

# Computational Concept and Architecture of Artificial Psyche

Anton Kolonin

[akolonin@aigents.com](mailto:akolonin@aigents.com)

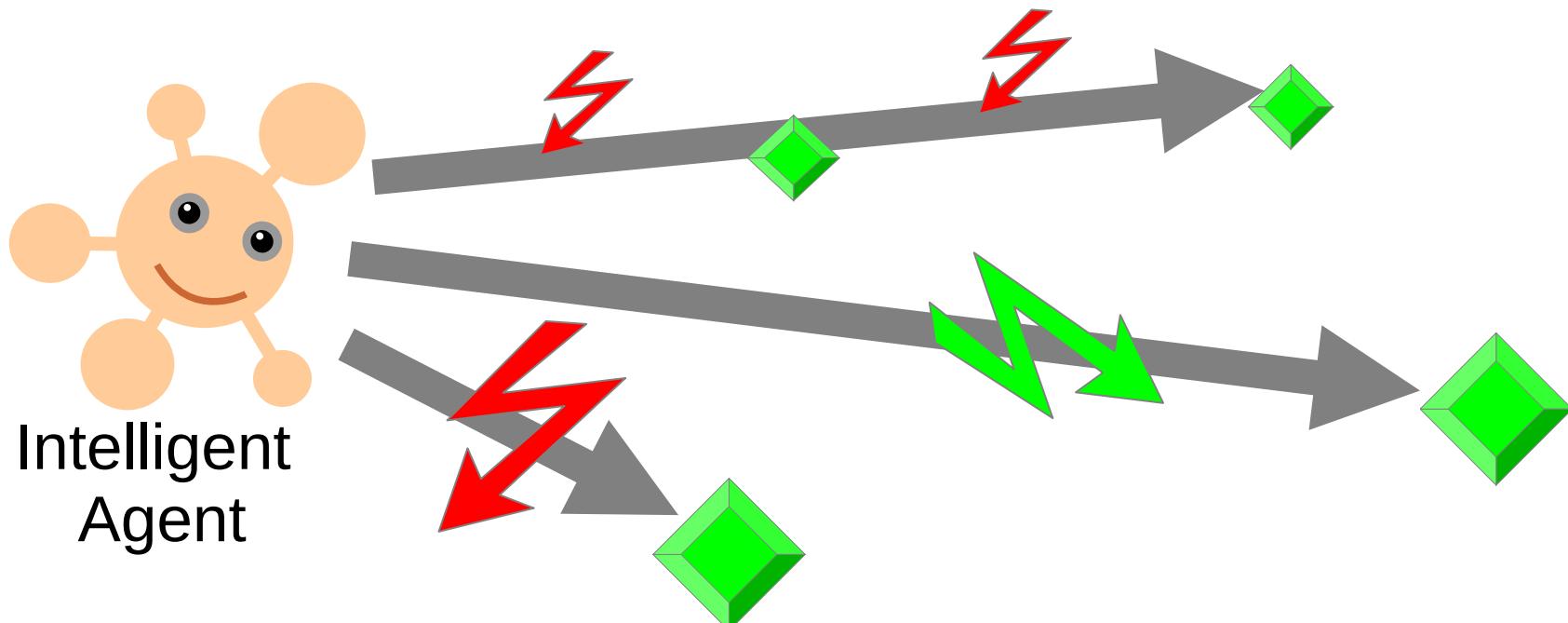
Telegram: akolonin



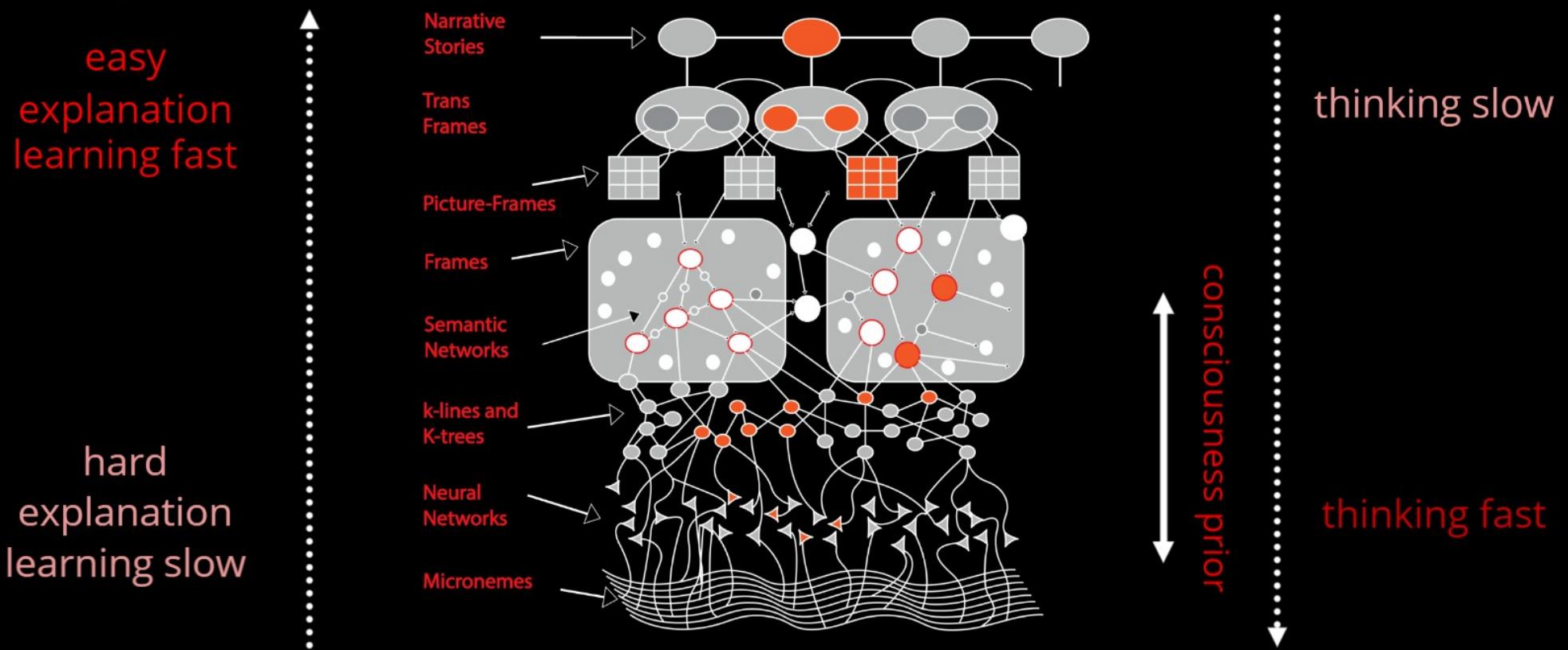
<https://agirussia.org>

# General Intelligence:

Reaching complex **goals** in different complex environments, using limited resources under uncertainty  
(Ben Goertzel + Pei Wang + Shane Legg + Marcus Hutter)

