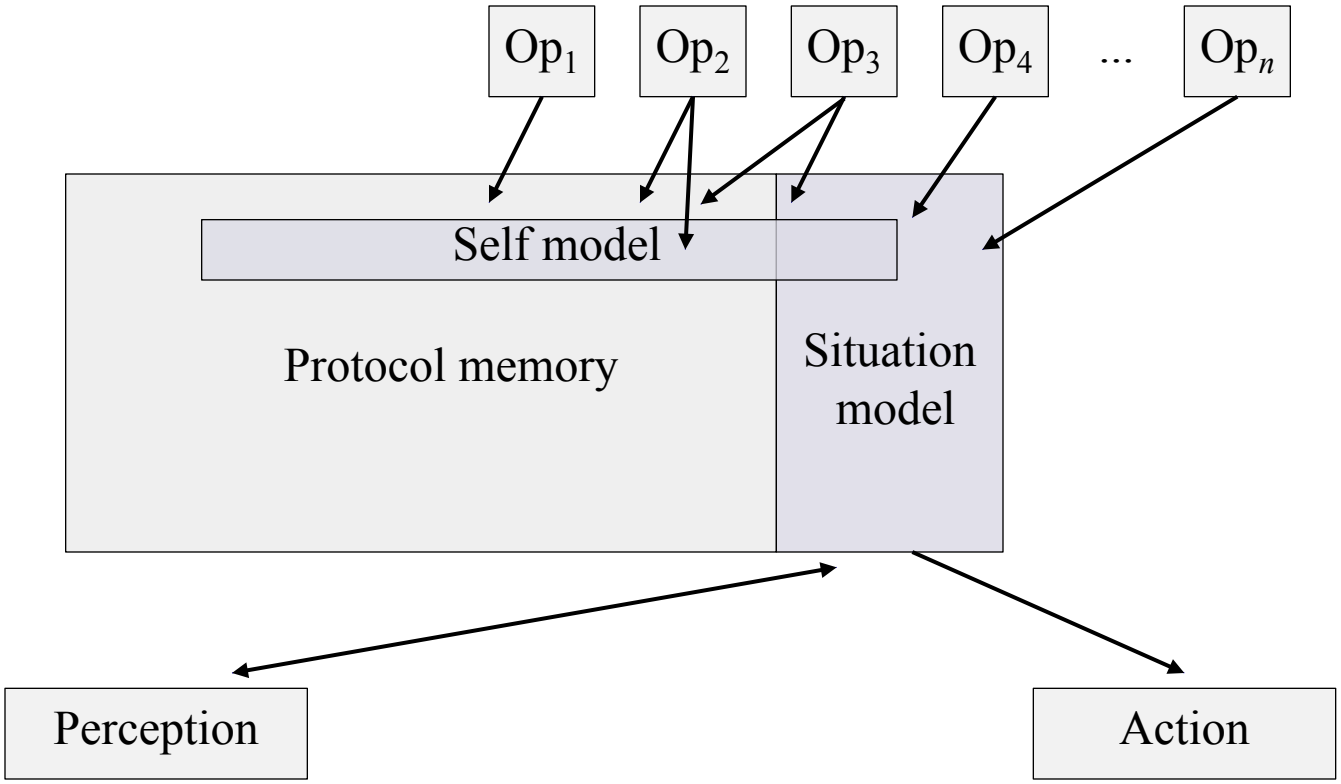
Components for Cognitive AI

- Model of current situation, and protocol of past situations



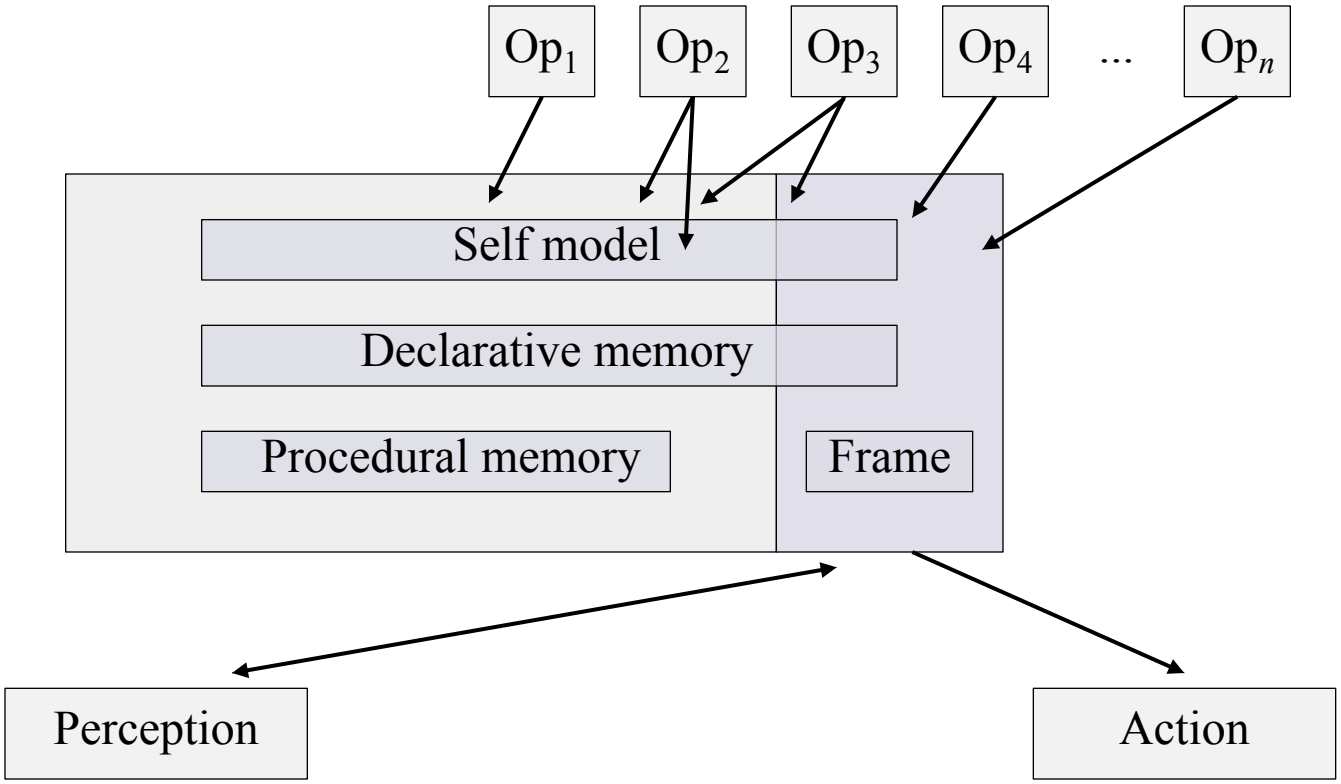
Components for Cognitive AI

- Model of self



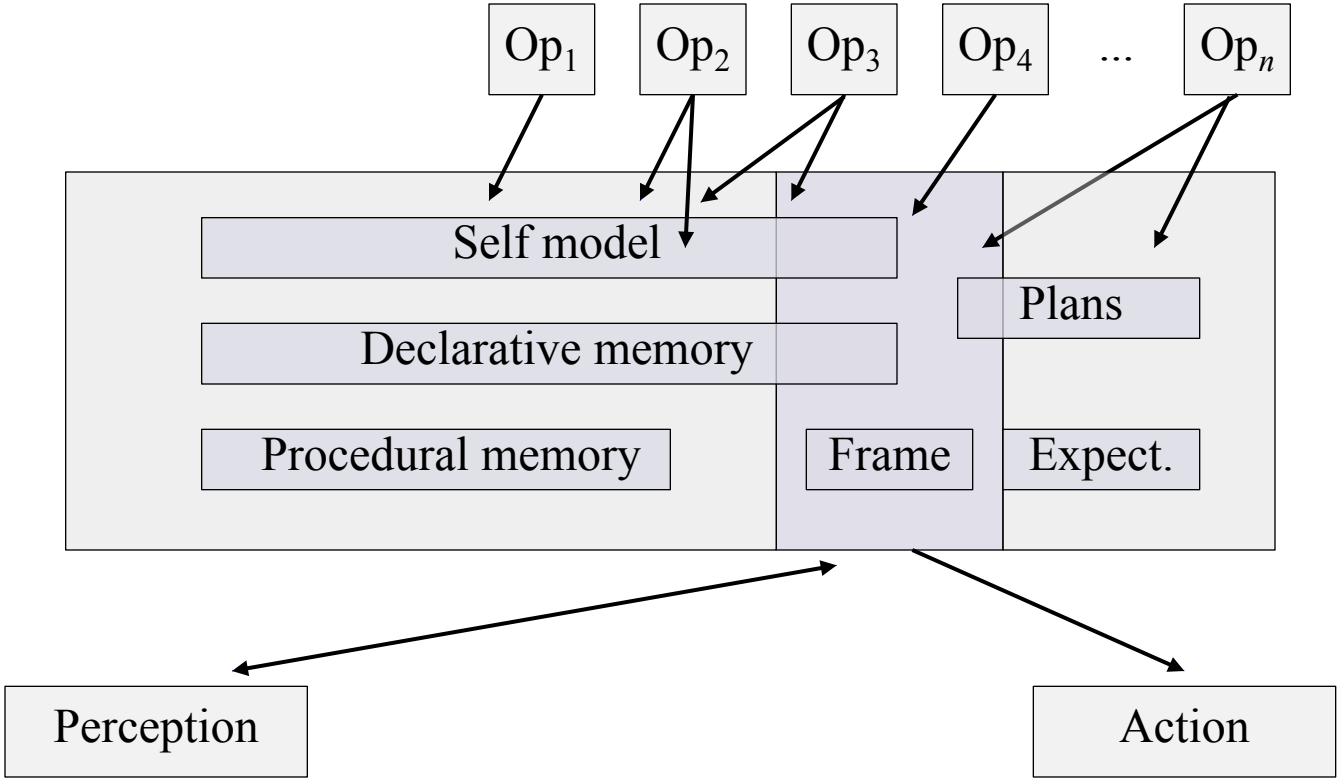
Components for Cognitive AI

- Abstractions of objects, episodes and types



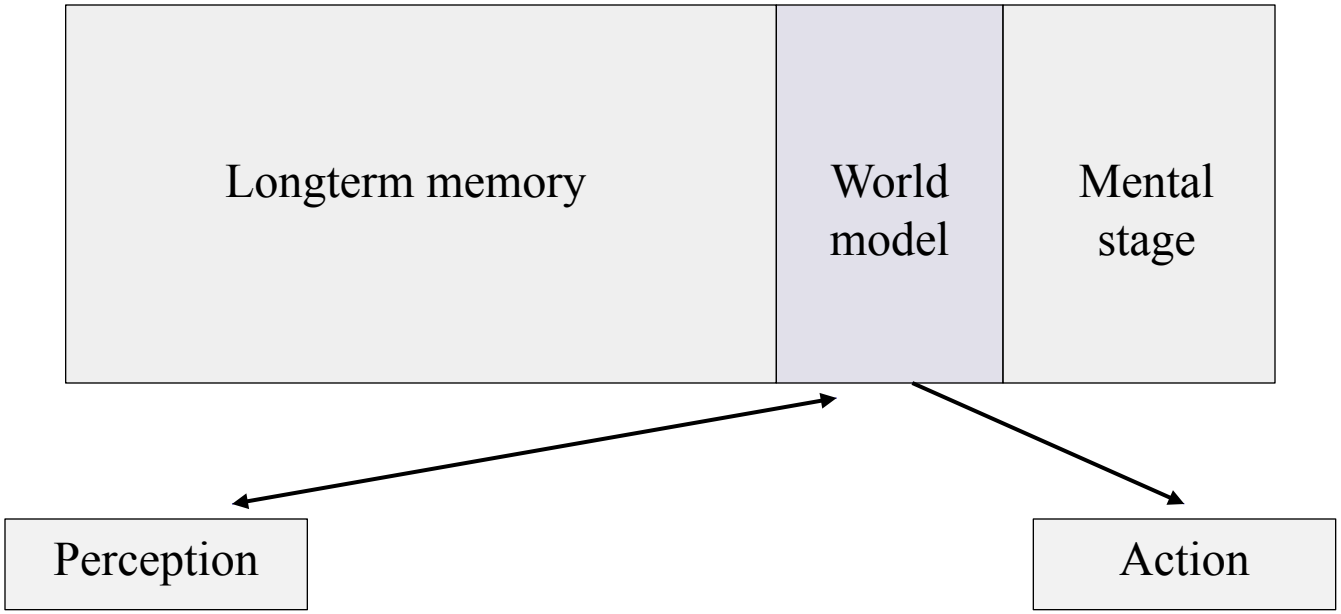
Components for Cognitive AI

- Anticipation of future developments



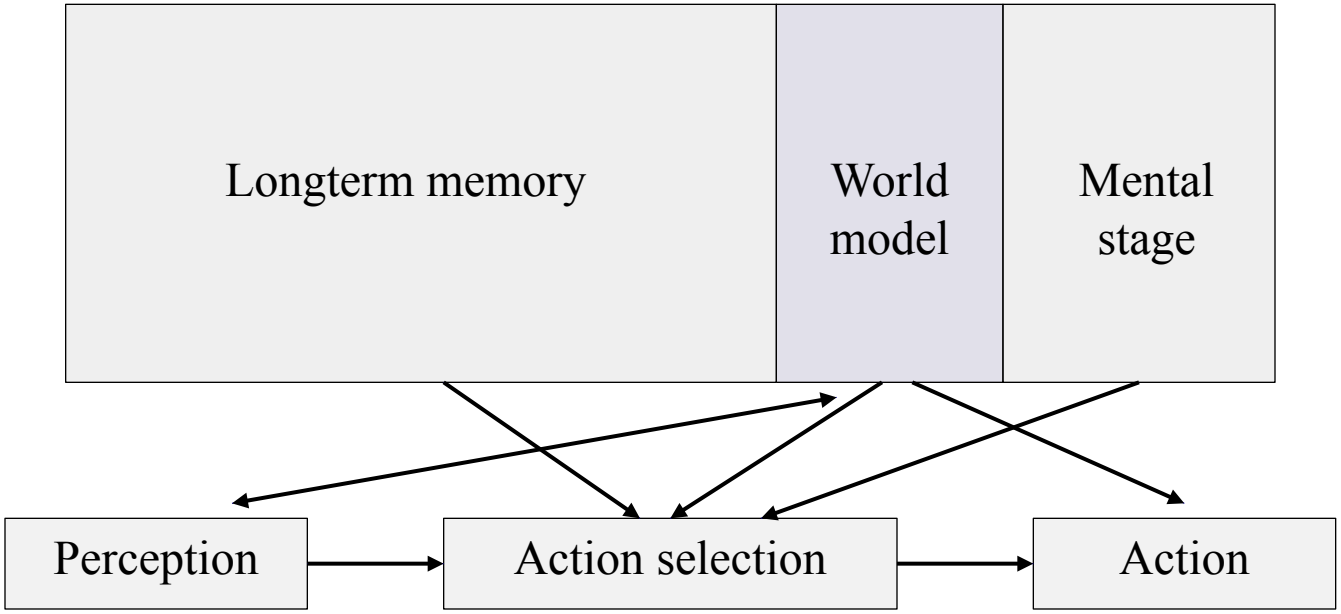
Components for Cognitive AI

- Action selection and executive control



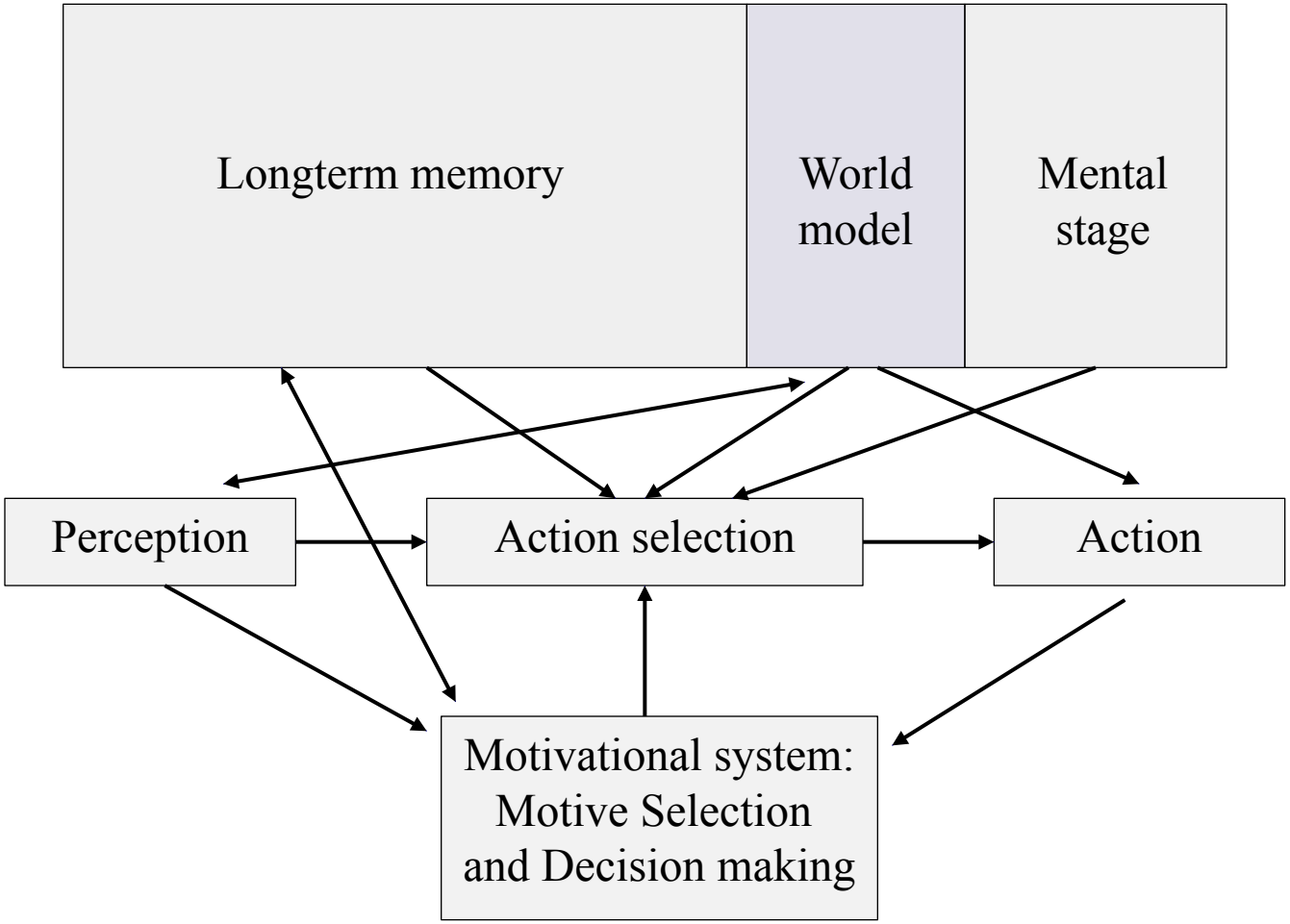
Components for Cognitive AI

- Action selection and executive control



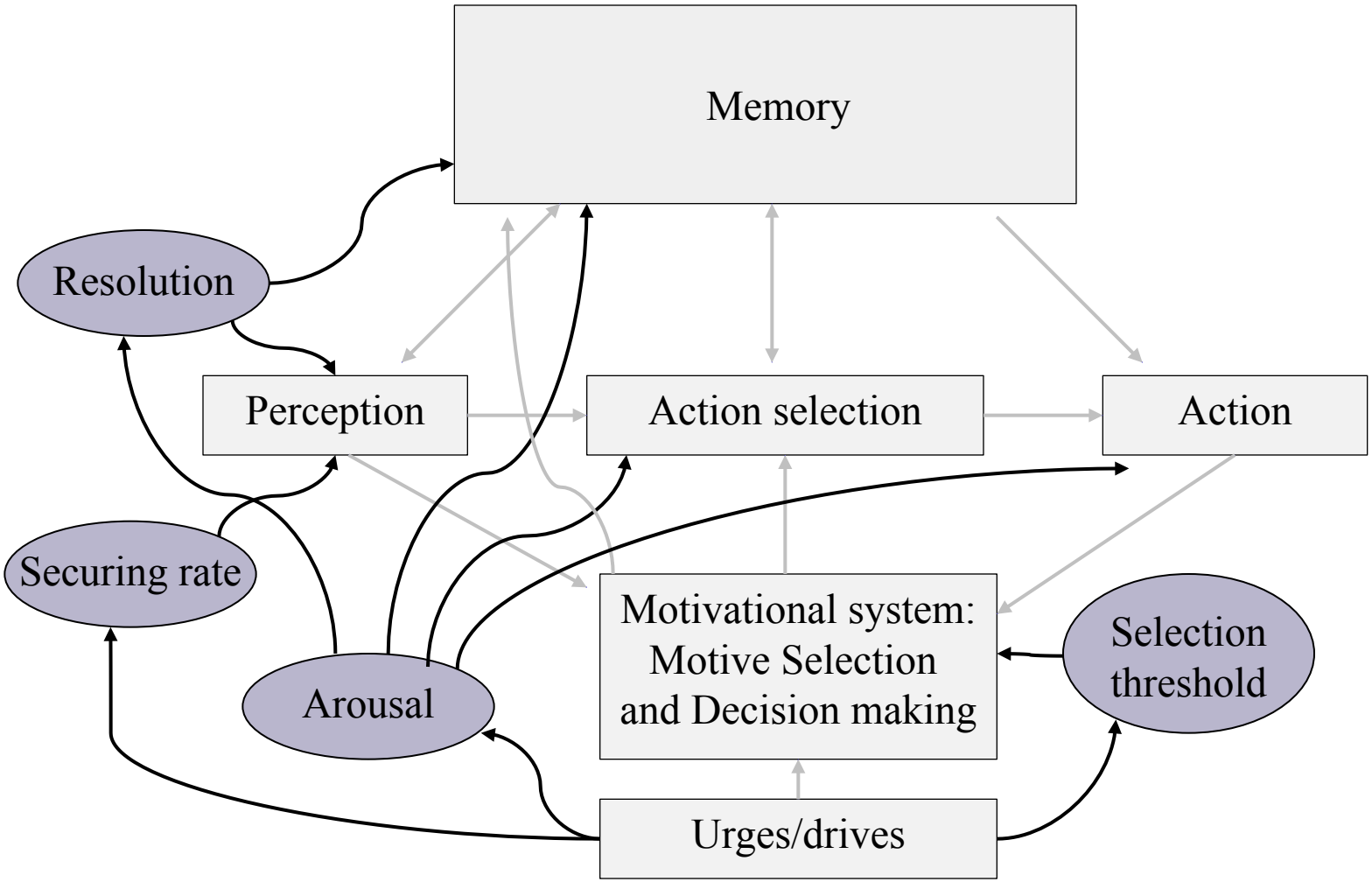
Components for Cognitive AI

- Universal motivation: autonomous identification of goals

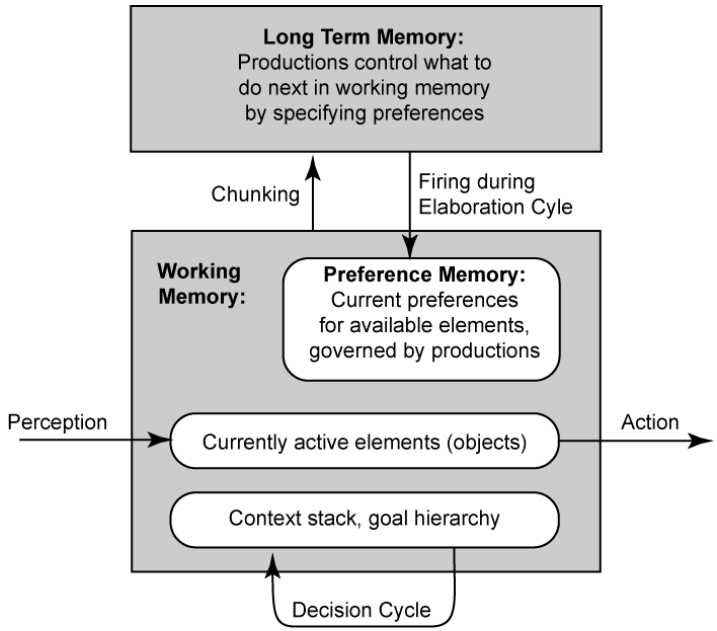


Components for Cognitive AI

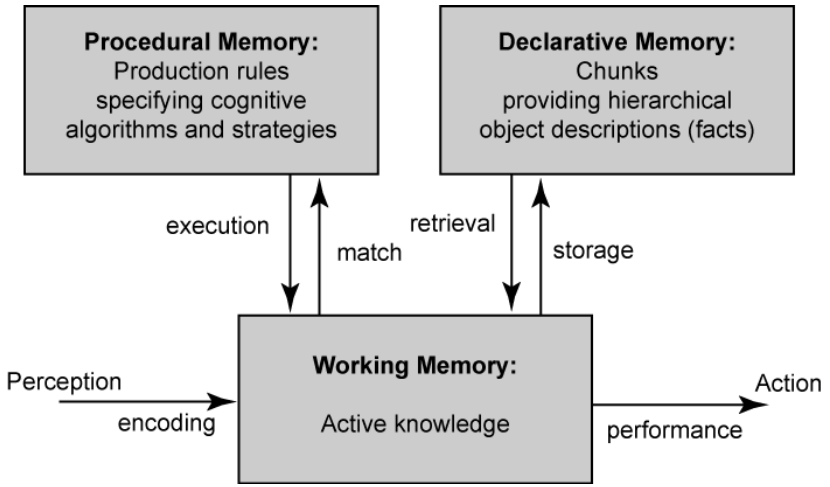
- Emotional modulation and affect



Cognitive Architectures



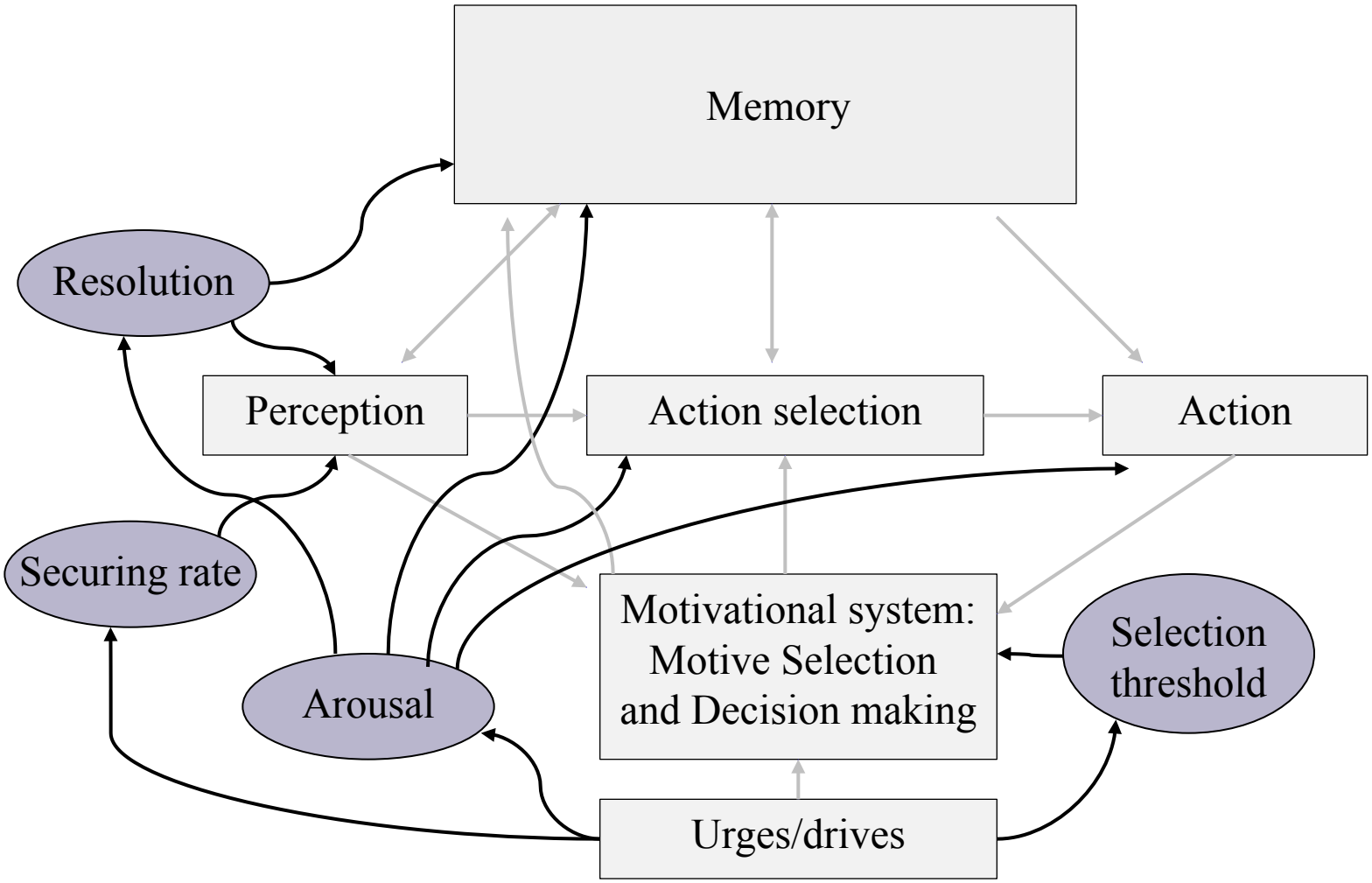
