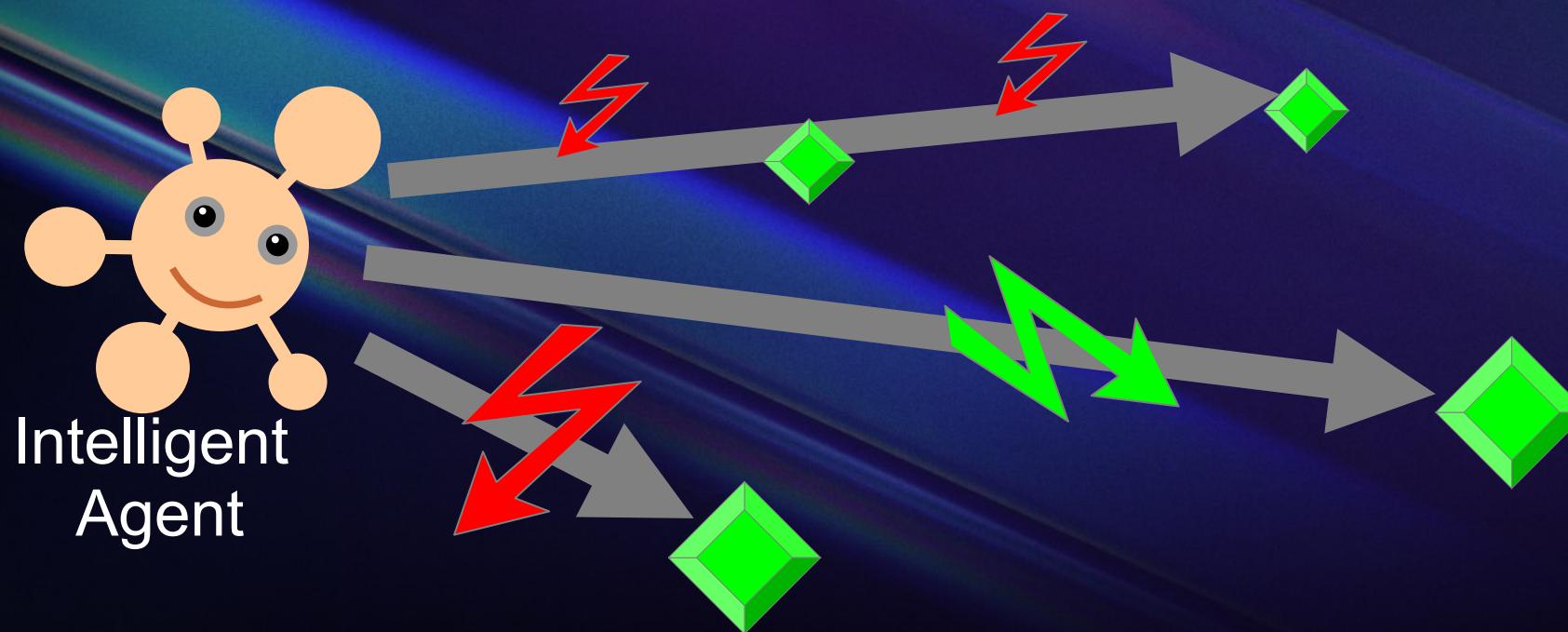


Computational Concept and Cognitive Architecture of Artificial Psyche

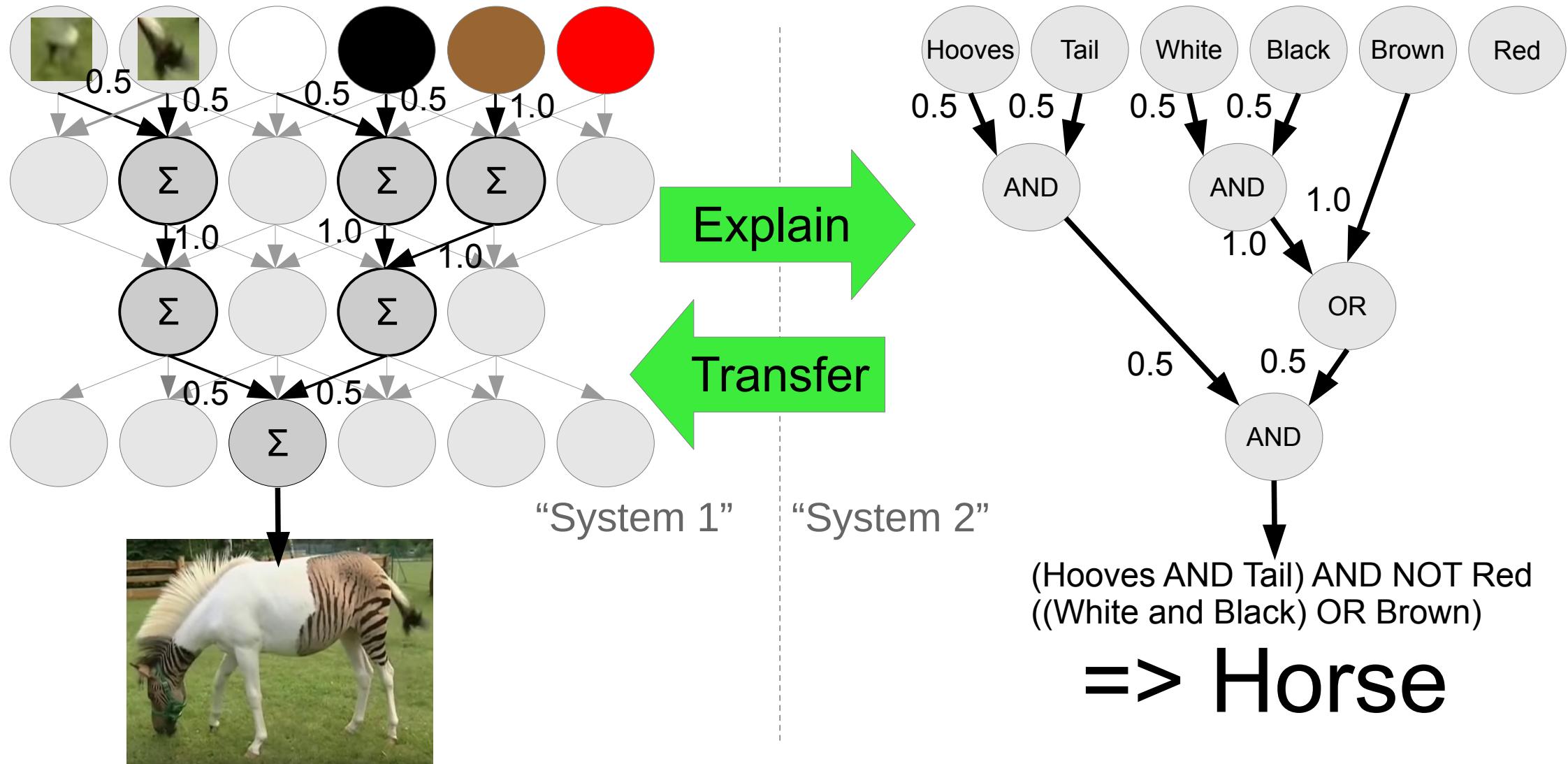
Anton Kolonin, Ph.D.
Novosibirsk State University <https://nsu.ru>
Russian AGI Community <https://agirussia.org>
Aigents <https://aigents.com>

General Intelligence:

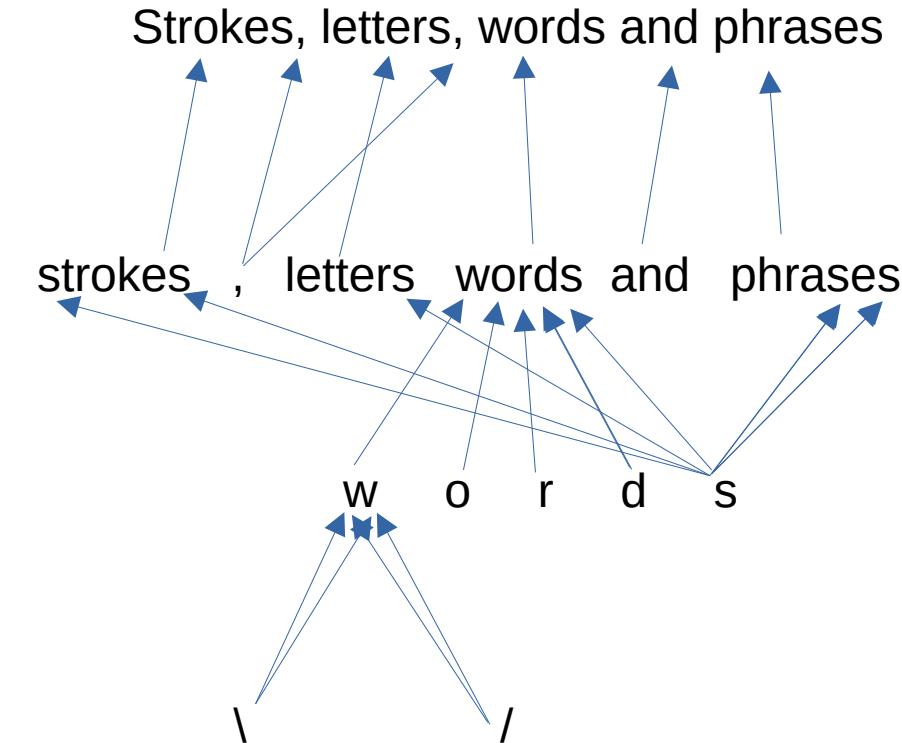
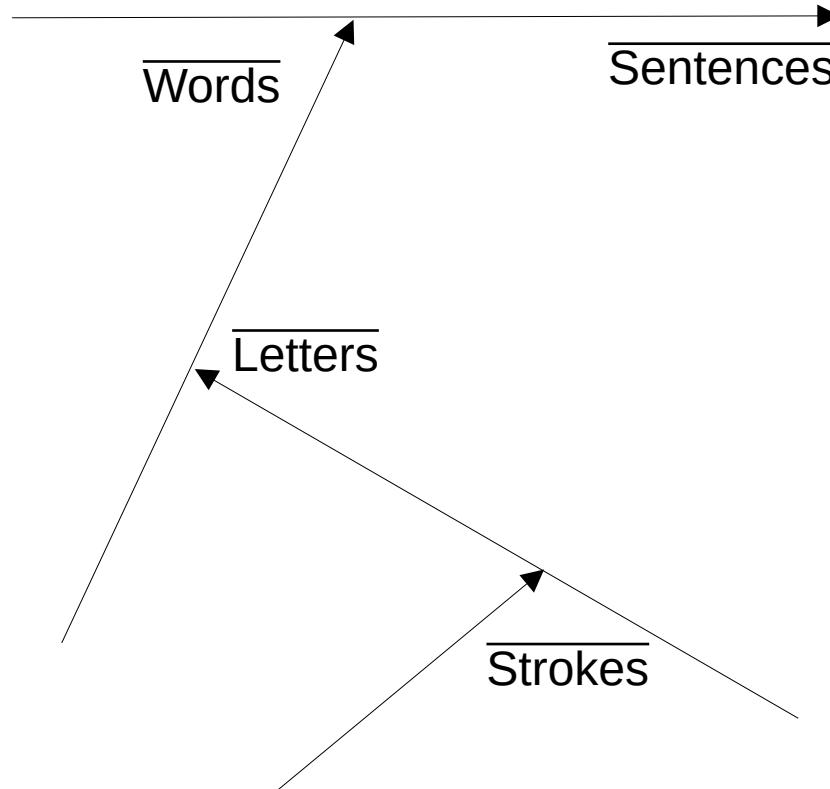
Reaching complex goals **in different** complex environments, **using limited resources** under uncertainty
(Ben Goertzel + Pei Wang + Shane Legg + Marcus Hutter)



Neuro-Symbolic Integration for Interpretable AI



Functional equivalence of tensor and graph (symbolic) models



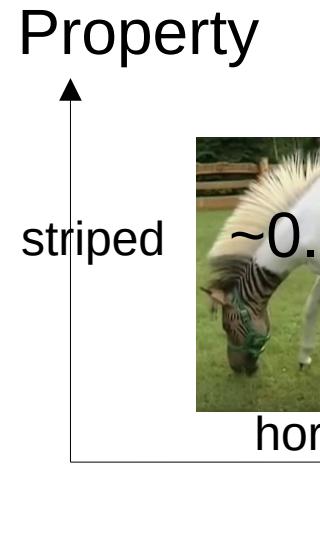
Functional equivalence of tensor and (symbolic) graph models

Truth-Value Tensor
(NARS/PLN)

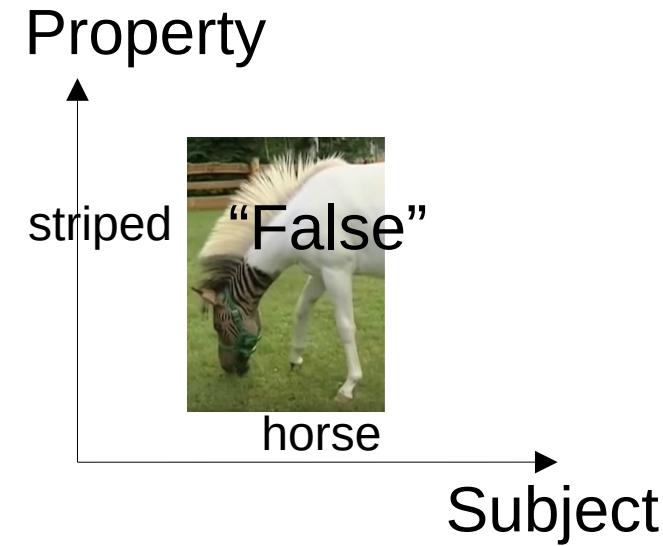


Life-long
learning?

Numerical Tensor
(ANN/Bayesian Logic)

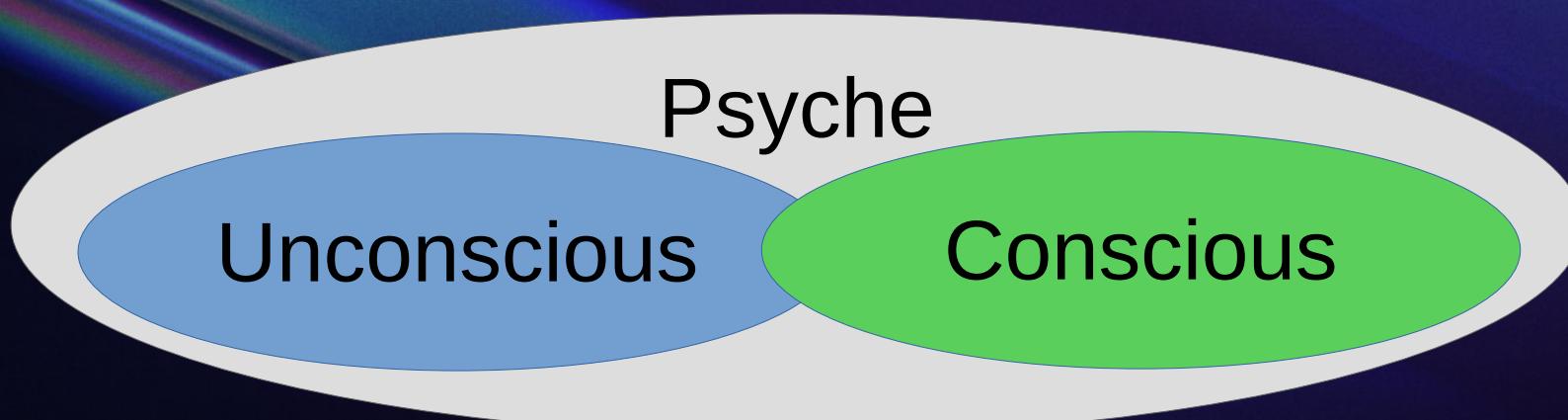


Boolean Tensor
(Boolean Logic)