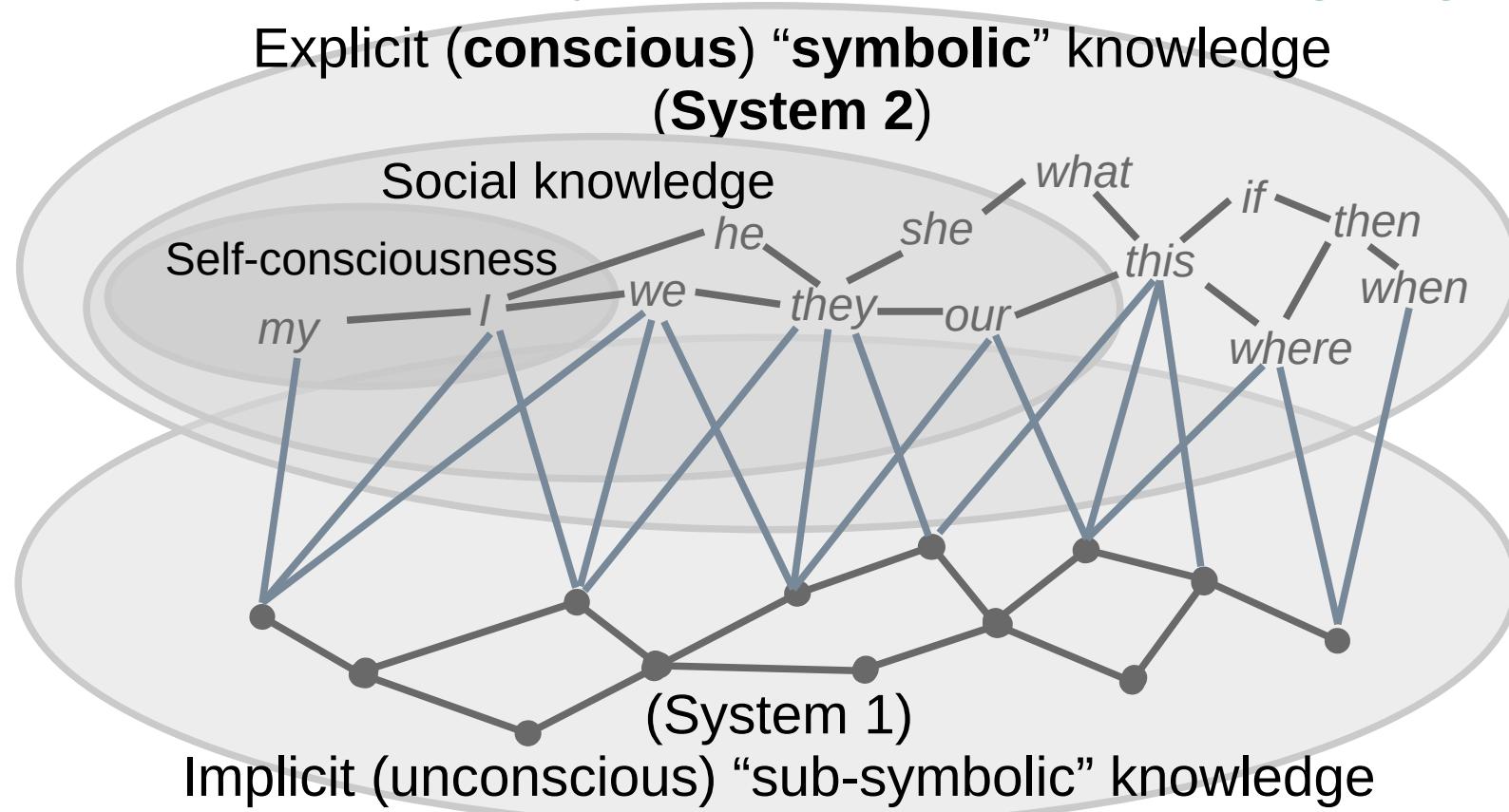


# “Fast and Slow Thinking” – Daniel Kahneman



<https://towardsdatascience.com/explainable-ai-vs-explaining-ai-part-1-d39ea5053347>

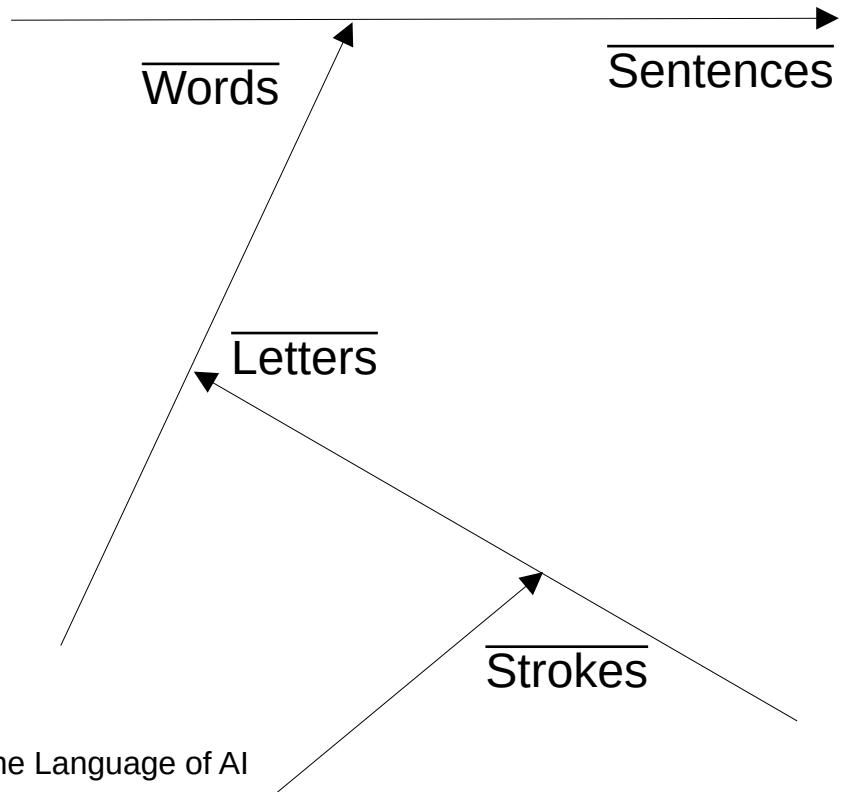
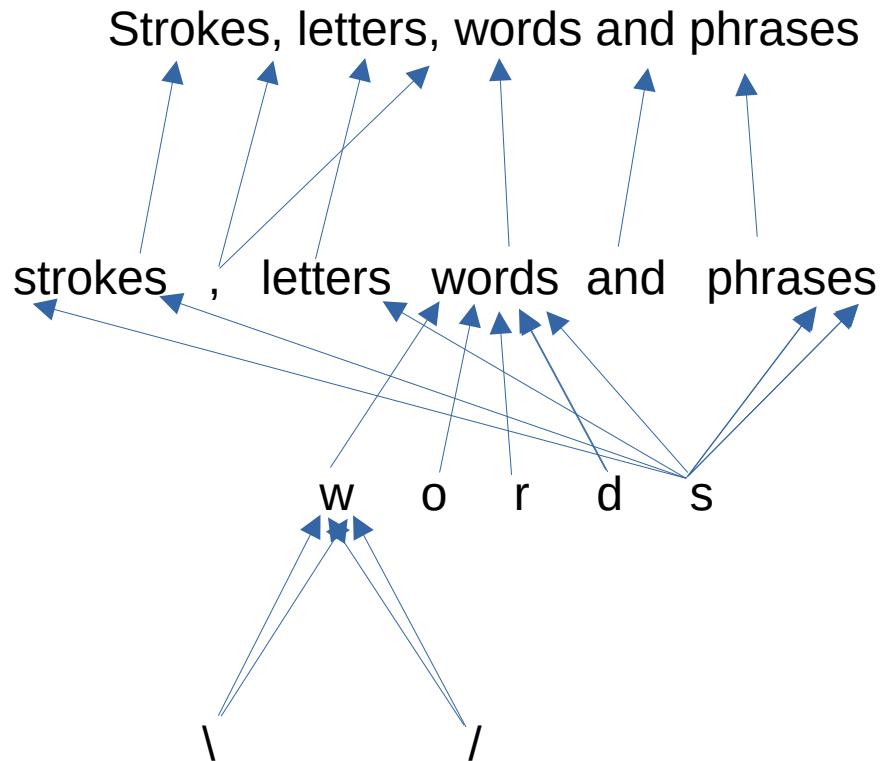
# Medium: “neuro-symbolic” “knowledge graph”



