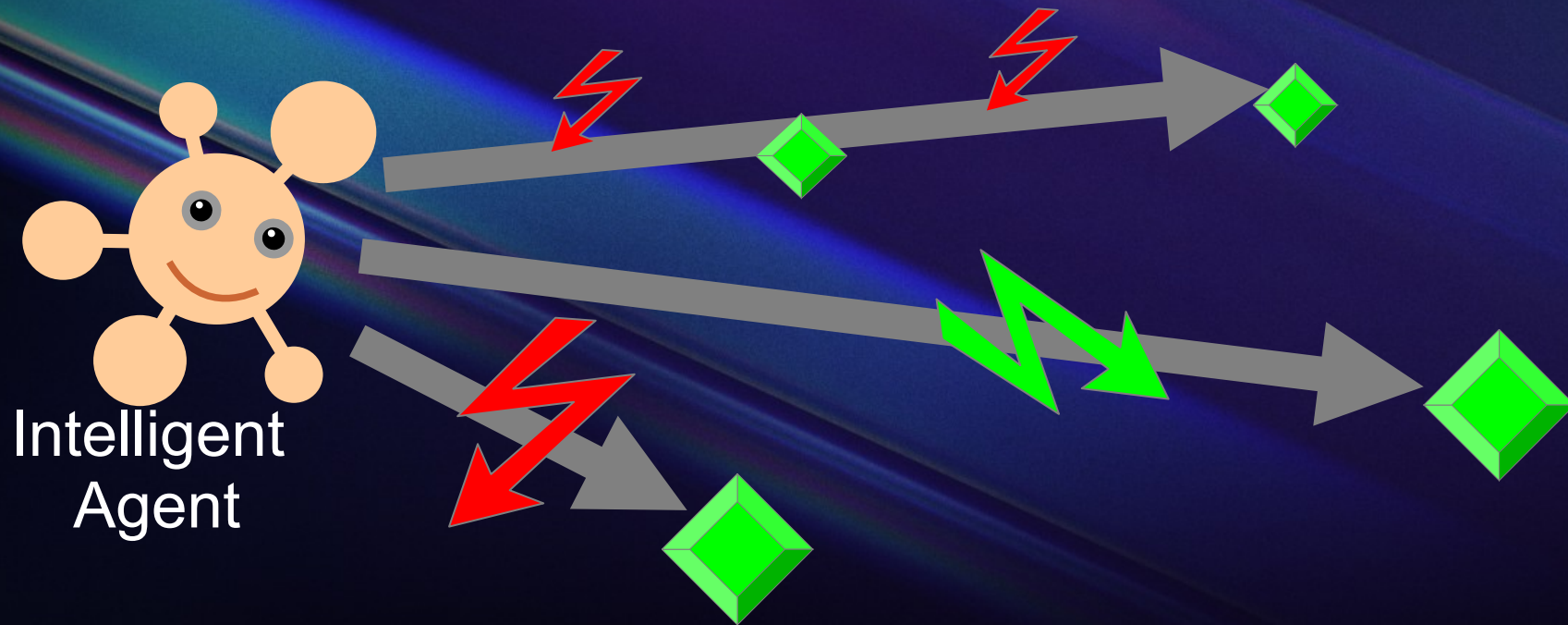


Computational Concept and Cognitive Architecture of Artificial Psyche

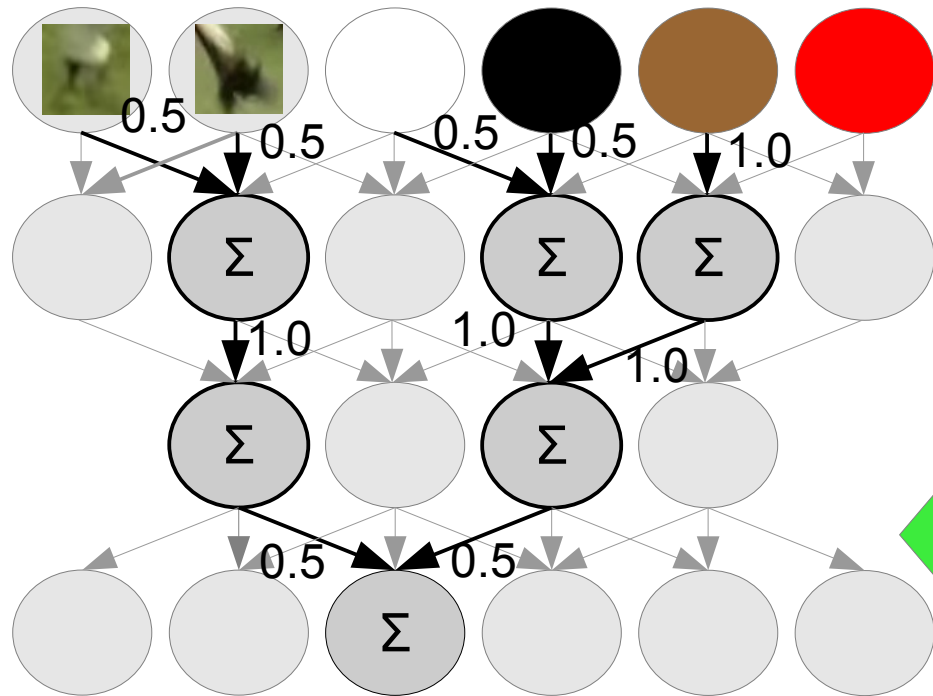
Anton Kolonin, Ph.D.
Novosibirsk State University <https://nsu.ru>
Russian AGI Community <https://agirussia.org>
Aigents <https://aigents.com>

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

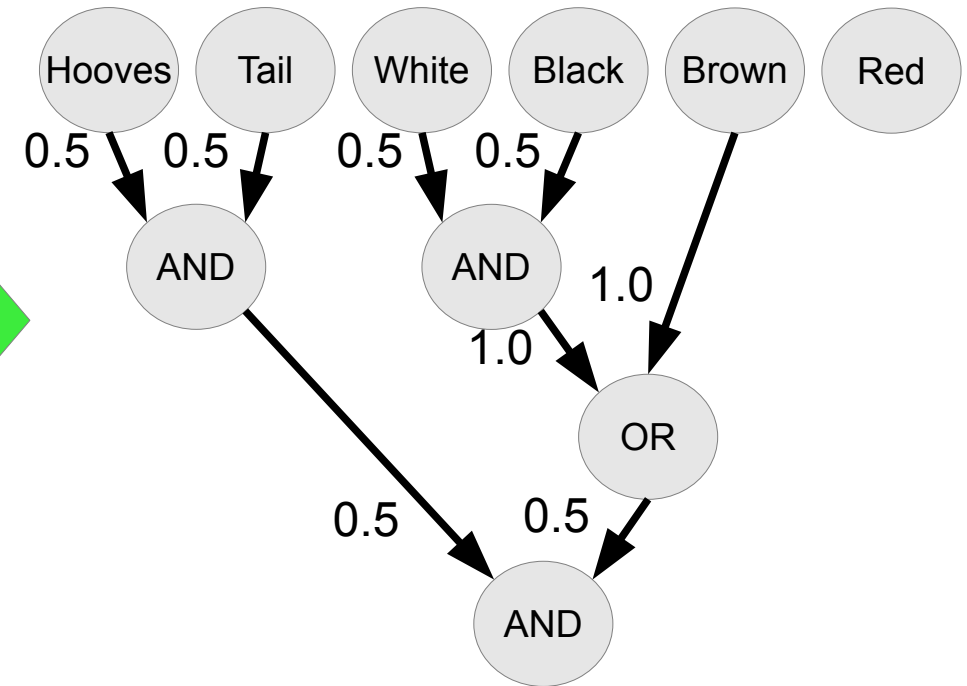


"System 1"