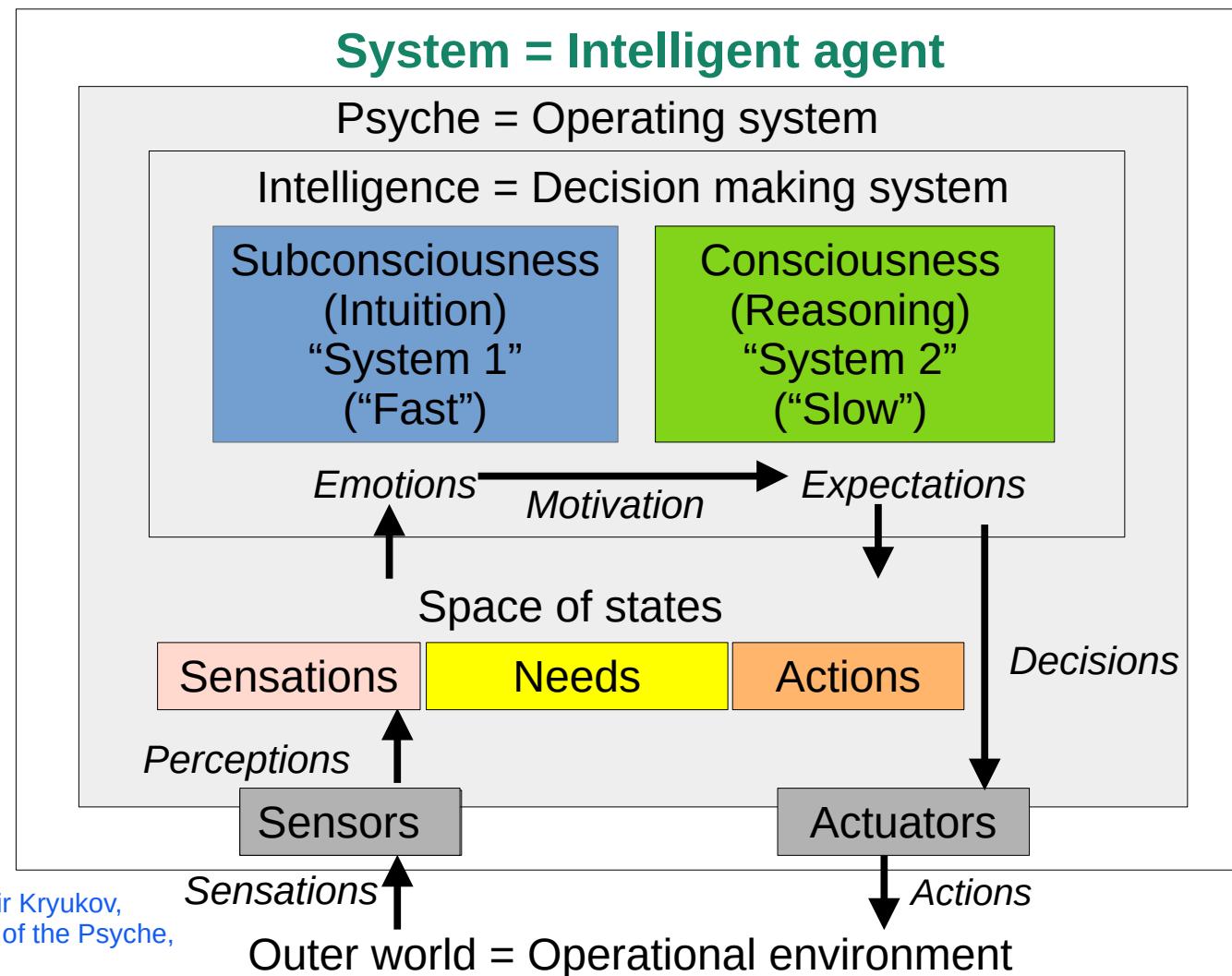


The psyche /'saɪki/ is currently used to describe the totality of the **human mind**, **conscious** and **unconscious**.

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



Computational Concept and Cognitive Architecture of Artificial Psyche



Computational Concept and Cognitive Architecture of Artificial Psyche

Psyche = Operating system

Intelligence = Decision making system

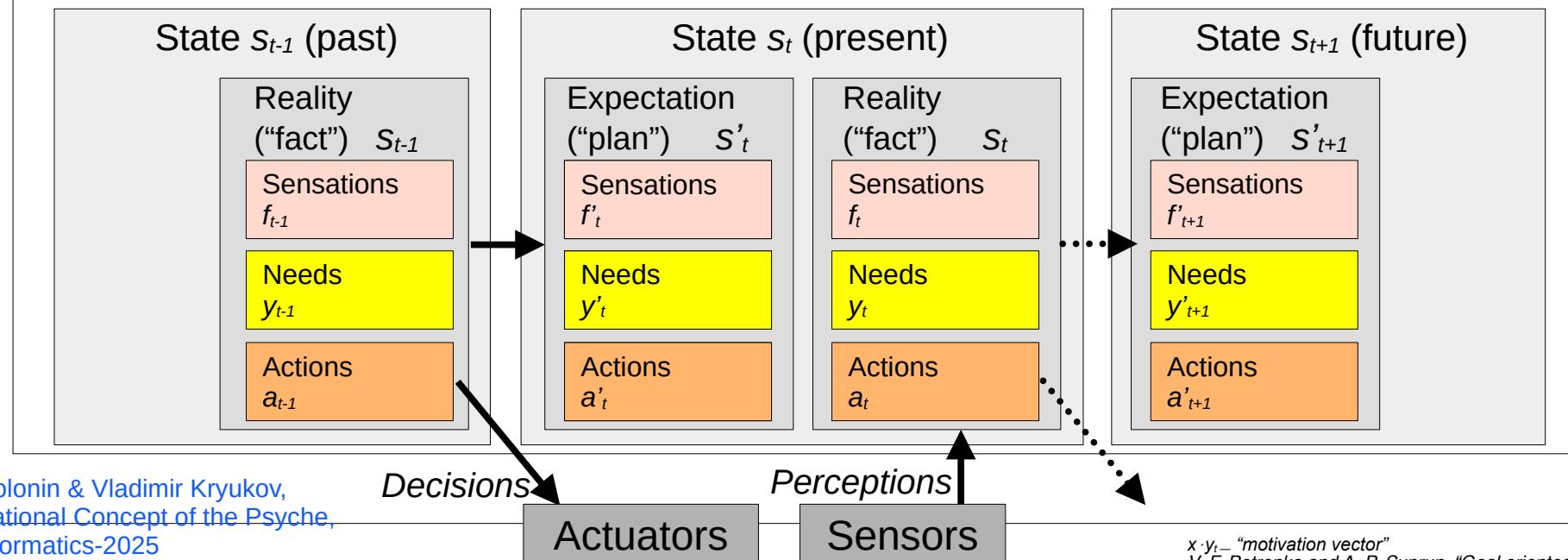
Models s ("invariants") of states with utilities U and probabilities P of transitions

$$U(\{S_T\}_{T \in \{-T, -1\}}, S'_0) = L(x \cdot (y_t - y_{t-1}), (s'_t - s_t), E(a_{t-1})) \quad s'_t = \text{argmax}_s(U(\{S_T\}_{T \in \{t-T, t-1\}}, S'_t), P(\{S_T\}_{T \in \{t-T, t-1\}}, S'_t))$$