<https://www.amazon.com/Thinking-Fast-Slow-Daniel-Kahneman/dp/0374533555>

<https://amit02093.medium.com/atomspace-hyper-graph-information-retrieval-system-450cab9d751e>

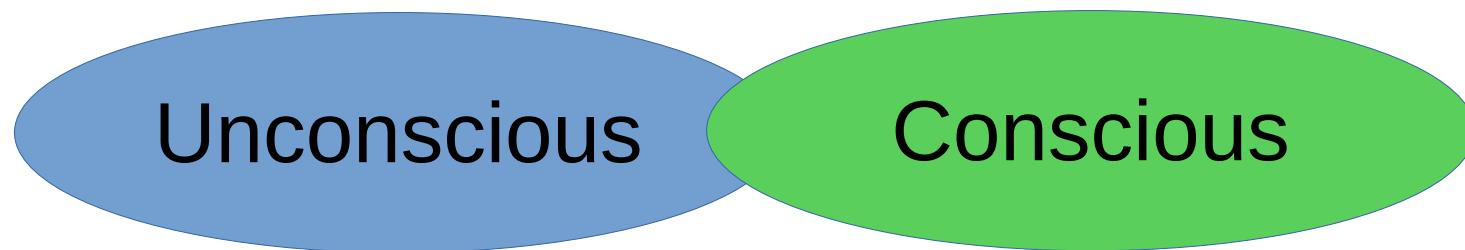
# Functional equivalence of graph and neural network tensor models



Pedro Domingos, Tensor Logic: The Language of AI  
<https://arxiv.org/pdf/2510.12269>

The psyche /'saɪki/ is currently used to describe the totality of the **human mind**, **conscious** and **unconscious**.

<https://en.wikipedia.org/wiki/Psyche>



# System = Intelligent agent

Psyche = Operating system

Intelligence = Decision making system

Subconsciousness  
(Intuition)  
“System 1”  
 (“Fast”)

Consciousness  
(Reasoning)  
“System 2”  
 (“Slow”)

Emotions → Motivation → Expectations

Space of states  
Sensations      Needs      Actions

Perceptions

Sensors

Decisions

Actuators

Sensations

Actions

Outer world = Operational environment

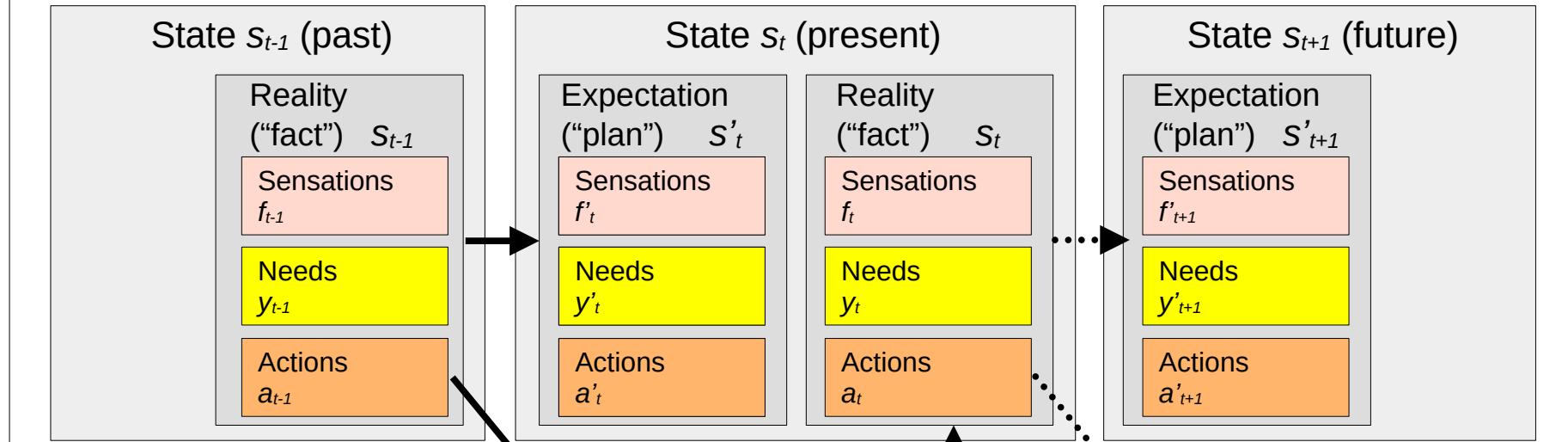
# Psyche = Operating system

Intelligence = Decision making system

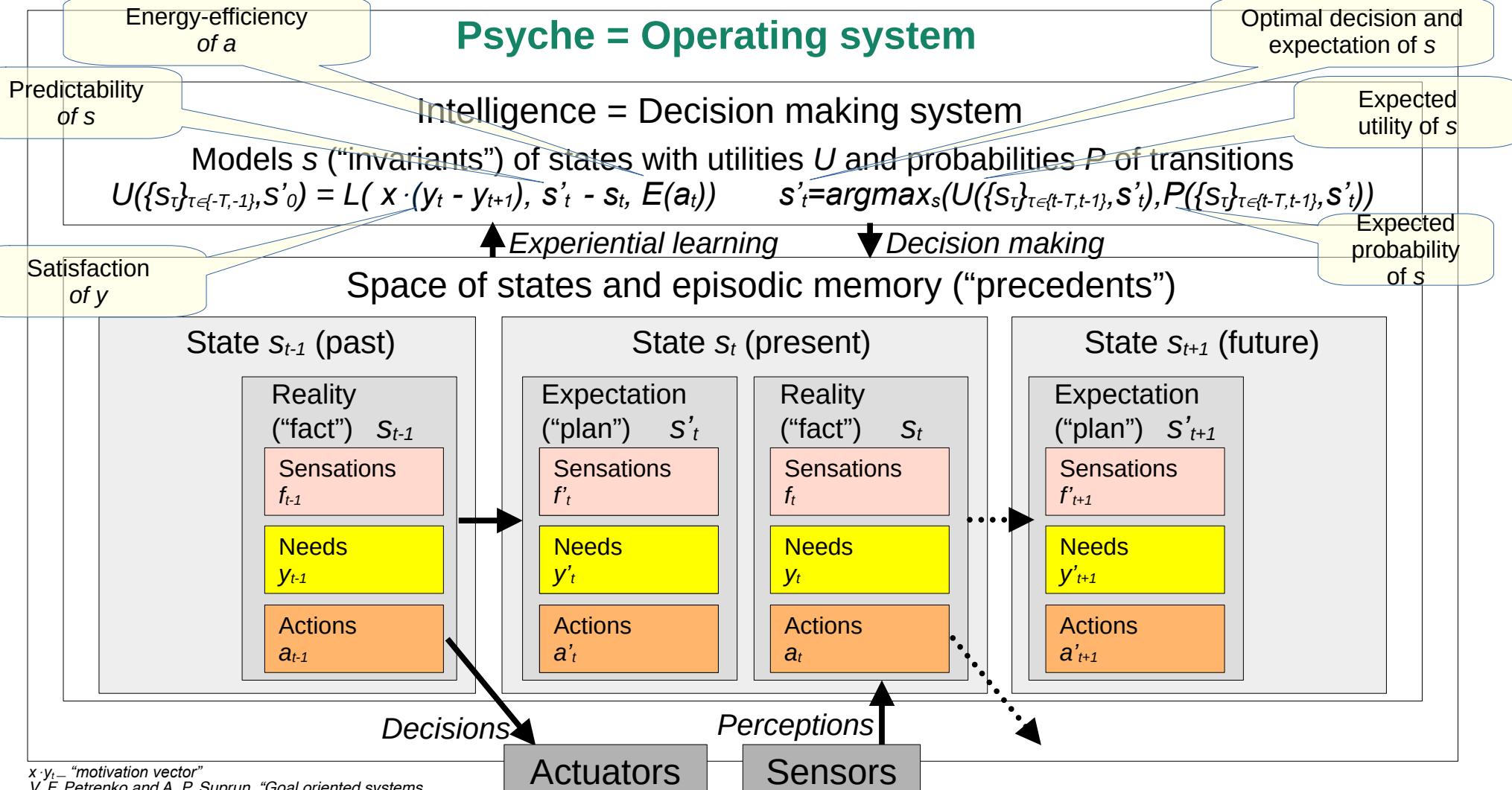
Models  $s$  ("invariants") of states with utilities  $U$  and probabilities  $P$  of transitions  
 $U(\{S_t\}_{t \in \{-T, -1\}}, S'_0) = L(x \cdot (y_t - y_{t+1}), s'_t - s_t, E(a_t))$        $s'_t = \text{argmax}_s(U(\{S_t\}_{t \in \{-T, t-1\}}, s'_t), P(\{S_t\}_{t \in \{-T, t-1\}}, s'_t))$