Soar



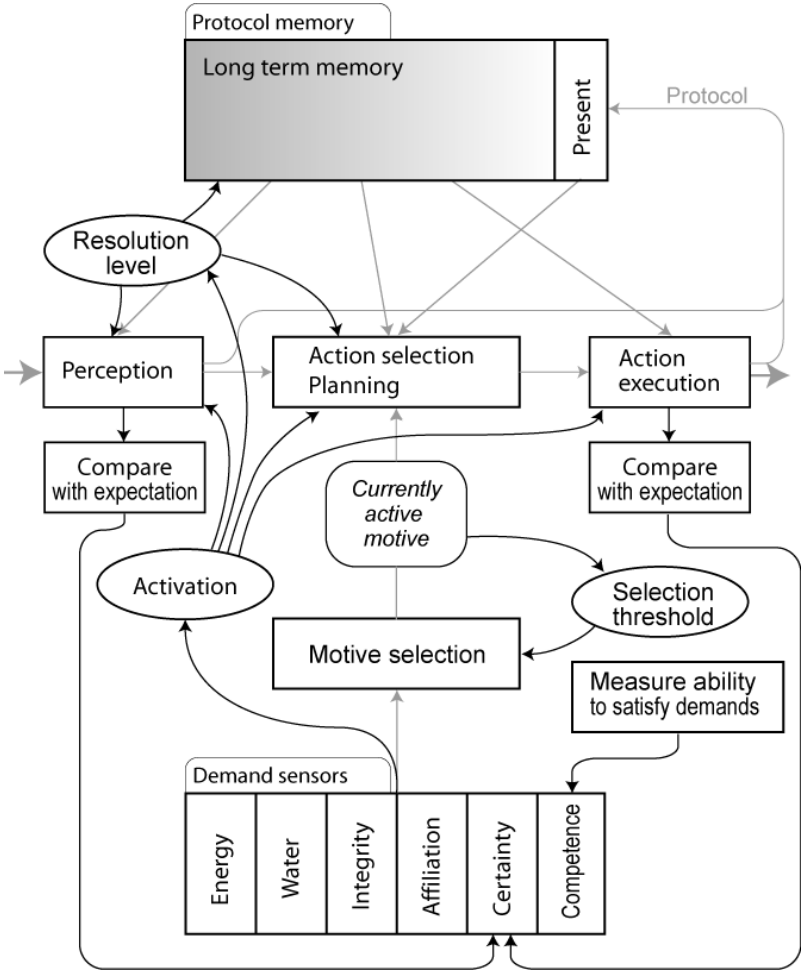
ACT-R

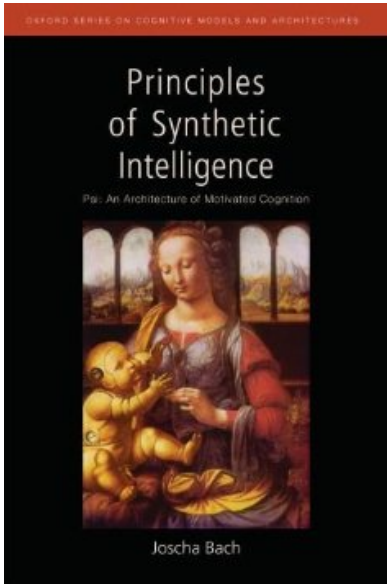
"Classical Cognitive Architectures" tend to focus on cognition as an isolated problem solving capability.

Components for Cognitive AI



MicroPsi architecture



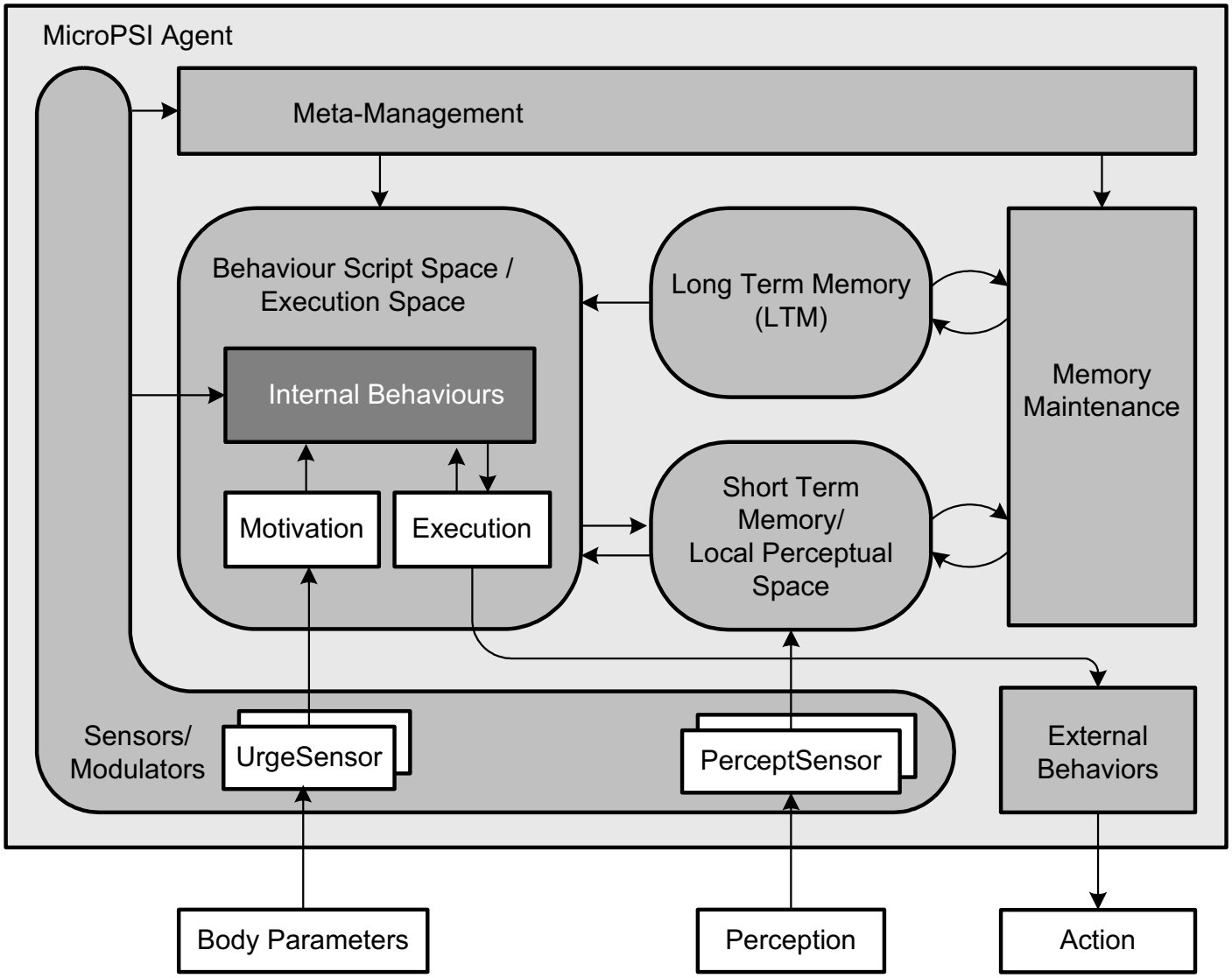


PSI theory

Principles of Synthetic Intelligence

(Dörner 1999; Bach 2003, 2009)

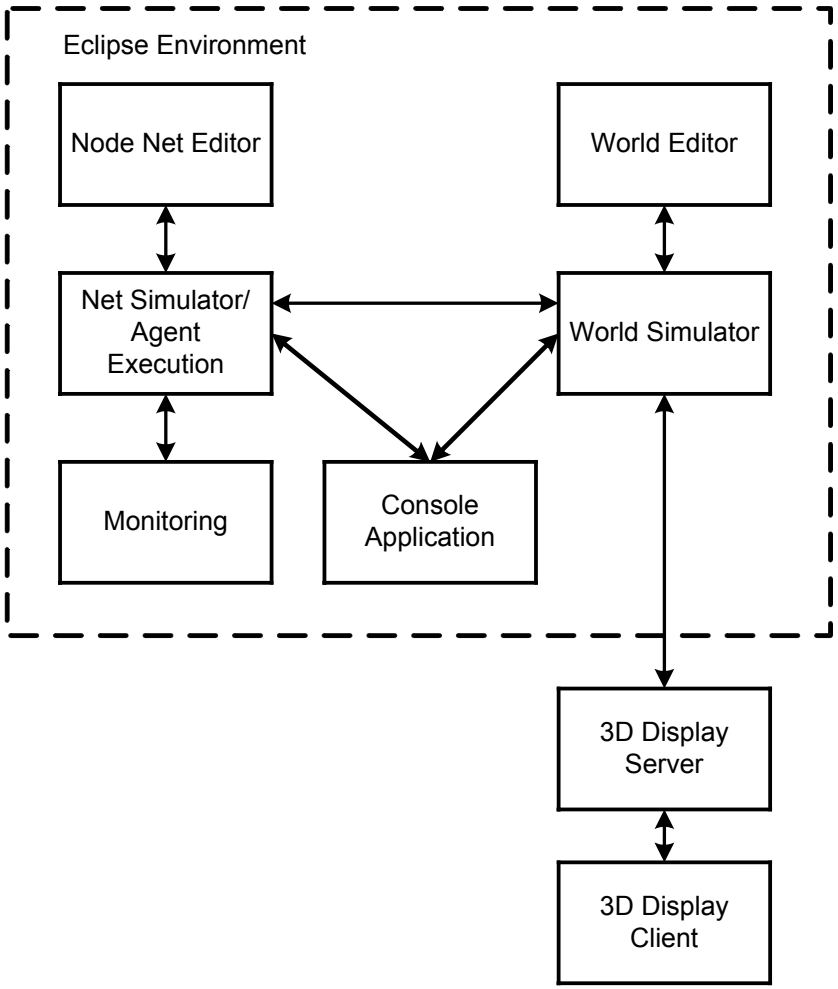
MicroPsi Architecture–simplified



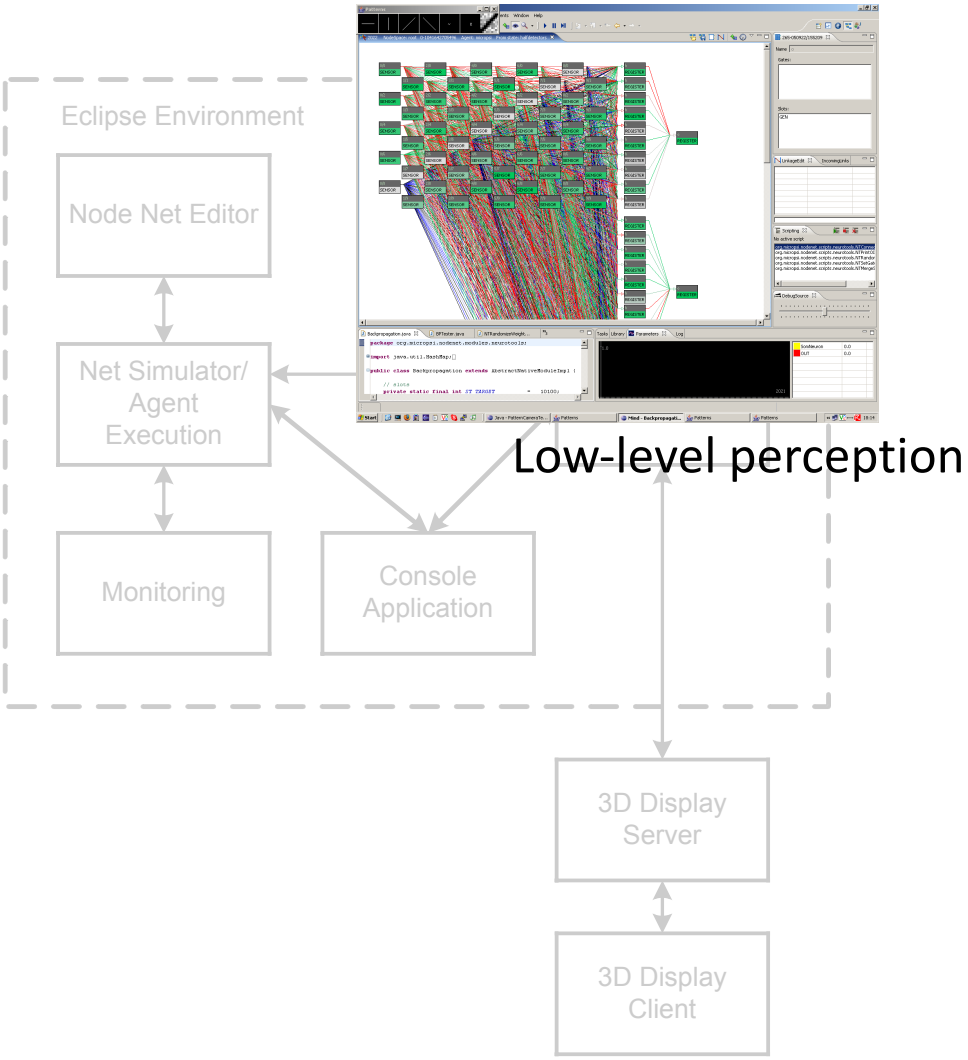
Agent Functionality

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

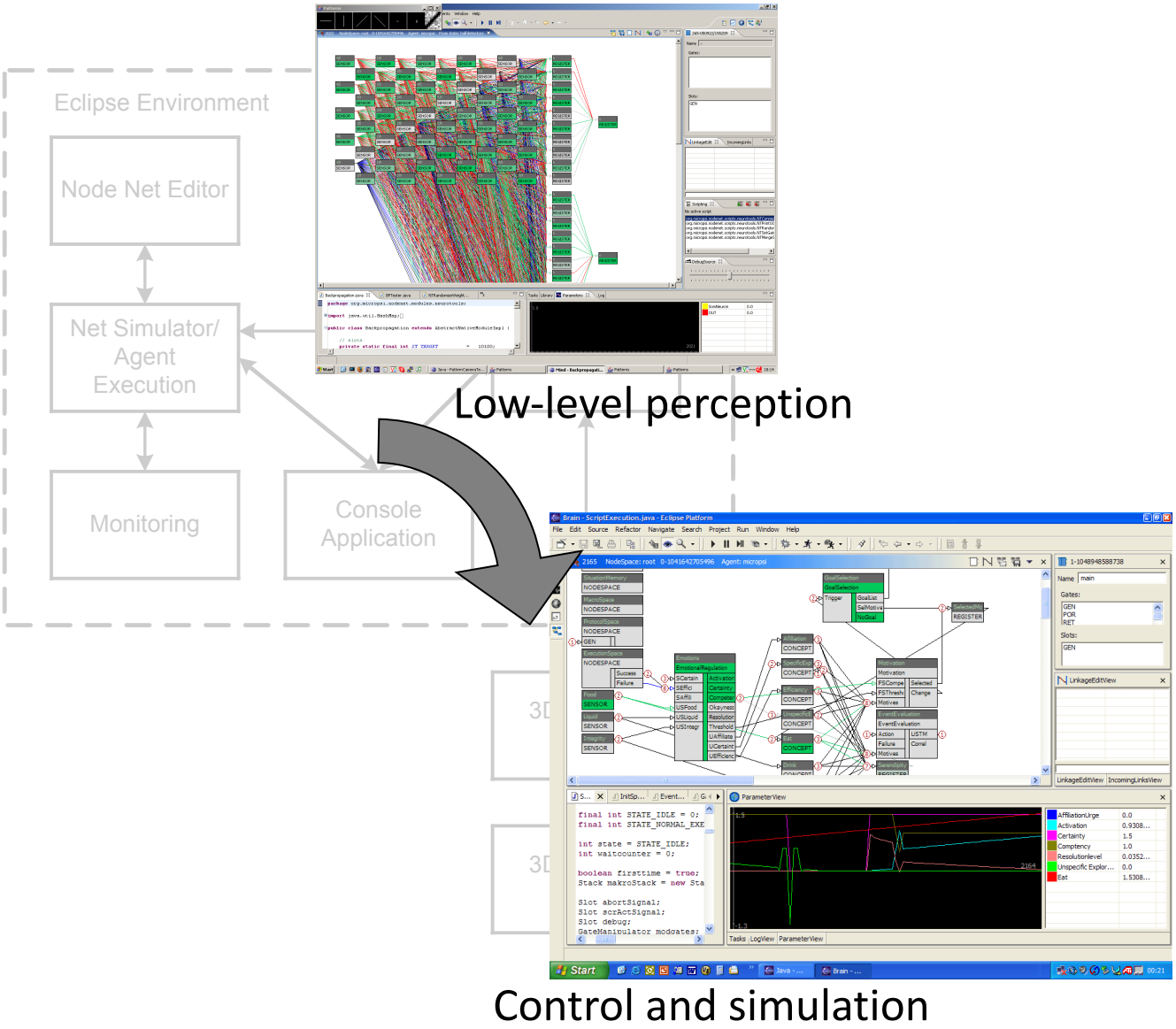
Implementation: MicroPsi (Bach 03, 05, 04, 06)



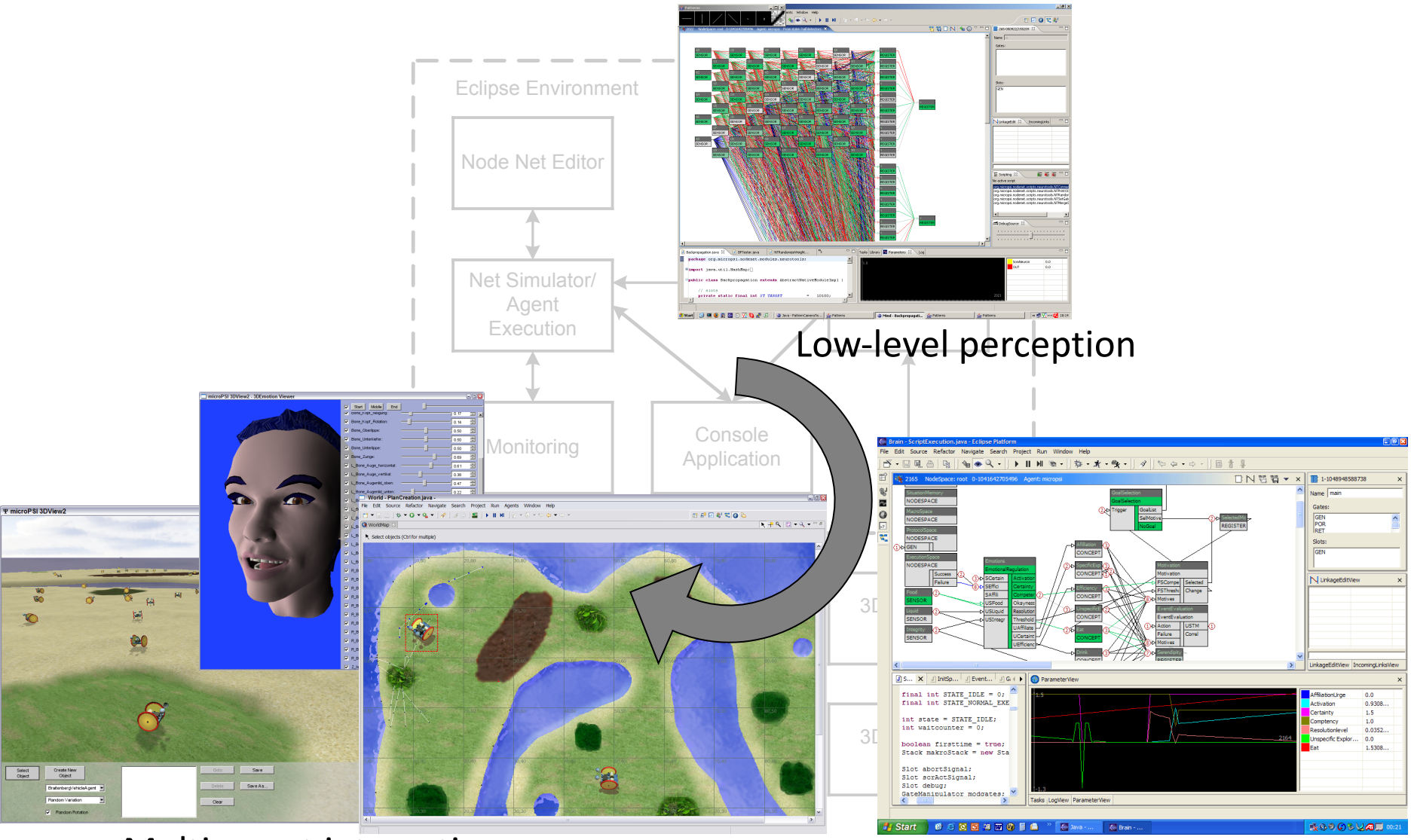
Implementation: MicroPsi (Bach 03, 05, 04, 06)



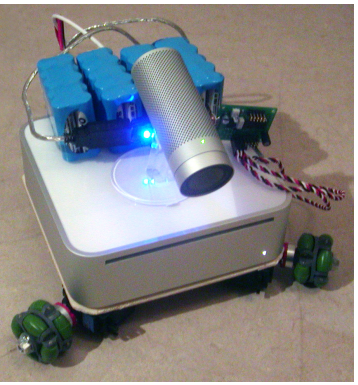
Implementation: MicroPsi (Bach 03, 05, 04, 06)



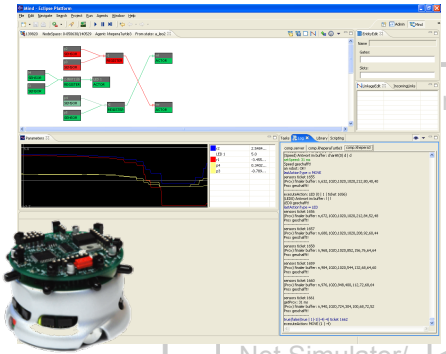
Implementation: MicroPsi (Bach 03, 05, 04, 06)

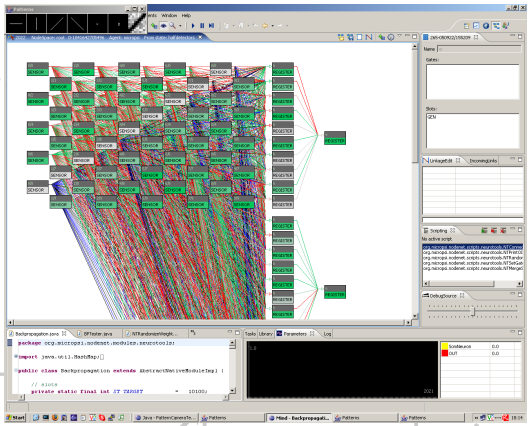


Implementation: MicroPsi (Bach 03, 04, 05, 06)