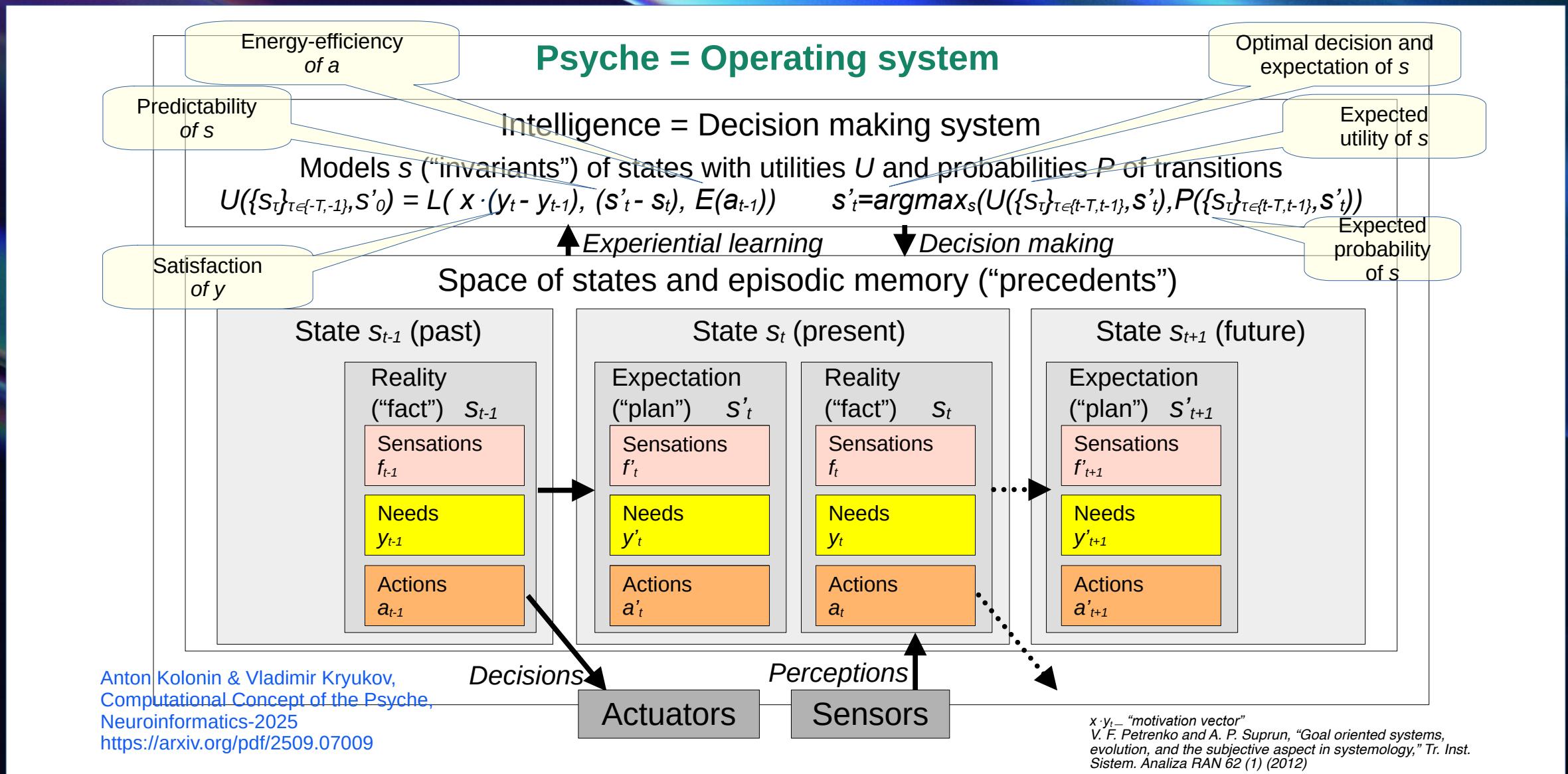
↑Experiential learning

↓Decision making

Space of states and episodic memory ("precedents")