↑Experiential learning      ↓Decision making

Space of states and episodic memory ("precedents")



# Psyche = Operating system

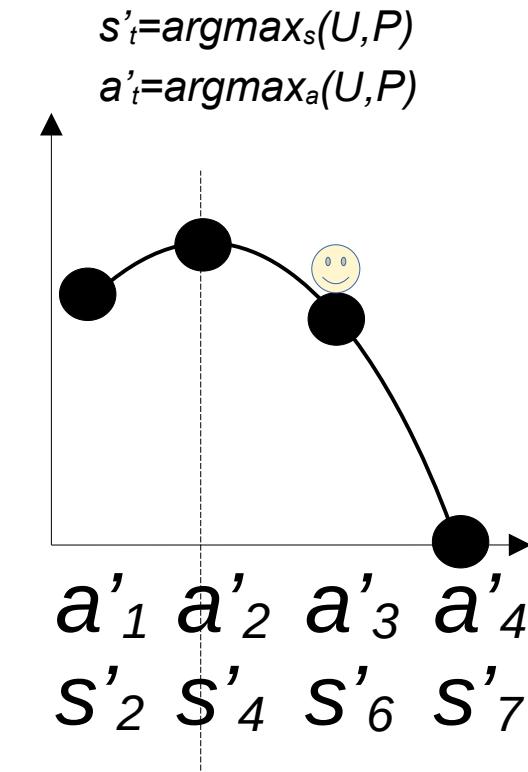


$x \cdot y_t$  – "motivation vector"

V. F. Petrenko and A. P. Suprun, "Goal oriented systems, evolution, and the subjective aspect in systemology," Tr. Inst. Sistem. Analiza RAN 62 (1) (2012)

# Decision making as operational risk management

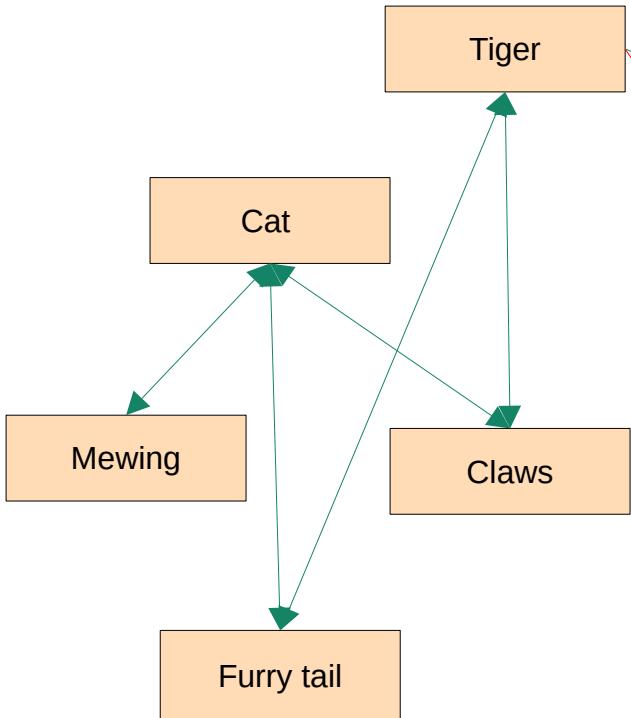
$s_t$	$s'_{t+1}$	$s'_{t+1}$			$U$	$P$	$\sum U^*P$
		$a'$	$y'$	$f'$			
$s_1$	$s'_2$	$a'_1$	$y'_1$	...	1.0	0.5	<u>0.7</u>
$s_1$	$s'_3$	$a'_1$	$y'_2$	...	0.4	0.5	
$s_1$	$s'_4$	$a'_2$	$y'_3$	...	1.0	0.8	<u>0.8</u>
$s_1$	$s'_5$	$a'_2$	$y'_4$	...	0.0	0.2	
$s_1$	$s'_6$	$a'_3$	$y'_5$	...	0.6	1.0	<u>0.6</u>
$s_1$	$s'_7$	$a'_4$	$y'_6$	...	0.0	1.0	<u>0.0</u>



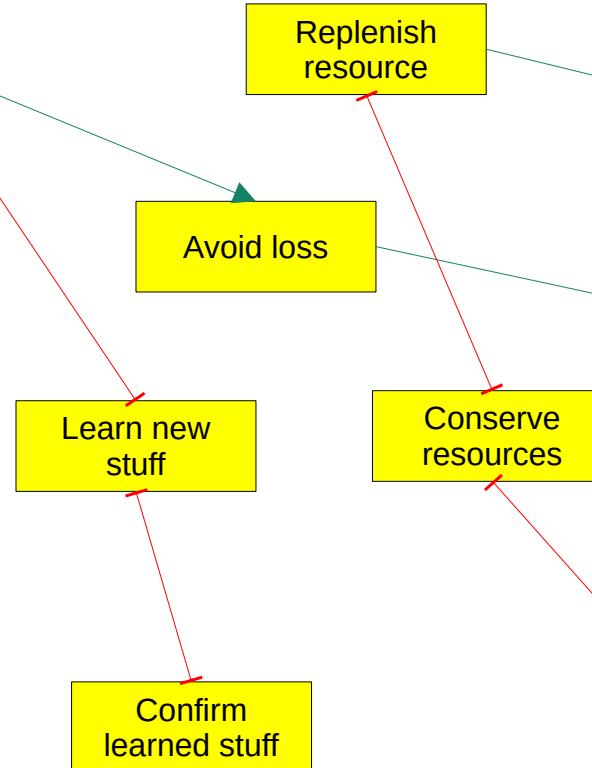
Tversky & Kahneman:  
most people choose  $a'_3$  и  $s'_6$   
("smaller profit with greater reliability")