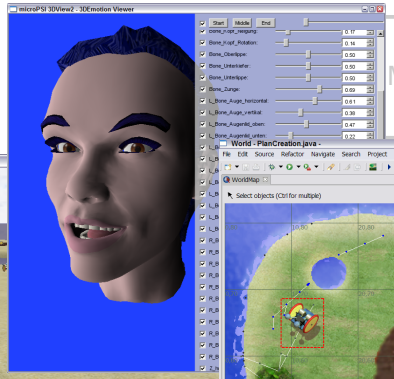


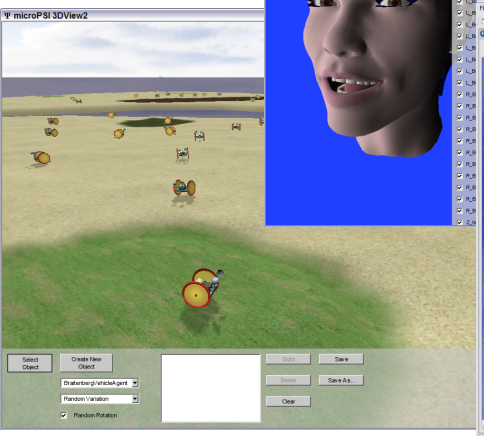
Robot control

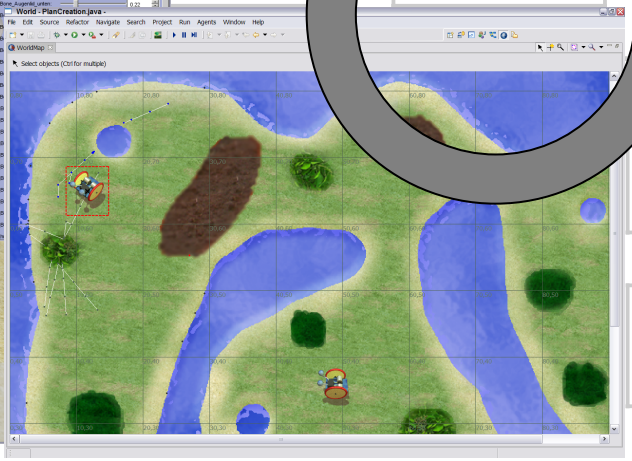




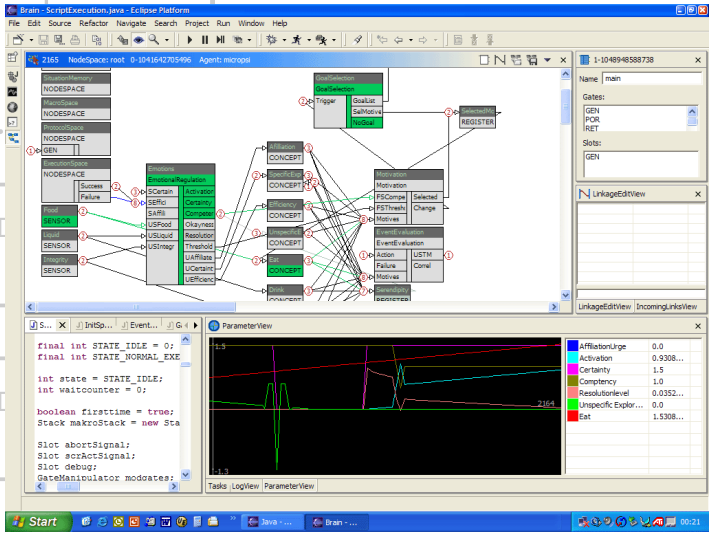
Low-level perception





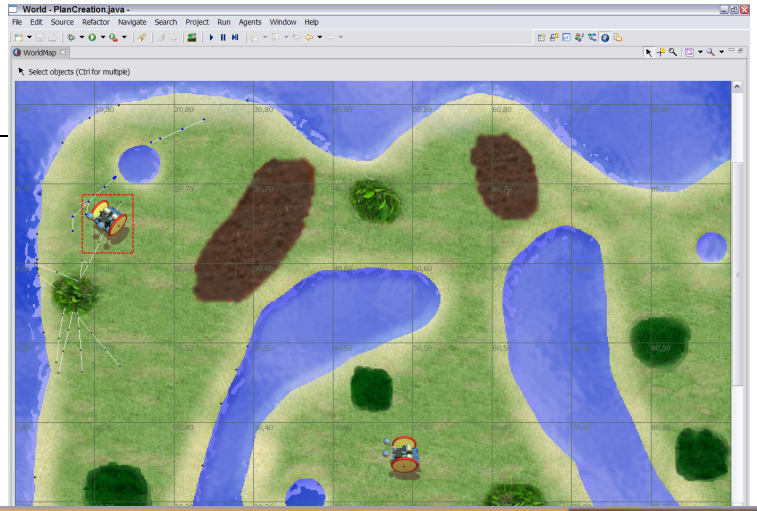


Multi-agent interaction

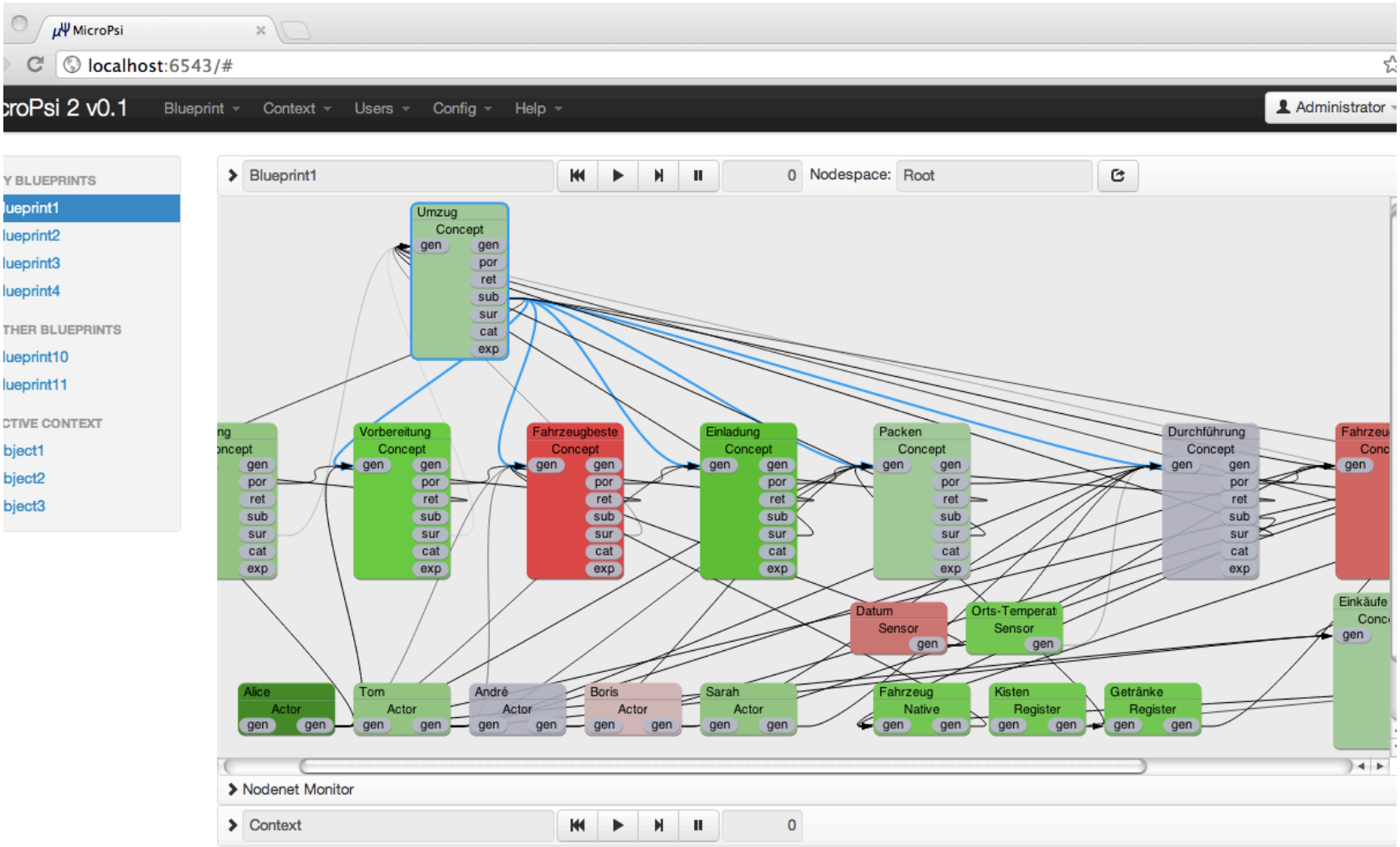


Control and simulation

Different simulation environments



Implementation: MicroPsi 2 (Bach, Welland, Vuine, Herger 12, 14, …)



Cognitive Artificial Intelligence

Methods should focus on components and performances necessary for intelligence:

- **Universal Representations:**
  Dynamic model of environment, possible worlds, and agent
- **(Semi-) Universal Problem Solving:**
  Learning, Planning, Reasoning, Analogies, Action Control, Reflection ...
- **Universal Motivation:**
  Polythematic, adaptive goal identification
- **Emotion and affect**
- **Whole, testable architectures**

Modeling Motivation

Modeling Motivation in a Cognitive Architecture

- *General intelligence needs General Motivation*
- Motivational system structures cognition
- Motivational dynamics: physiological, social and cognitive drives
- Intention selection and action control
- Motivation vs. affect

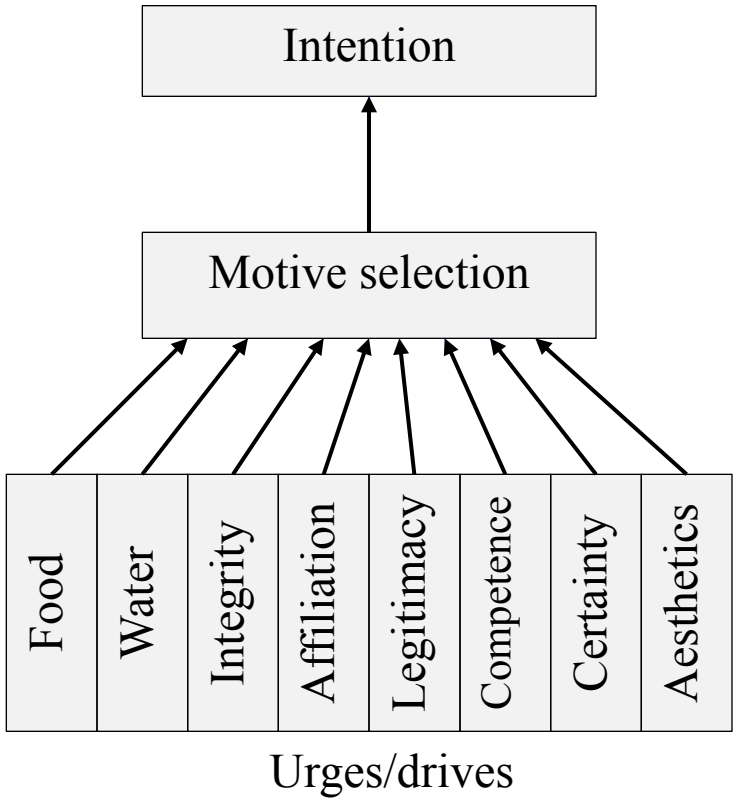
Motivation vs. emotion

- Motivation:
  - reflects needs
  - gives rise to goals and directed behavior
  - does not have to be associated with emotions

- Emotion:
  - modulates perception, cognition, action
  - receives valence from motivation
  - receives objects from cognition
  - leads to affective expression

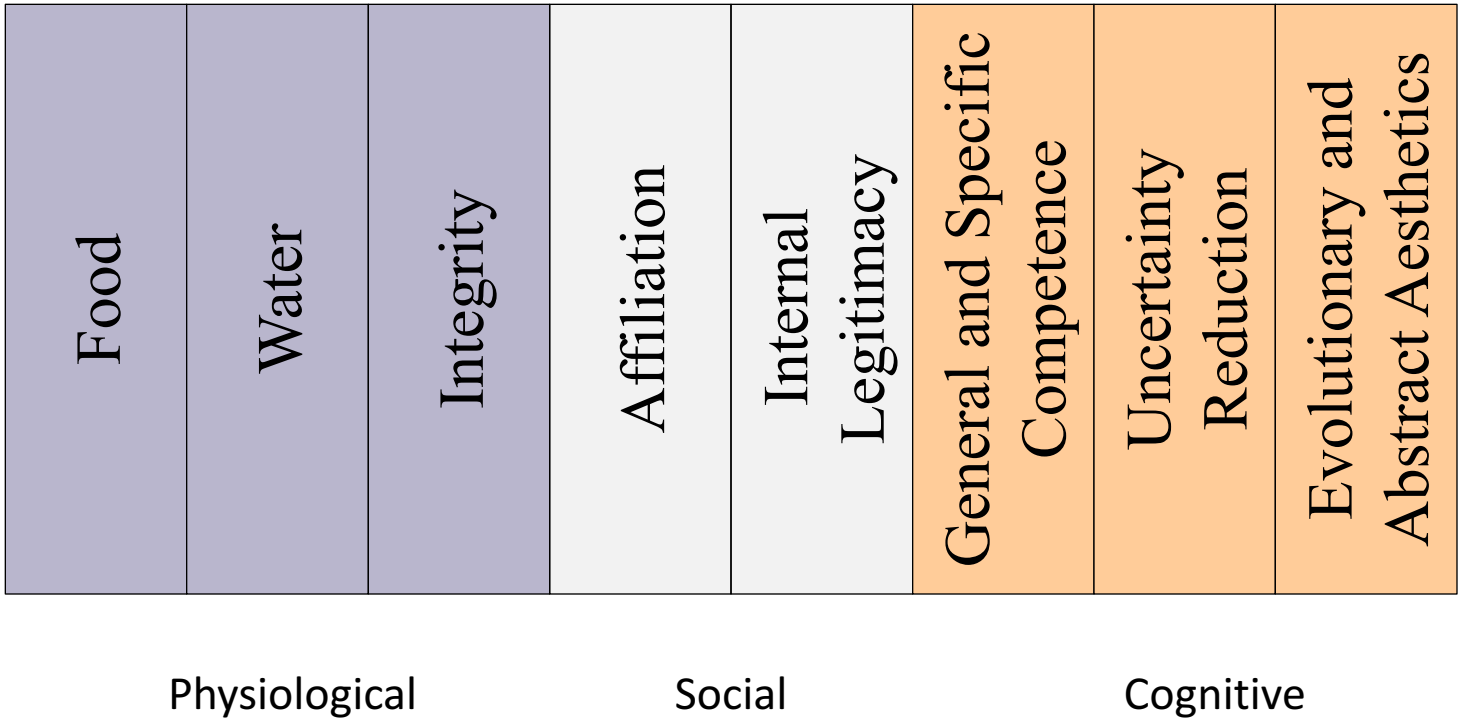
Motivational System

- 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

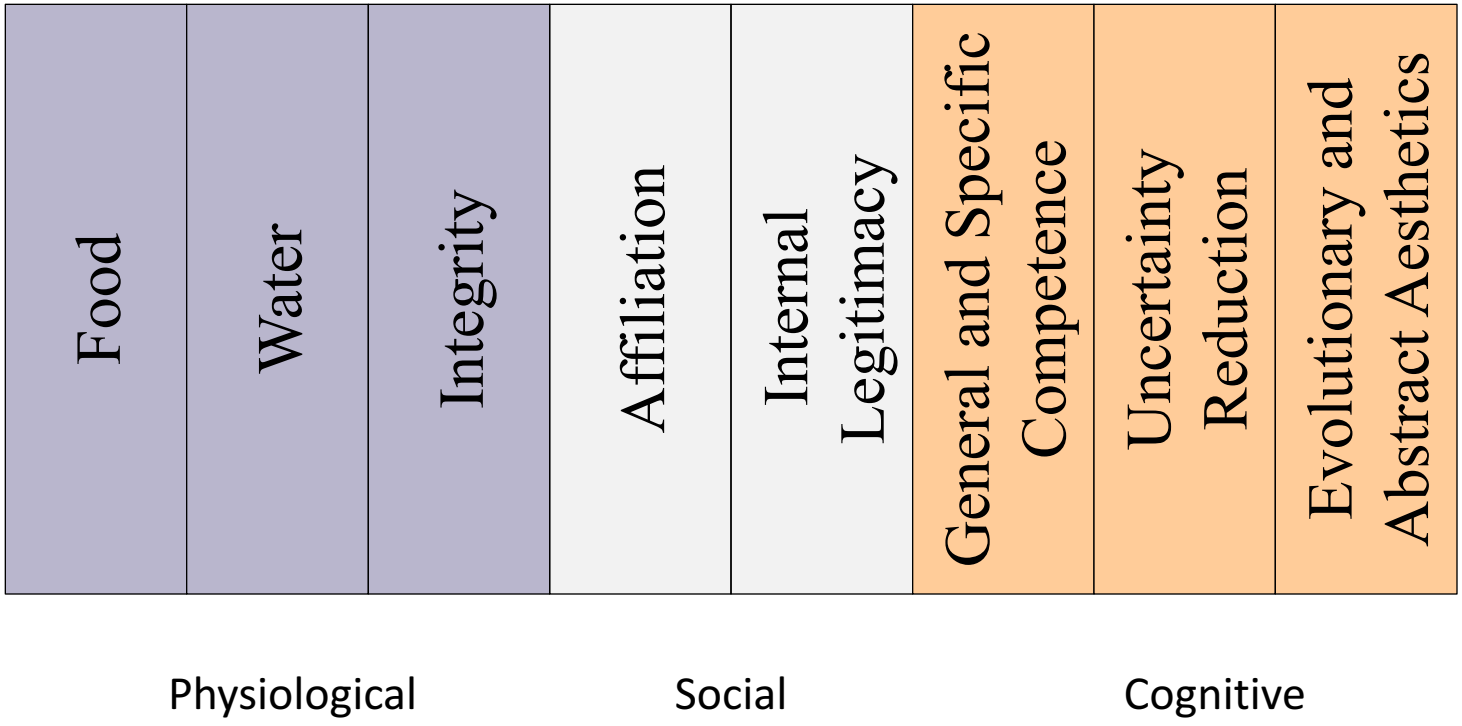
- *external* legitimacy: group acceptance
- *internal* legitimacy: "honor", conformance to internalized social norms

Cognitive Drives

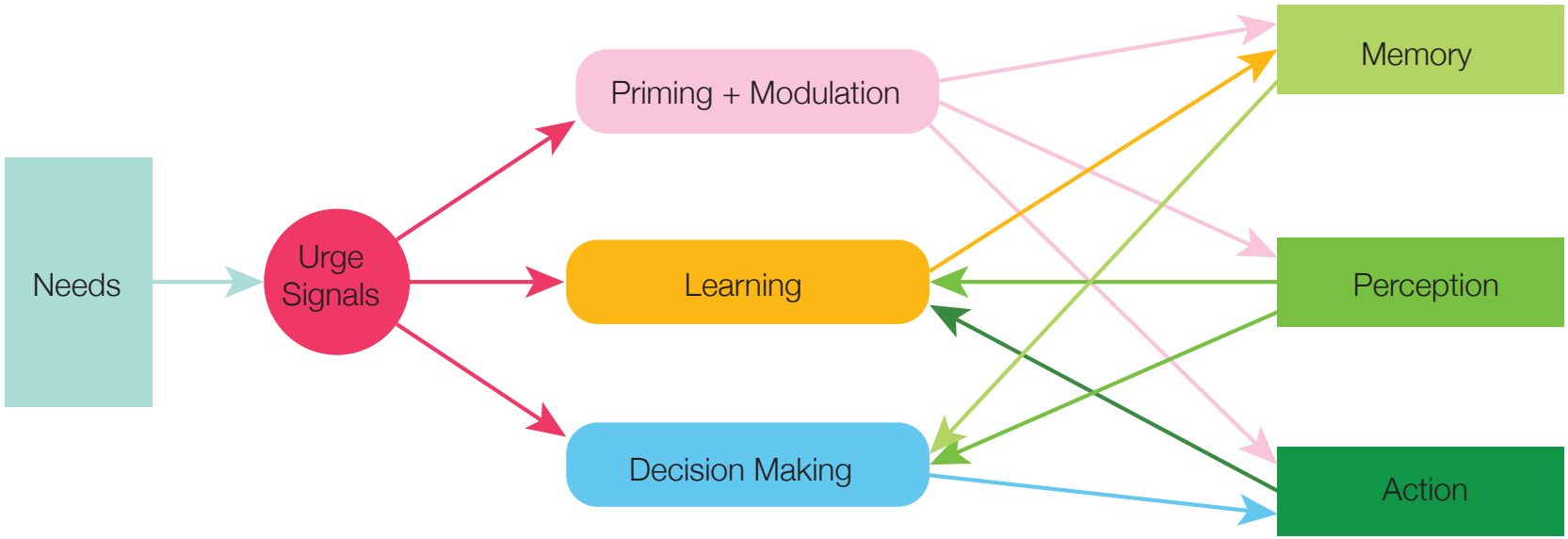
- Competence
  - epistemic (problem specific)
  - general (ability to satisfy demands)
  - effect oriented
- Uncertainty reduction
  - novelty seeking
- Aesthetics
  - evolutionary preferences (stimulus oriented)
  - abstract (representation oriented)

Motivational System

All possible goals correspond to (at least one) demand



From Needs to Behavior



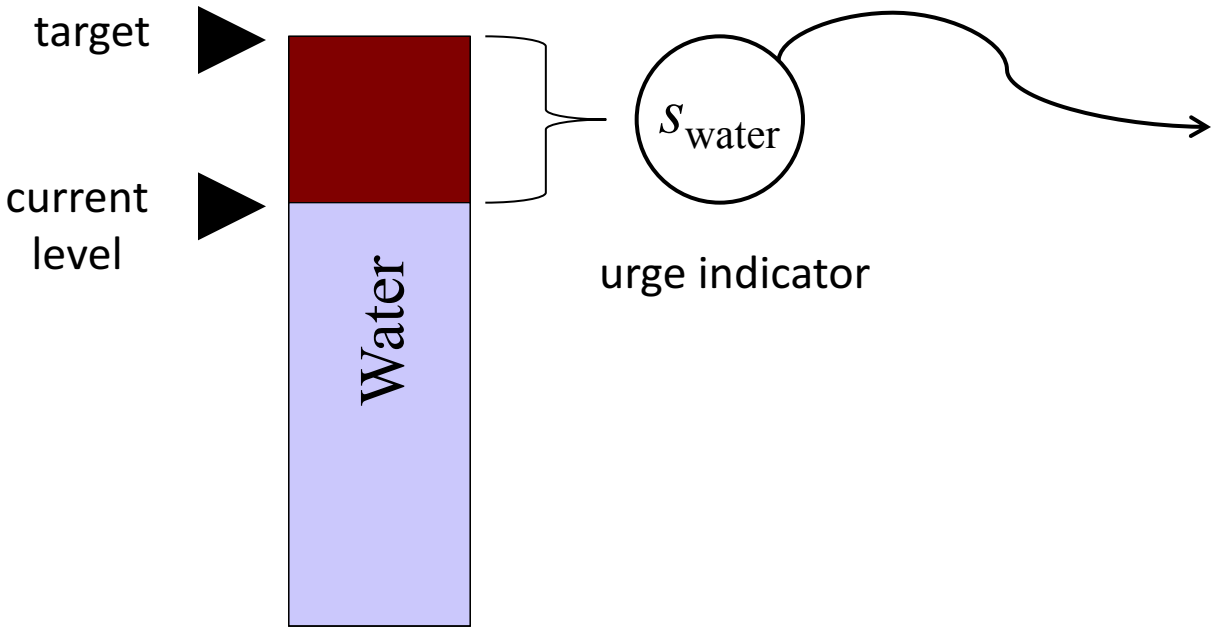
Pleasure and distress:

- *Change* of a demand is reflected in *pleasure* or *distress signal*
- Strength is *proportional* to amount of change

- Pleasure and distress signals deliver ***reinforcement*** values for behavioral procedures and episodic sequences and define ***appetitive*** and ***aversive*** goals.