Explain

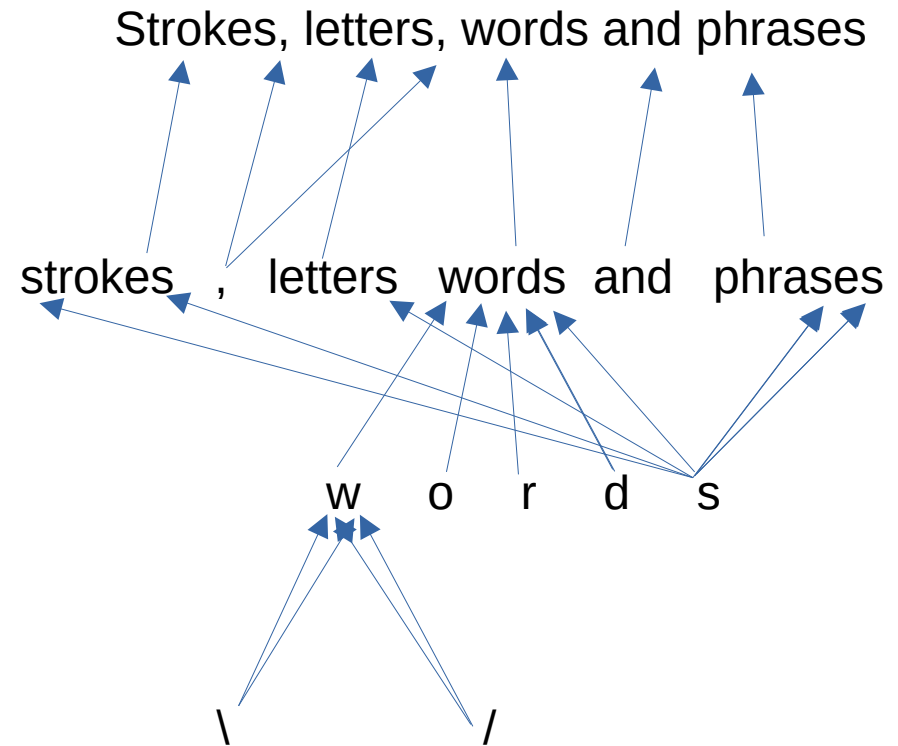
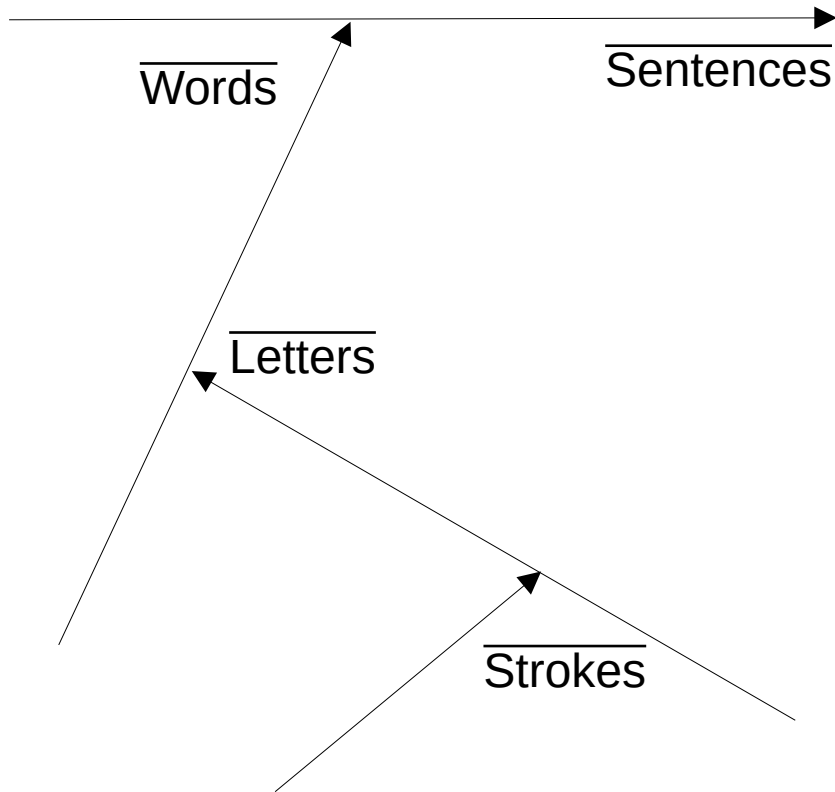
Transfer



(Hooves AND Tail) AND NOT Red
((White and Black) OR Brown)

=> Horse

Functional equivalence of tensor and graph (symbolic) models



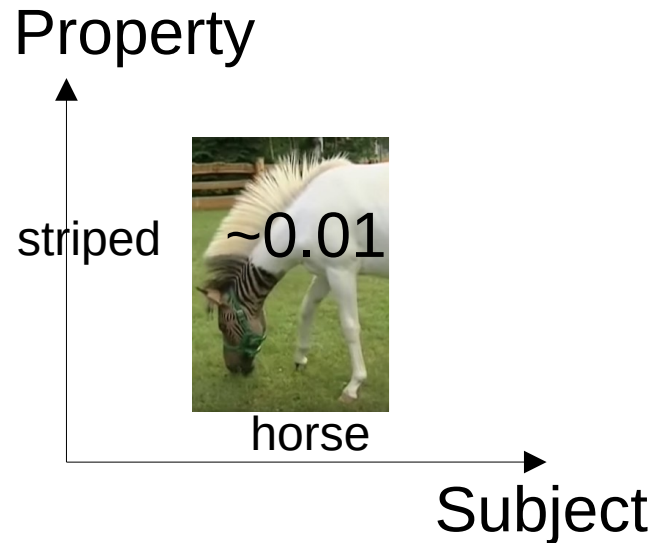
Functional equivalence of tensor and (symbolic) graph models

Truth-Value Tensor (NARS/PLN)

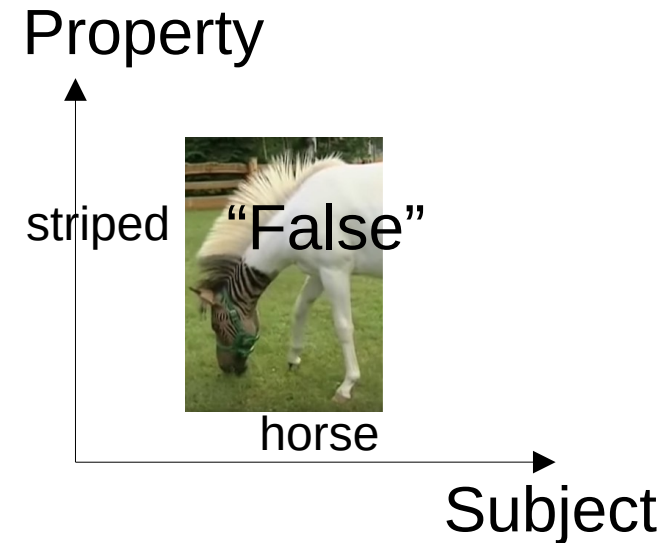


Life-long
learning?

Numerical Tensor (ANN/Bayesian Logic)

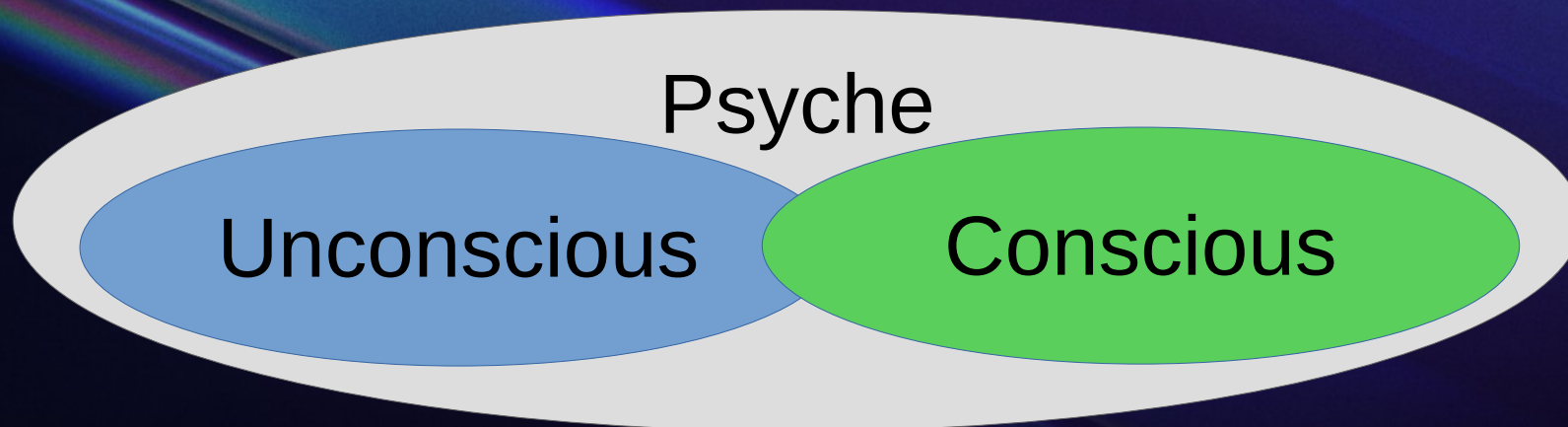


Boolean Tensor (Boolean Logic)