Computational Concept and Cognitive Architecture of Artificial Psyche

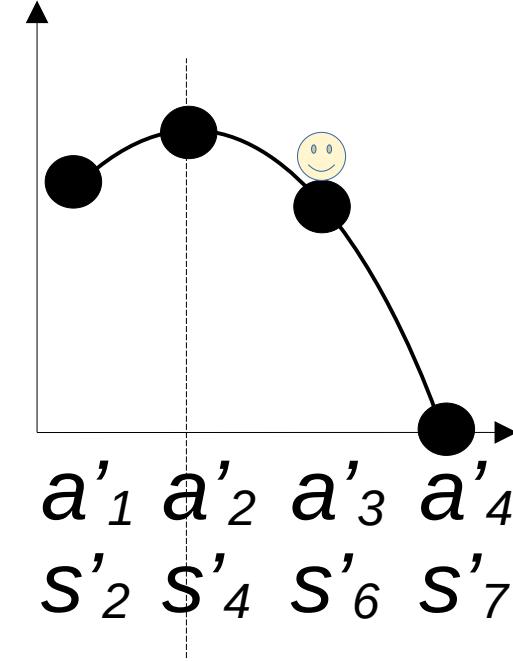


Decision making as operational risk management

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

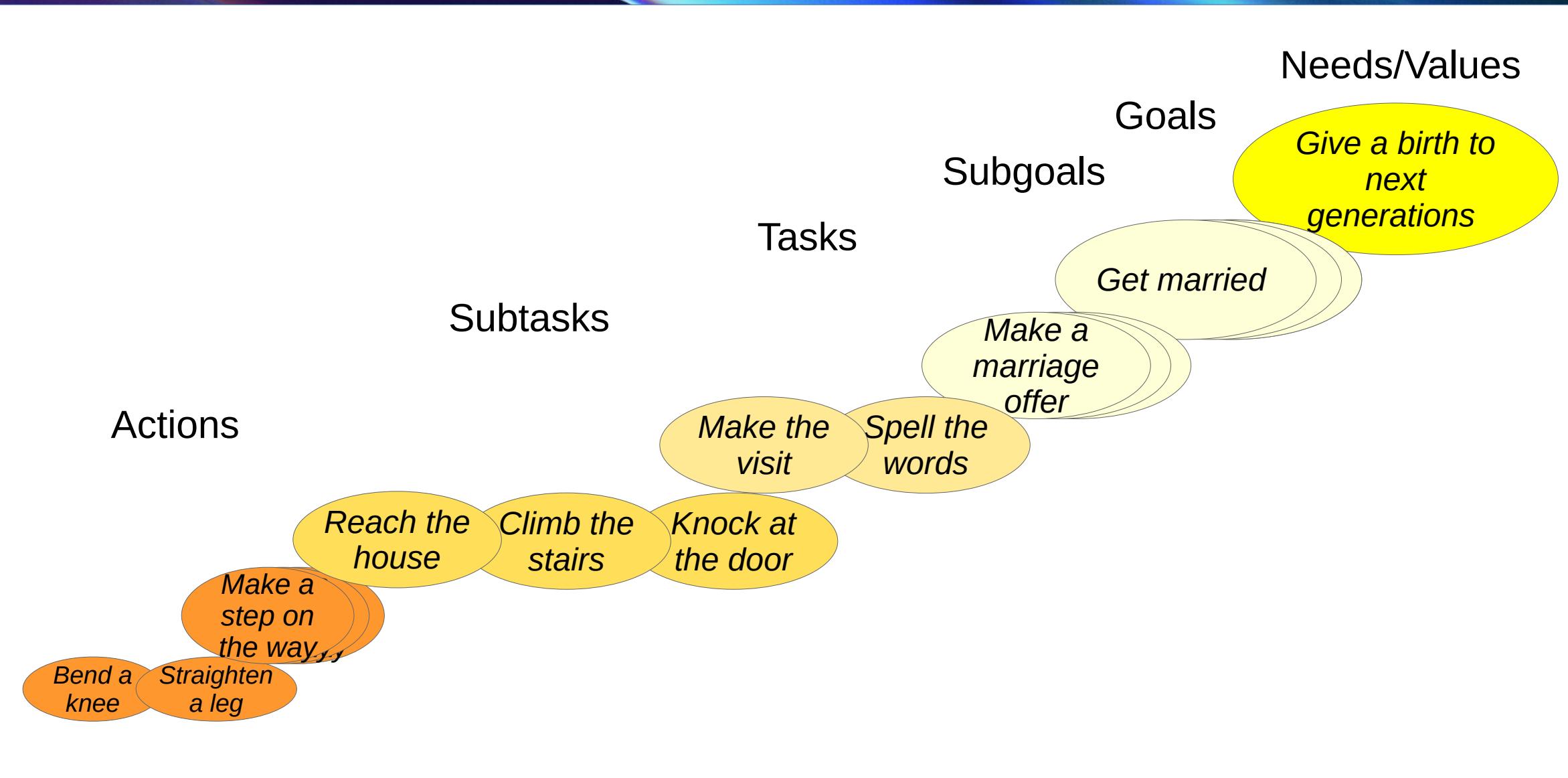
$$s'_t = \text{argmax}_s(U, P)$$

$$a'_t = \text{argmax}_a(U, P)$$



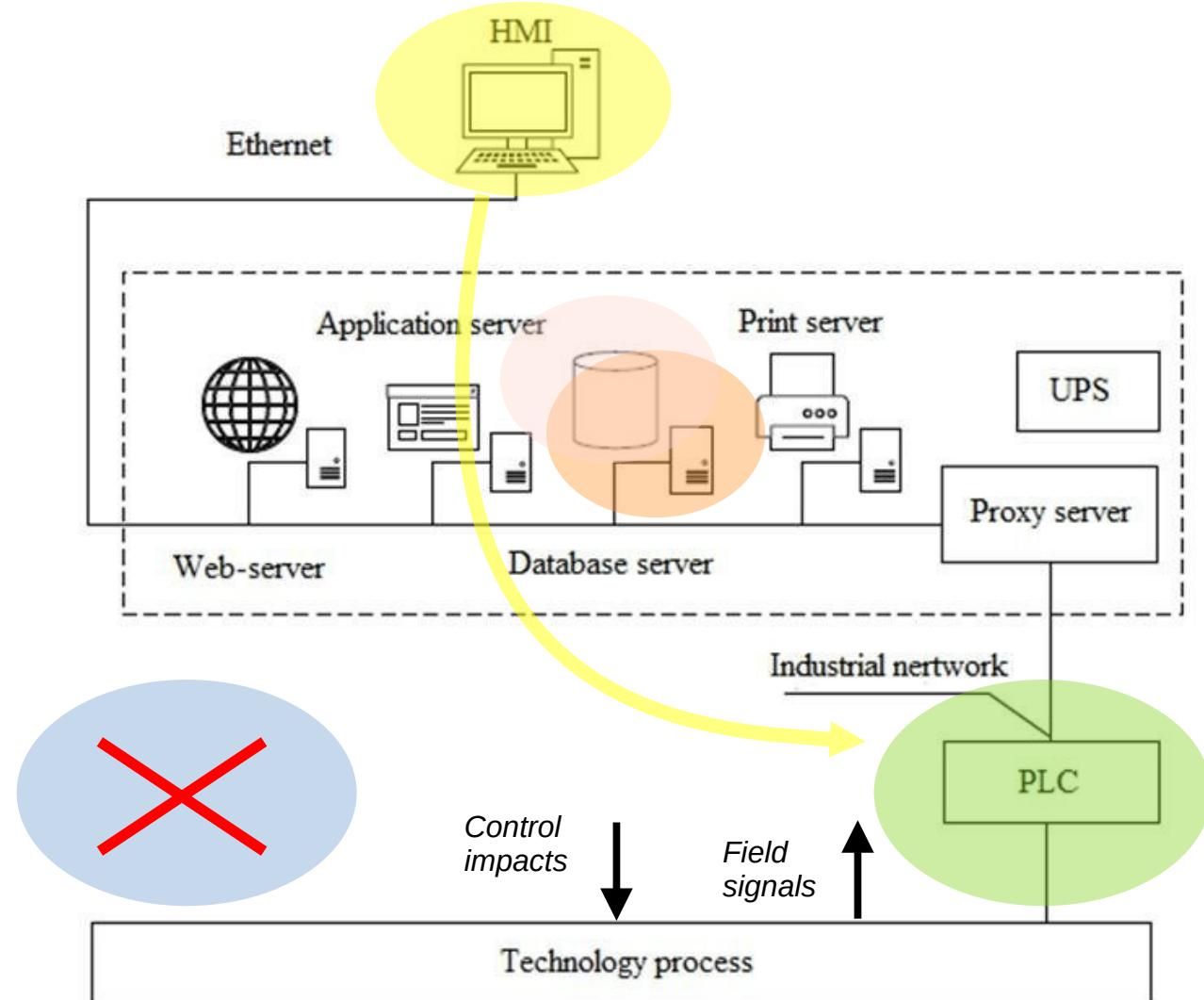
Tversky & Kahneman:
most people choose a'_3 и s'_6
(“smaller profit with
greater reliability”)

Hierarchy of needs/values/goals/subgoals/tasks/subtasks



Application cases

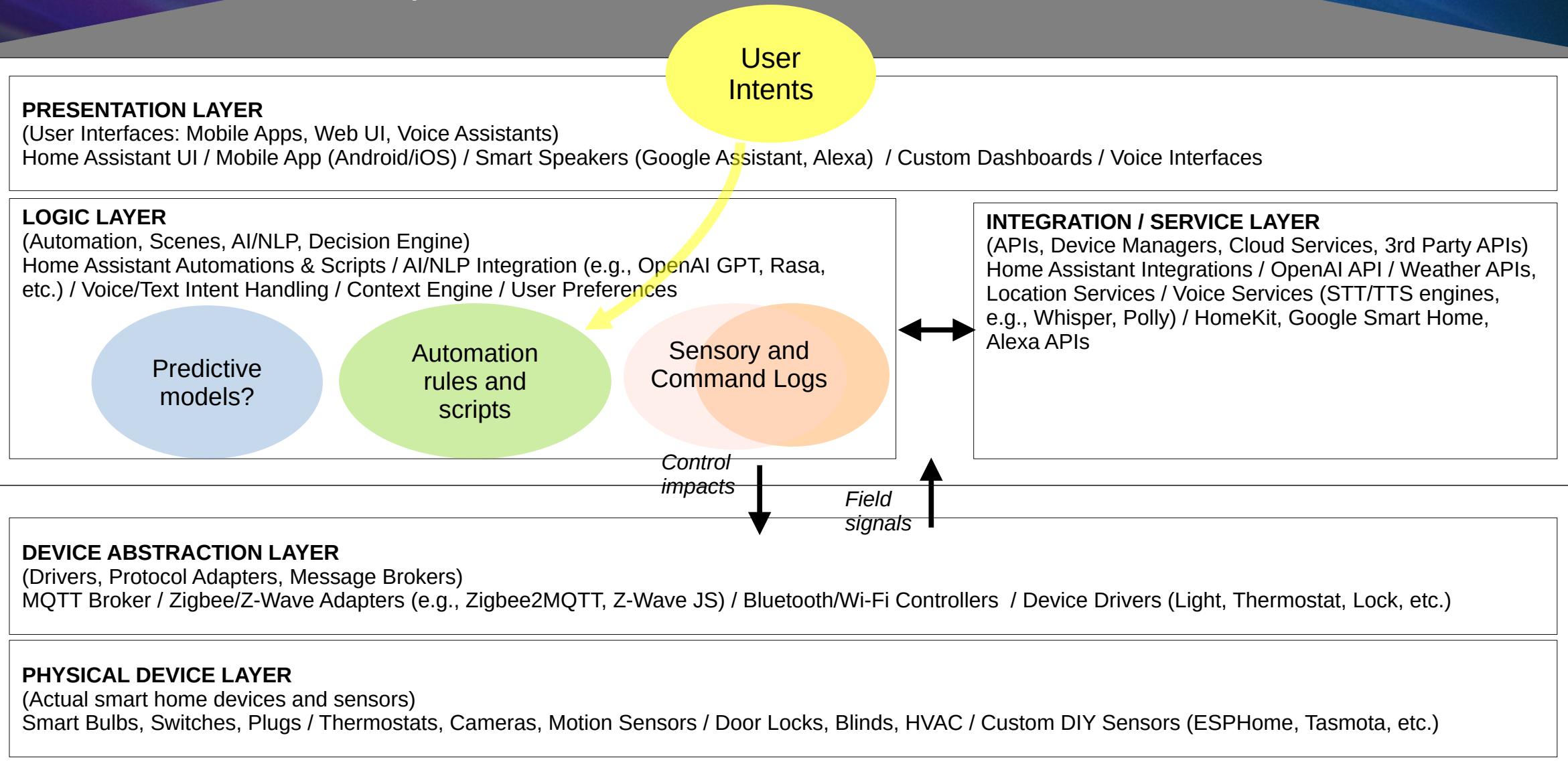
A) Automated Process Control Systems (APCS)



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

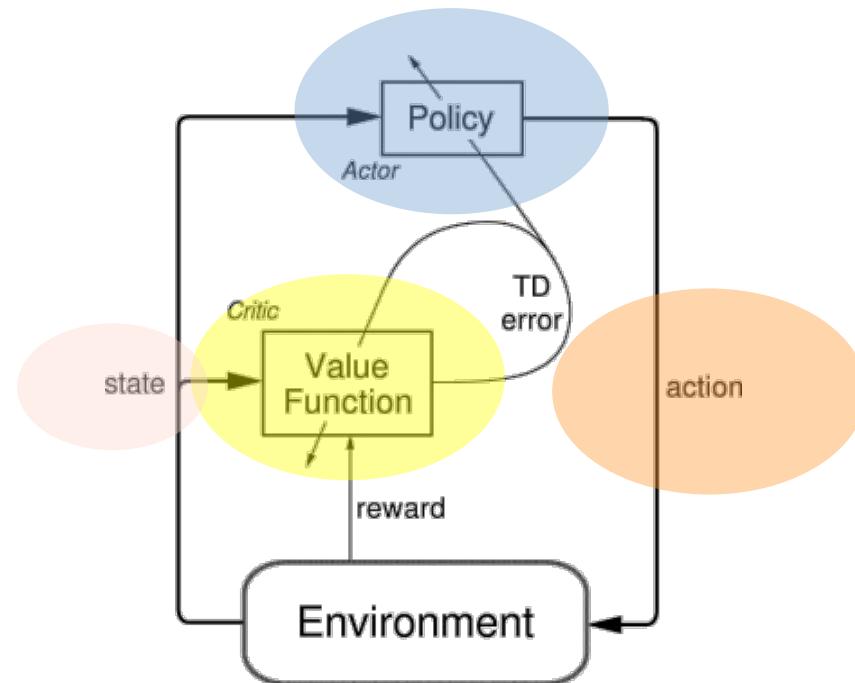
Application cases

B) “Smart Home”/“Smart House”



