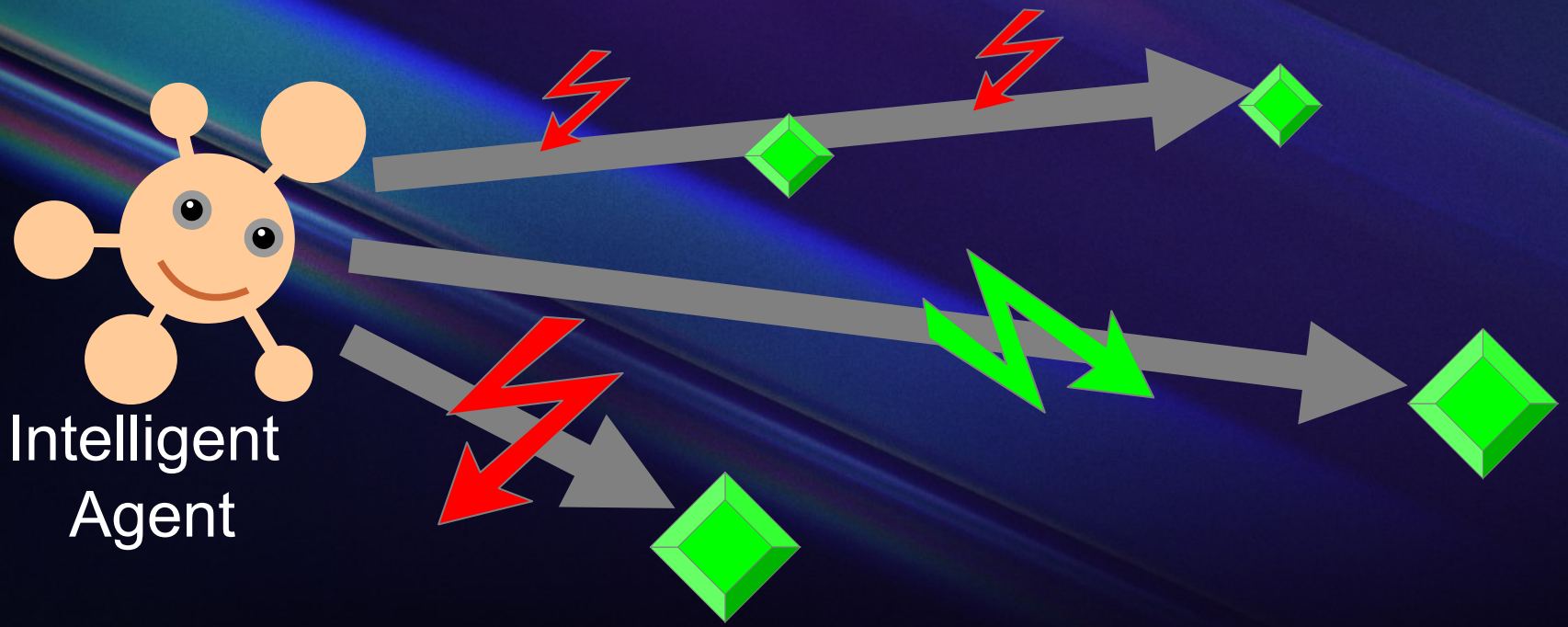
# Computational Concept and Cognitive Architecture of Artificial Psyche

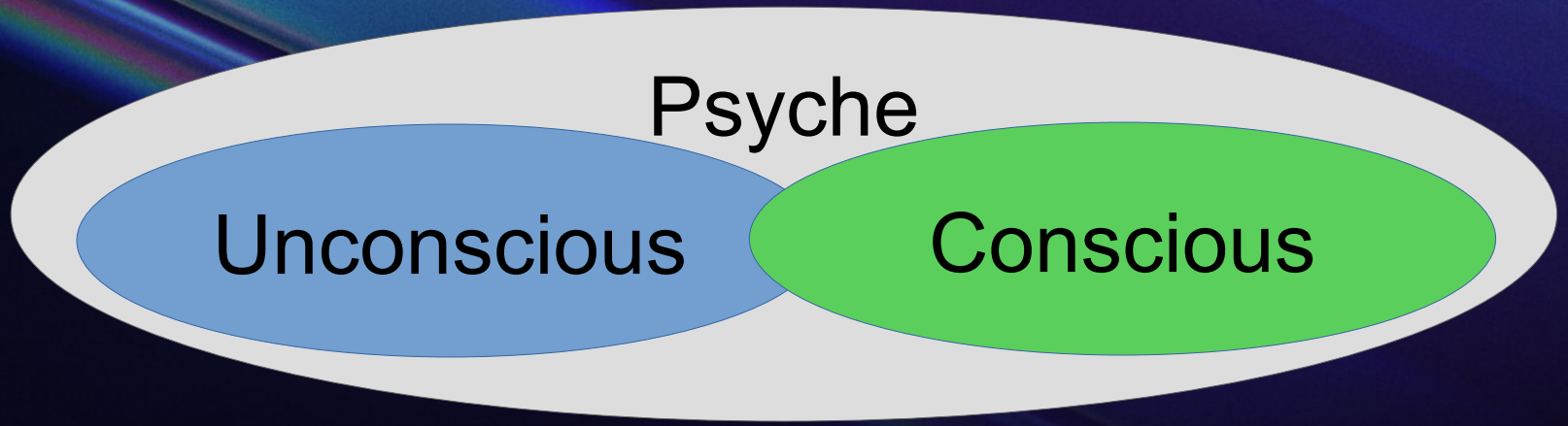
Anton Kolonin, Ph.D.
Novosibirsk State University https:/nsu.ru
Russian AGI Community https://agirussia.org
Aigents https://aigents.com