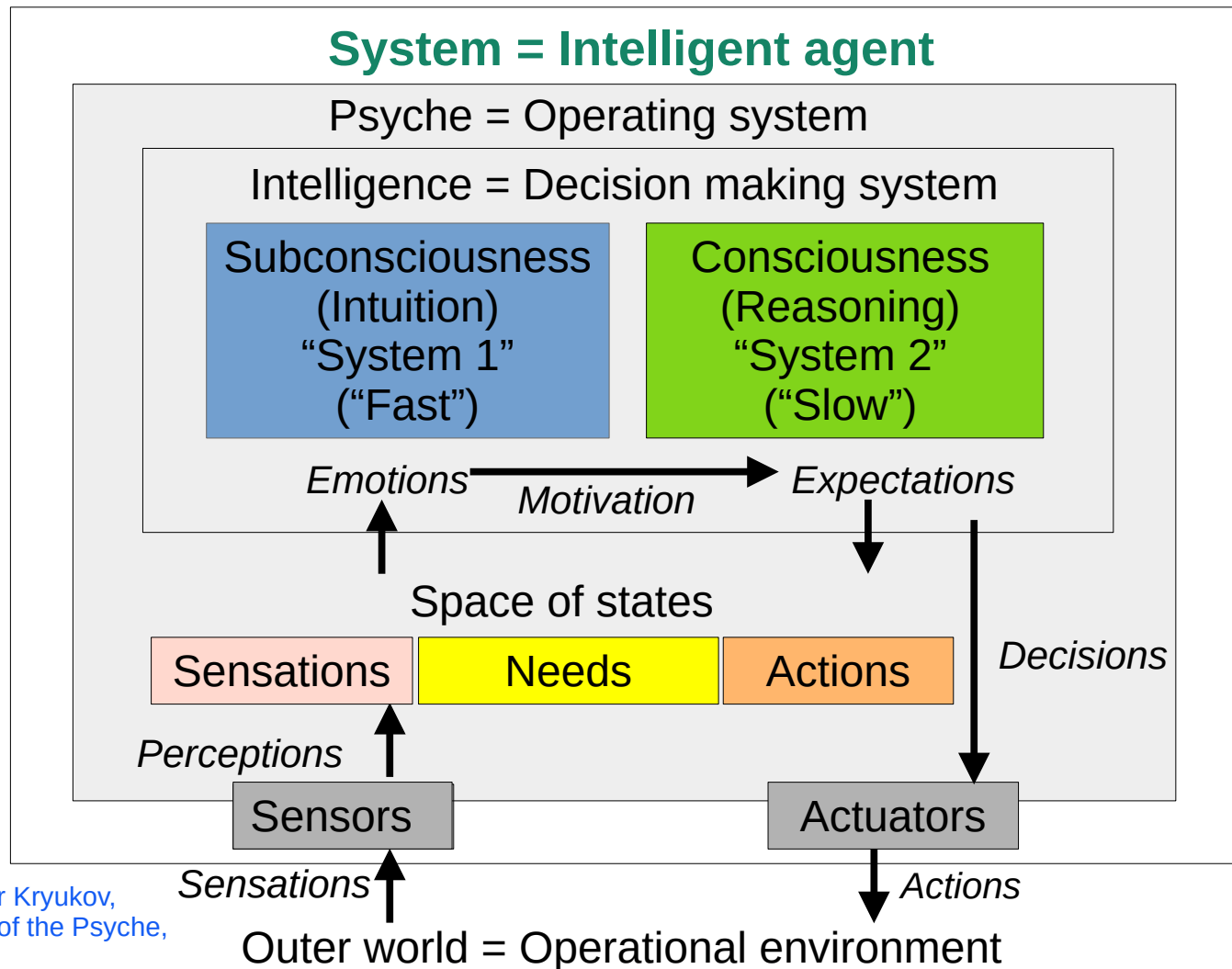


The psyche /'saɪki/ is currently used to describe the totality of the human mind, conscious and unconscious.

<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

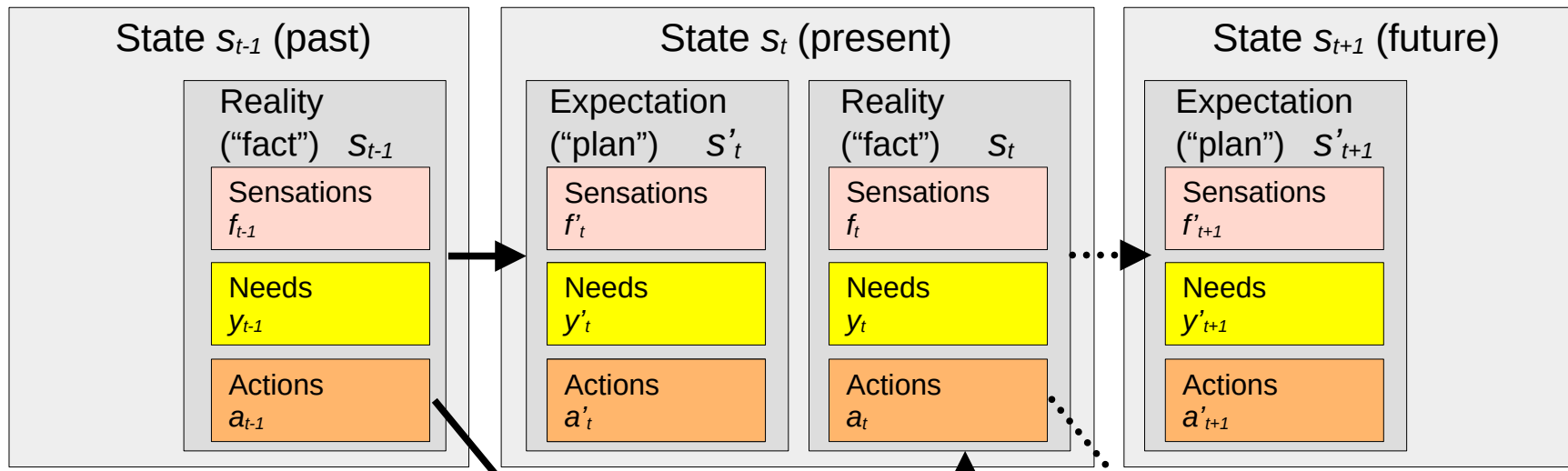
Models s ("invariants") of states with utilities U and probabilities P of transitions

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

↑ *Experiential learning*

↓ *Decision making*

Space of states and episodic memory ("precedents")