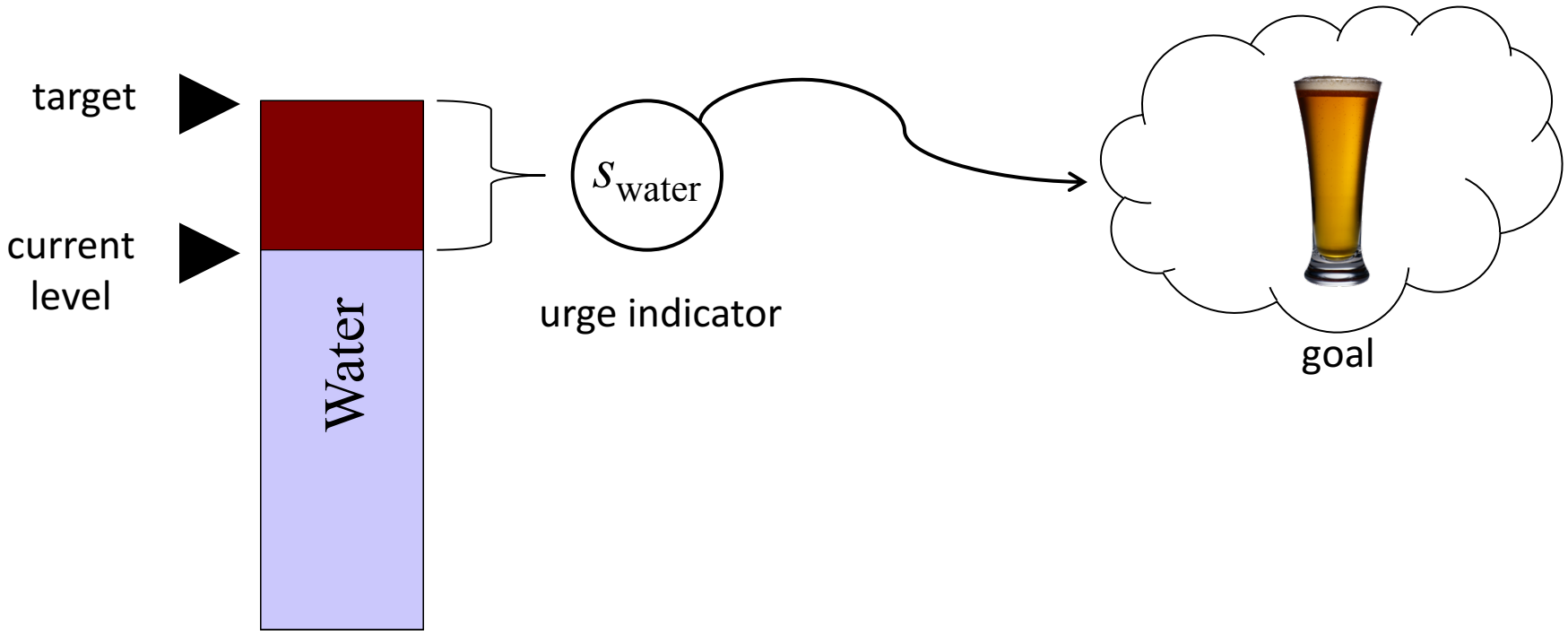
Motivational System

- drive = demand + urge indicator



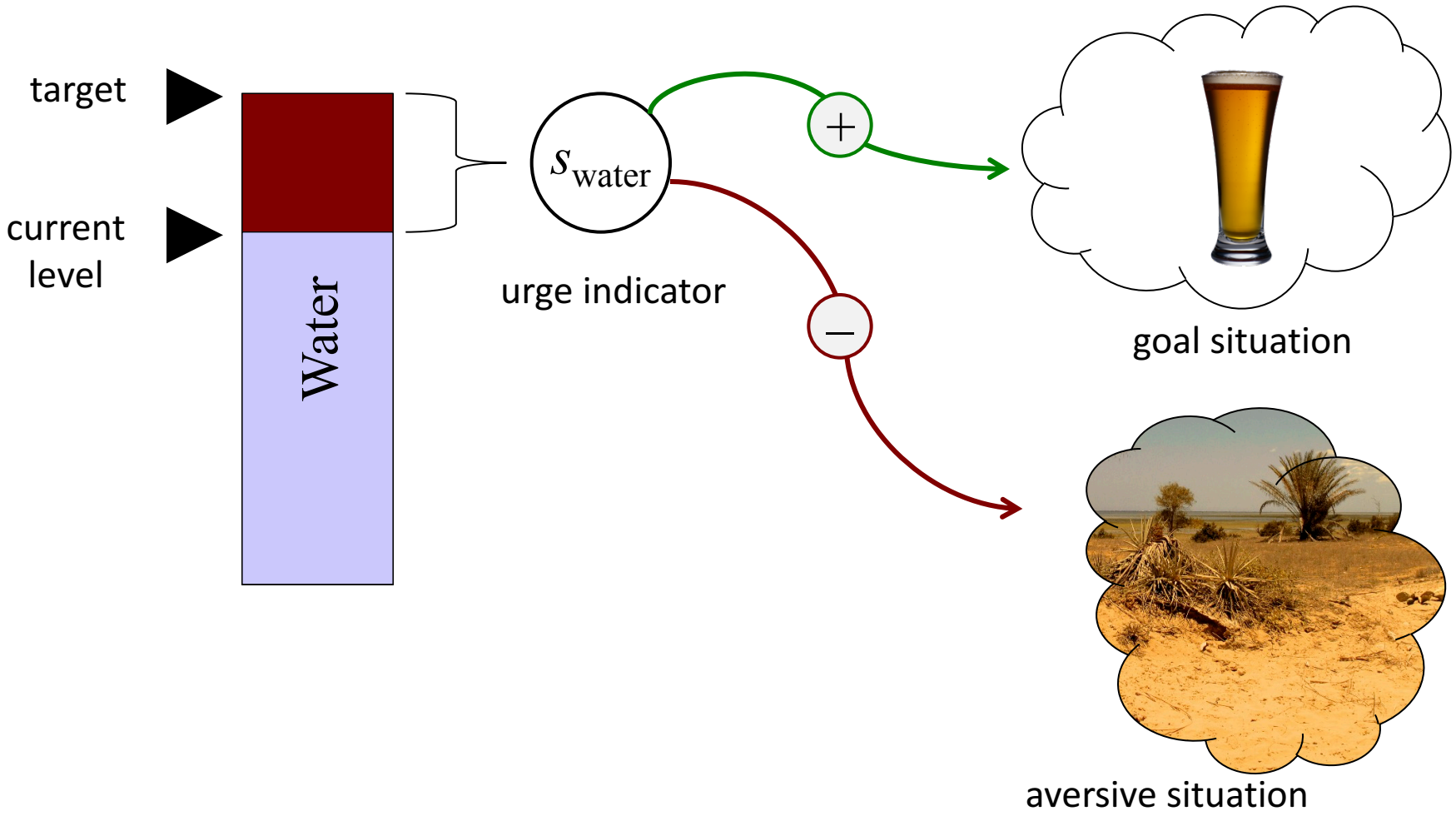
Motivational Learning

- motive = urge + goal situation



Motivational Learning

- motive = urge + goal situation



Goals

- Goal: situation or action that affords to satisfy a need
- Aversive goal: situation or action that frustrate a need
- All behavior is directed on satisfying an appetitive goal or avoiding an aversive goal
- Needs are predefined, goals are learned

Physiological needs

- Thirst
- Hunger
- Rest
- Warmth
- Libido
- …

→ Survival as emergent property

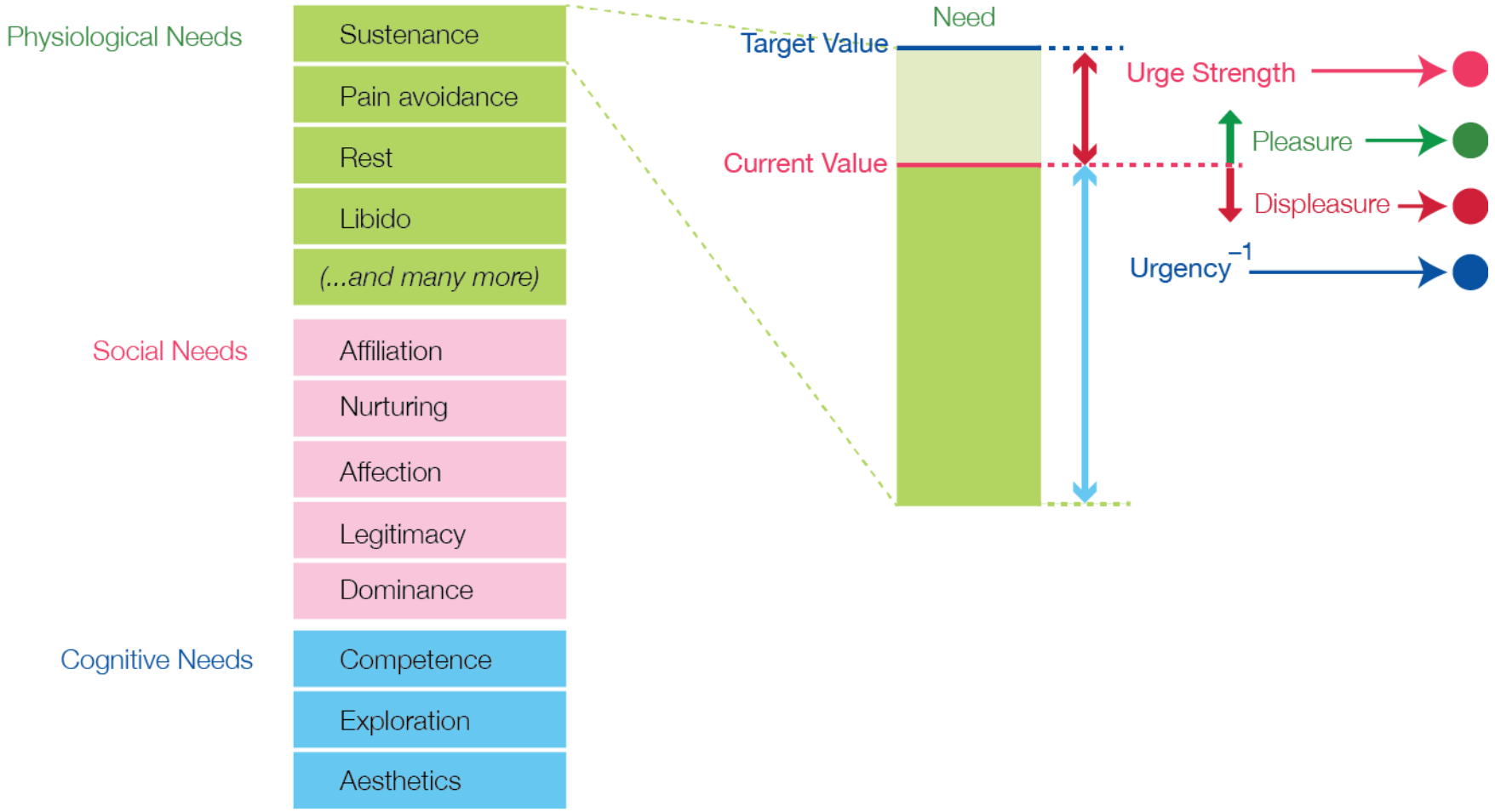
Social needs

- Affiliation (Attention from others, external legitimacy)
- Internal legitimacy
- Nurturing (caring for others)
- Affection
- Dominance

Cognitive needs

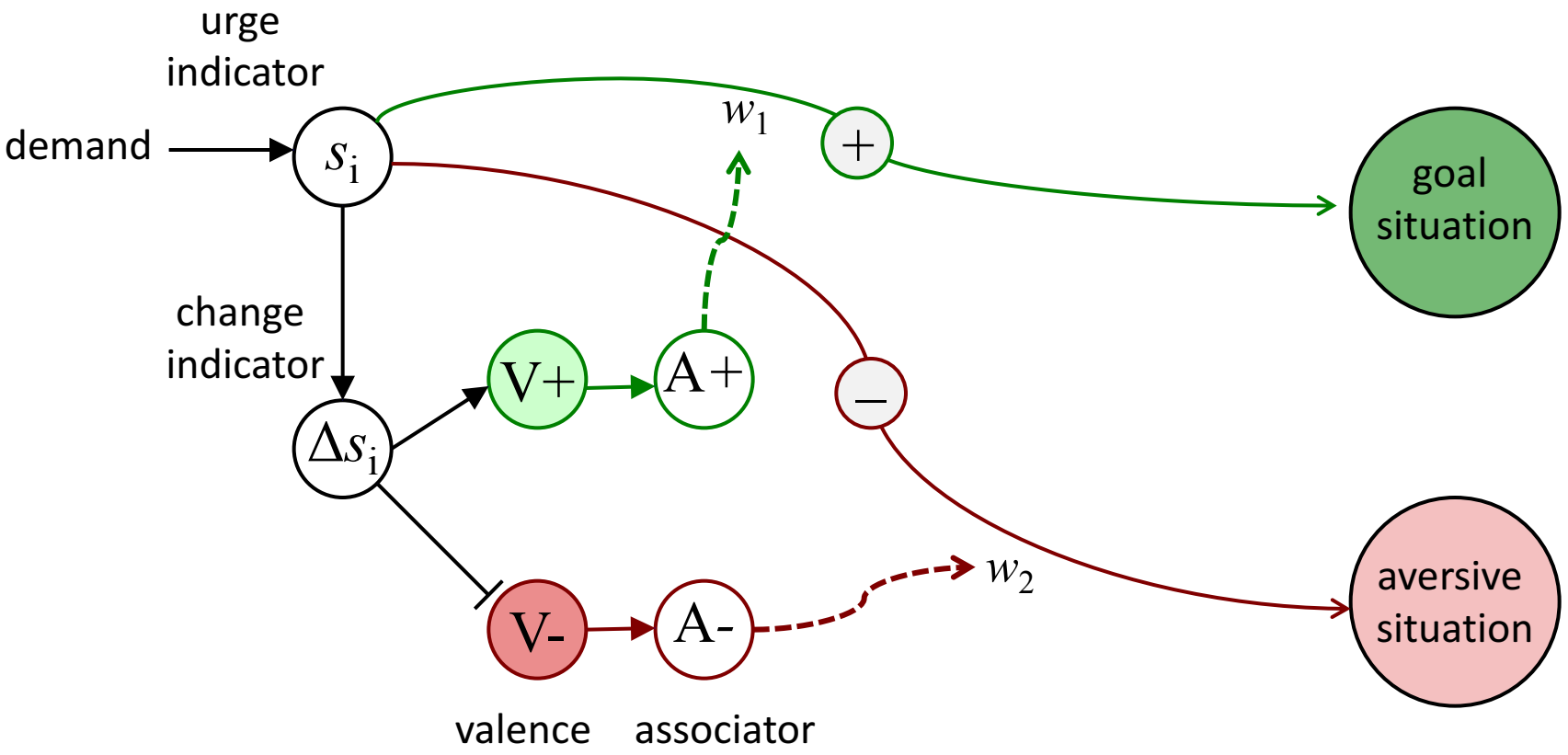
- Competence:
  - Skill acquisition (epistemic competence)
  - Coping/control ability (general competence)
  - Effect generation
- Uncertainty reduction:
  - Exploration
- Aesthetics:
  - Stimulus oriented
  - Structure oriented (abstract aesthetics)

Needs and urges



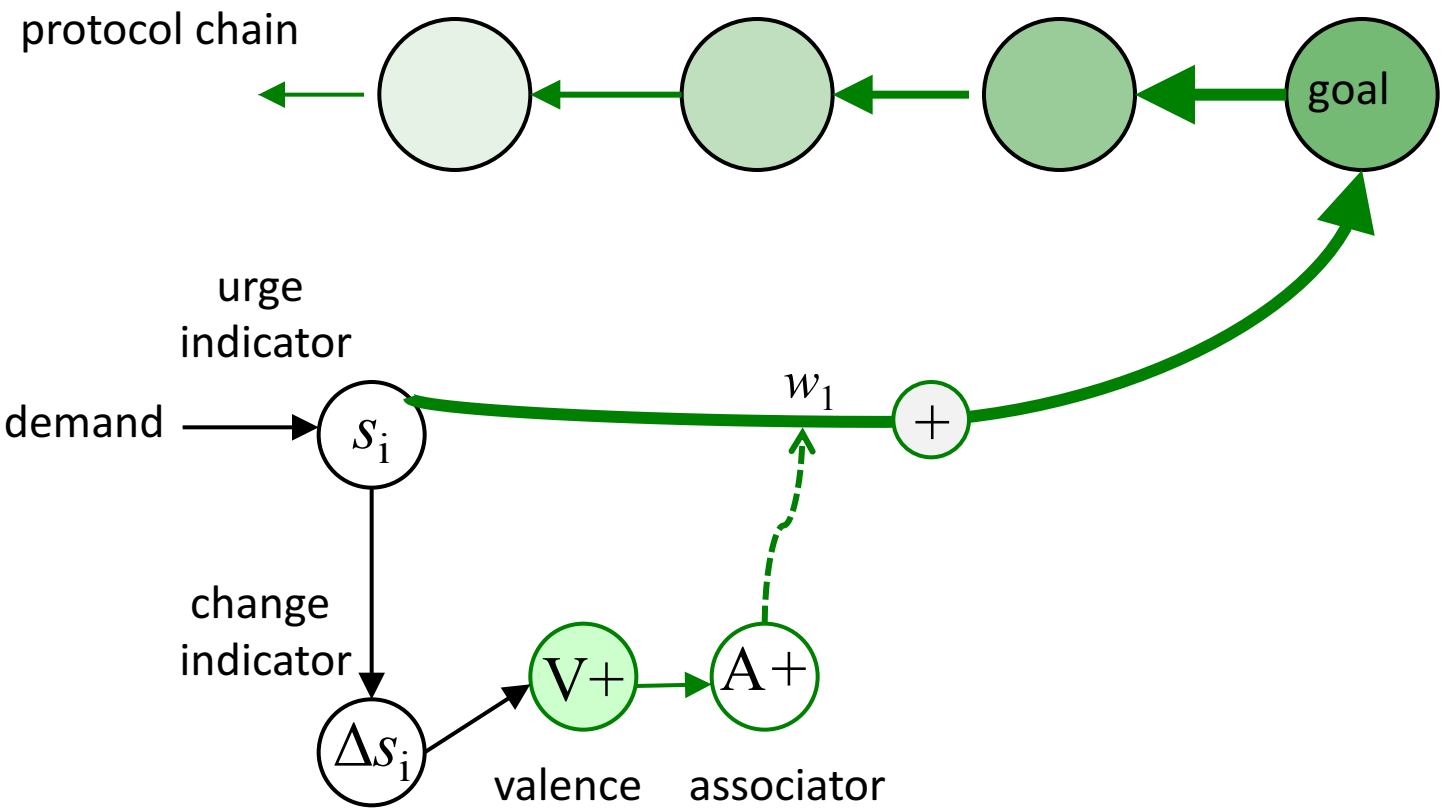
Motivational Learning

- association by learning:



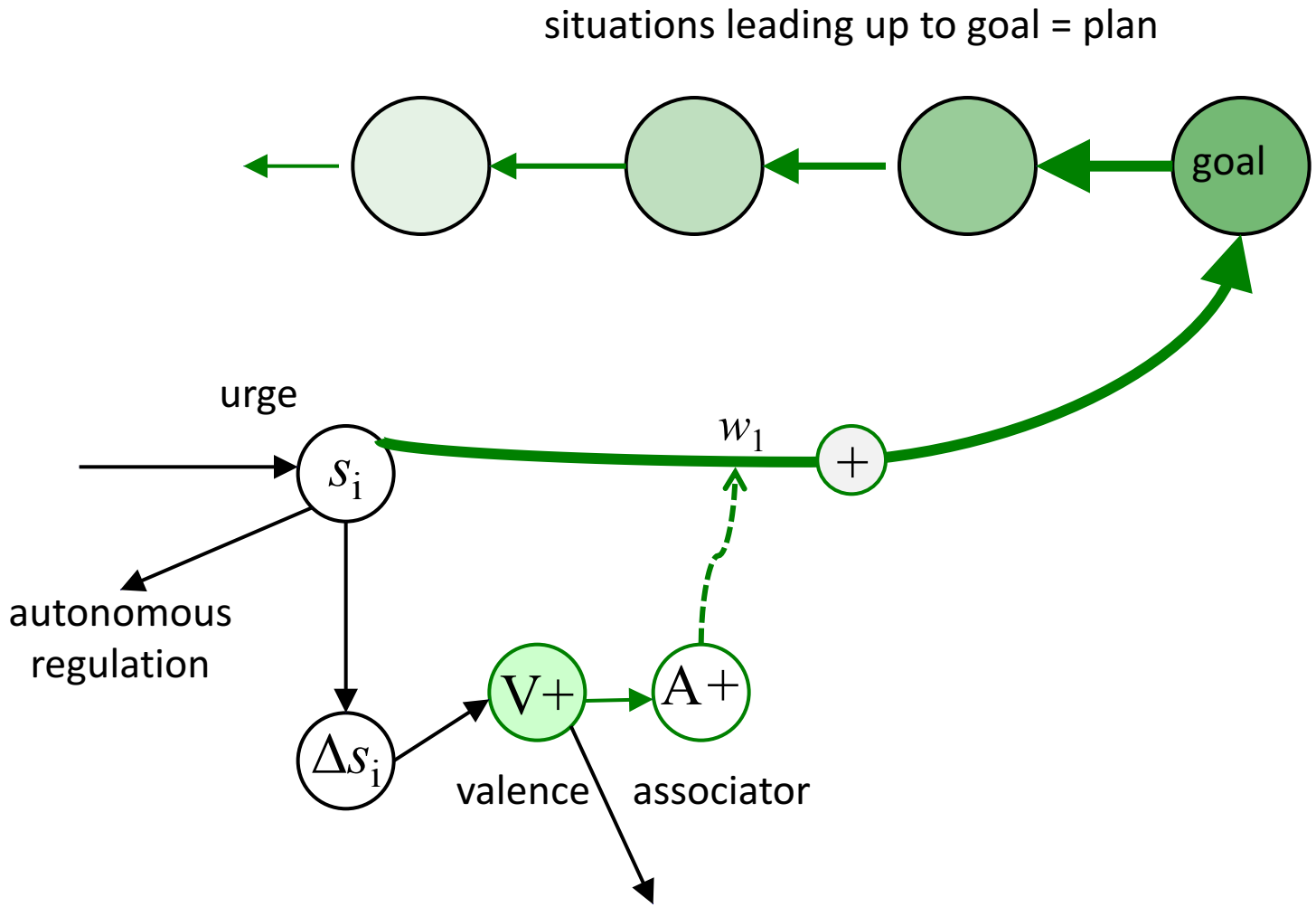
Motivational Learning

- retrogradient reinforcement



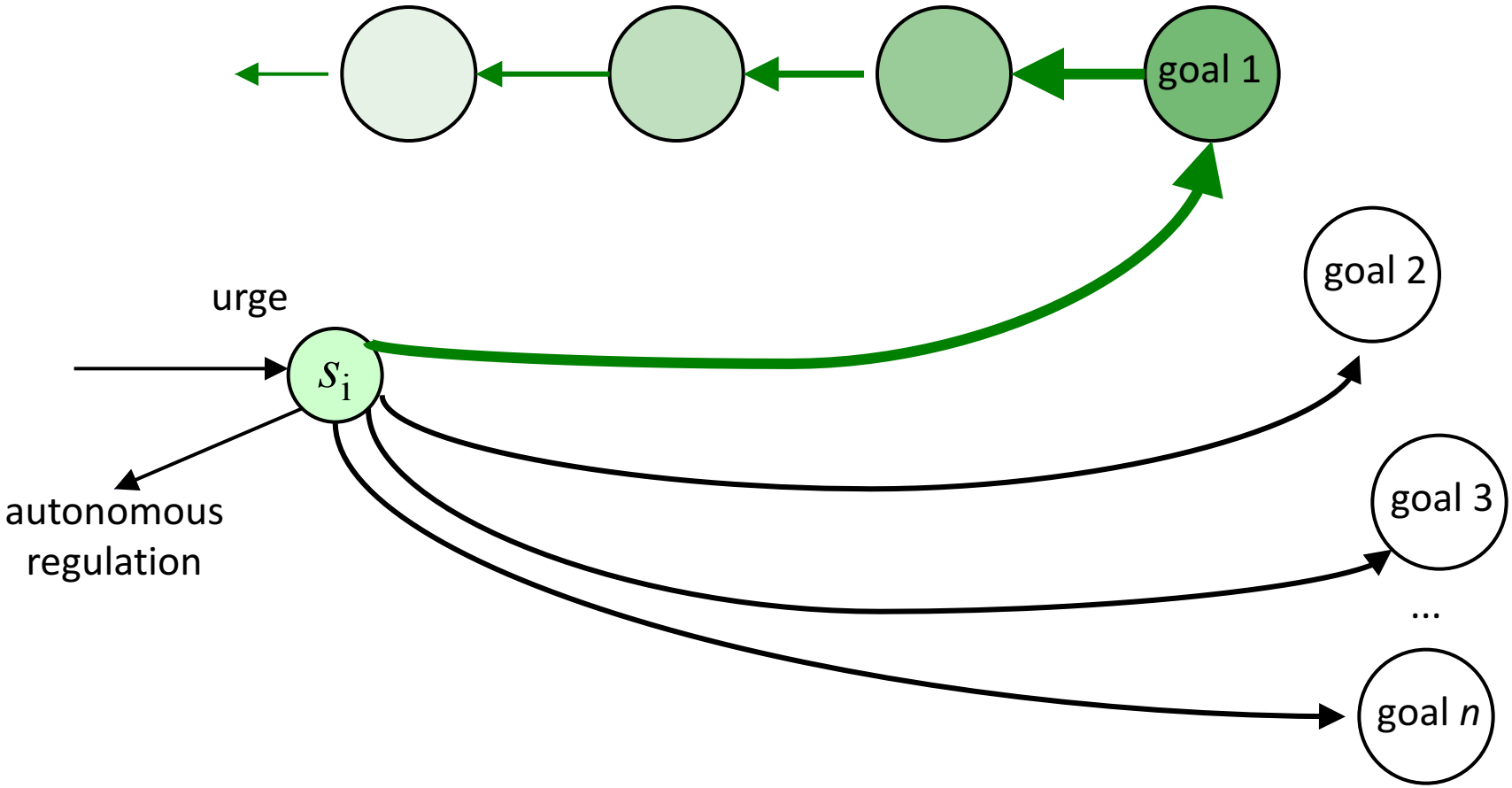
Motivational Learning

Motivator:

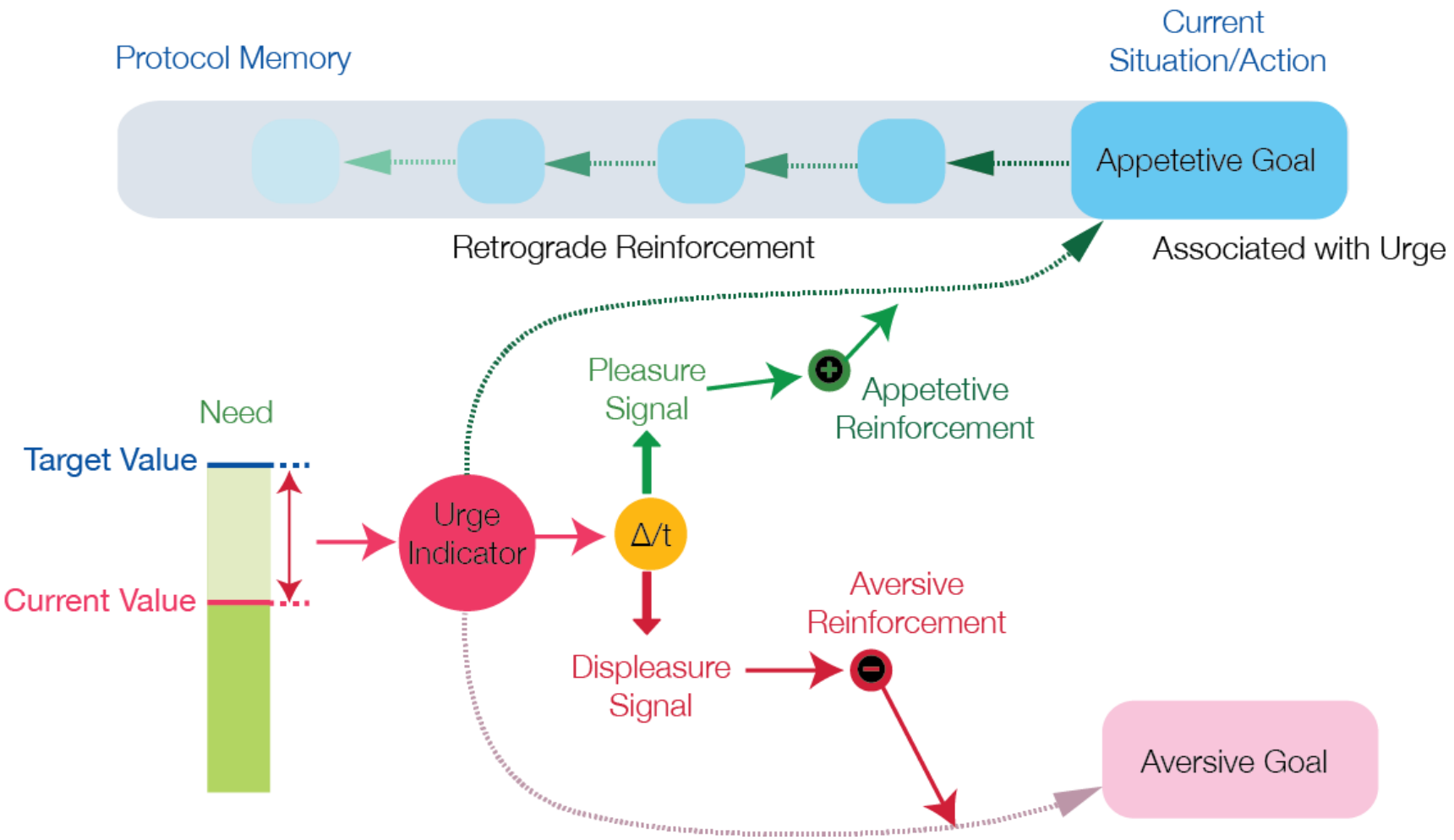


Motivational Learning

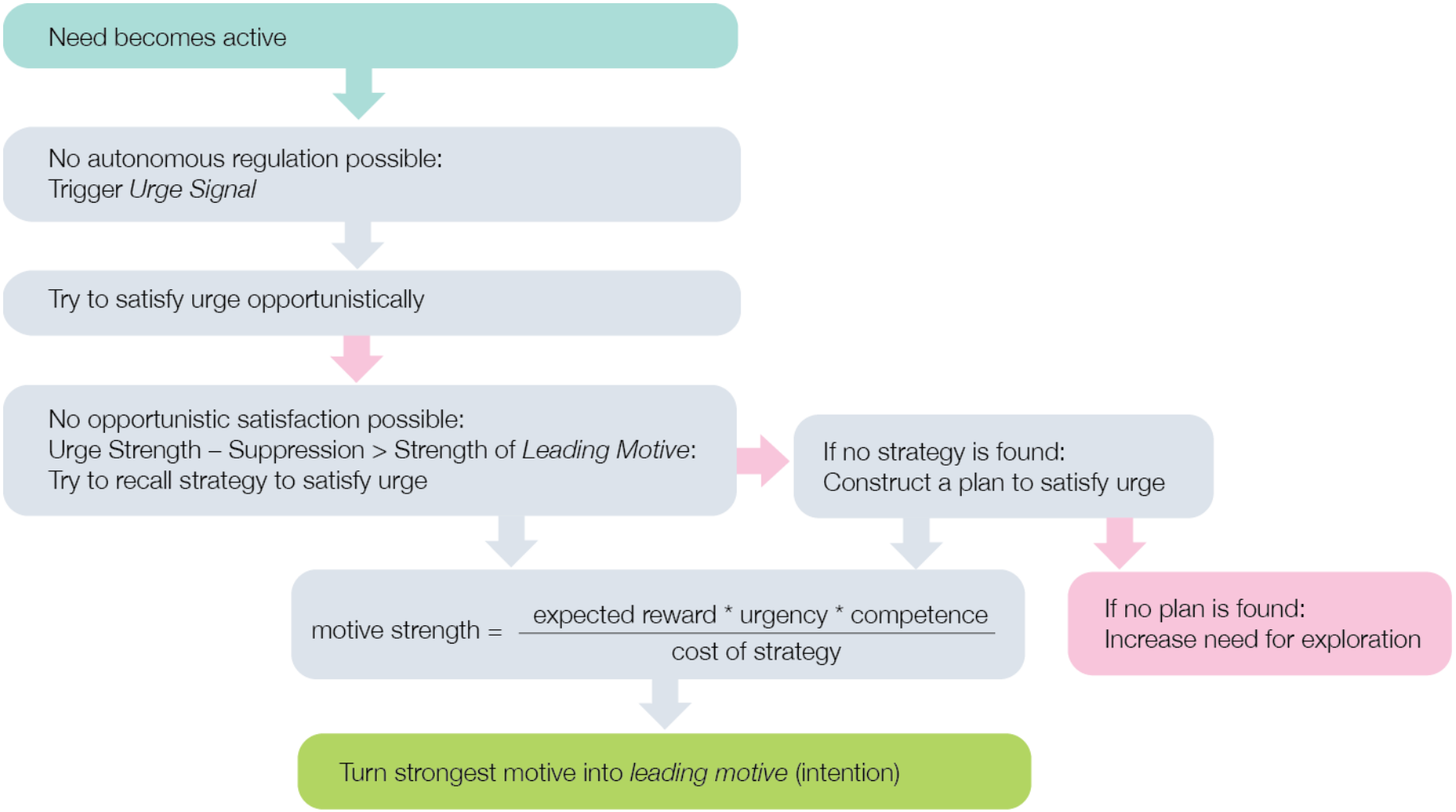
Intention:



Motivational learning



Motive selection



Need parameters

- Strength: relative importance
- Decay: rate of replenishment
- Gain: effect of satisfaction
- Loss: effect of frustration

- different configuration of need parameters = different personality traits

Anticipated needs

- Actual reward determines reinforcement (striatum, basal ganglia)
- Anticipated reward determines action (prefrontal dopamine)

- Valenced reactions are not only caused by present rewards, but also by imagined/anticipated rewards (expectations, memories)

Implementation of motivation

$$Need := \left\{ \begin{array}{c} type \in \{physiological, social, cognitive\}, \\ value_{\,t} \in [0,1], \\ value_{\,t0} \in [0,1], \\ urge_{\,t} \in [0,1], \\ urgency_{\,t} \in [0,1], \\ pleasure_{\,t} \in [0,1], \\ pain_{\,t} \in [0,1], \\ weight \in \mathbb{R}^{+}, \\ decay\ time \in [-1,1], \\ gain \in [0,1], \\ loss \in [0,1], \\ satisfaction^{img} \in [0,1], \\ frustration^{img} \in [0,1], \\ pleasure\ sensitivity \in \mathbb{R}^{+}, \\ pain\ sensitivity \in \mathbb{R}^{+}, \\ pleasure\ decay\ time \in \mathbb{R}^{+}, \\ pain\ decay\ time \in \mathbb{R}^{+}, \\ pleasure\ sensitivity^{img} \in [0,1], \\ pain\ sensitivity^{img} \in [0,1] \end{array} \right\}$$

Urge strength and urgency

- $urge_t = weight\,[1 - value_{t-1}]_0^{1^2}$
- $urgency_t = weight\,\left[\frac{k - remaining\ time_t}{k}\right]_0^{1^2}$

Value and decay of a need

- $$value_t = \left[\begin{array}{c} \text{decay}(value_{t-1}) \\ +gain \times \delta_t^{consume} \\ +loss \times \delta_t^{aversive} \\ +gain \times satisfaction^{img}\ \delta_t^{img\ consume} \\ +loss \times frustration^{img}\ \delta_t^{img\ aversive} \end{array} \right]_0^1$$

- $$\text{decay}(v_{t-1}) := \sigma\left(\sigma^{-1}(v_{t-1}) + \frac{\text{duration}(t,t-1)}{decay\ time}\right)$$

Pleasure and pain associated with a need

- $pleasure_t = \begin{bmatrix} \text{decay}(pleasure_{t-1}, pleasure\ decay\ time) \\ +gain \times pleasure\ sensitivity\ \times \delta_t^{consume} \\ +gain \times pleasure\ sens.^{img} \times\ \delta_t^{\ img\ consume} \end{bmatrix}_0^1$

- $pain_t = \begin{bmatrix} \text{decay}(pain_{t-1}, pain\ decay\ time) \\ +loss \times pain\ sensitivity \times \delta_t^{aversive} \\ +loss \times pain\ sensitivity^{img} \times\ \delta_t^{\ img\ aversive} \\ +\text{pain from depletion}(\text{value}_{t-1}) \end{bmatrix}_0^1$

- $\text{pain from depletion}(value) \coloneqq \left[\left(1 - \frac{value}{\theta}\right)\right]_0^{1^2}$

Events and consumptions

- $Event \coloneqq \left\{ \begin{array}{c} consumption \in Consumptions, \\ expected\ reward \in [-1, 1], \\ certainty \in (0, 1], \\ skill \in [0, 1], \\ remaining\ time_t \in \mathbb{R}^+ \end{array} \right\}$

- $Consumption \coloneqq \left\{ \begin{array}{c} need \in Needs, \\ reward_t \in \mathbb{R}, \\ total\ reward \in \mathbb{R}, \\ reward\ duration \in \mathbb{R}^+, \\ max\ reward \in \mathbb{R}, \\ discount \in [0, 1] \end{array} \right\}$

Reward signal and reward summation

- $signal(t) := \ t\, e^{-\frac{1}{2}t^2}$
- $t_1 = (t - t_{onset})\ \dfrac{k}{reward\ duration}\ \mathrm{duration(t, t-1)}$
- $t_1 = (t - t_{onset})\ \dfrac{k}{reward\ duration}\ \mathrm{duration(t, t-1)}$
- $reward_t = \left[\dfrac{k \times total\ reward}{reward\ duration} \int_{t_1}^{t_2} t\, e^{-\frac{1}{2}t^2}\right]_{-max\ reward}^{max\ reward}$

$$= \left[\frac{k \times total\ reward}{reward\ duration}\left(e^{-\frac{1}{2}t_1^2} - e^{-\frac{1}{2}t_2^2}\right)\right]_{-max\ reward}^{max\ reward}$$

Actualized rewards change values of needs

- $\delta_t^{consume} = \left[reward_t^{consumption}\right]_0^{\infty}$
- $\delta_t^{aversive} = \left[reward_t^{consumption}\right]_{-\infty}^{0}$

Dealing with anticipated rewards

- $\delta_t^{img\ consume} = certainty \times skill \times \left[\frac{reward_t^{consumption}}{1+d \times remain.time_t}\right]_0^{\infty}$

- $\delta_t^{img\ aversive} = cert. \times (1 - skill) \times \left[\frac{reward_t^{consumption}}{1+d \times remain.time_t}\right]_{-\infty}^{0}$

(d is discount; hyperbolic discounting)

Cognitive Artificial Intelligence

Methods should focus on components and performances necessary for intelligence:

- **Universal Representations:**
  Dynamic model of environment, possible worlds, and agent
- **(Semi-) Universal Problem Solving:**
  Learning, Planning, Reasoning, Analogies, Action Control, Reflection ...
- **Universal Motivation:**
  Polythematic, adaptive goal identification
- **Emotion and affect**
- **Whole, testable architectures**

Modeling Emotion

Emotional expression

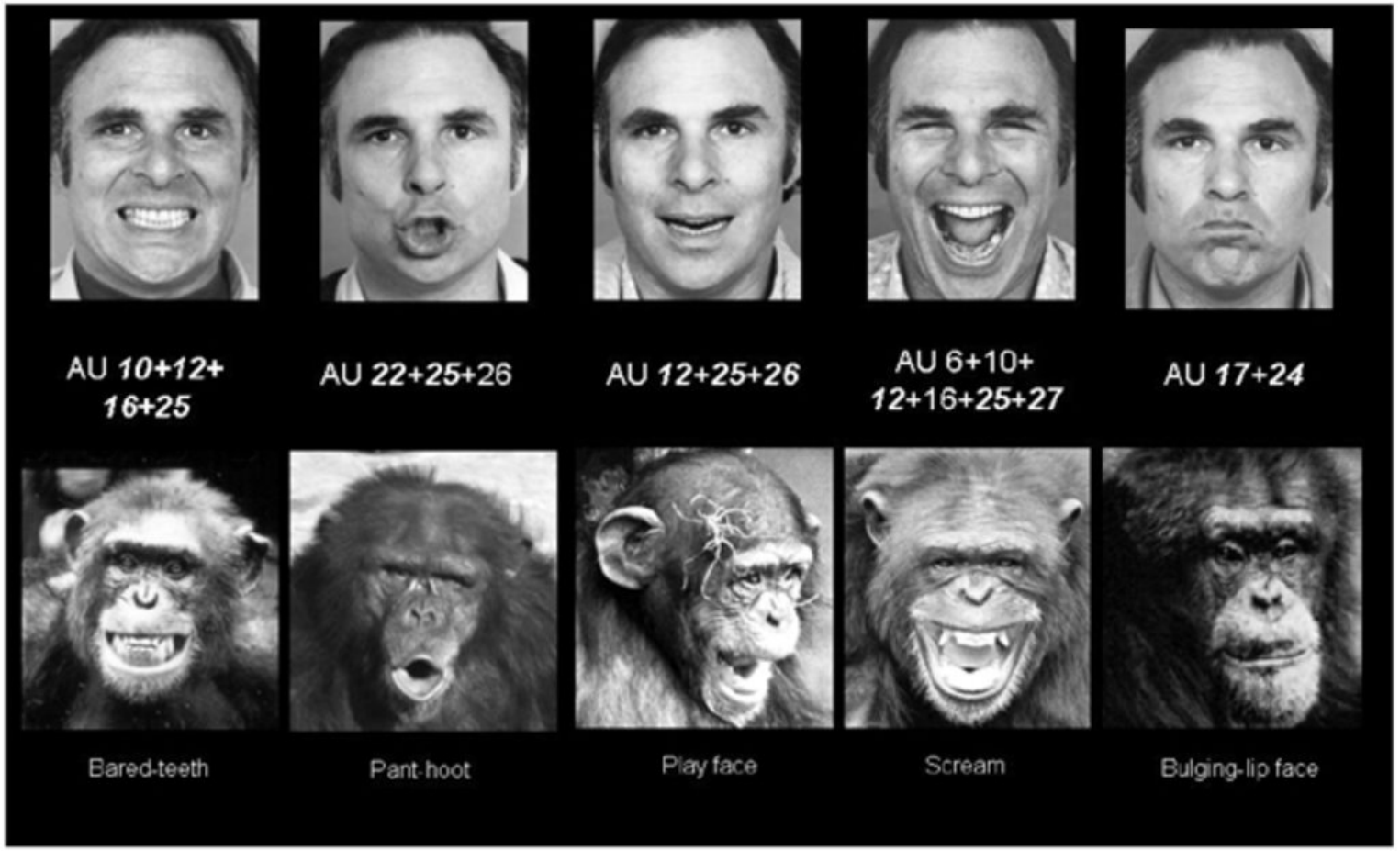
- Paul Ekman: Facial Action Coding

Emotional expression

- Paul Ekman: Facial Action Coding

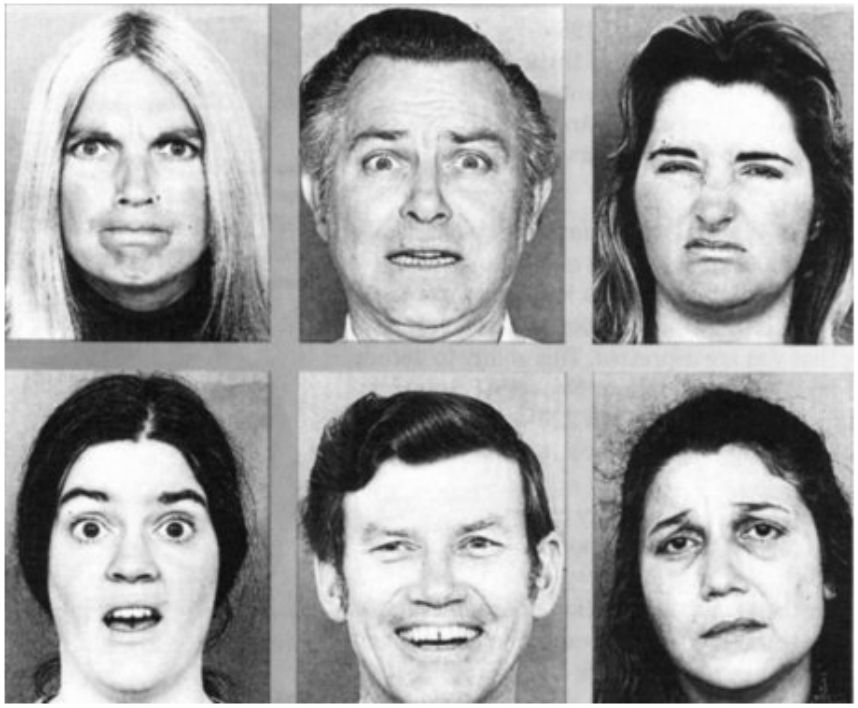
Emotional expression

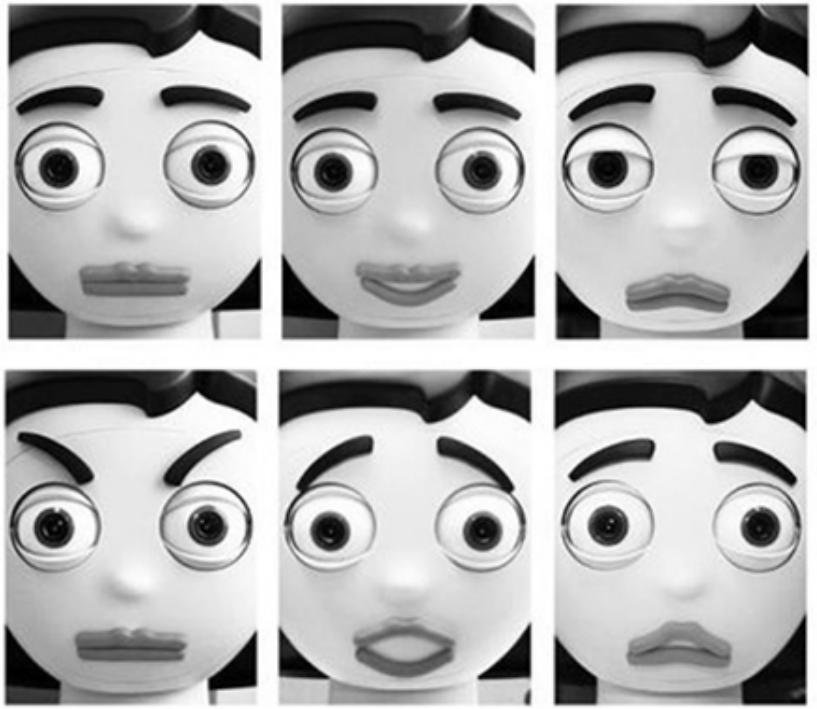
- Paul Ekman: Facial Action Coding



Affective computing

Recognize, process, simulate, express human affects

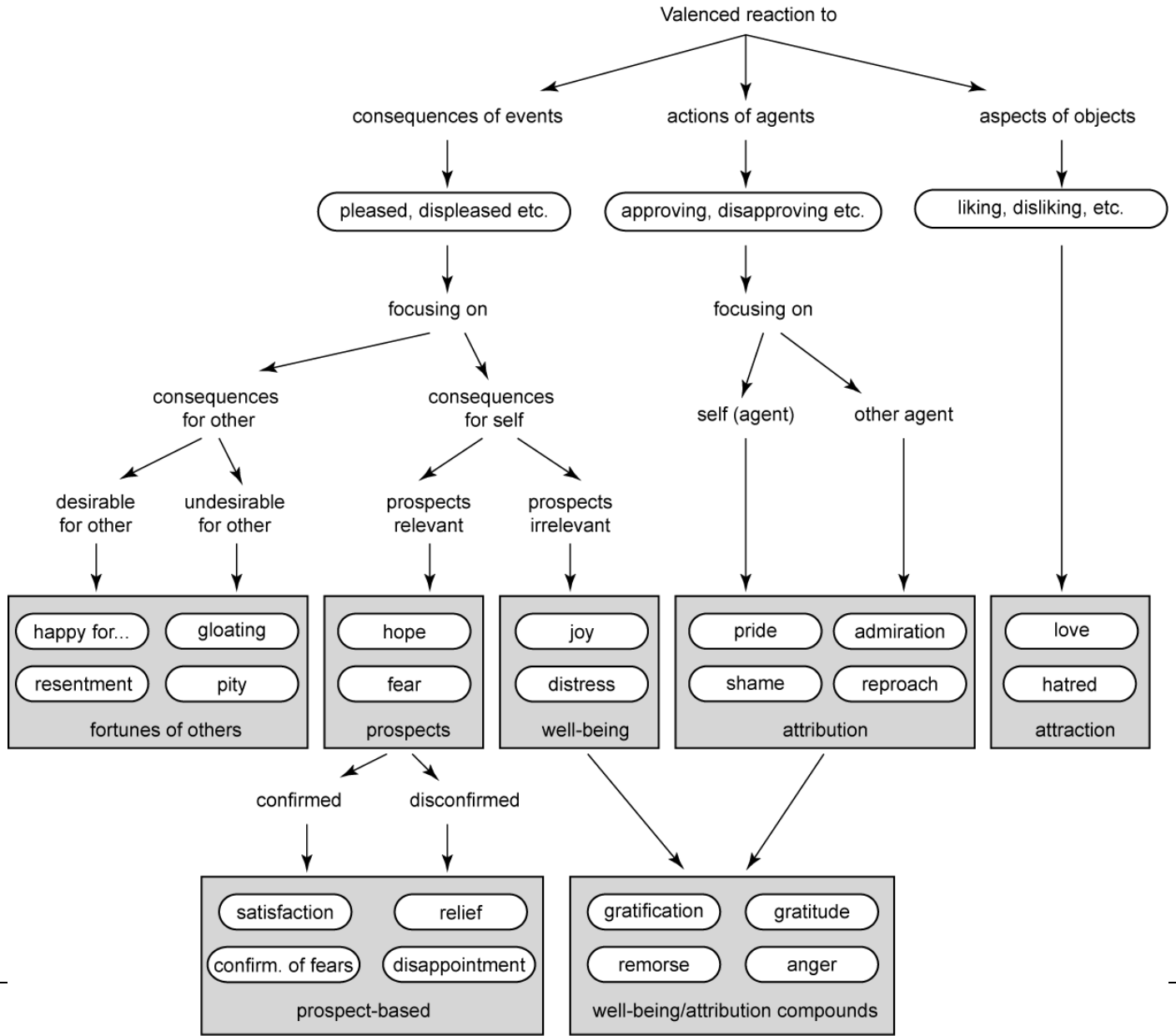




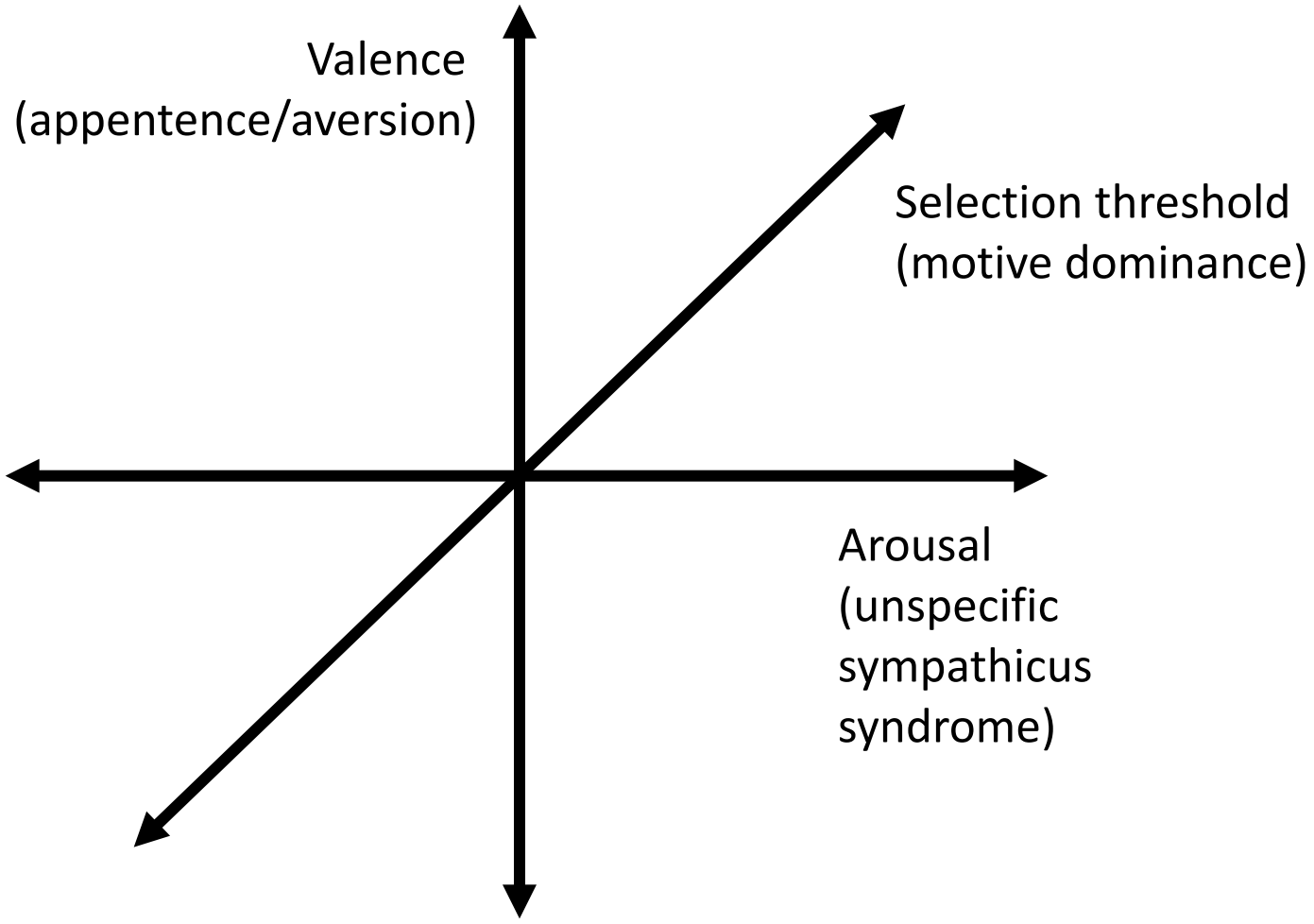
Appraisal models

- Magda Arnold, Richard Lazarus:
  - Emotion as cognitive appraisals of relations, motivation, cognition
- Klaus Scherer:
  - Stimulus evaluation checks
  - innate (sensory motor) $\rightarrow$ learned (schemas) $\rightarrow$ deliberate
  - relevance $\rightarrow$ implication $\rightarrow$ coping potential $\rightarrow$ normative signific.