# Mutual dependency of state variable subgraphs tensors

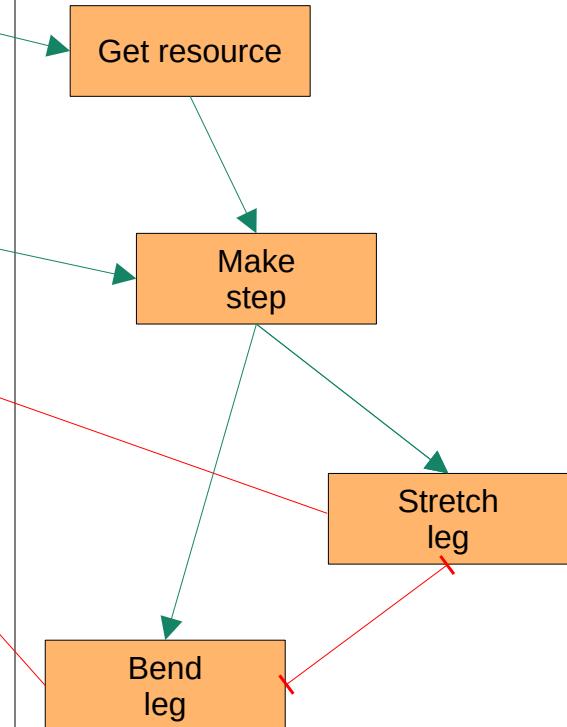
Percept feelings



Satisfied needs



Decided & committed actions



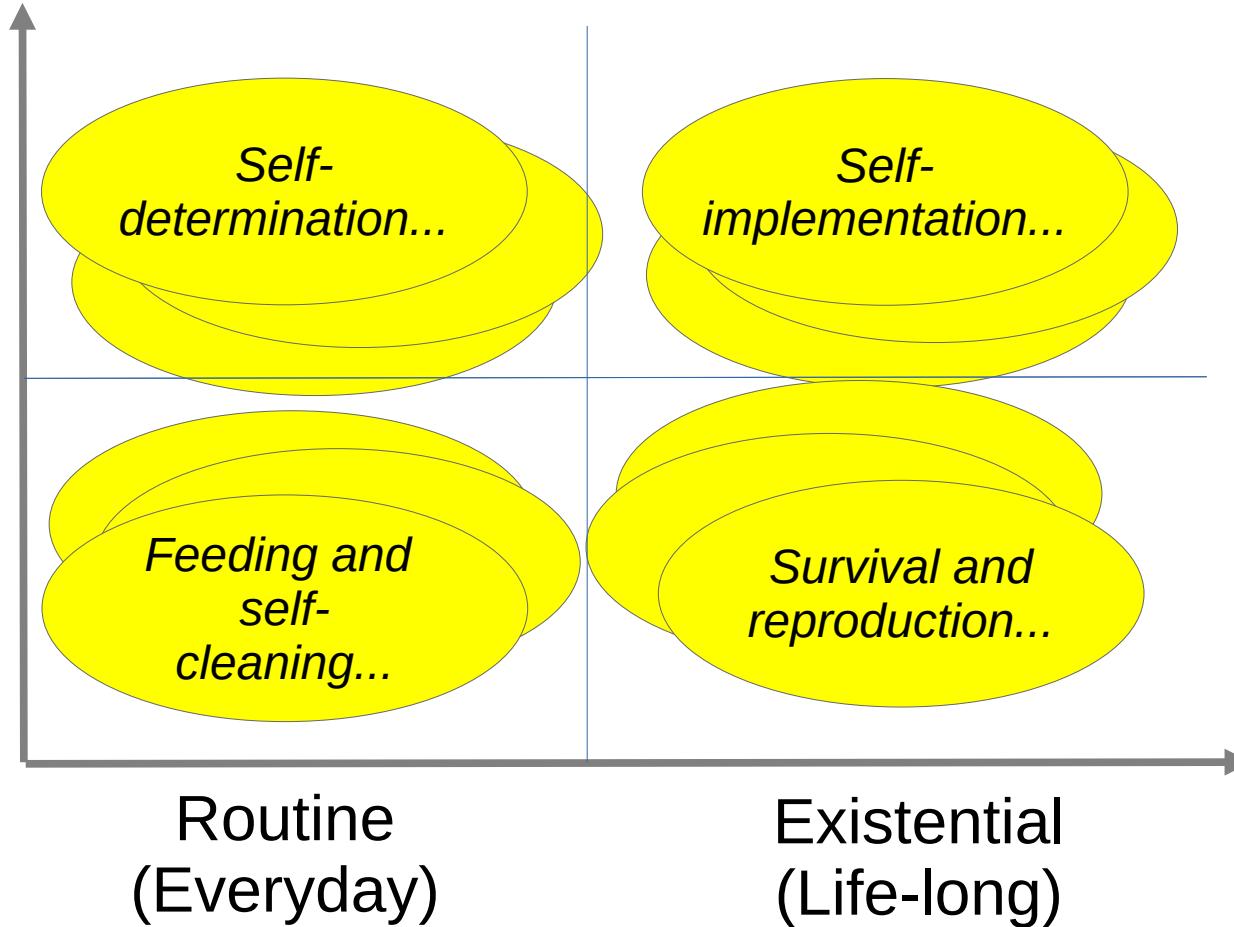
Anton Kolonin & Vladimir Kryukov,  
Computational Concept of the Psyche,  
Neuroinformatics-2025

<https://arxiv.org/pdf/2509.07009>

# Space of needs/goals/values

Psychological

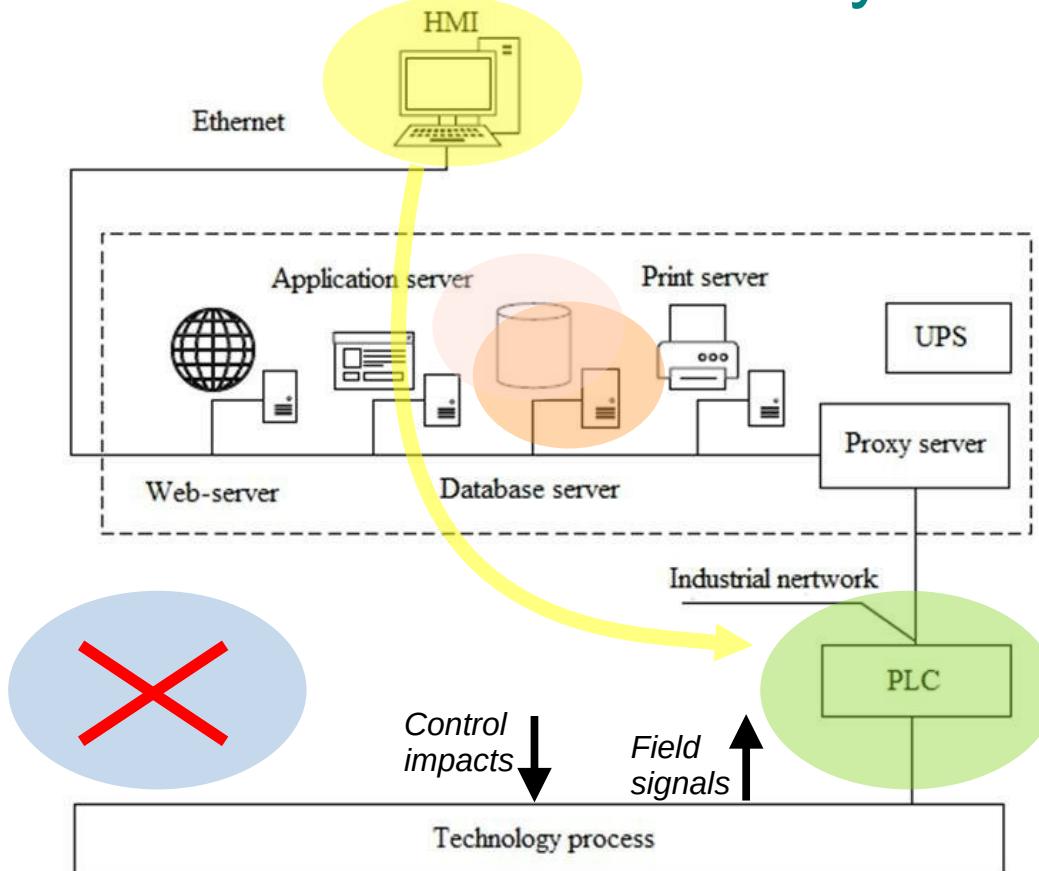
Physical  
(Physiological)