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

Decisions

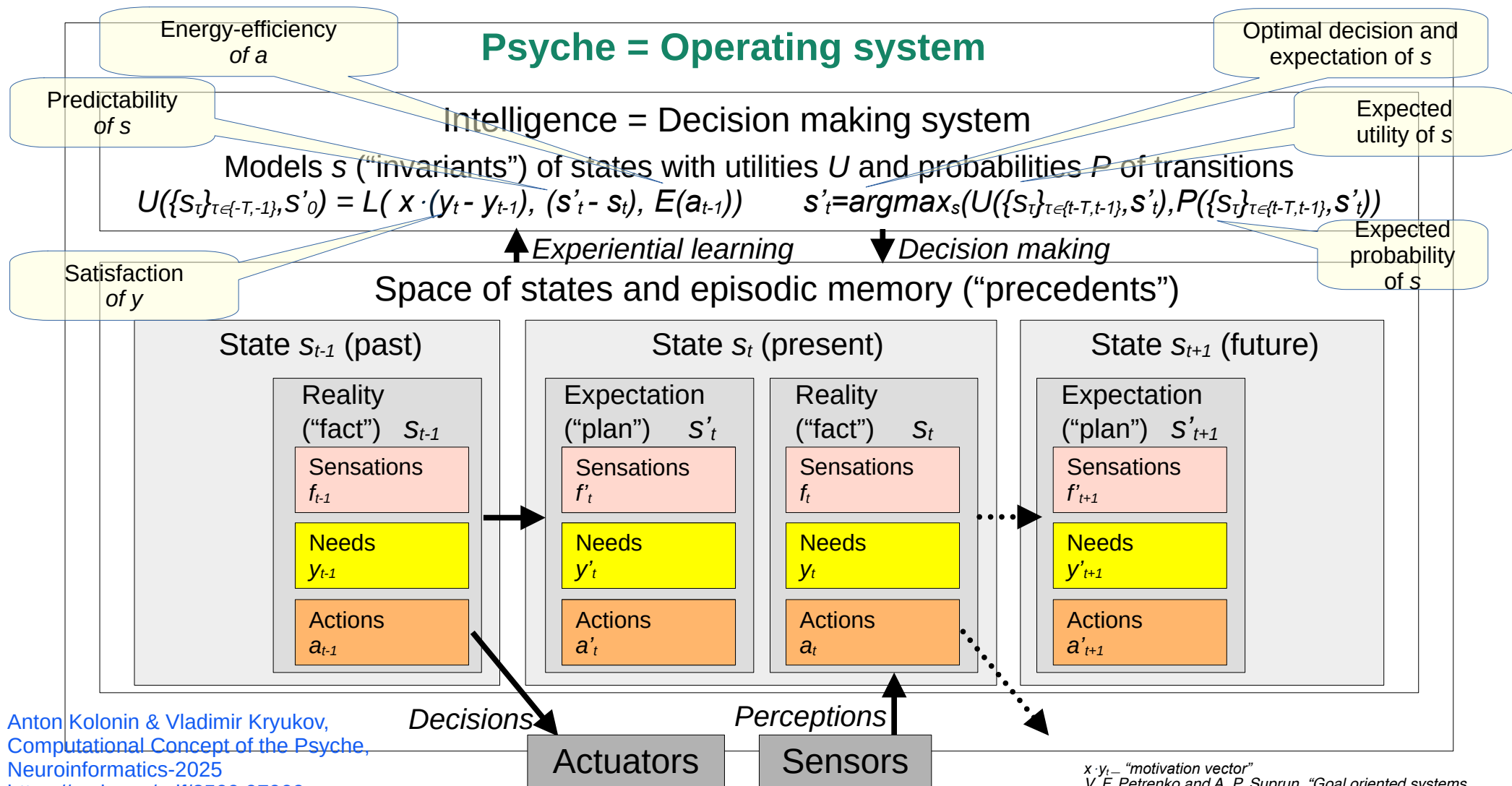
Perceptions

Actuators

Sensors

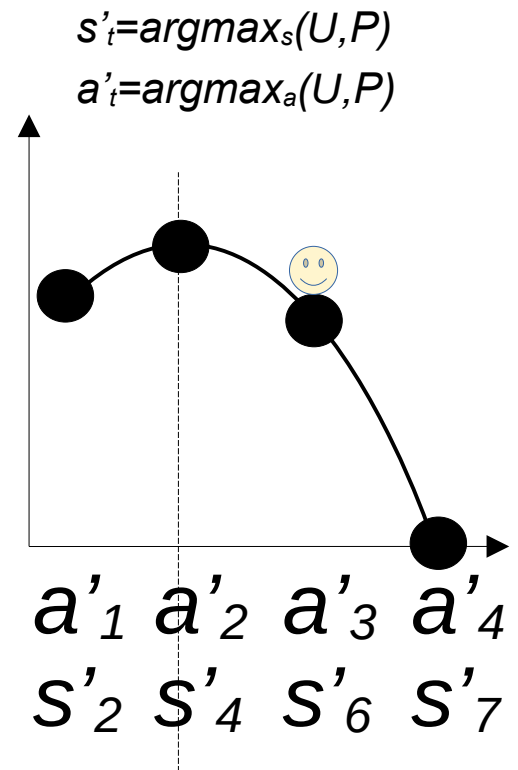
$x \cdot y_t$ – "motivation vector"
V. F. Petrenko and A. P. Suprun, "Goal oriented systems, evolution, and the subjective aspect in systemology," Tr. Inst. Sist. Anal. RAN 62 (1) (2012)

Computational Concept and Cognitive Architecture of Artificial Psyche



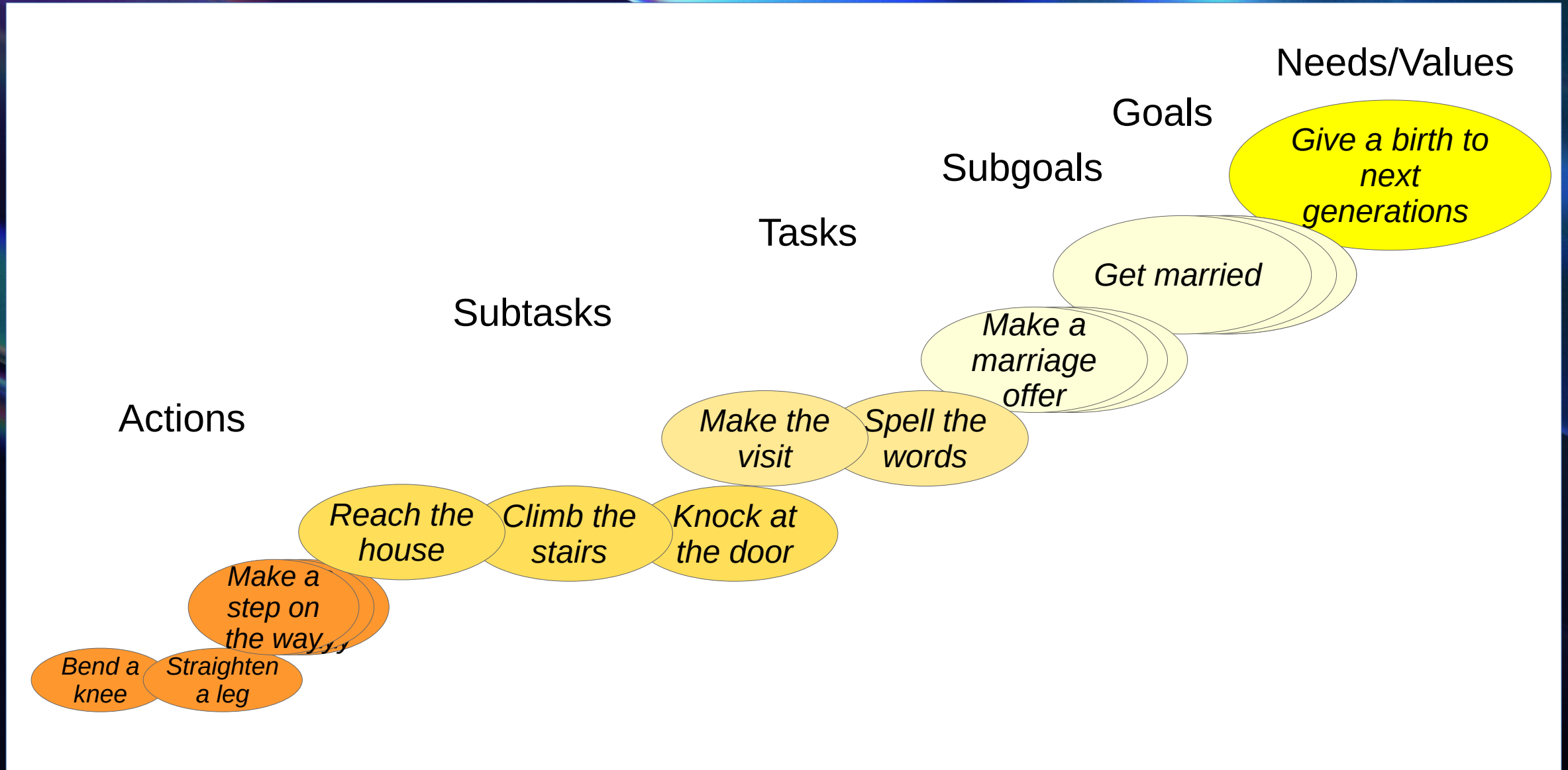
Decision making as operational risk management