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

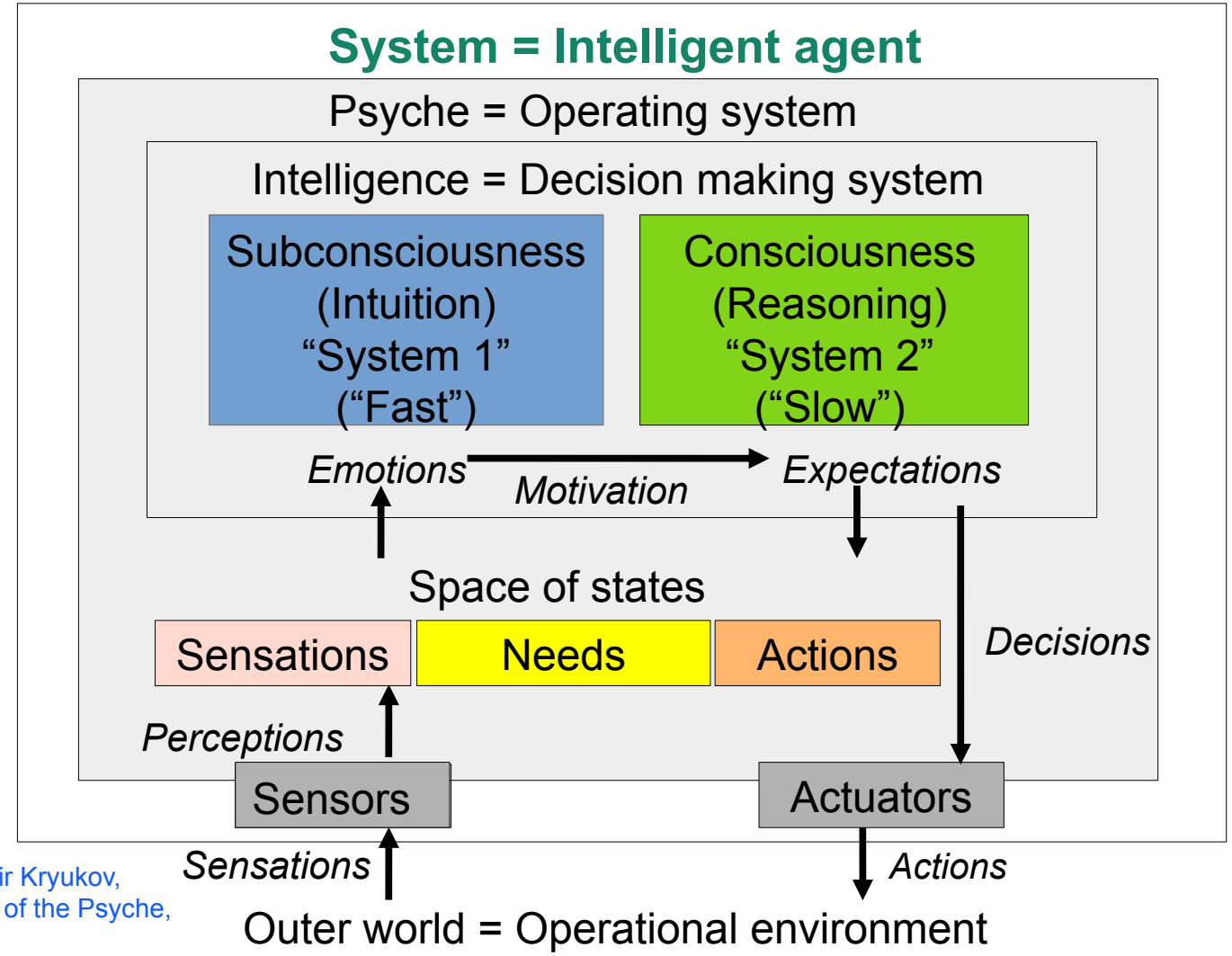


The psyche /ˈsaɪki/ is currently used to describe the totality of the **human mind**, conscious and unconscious.