According to the ideas of A. Maslow, any goal can be represented as a combination (superposition) of basic needs

# Application cases

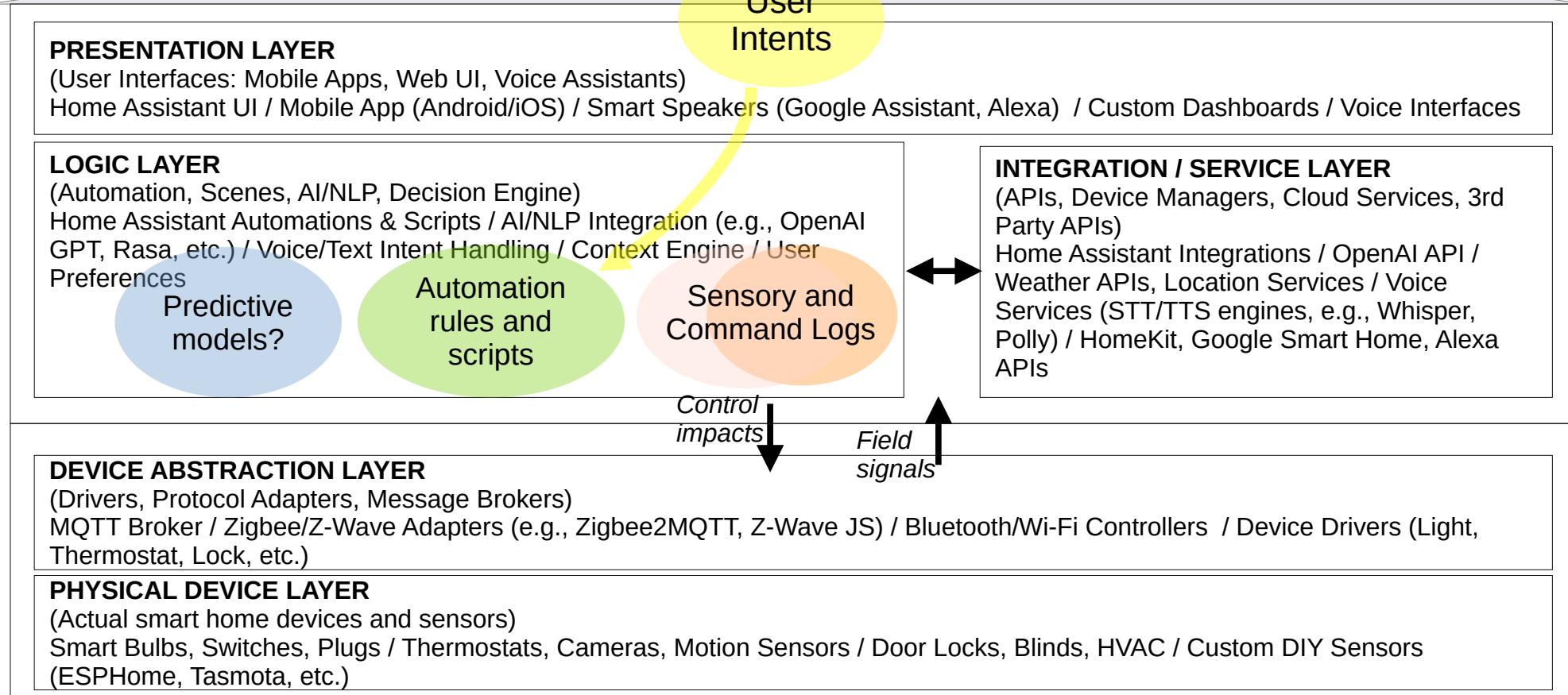
## A) Automated Process Control Systems (APCS)



Programmable logic  
**IEC 61131** (synchronous)  
**IEC 61499** (asynchronous)

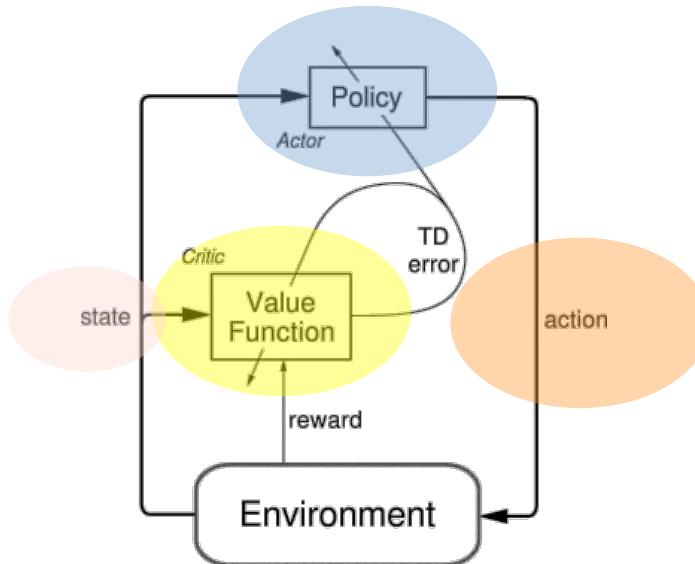
# Application cases

## B) “Smart Home”



# Implementation options

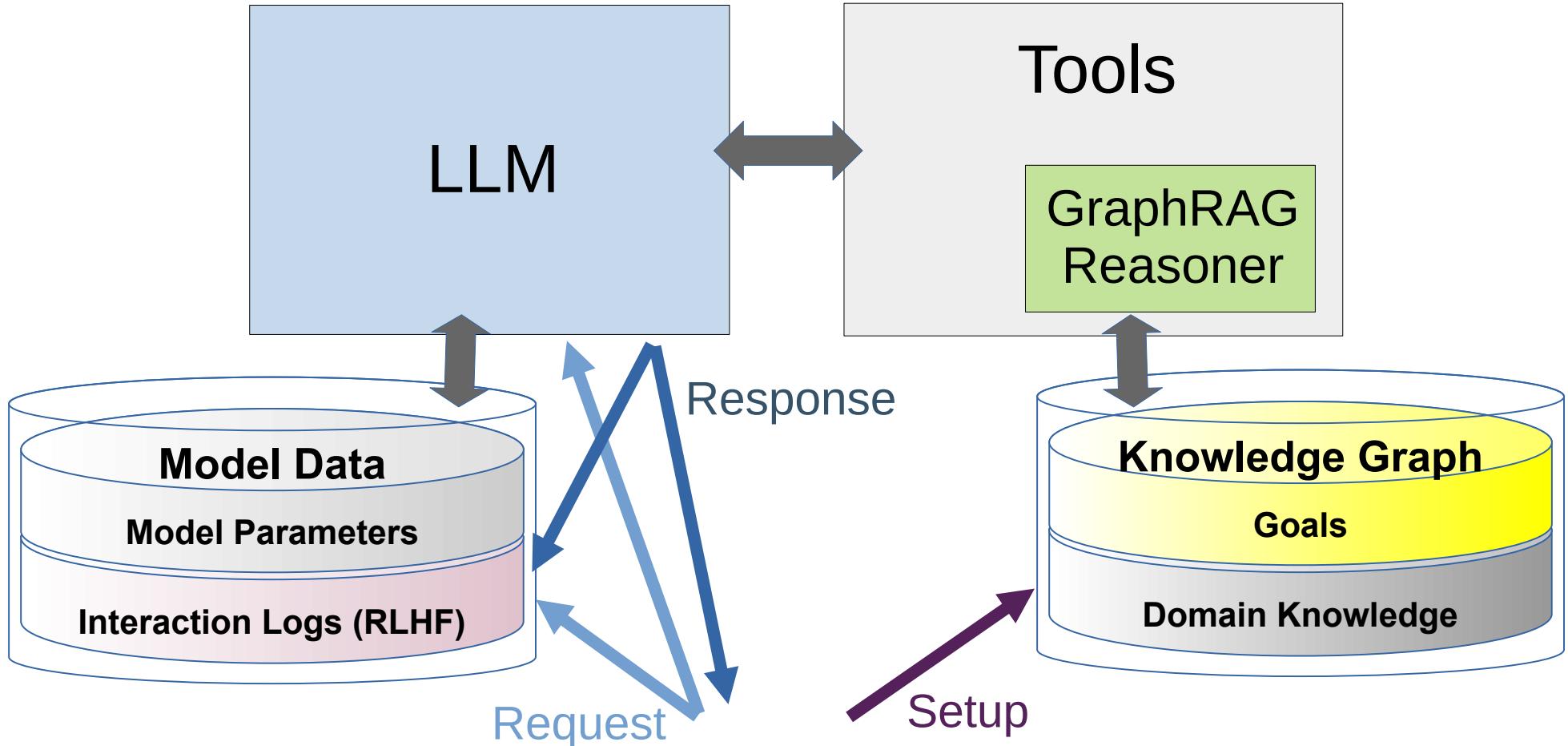
## 1) Reinforcement learning based on “actor-critic” model