Implementation options

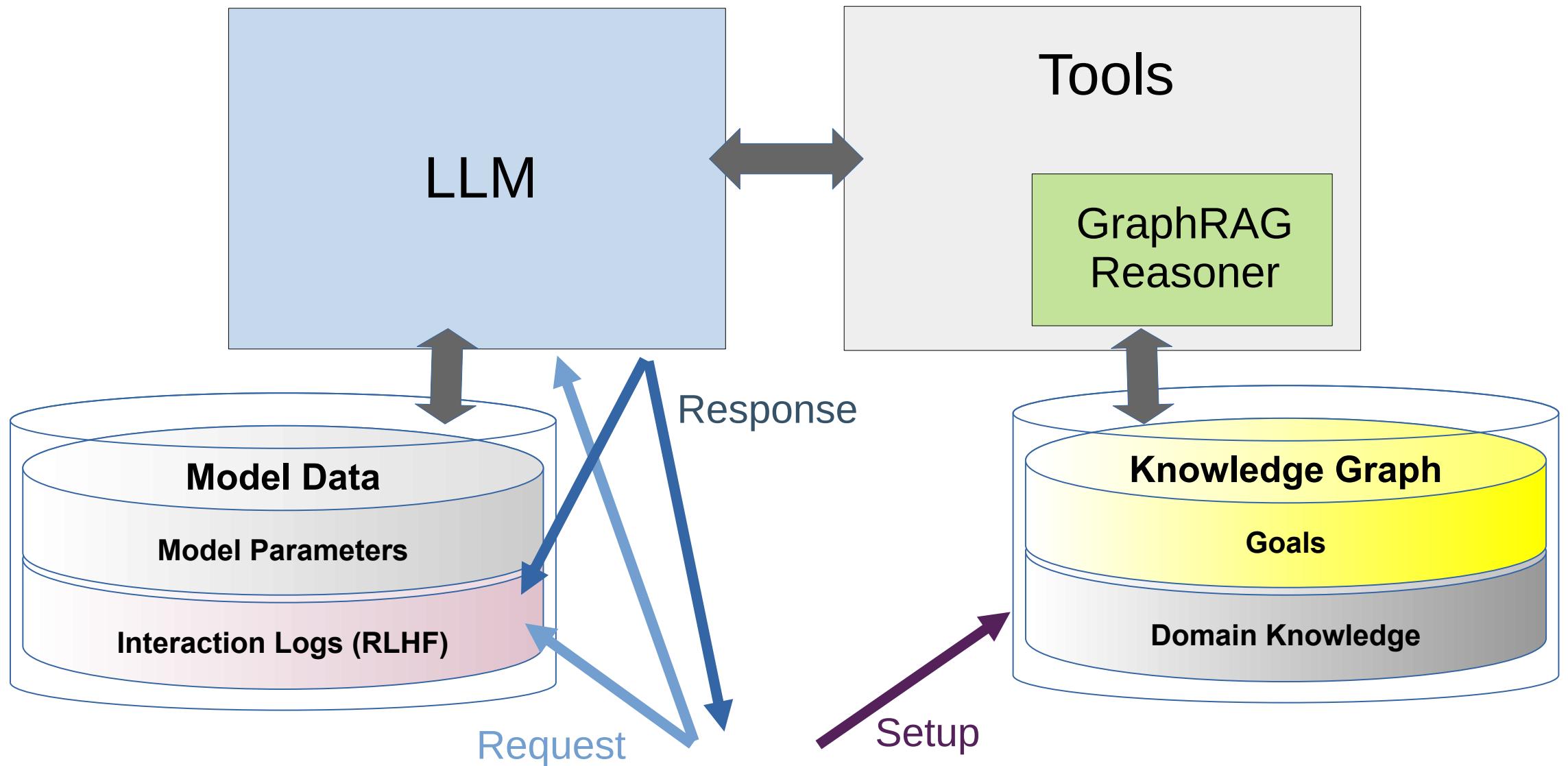
1) Reinforcement learning based on “actor-critic” model