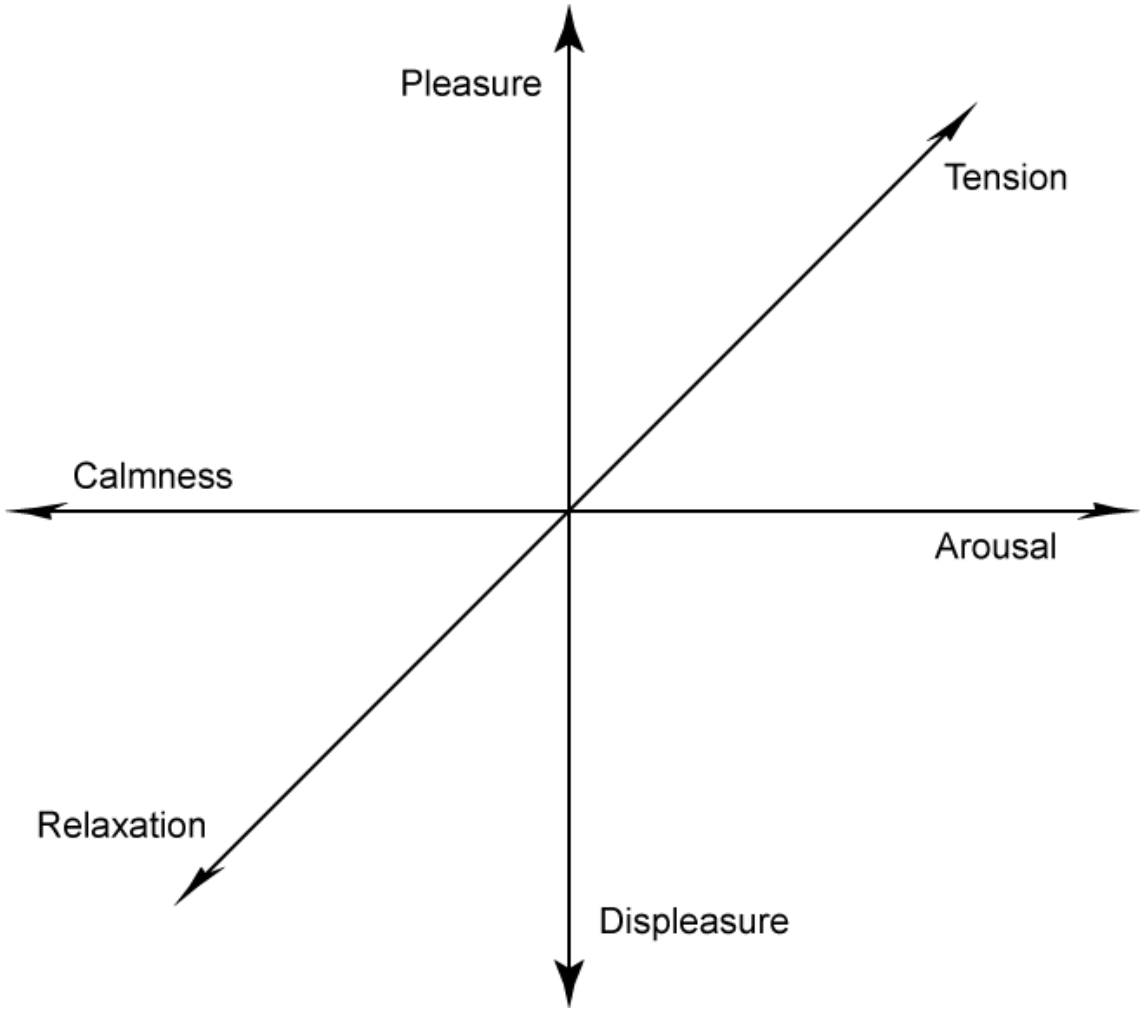
Conceptual analysis: Ortony, Clore, Collins 1988:



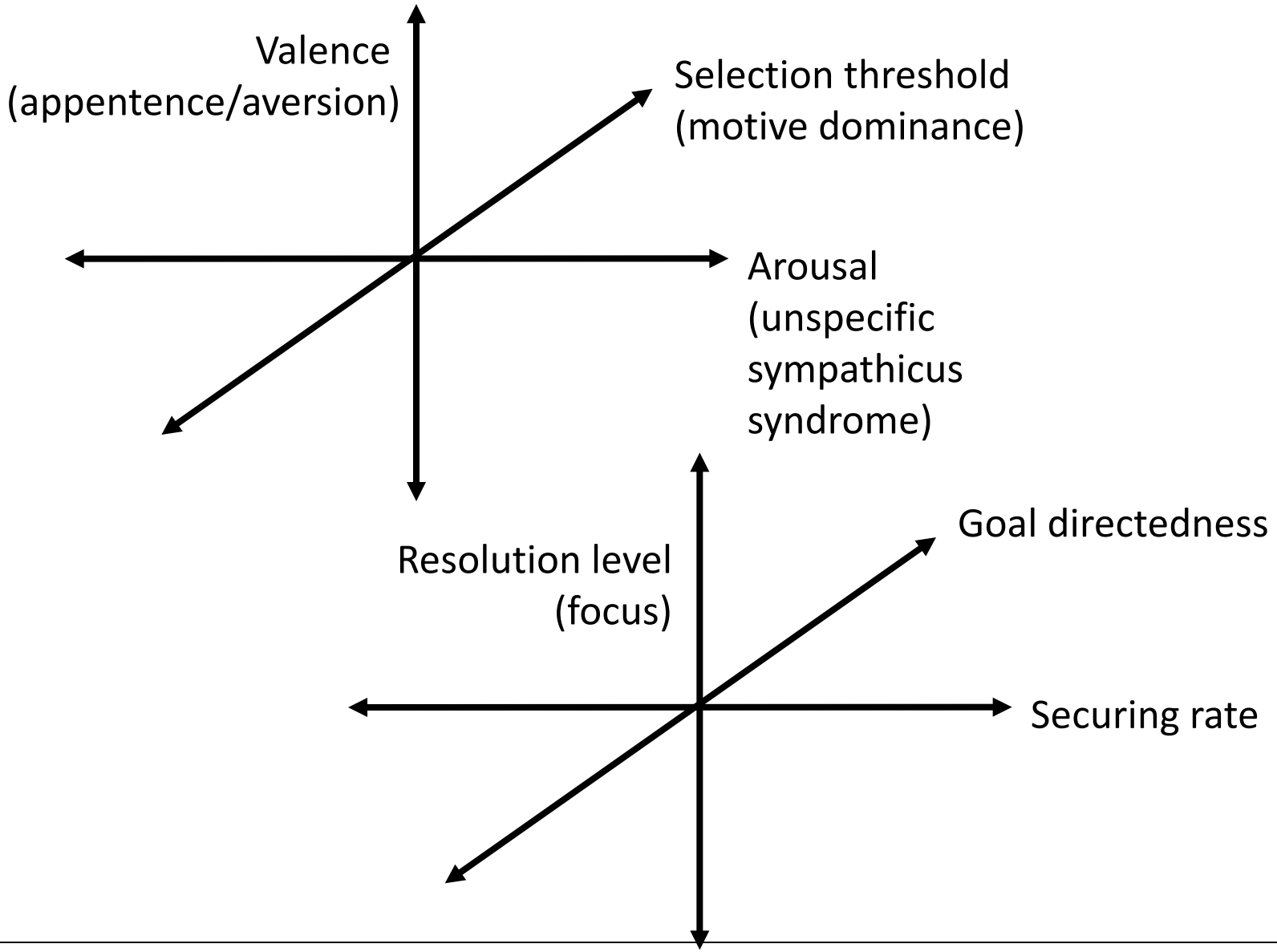
Affective dimensions in the PSI theory (Dörner 1999)



Compare: Affective dimensions (Wundt 1910)



Affective dimensions in the PSI theory



The Psi theory about emotion

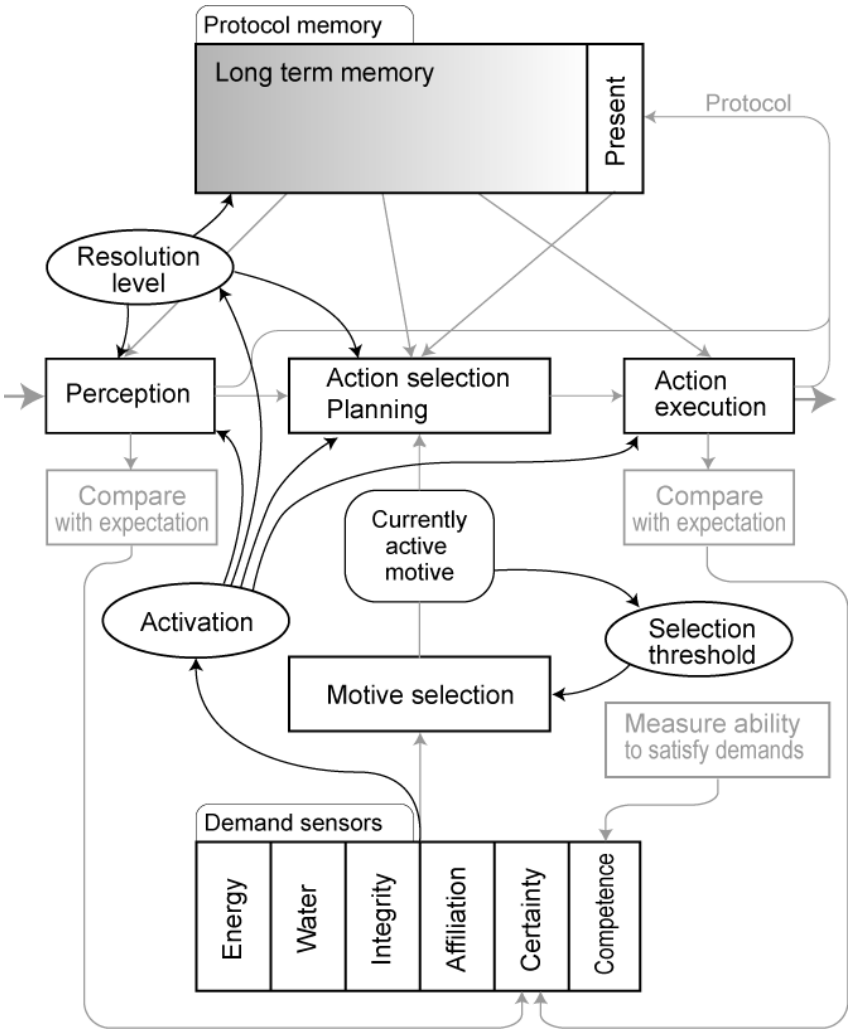
- Affect is seen as a *configuration* of a cognitive system
- Modulators of cognition:
  - arousal, selection threshold, securing threshold, resolution level
  - estimate of competence and certainty
  - pleasure/distress signals $\rightarrow$ valence
- Affective state is emergent property of modulation
- 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
  $\rightarrow$ modify perception, memory, planning and action selection

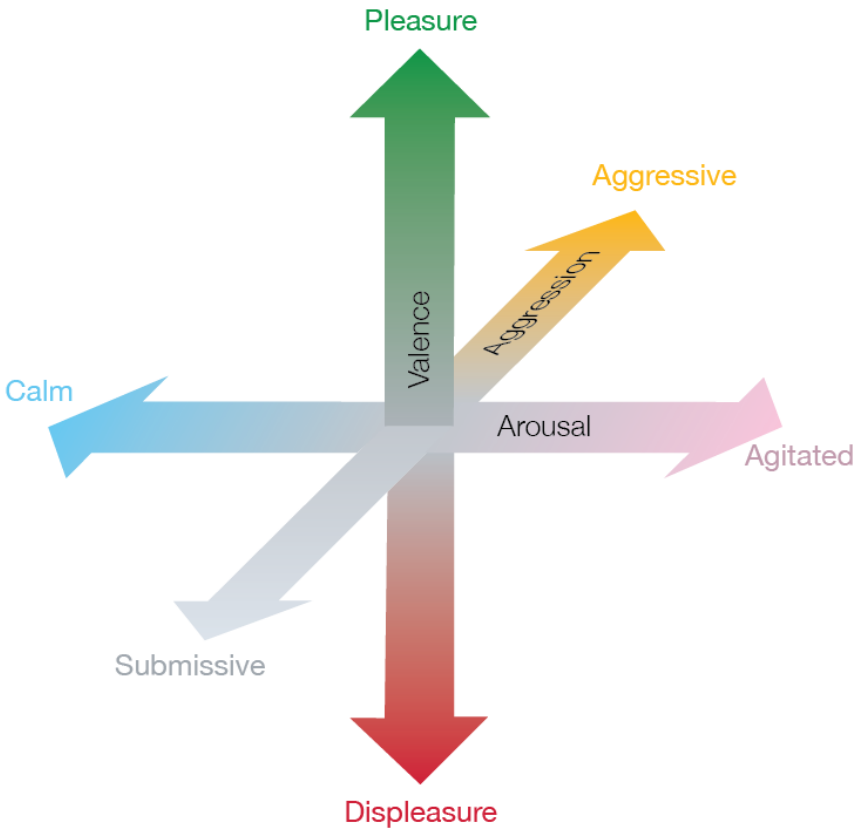
- Reduce complexity of cognitive processes

Modulation in PSI/MicroPsi



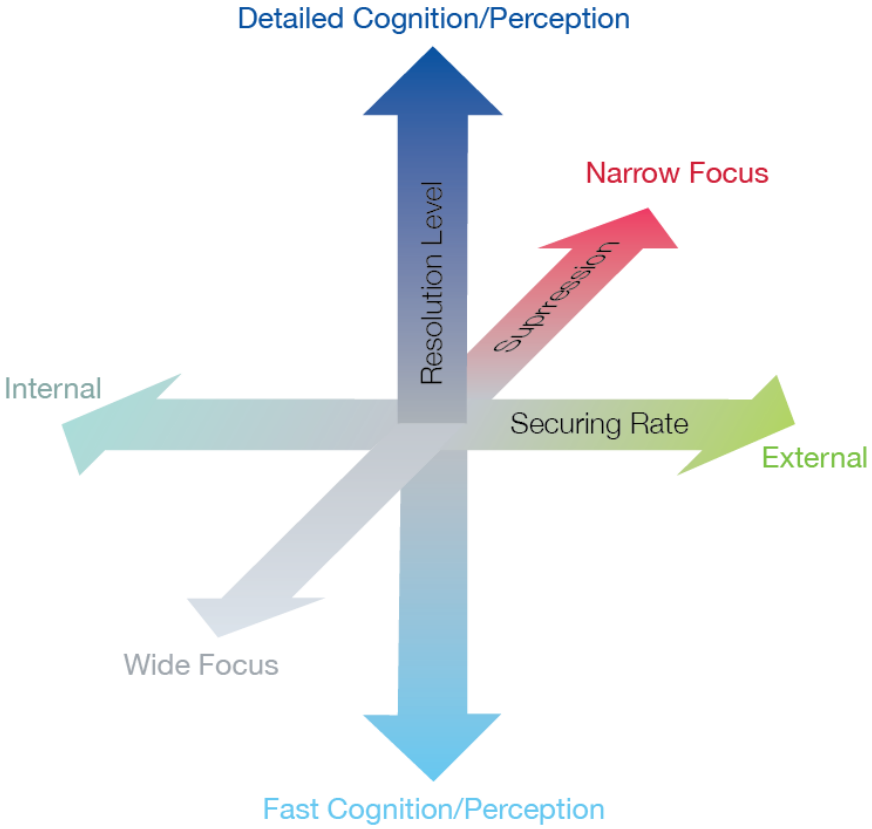
Primary modulators

Arousal	unspecific sympathicus syndrome
Valence	situation evaluation (good/bad)
Aggression	fight or flight

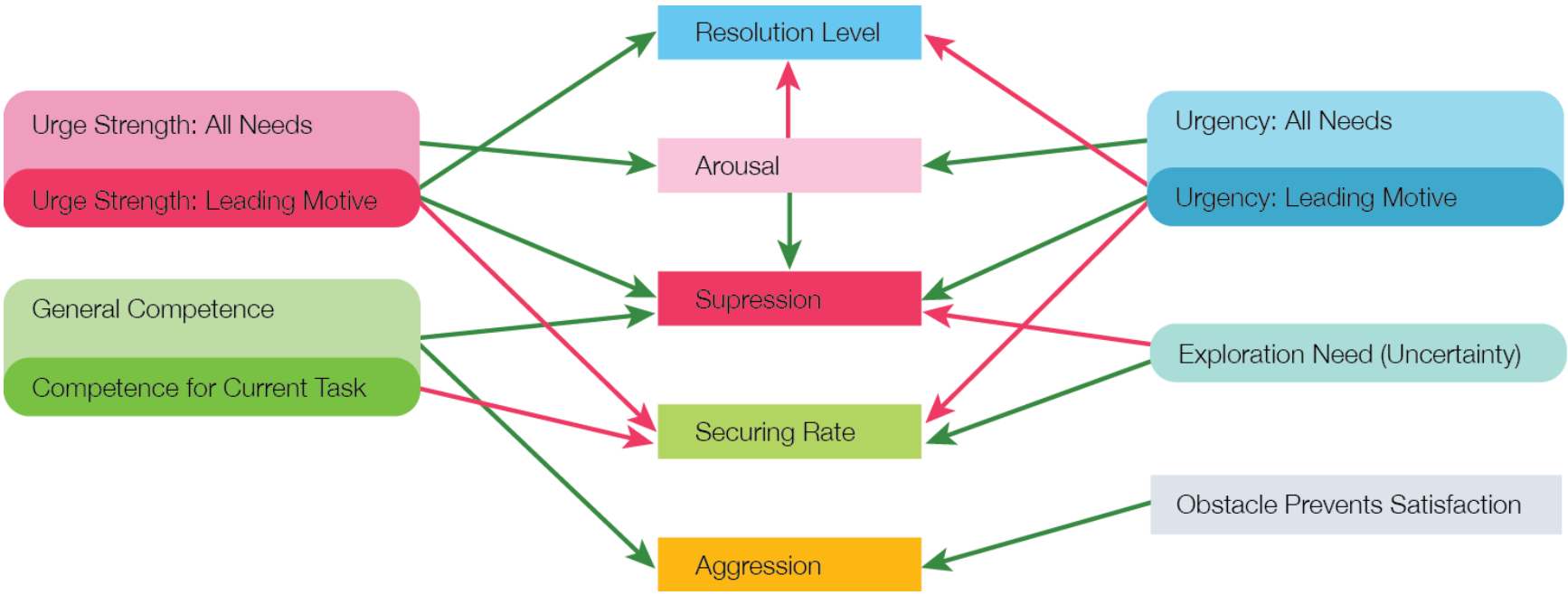


Attentional modulators

Resolution Level	width of focus
Supression	depth of focus; motive stability
Securing Rate	rate of checking the environment



Modulator dynamics



Modulator parameters

- Baseline
- Range
- Volatility
- Duration

- Different modulator parameter configurations = different temperaments

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)

Implementation of affective modulation

- $Modulator := \left\{ \begin{array}{c} min \in \mathbb{R} \\ max \in \mathbb{R}, \\ level_t \in [min, max], \\ baseline \in [min, max], \\ volatility \in \mathbb{R}^+, \\ decay\ time \in \mathbb{R}^+ \end{array} \right\}$

- $interval = \begin{cases} max - baseline, \text{if } level_{t-1} > baseline \\ baseline - min, \text{else} \end{cases}$

- $\delta_t = (target \times interval + baseline - level_{t-1})\ volatility$

Implementation of valence

- $\text{marginal sum}(V, limit) := \sum_{n=0}^{|V|} S_n \Big| S_n := \frac{limit - S_{n-1}}{limit} v_n$
- $limit = \max(\{\omega | \omega \in weight_{needs}\})$

- $combined\ pain = \text{marginal sum}(\{\text{weight}_{\text{need}} \times pain_{need}\})$

- $combined\ pleasure = \text{marginal sum}(\{\text{weight}_{\text{need}} \times pleasure_{need}\})$

- $target^{valence} = \frac{combined\ pleasure - combined\ pain}{limit}$

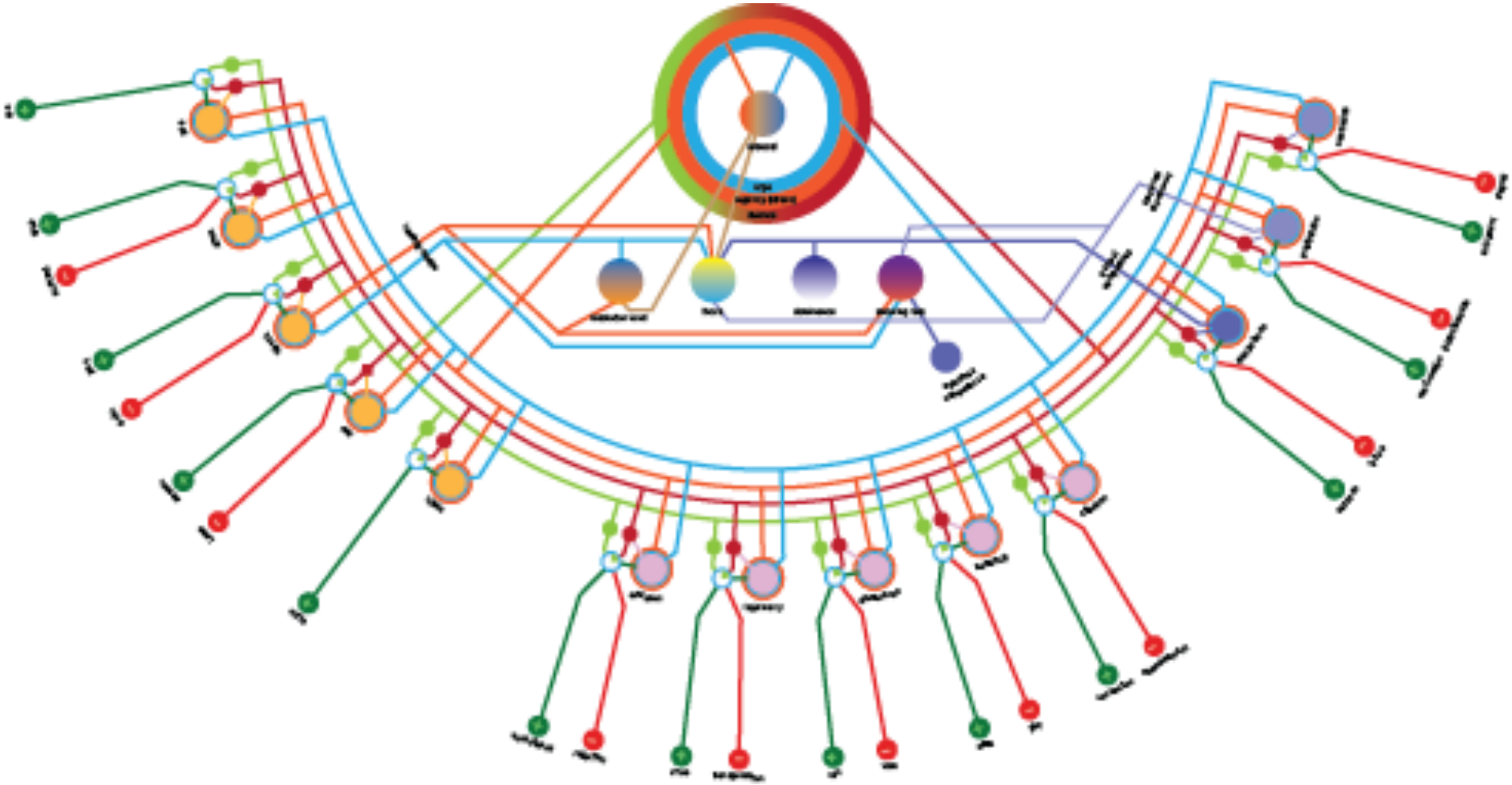
Implementation of arousal

- $combined\ urge = \frac{\text{marginal sum}(\{\text{weight}_{\text{need}} \times urge_{need}\})}{limit}$
- $combined\ urgency = \frac{\text{marginal sum}(\{\text{weight}_{\text{need}} \times urgency_{need}\})}{limit}$
- $target^{arousal} = combined\ urge + combined\ urgency - 1$

Implementation of aggression/regression

- $epistemic\ competence = skill_{current\ goal\ event}$
- $general\ competence = \sqrt{\text{value}^{\text{competence}} \times epistemic\ comp.}$
- $target^{aggression} = general\ comp. + epistemic\ comp. - 1$

Emotion viewer



Personality modeling

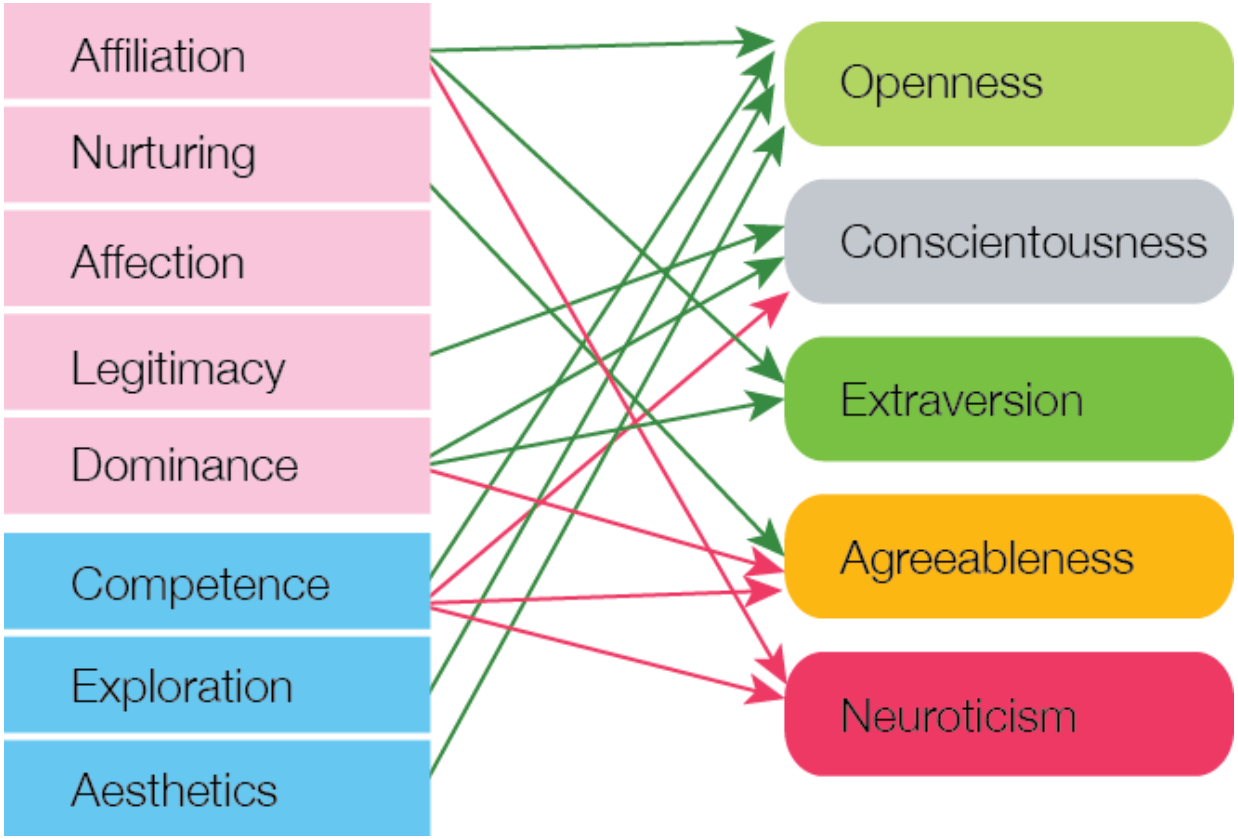
- Motivation parameters: Personality properties
- Modulation parameters: Temperament