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



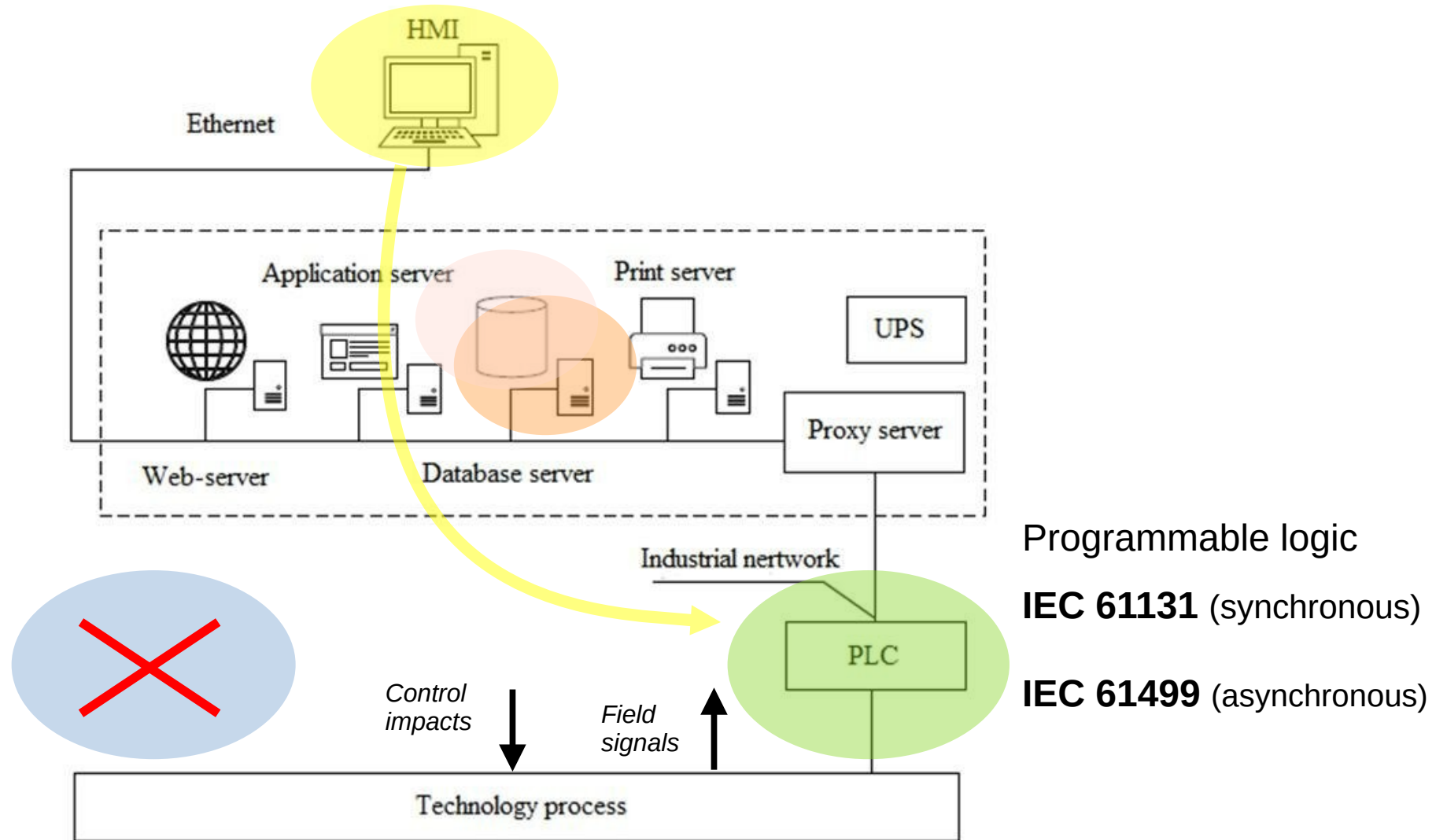
Tversky & Kahneman:
 most people choose a'_3 и s'_6
 ("smaller profit with
 greater reliability")

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



Application cases

A) Automated Process Control Systems (APCS)



Application cases

B) “Smart Home”/“Smart House”

User
Intents

PRESENTATION LAYER

(User Interfaces: Mobile Apps, Web UI, Voice Assistants)

Home Assistant UI / Mobile App (Android/iOS) / Smart Speakers (Google Assistant, Alexa) / Custom Dashboards / Voice Interfaces

LOGIC LAYER

(Automation, Scenes, AI/NLP, Decision Engine)

Home Assistant Automations & Scripts / AI/NLP Integration (e.g., OpenAI GPT, Rasa, etc.) / Voice/Text Intent Handling / Context Engine / User Preferences

Predictive
models?

Automation
rules and
scripts

Sensory and
Command Logs

INTEGRATION / SERVICE LAYER

(APIs, Device Managers, Cloud Services, 3rd Party APIs)
Home Assistant Integrations / OpenAI API / Weather APIs,
Location Services / Voice Services (STT/TTS engines,
e.g., Whisper, Polly) / HomeKit, Google Smart Home,
Alexa APIs

Control
impacts

Field
signals

DEVICE ABSTRACTION LAYER

(Drivers, Protocol Adapters, Message Brokers)

MQTT Broker / Zigbee/Z-Wave Adapters (e.g., Zigbee2MQTT, Z-Wave JS) / Bluetooth/Wi-Fi Controllers / Device Drivers (Light, Thermostat, Lock, etc.)

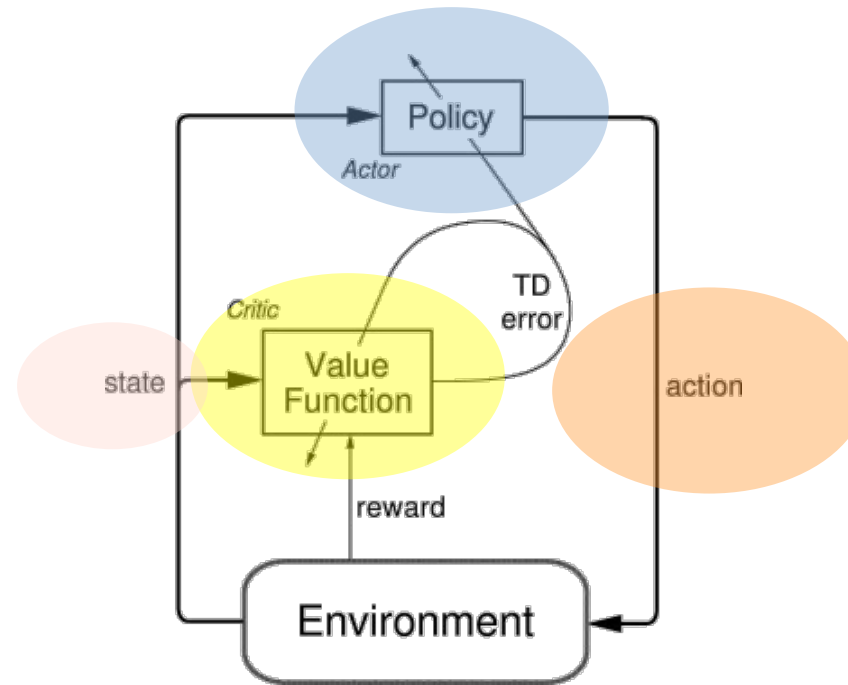
PHYSICAL DEVICE LAYER

(Actual smart home devices and sensors)

Smart Bulbs, Switches, Plugs / Thermostats, Cameras, Motion Sensors / Door Locks, Blinds, HVAC / Custom DIY Sensors (ESPHome, Tasmota, etc.)