<https://medium.com/intro-to-artificial-intelligence/the-actor-critic-reinforcement-learning-algorithm-c8095a655c14>

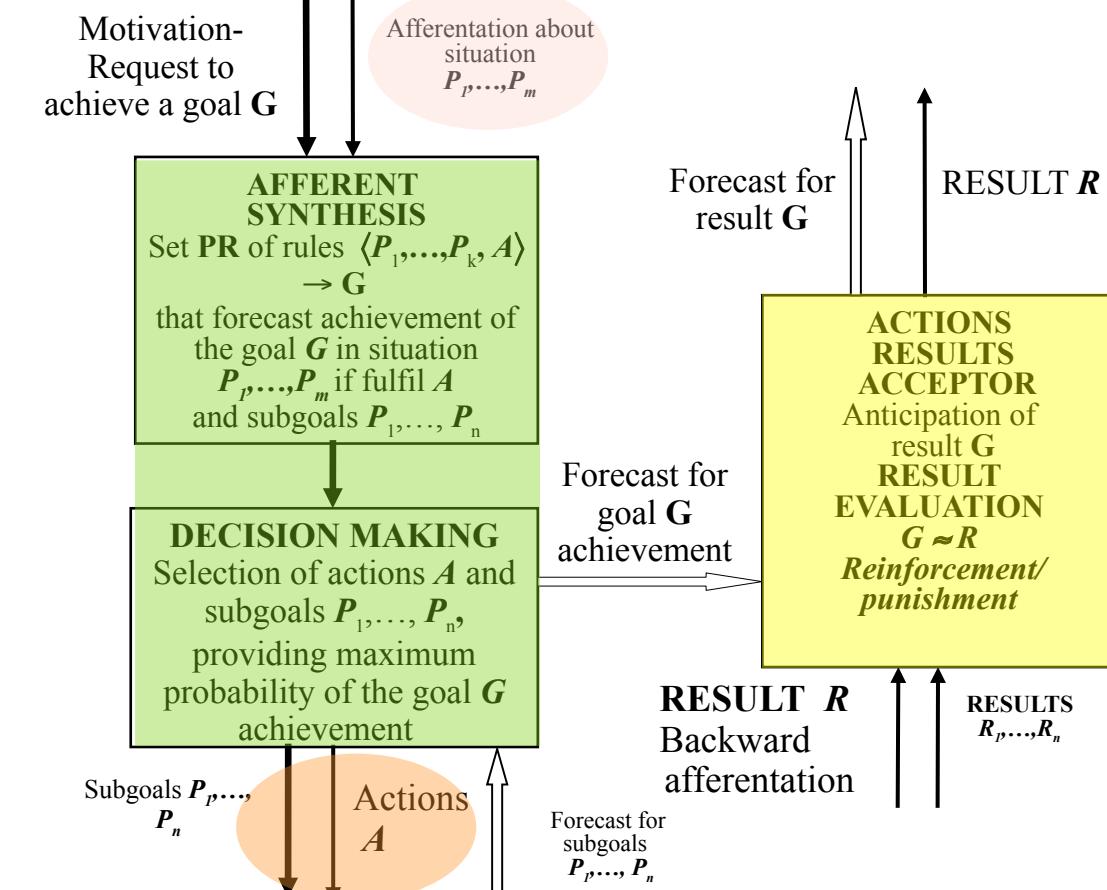
# Implementation options

## 2) Cognitive architecture based on LLM and (active) GraphRAG



# Implementation options

## 3) Cognitive architecture based on probabilistic logic



$$\text{Prob}(G | P_1, \dots, P_k, P_1, \dots, P_n, A) = \text{Prob}(\text{rule}) \cdot \text{Prob}(P_1) \cdot \dots \cdot \text{Prob}(P_n)$$

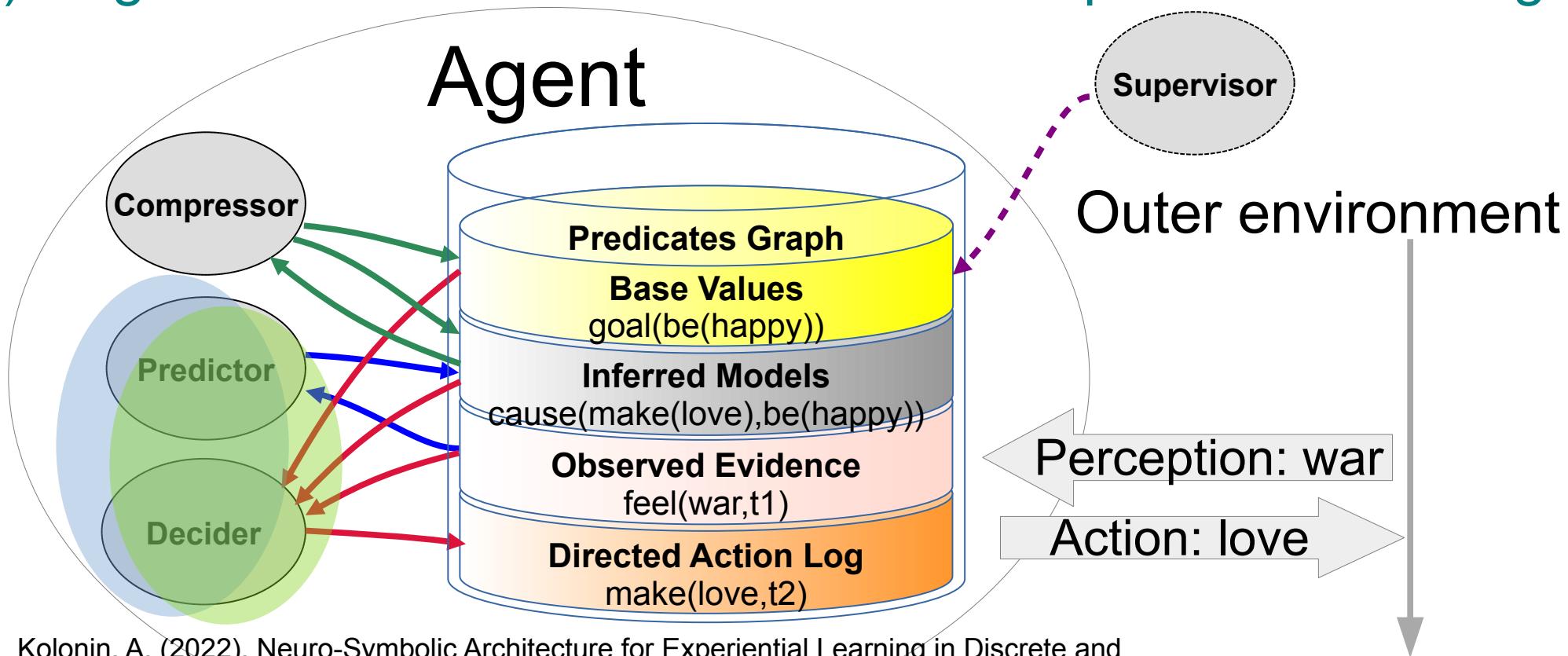
Evgenii Vityaev, Alexander Demin:  
Adaptive Control of Modular Robots // Conference Paper in Advances in Intelligent Systems and Computing, Conference: First International Early Research Career Enhancement School on Biologically Inspired Cognitive Architectures, Springer, August 2018

Evgenii E. Vityaev: Purposefulness as a Principle of Brain Activity // Anticipation: Learning from the Past, (ed.) M. Nadin. Cognitive Systems Monographs, V.25, Chapter No.: 13. Springer, 2015, pp. 231-254.