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

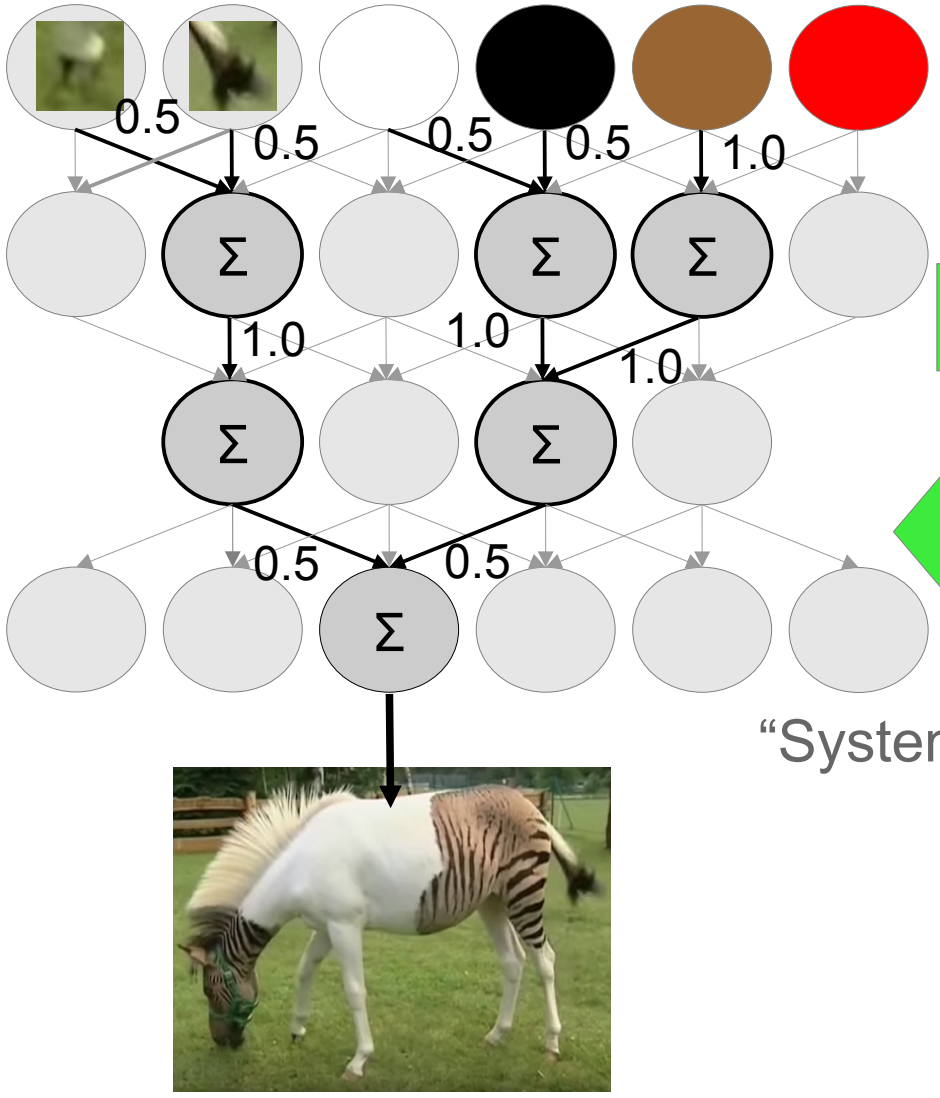


# Computational Concept and Cognitive Architecture of Artificial Psyche




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