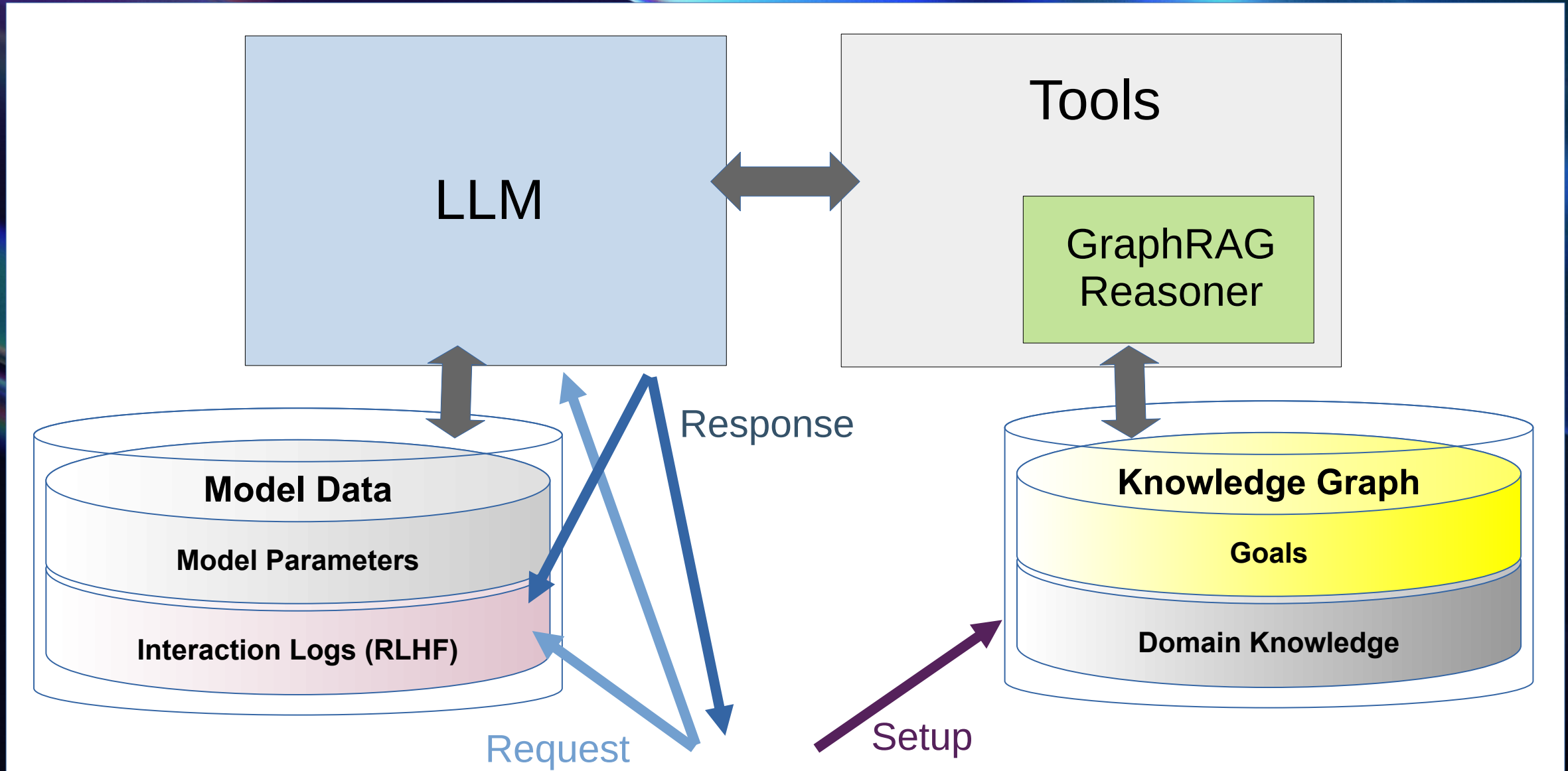
Implementation options

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



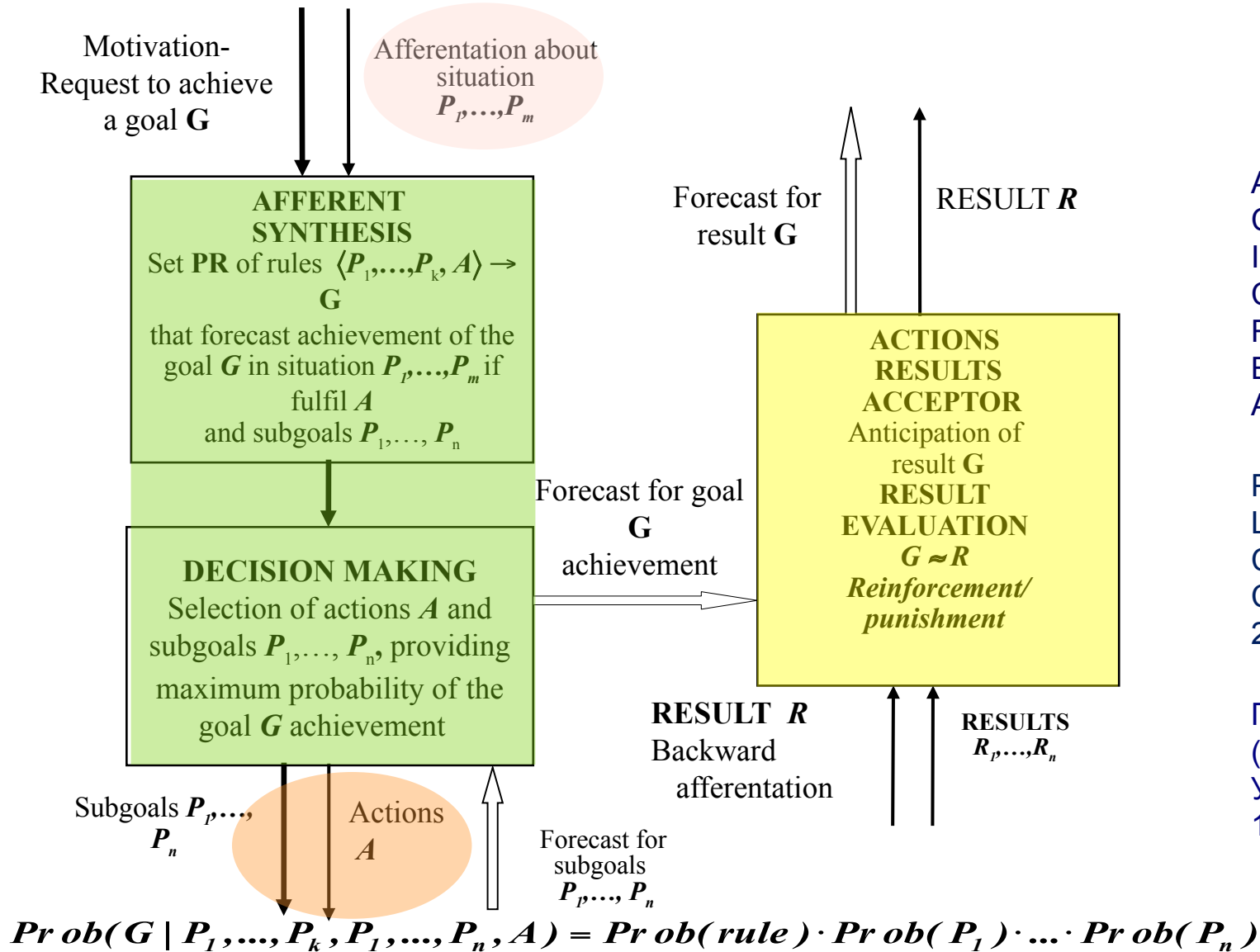
Implementation options

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



Implementation options

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



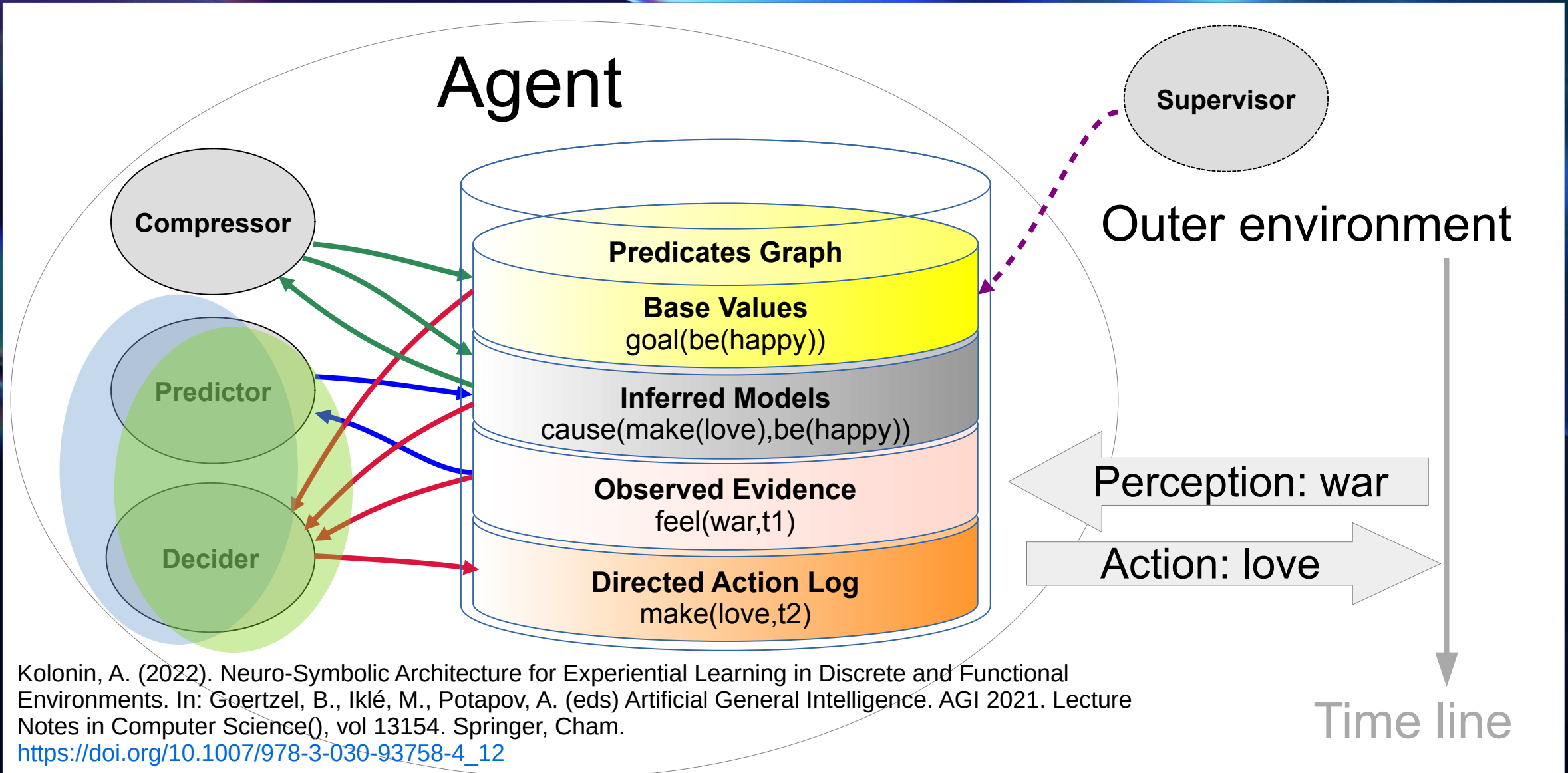
Evgenii Vityaev, Alexander Demin:
Adaptive Control of Modular Robots // Conference Paper in Advances in Intelligent Systems and Computing, Conference: First International Early Research Career Enhancement School on Biologically Inspired Cognitive Architectures, Springer, August 2018

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.

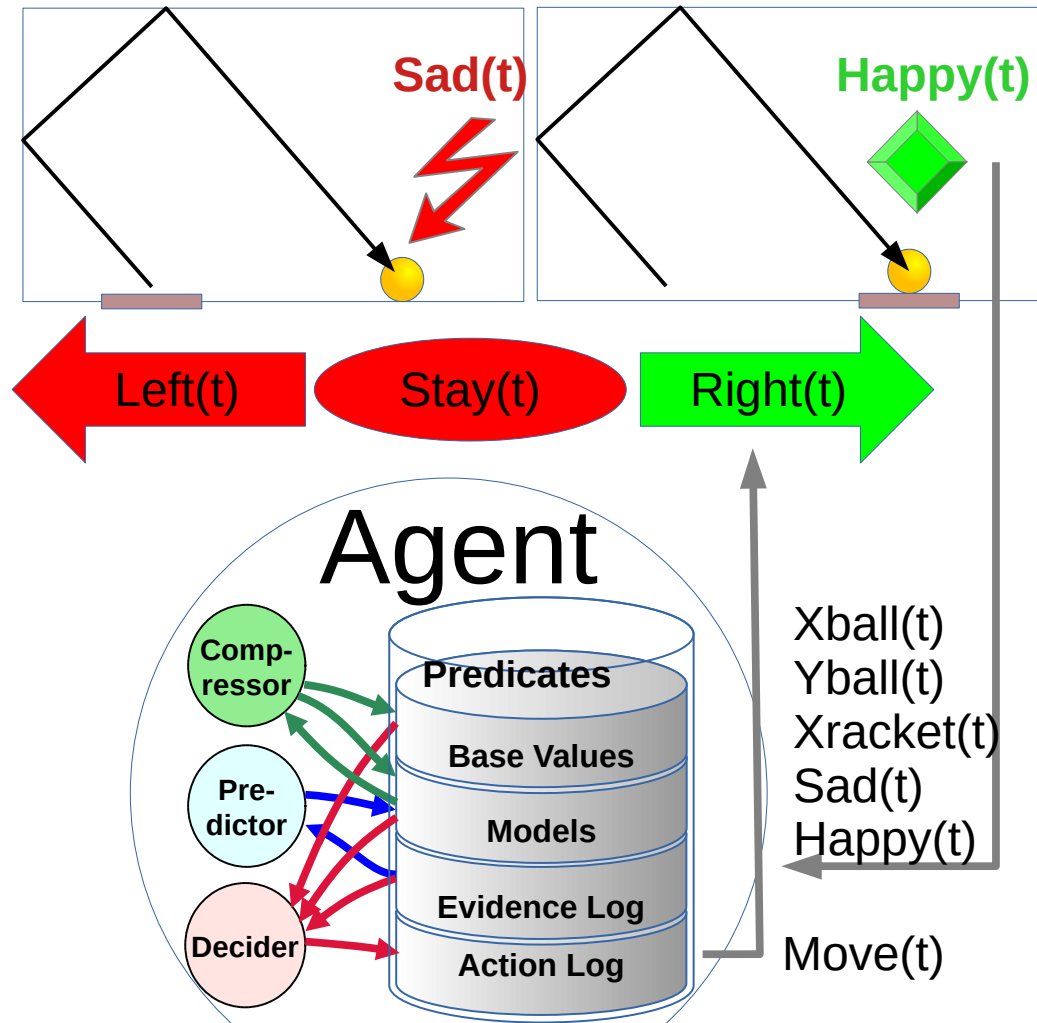
Витяев Е.Е. Логика работы мозга. Подходы к моделированию мышления. (сборник под ред. д.ф.-м.н. В.Г. Редько). УРСС Эдиториал, Москва, 2014г., стр. 120-153.