Individual Variations by Parameterizing

Possible grounding of personality properties (FFM):

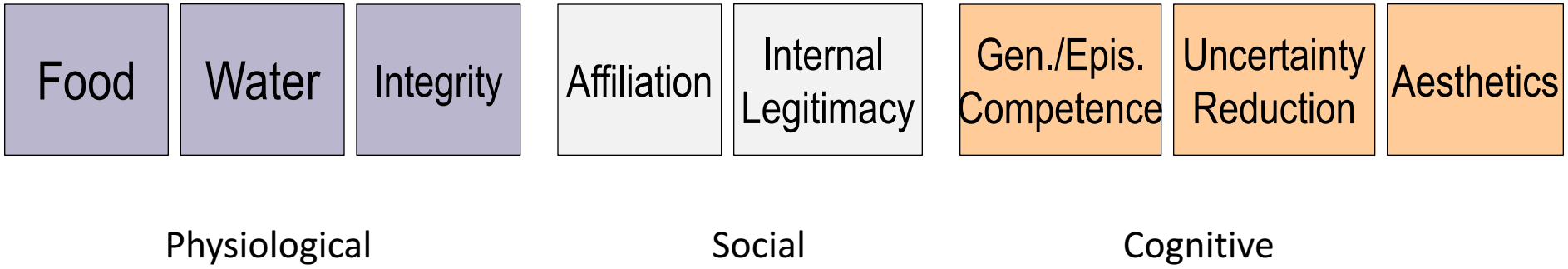
- **Openness**: appreciation of art and new ideas, curiousity
- **Conscientiousness**: rulefollowing vs. chaotic
- **Extraversion**: tendency to seek stimulation by environment and others
- **Agreeableness**: tendency for cooperativeness and compassion
- **Neuroticism**: emotional stability, effect of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

Needs and Big Five



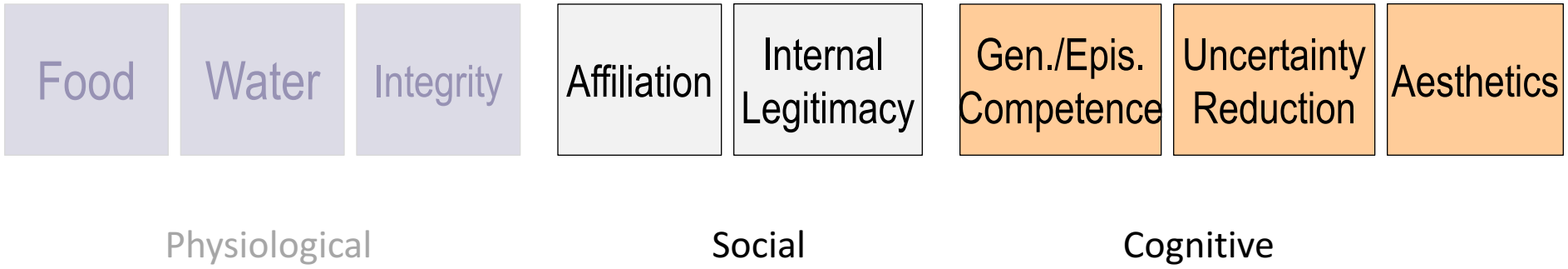
Example: Mapping to FFM (Big Five)

Demand dynamics:



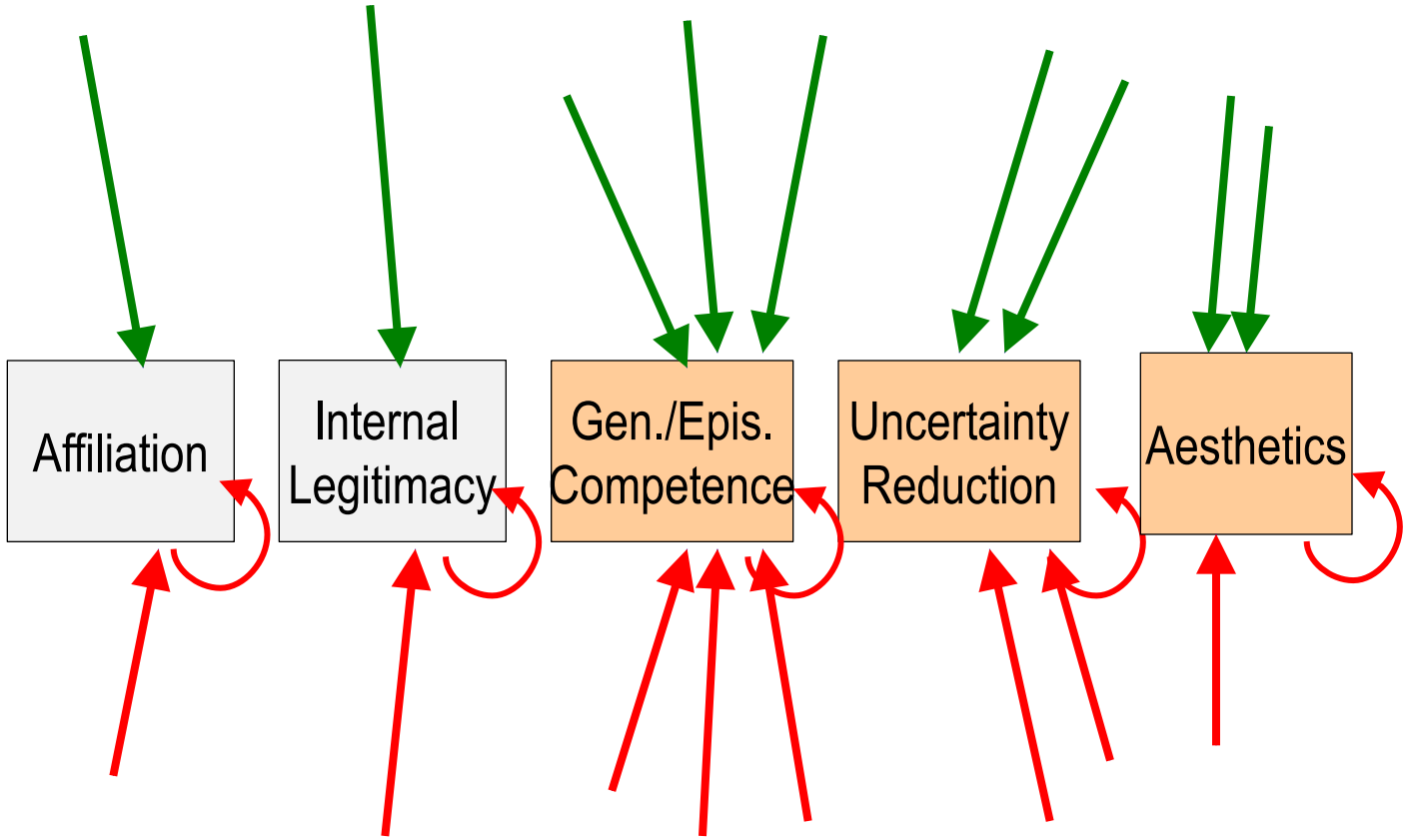
Example: Mapping to FFM (Big Five)

Demand dynamics:



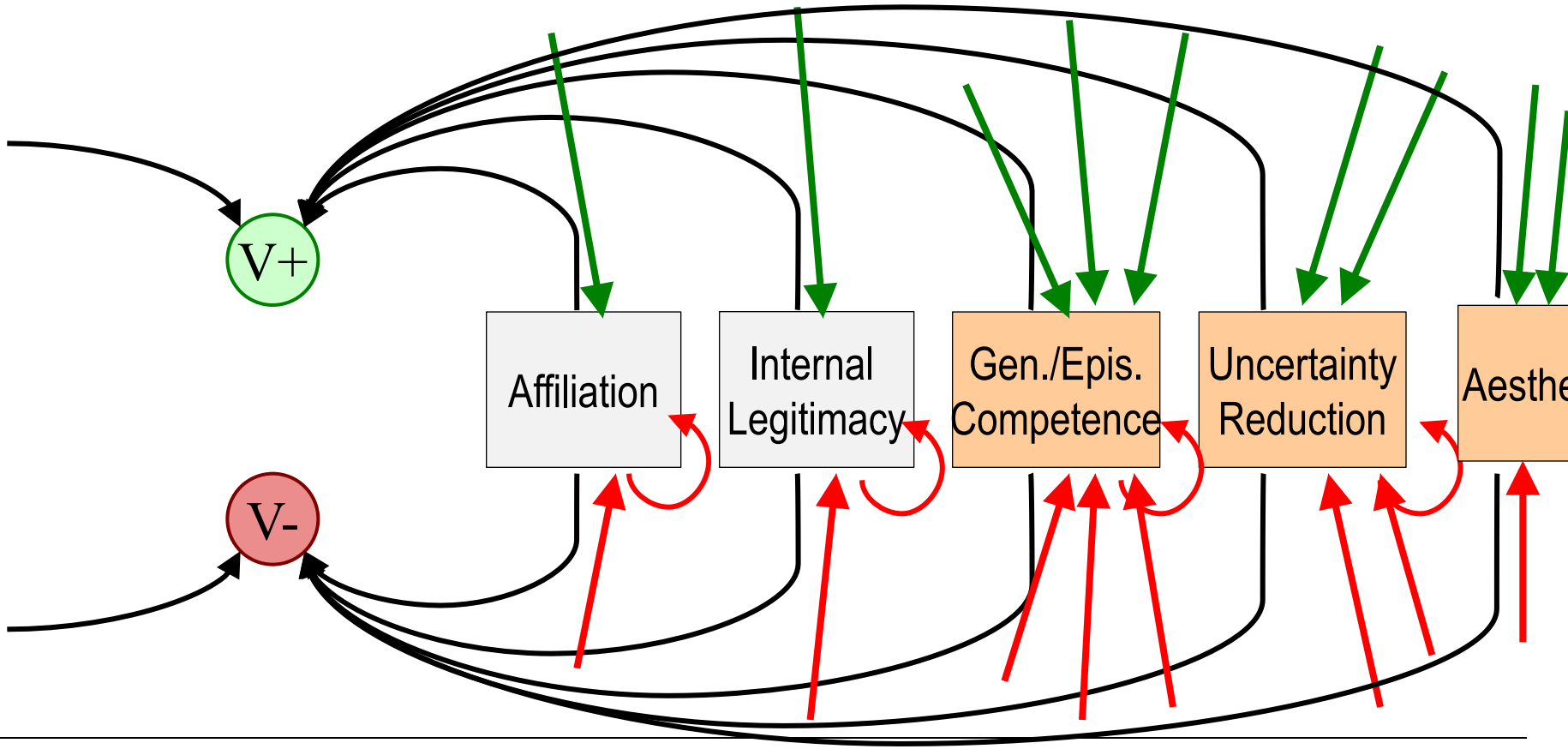
Example: Mapping to FFM (Big Five)

Demand dynamics:



Example: Mapping to FFM (Big Five)

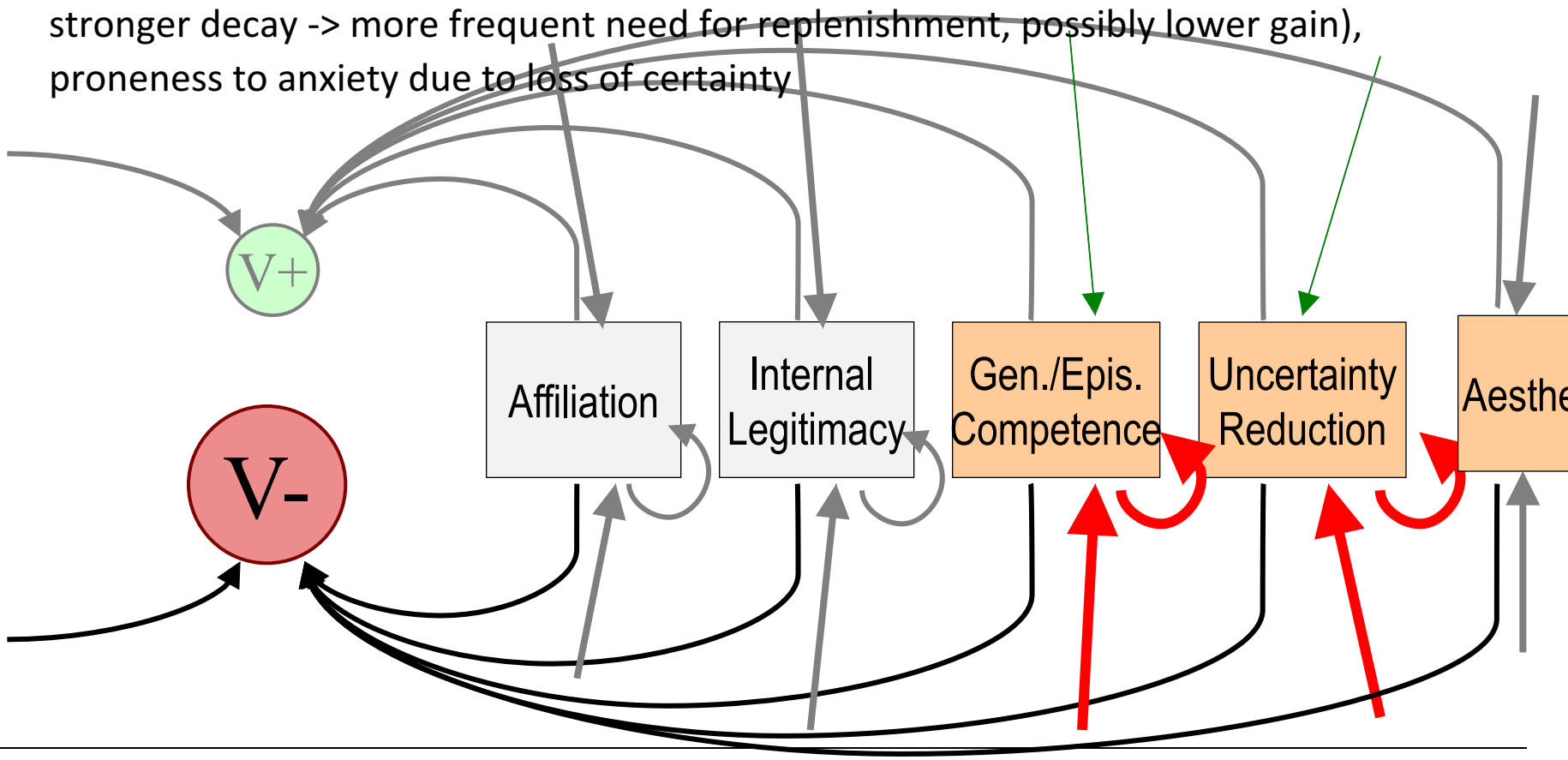
Valence: Pleasure/Pain signals



Example: Mapping to FFM (Big Five)

Neuroticism: stronger experience of negative emotions, lower emotional stability

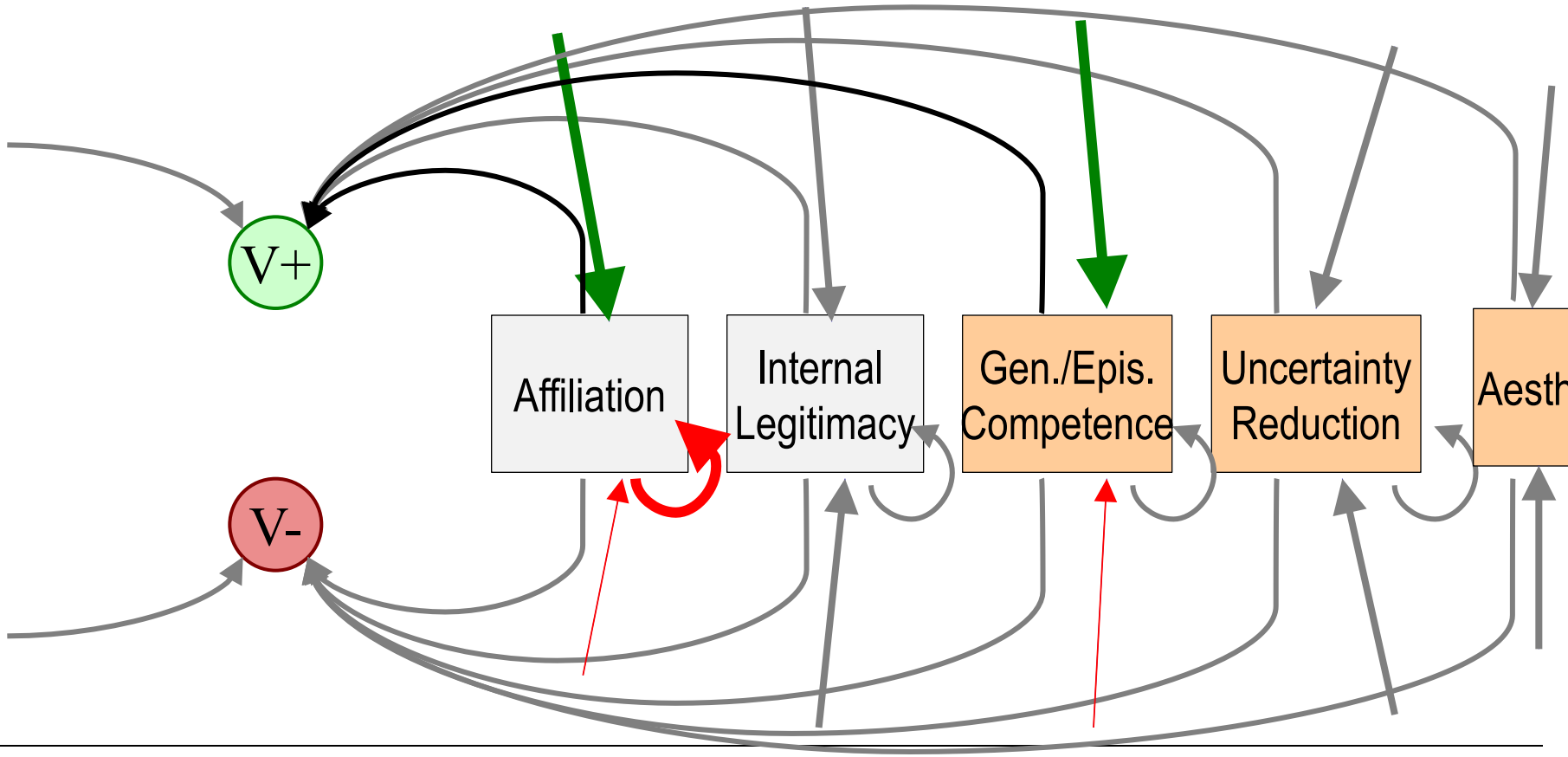
(strong negative reward/stronger loss of competence, certainty;
stronger decay -> more frequent need for replenishment, possibly lower gain),
proneness to anxiety due to loss of certainty



Example: Mapping to FFM (Big Five)

Extraversion: surgency, activity in social relations, expressivity

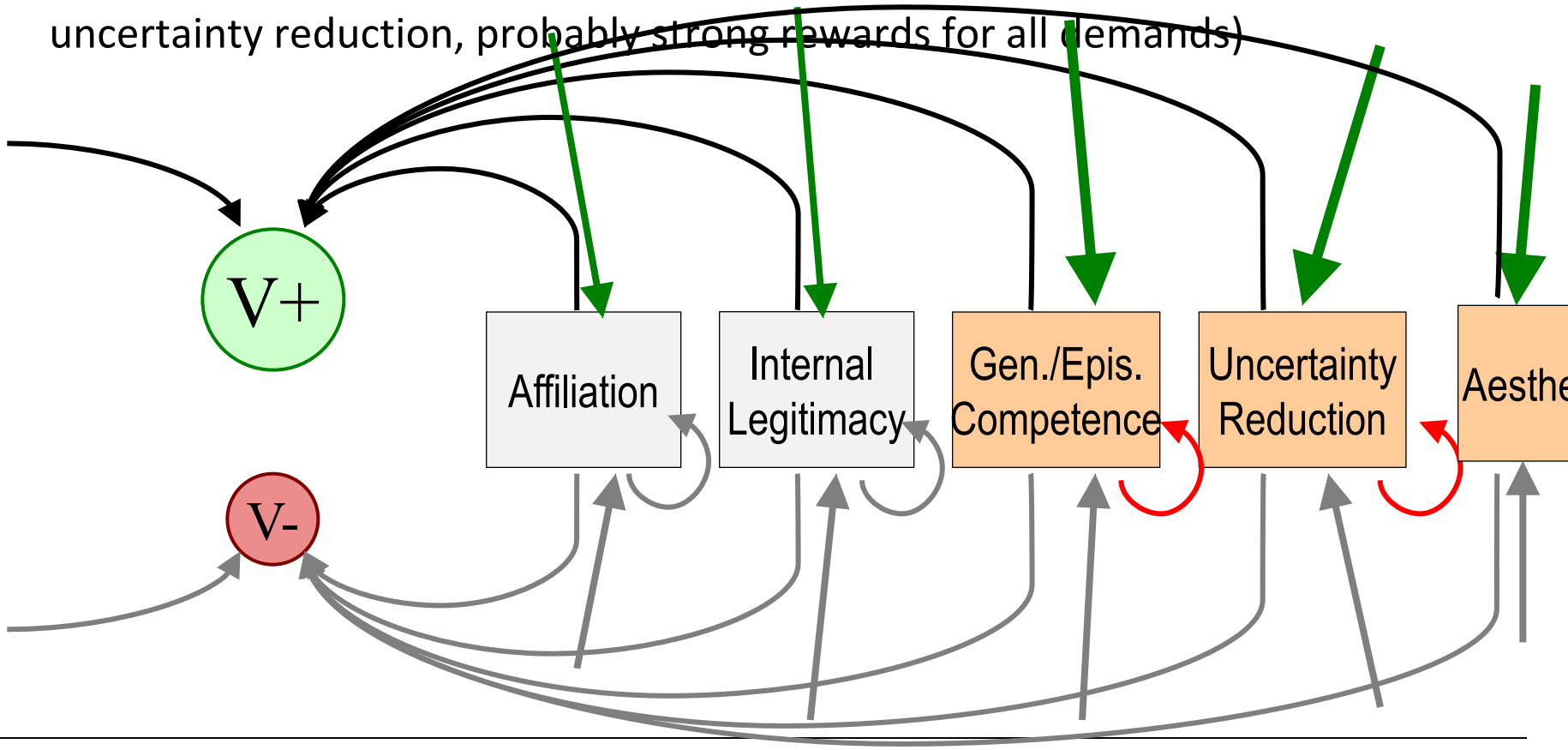
(strong gain for affiliation and competence, high decay of affiliation)



Example: Mapping to FFM (Big Five)

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

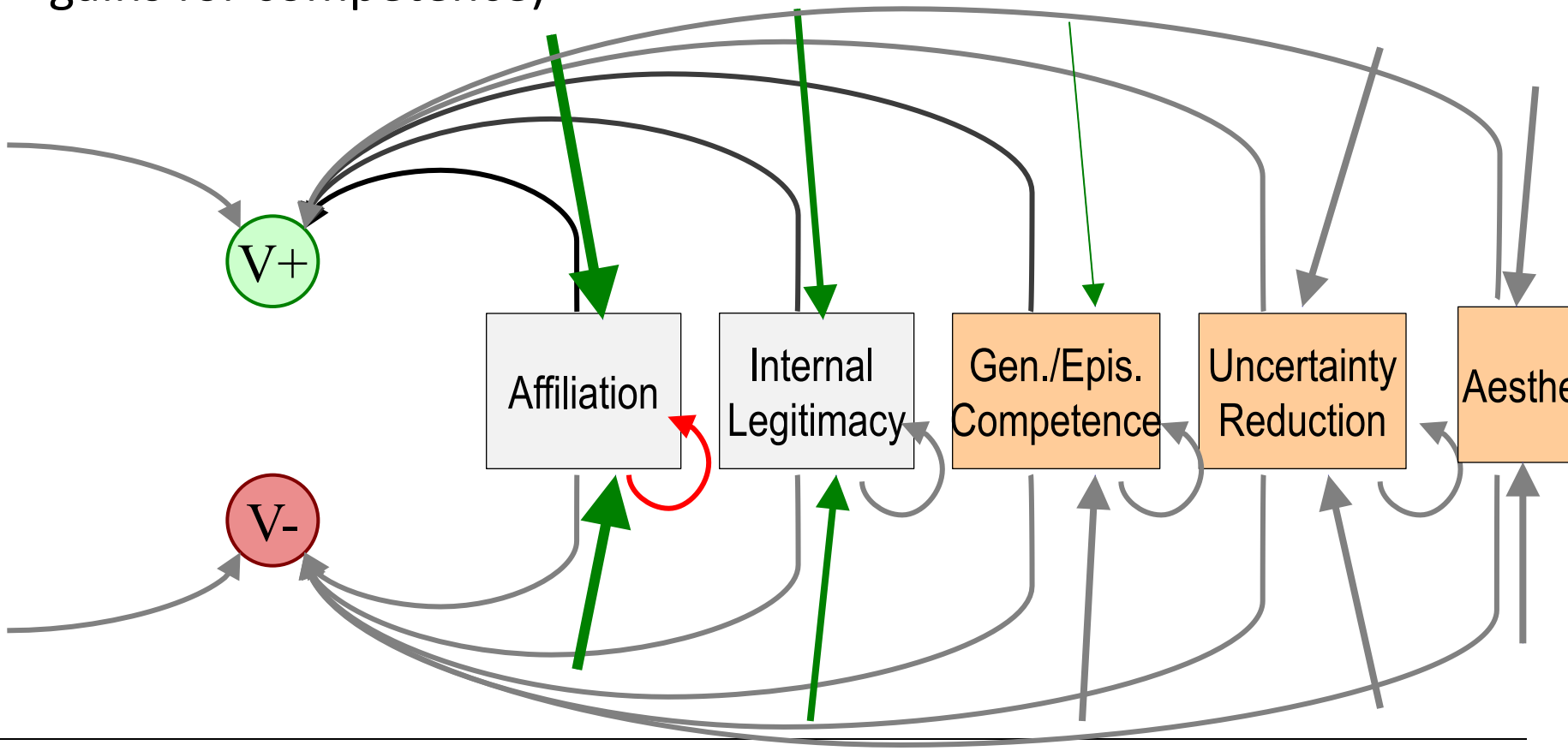
(strong gain for uncertainty reduction, high epistemic competence for uncertainty reduction, probably strong rewards for all demands)



Example: Mapping to FFM (Big Five)