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

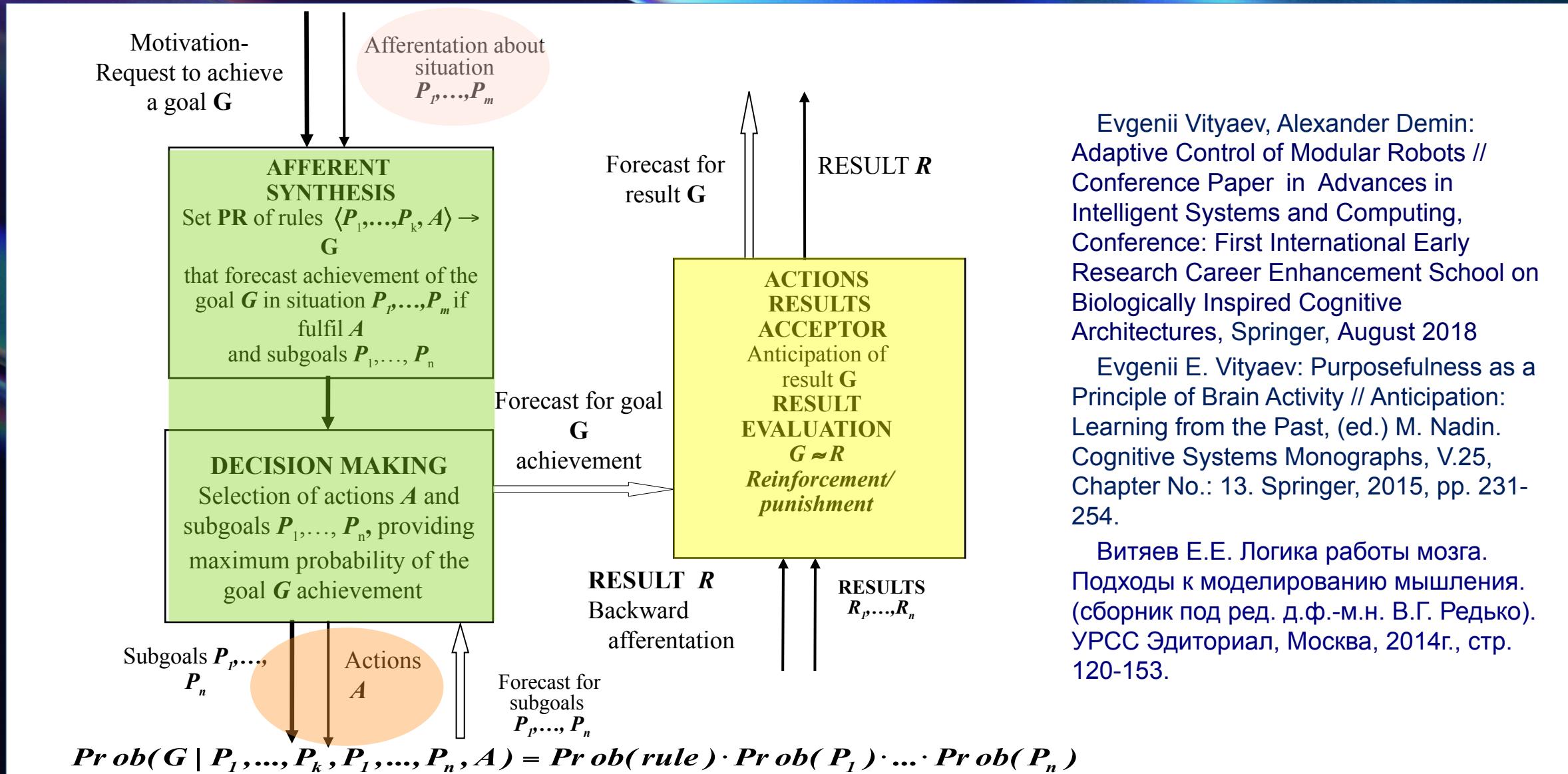
Implementation options

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



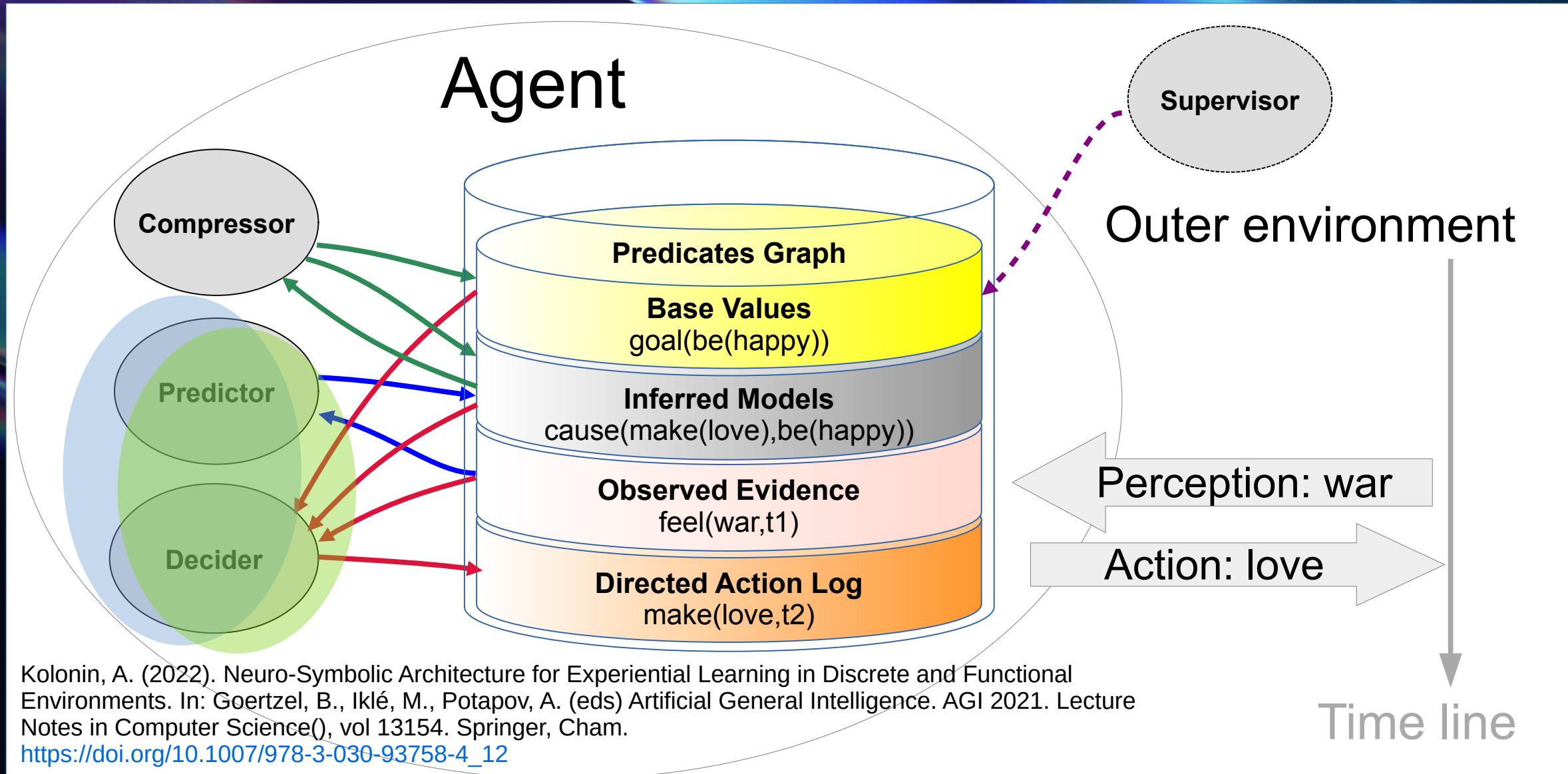
Implementation options

3) Cognitive architecture based on probabilistic logic (“Discovery”)



Implementation options

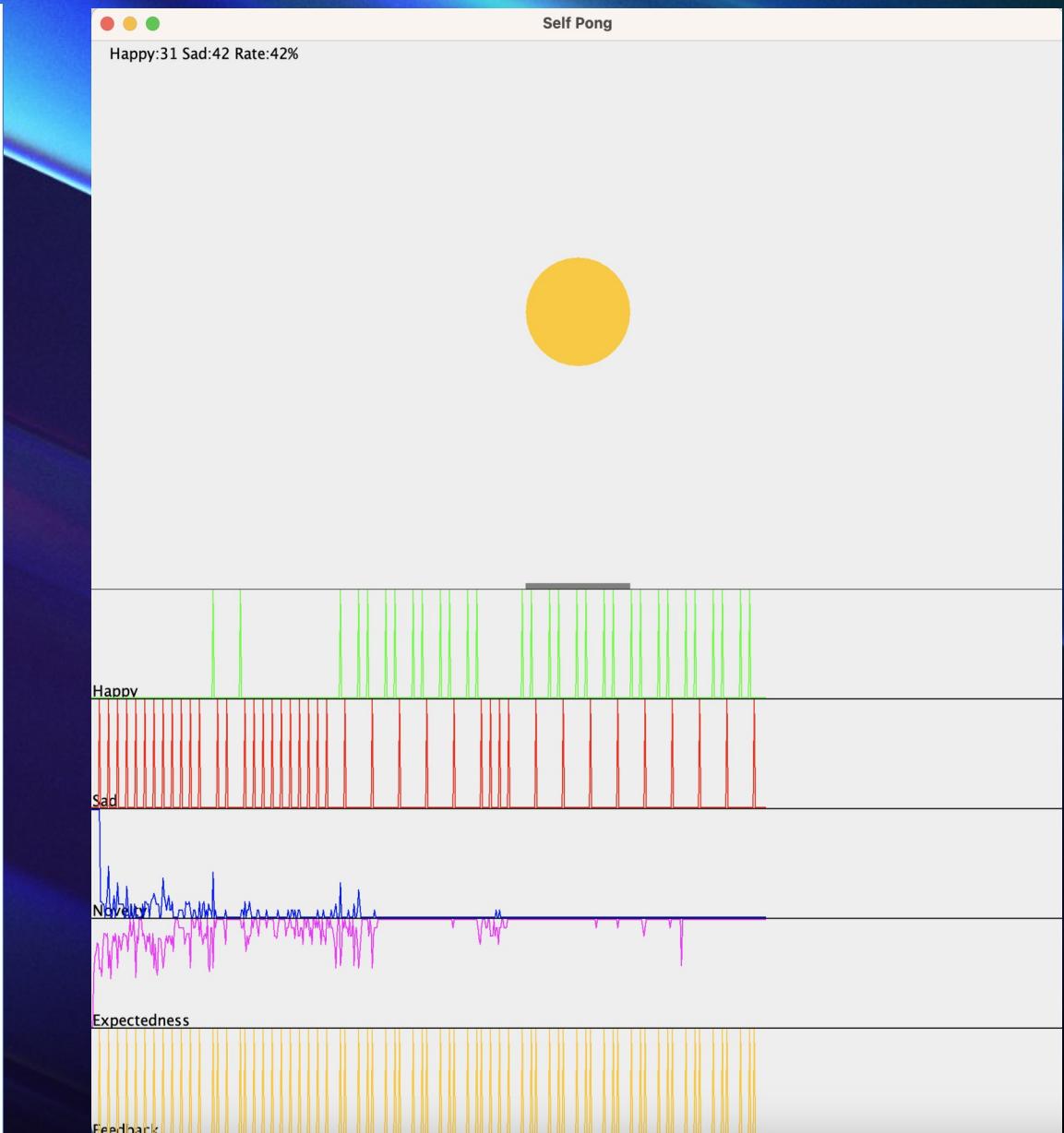
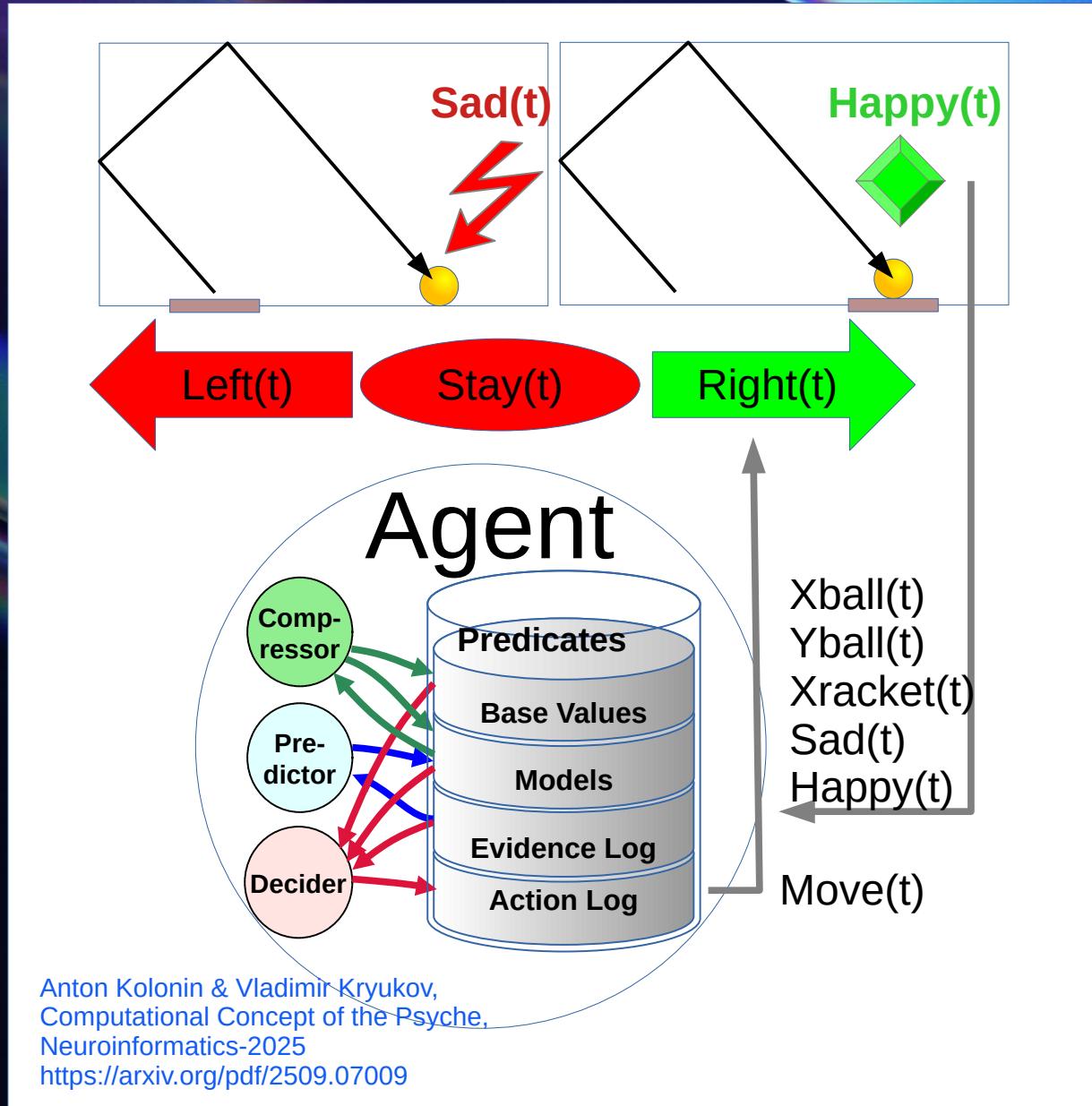
4) Cognitive architecture of value-based experiential learning