Implementation options

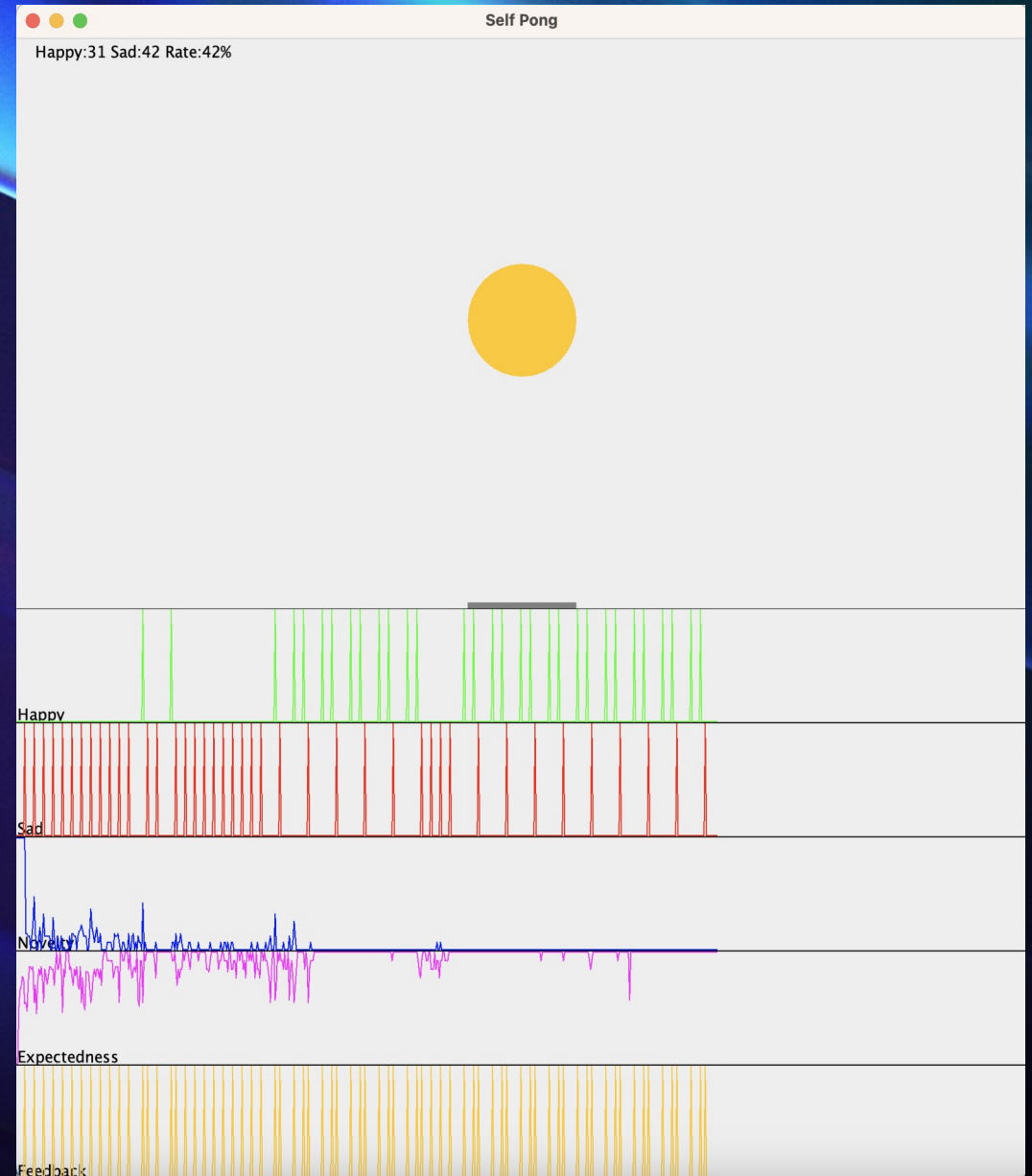
4) Cognitive architecture of value-based experiential learning



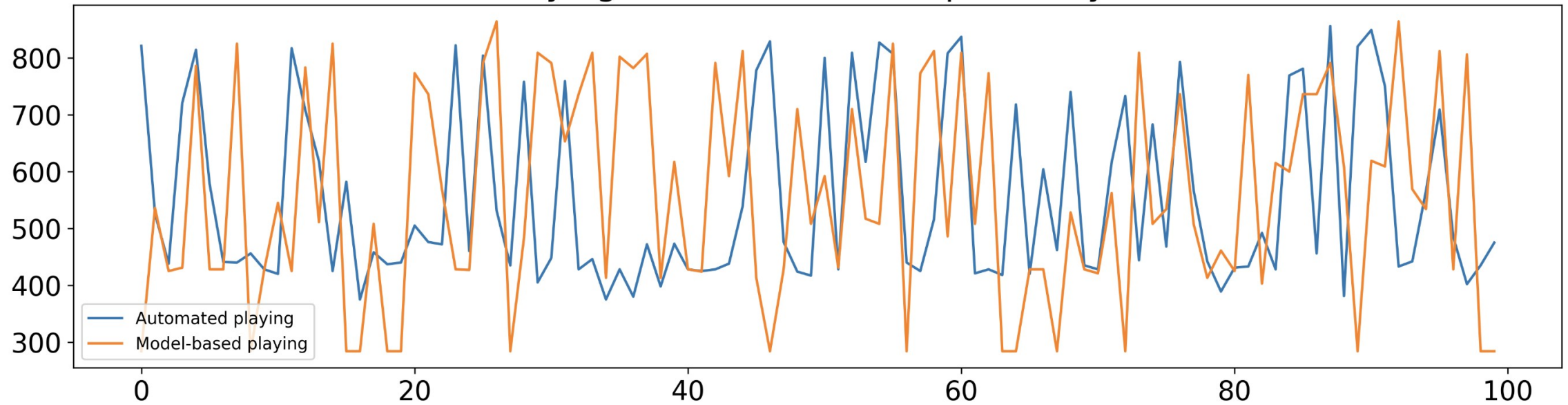
Cognitive architecture of value-based experiential learning



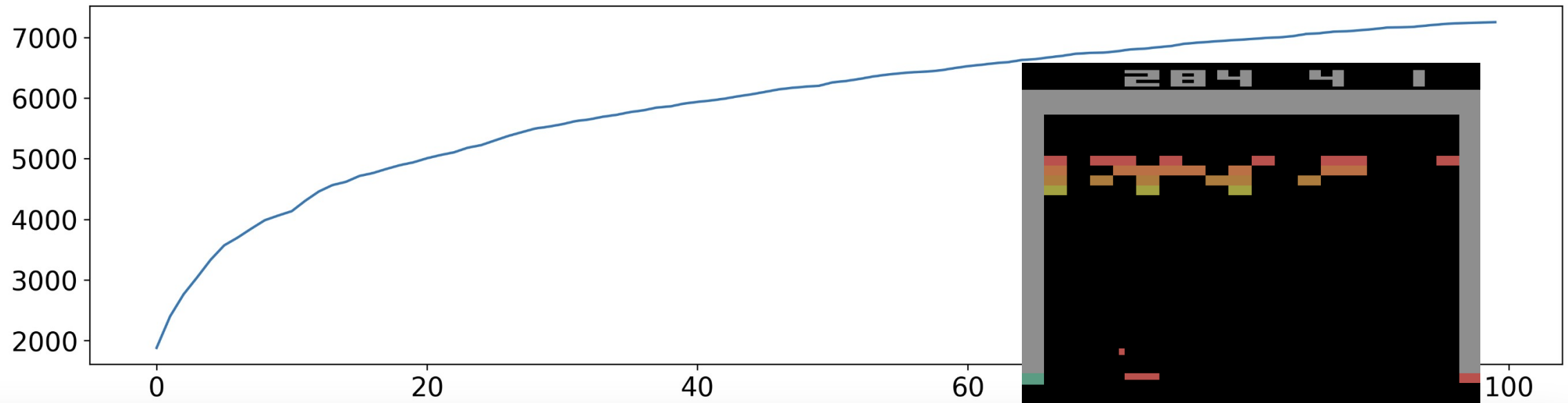
Anton Kolonin & Vladimir Kryukov,
Computational Concept of the Psyche,
Neuroinformatics-2025
<https://arxiv.org/pdf/2509.07009>



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