Витяев Е.Е. Логика работы мозга. Подходы к моделированию мышления. (сборник под ред. д.ф.-м.н. В.Г. Редько). УРСС Эдиториал, Москва, 2014г., стр. 120-153.

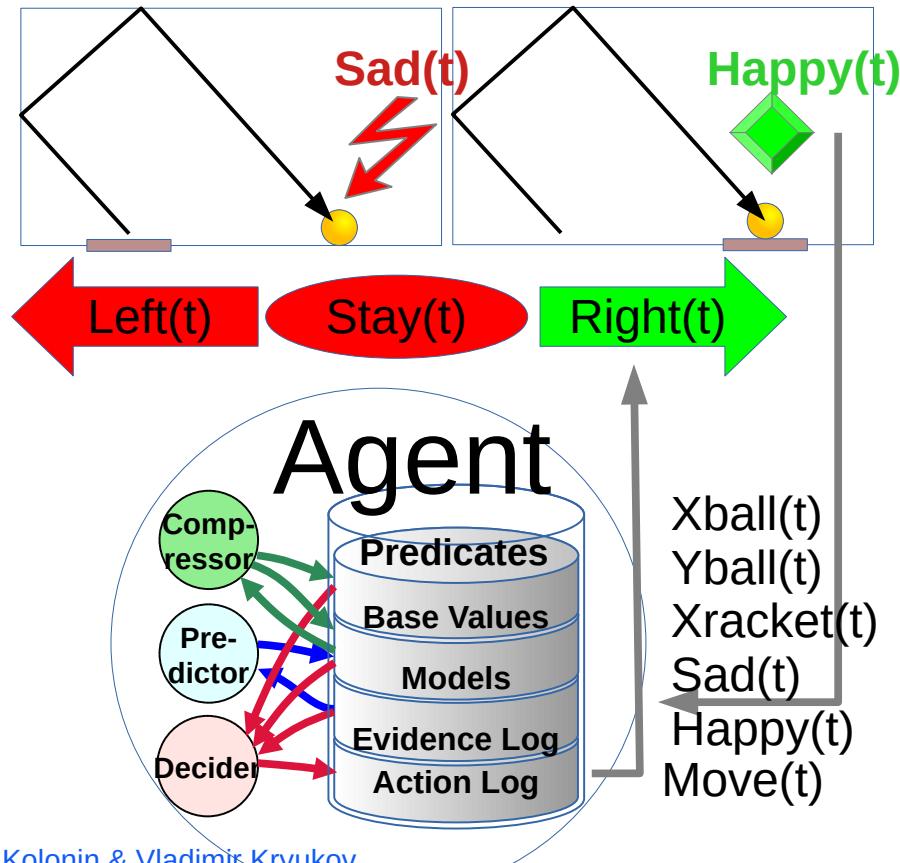
# Implementation options

## 4) Cognitive architecture of value-based experiential learning

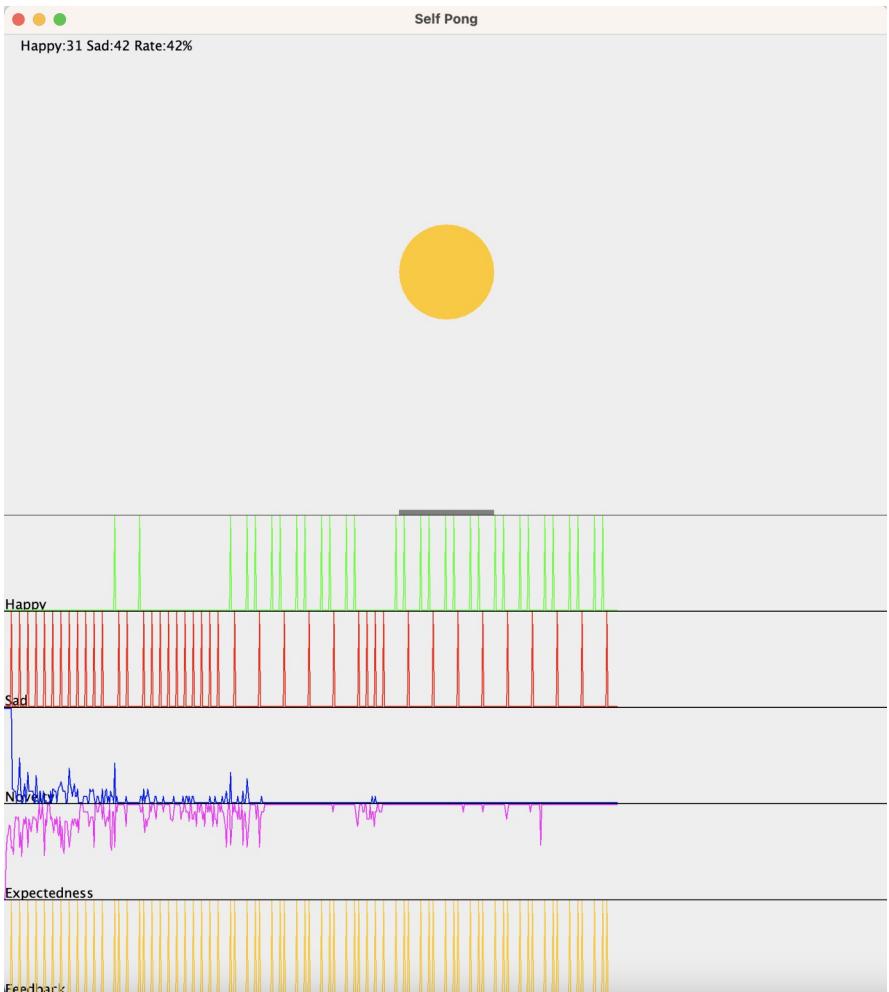


Kolonin, A. (2022). Neuro-Symbolic Architecture for Experiential Learning in Discrete and Functional Environments. In: Goertzel, B., Iklé, M., Potapov, A. (eds) Artificial General Intelligence. AGI 2021. Lecture Notes in Computer Science(), vol 13154. Springer, Cham.  
[https://doi.org/10.1007/978-3-030-93758-4\\_12](https://doi.org/10.1007/978-3-030-93758-4_12)

# Cognitive architecture of value-based experiential learning



Anton Kolonin & Vladimir Kryukov,  
Computational Concept of the Psyche,  
Neuroinformatics-2025  
<https://arxiv.org/pdf/2509.07009>



# Thank you for attention! Questions?

Anton Kolonin

[akolonin@aigents.com](mailto:akolonin@aigents.com)

Telegram/GitHub: akolonin

Workshop recording  
on the subject



Anton Kolonin & Vladimir Kryukov,  
Computational Concept of the  
Psyche, Neuroinformatics-2025