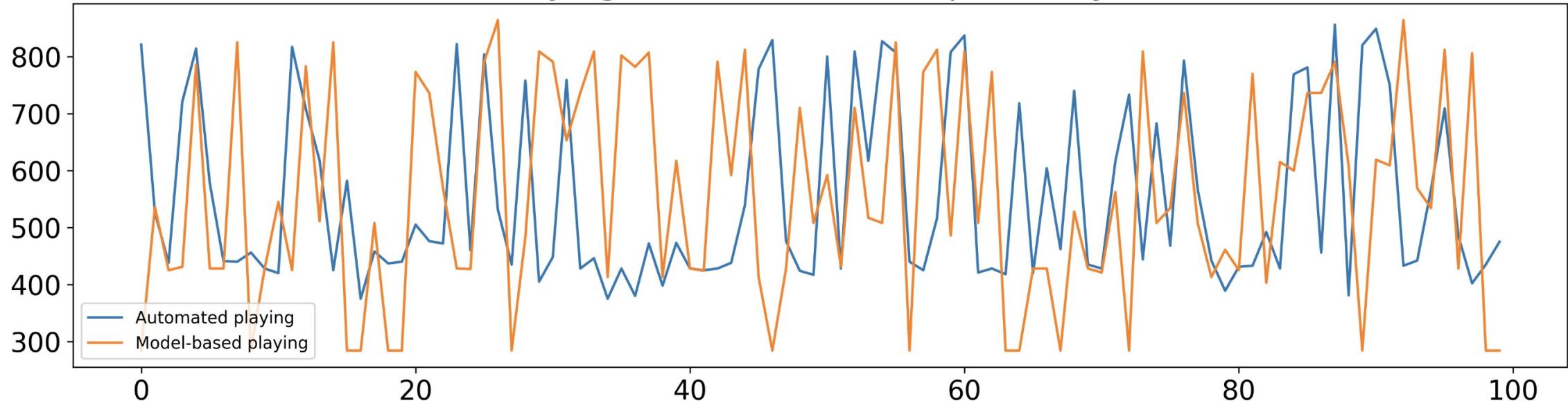
Kolonin, A. (2022). Neuro-Symbolic Architecture for Experiential Learning in Discrete and Functional Environments. In: Goertzel, B., Iklé, M., Potapov, A. (eds) Artificial General Intelligence. AGI 2021. Lecture Notes in Computer Science(), vol 13154. Springer, Cham.

https://doi.org/10.1007/978-3-030-93758-4_12

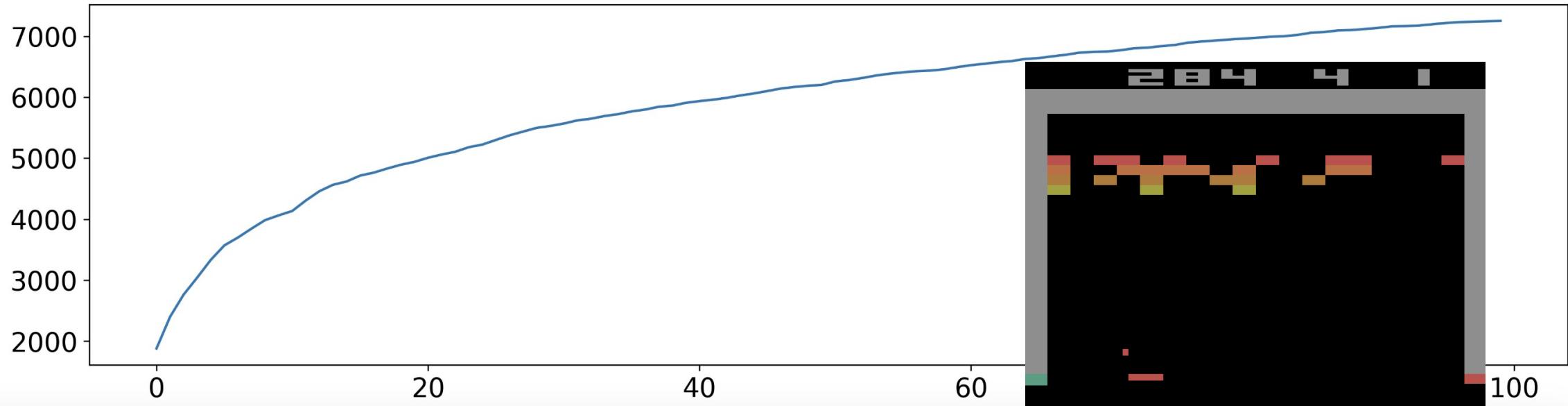
Cognitive architecture of value-based experiential learning



Playing Atari Breakout in Open AI Gym



N of States in the Model



Thank you for attention! Questions?

Anton Kolonin

akolonin@aigents.com

Telegram/GitHub: akolonin

Workshop recording
on the subject



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