Agreeableness:

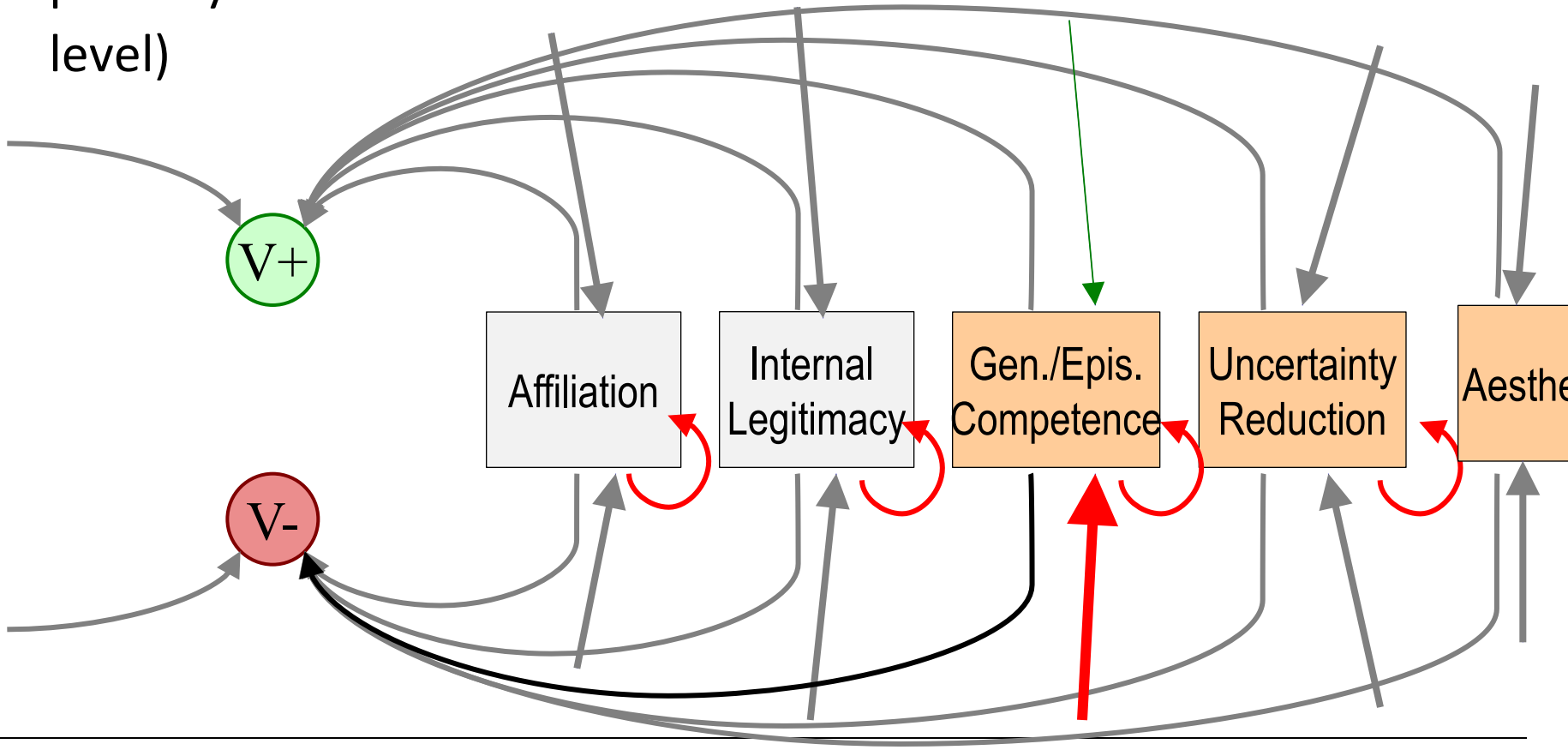
(strong positive and negative reward for affiliation, lower gains for competence)



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)



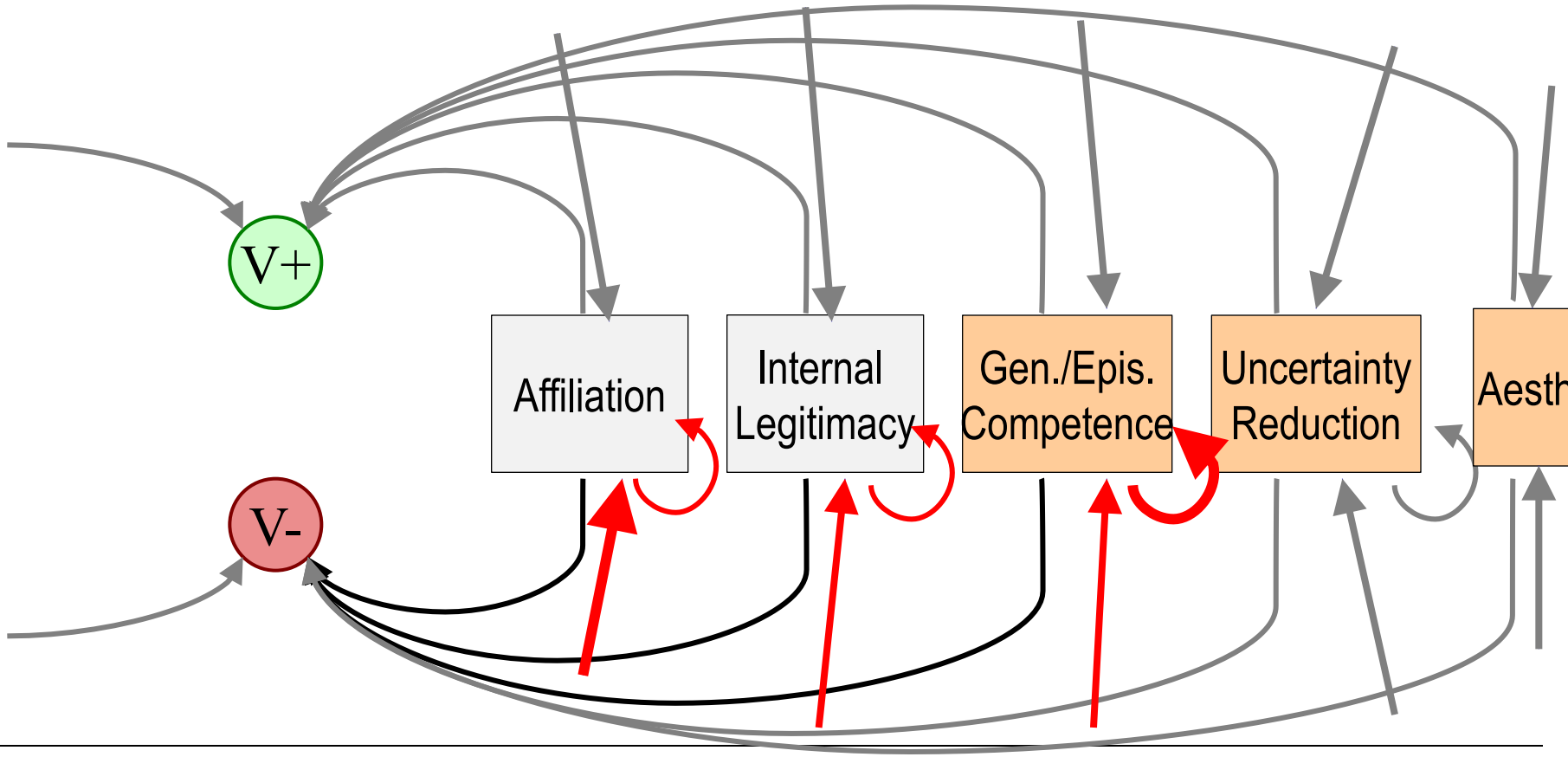
Example: Mapping to FFM (Big Five)

- Why not one free variable per FFM dimension?

FFM does not tell the complete story

Shyness != Introversion

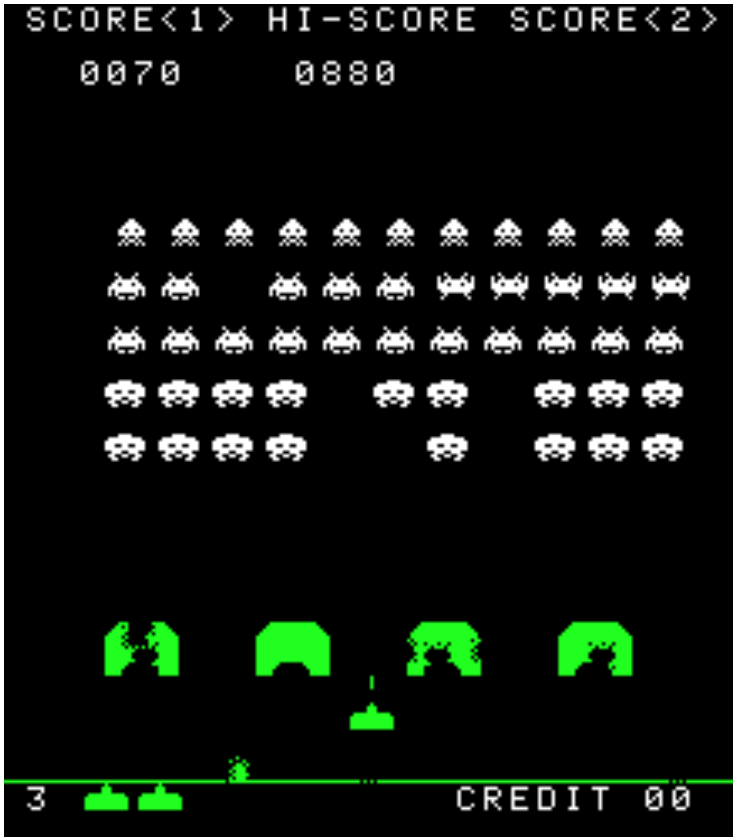
(high loss of affiliation, low competence)



How can we evaluate a model of motivation?

- Games!

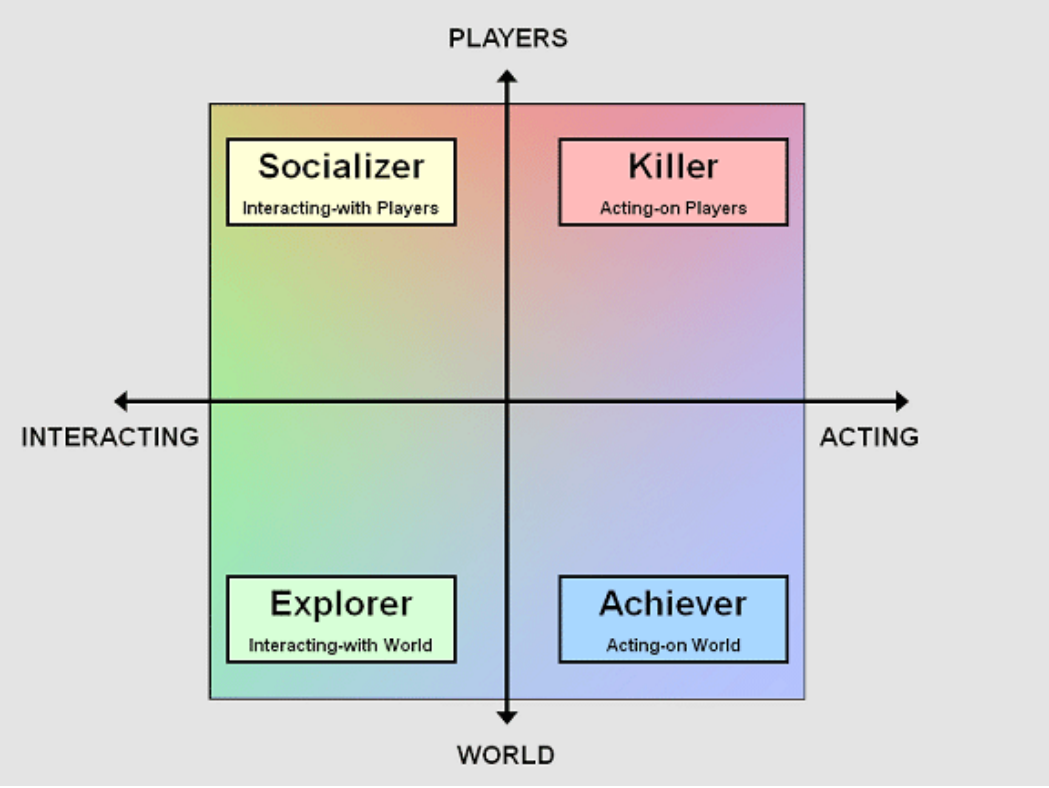
Space Invaders (1978, Tomohiro Nishikado)



World of Warcraft (Rob Pardo, Jeff Kaplan et al. 2004)

Player personality types

Richard Bartle (1996): “Hearts, Clubs, Diamonds, Spades: Players Who suit MUDs”

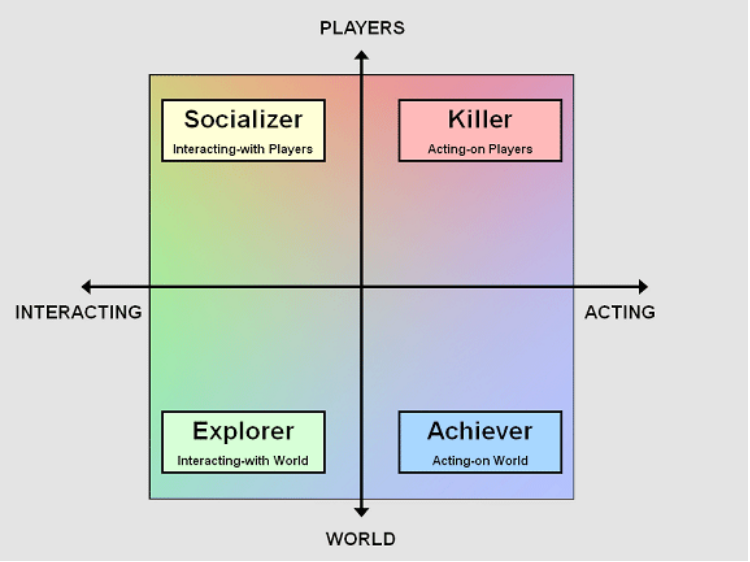


Motivation and personality

- Personality properties can be modeled as motivational variability

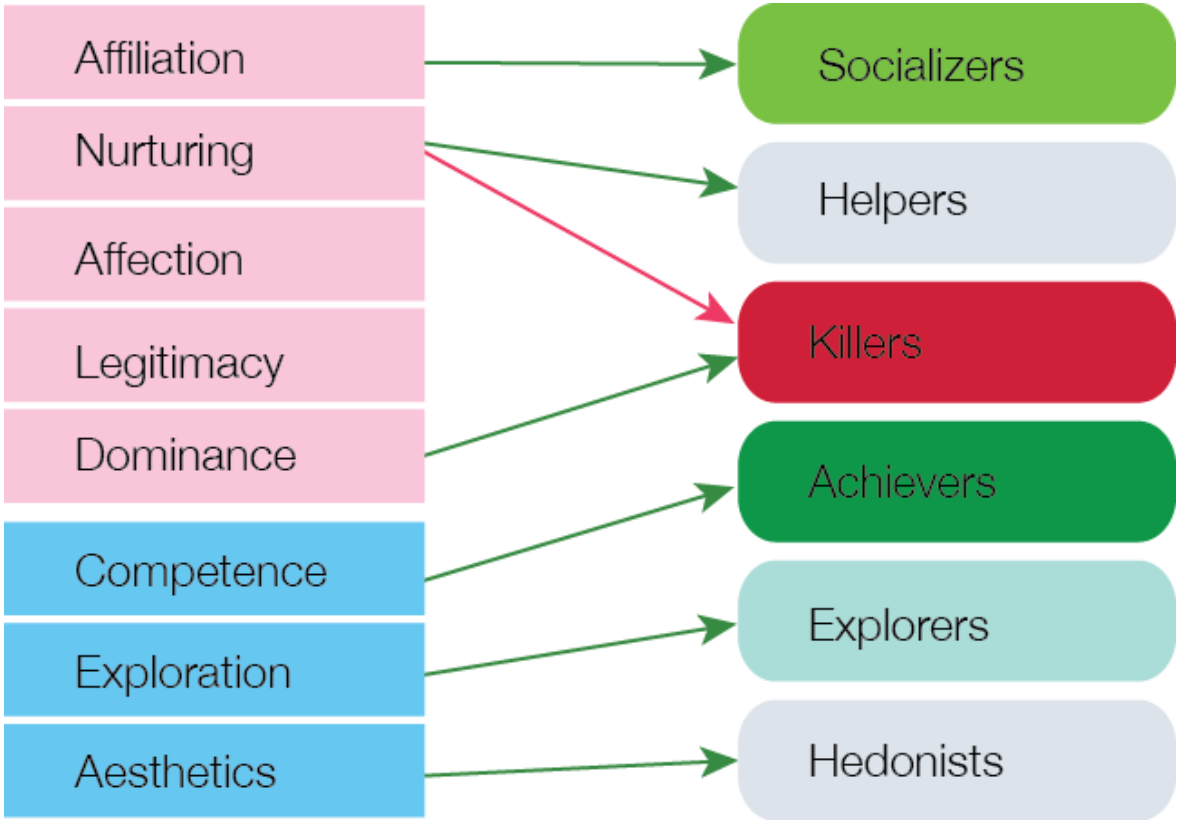
Affiliation

Uncertainty reduction



Competence

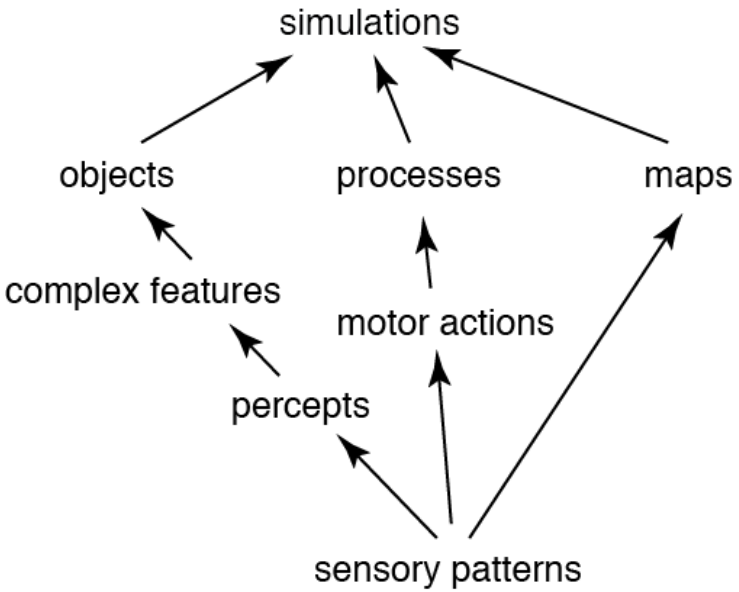
Needs and player types (with S. Tekovsky)



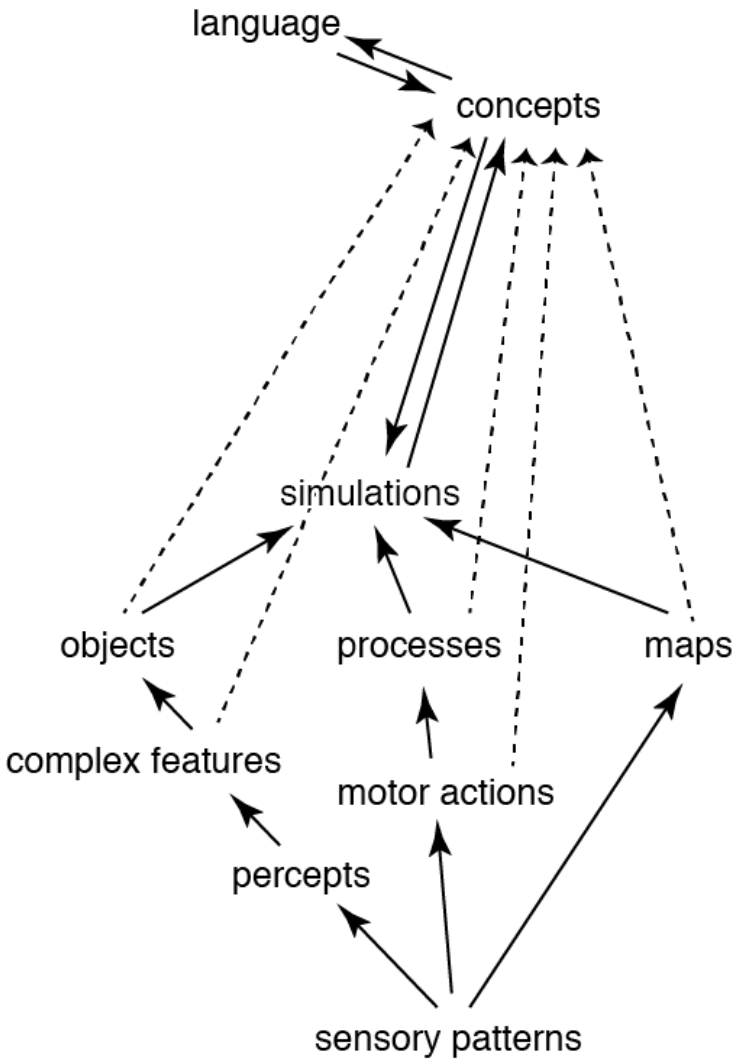
Emotion and the Self

Neocortex as a modeling system

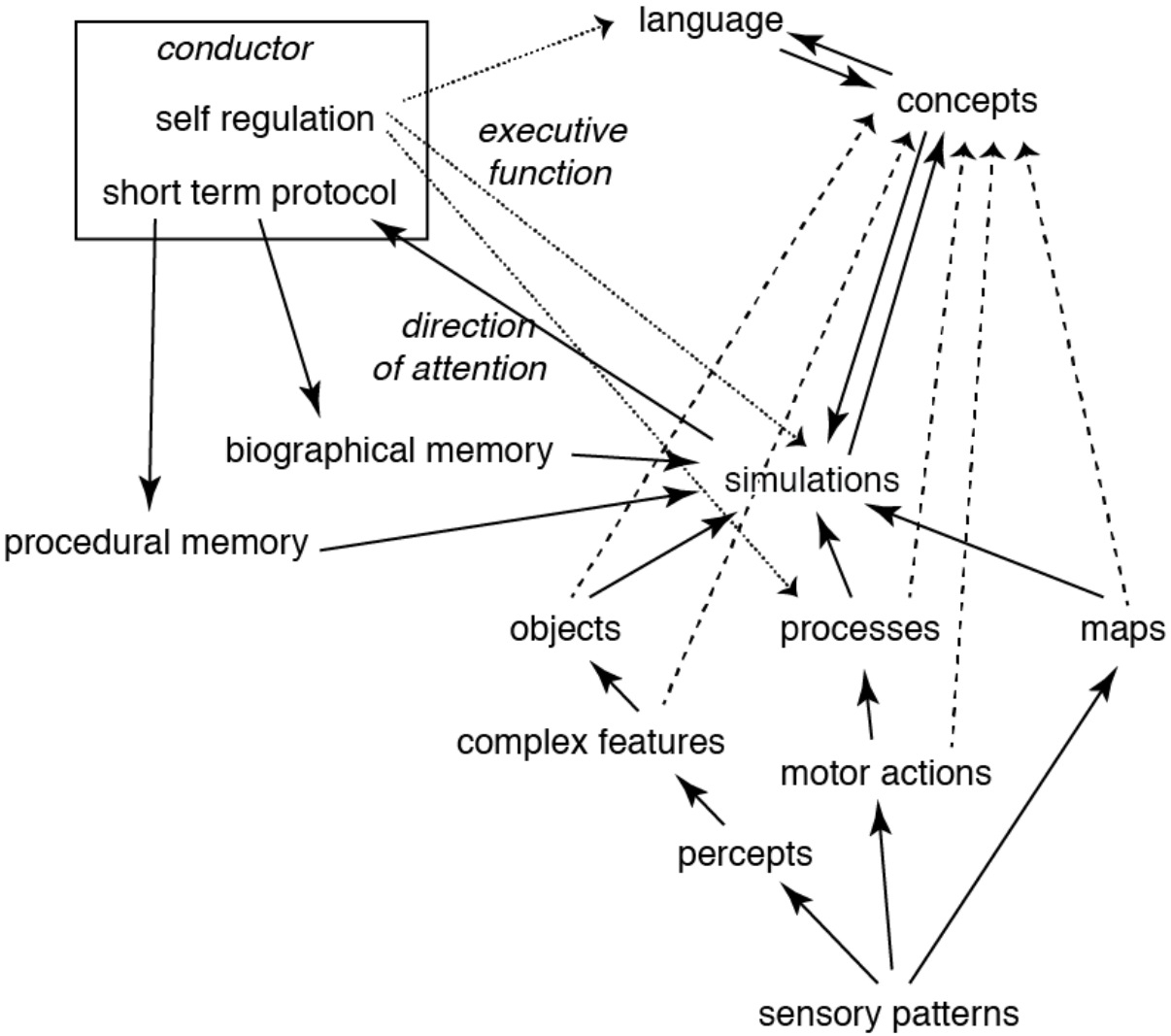
- cortical modeling is a predictive model
  of sensory patterns,
  as a dynamic simulation of world and agent



Neocortex as a modeling system



Neocortex as a modeling system



Neocortex as a modeling system

- Primary model: environment and agent
- Secondary model: interaction of agent with environment
- Tertiary model: functioning of secondary modeling

→Self

- Consciousness as a model of attention

Integration of emotion and motivation

- Control and self regulation depends on learned functional representations in cortical structure
- Attentional biases

- Attention is mechanism for directed learning
- Experiential access via attentional protocol

- Structure of self determines experience of emotion

Social emotions

- Object of emotion requires representation and motivational relevance
- Social urges: affiliation, romantic affect, libido, dominance are transactional
- Love: non-transactional emotion

- Love requires shared purpose, via legitimacy

Acknowledgements

Work on MicroPsi2 is collaborative effort:

- **Ronnie Vuine, Dominik Welland, Priska Herger, Jonas Kemper** are contributors to the current version
- Architecture/concepts have been inspired by Dietrich Dörner, Aaron Sloman, Marvin Minsky, Stan Franklin and many others
- Support from Humboldt University of Berlin, University of Osnabrück (Institute for Cognitive Science), Berlin School of Mind and Brain, Harvard Program of Evolutionary Dynamics, MIT Media Lab

Thank you!

Interesting questions:

- Is recursive function approximation plus motivational system sufficient for general intelligence?
- Could we functionally recreate human-like minds with our model?
- How does a motivated/emotional system evolve when it can modify itself?