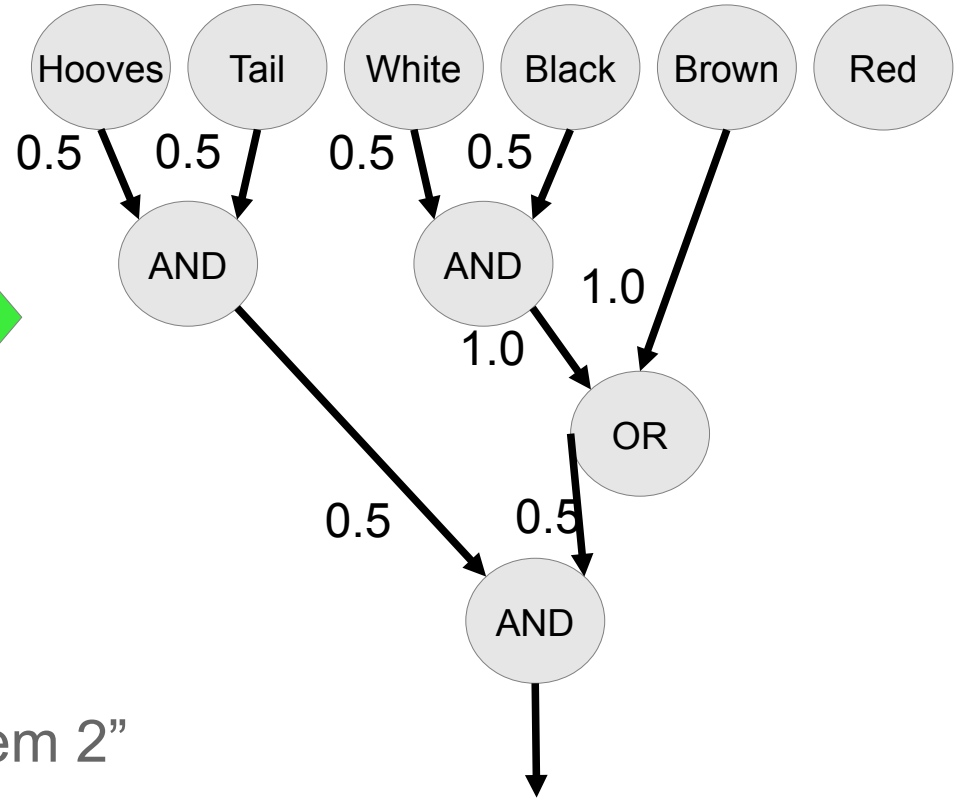
# Neuro-Symbolic Integration for Interpretable AI






Explain

Transfer



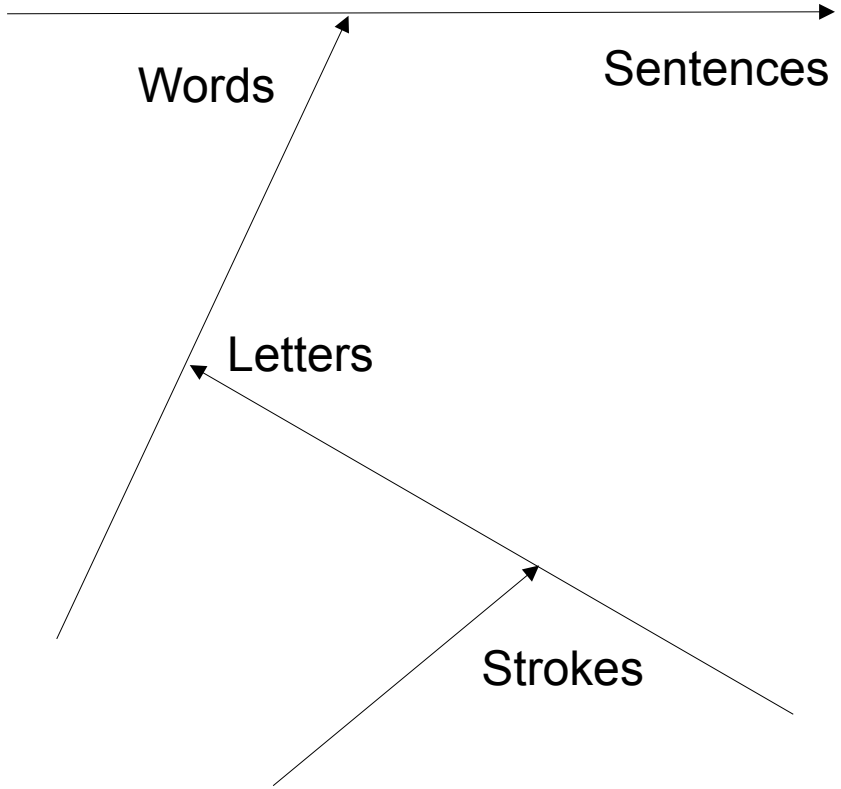

"System 1" "System 2"

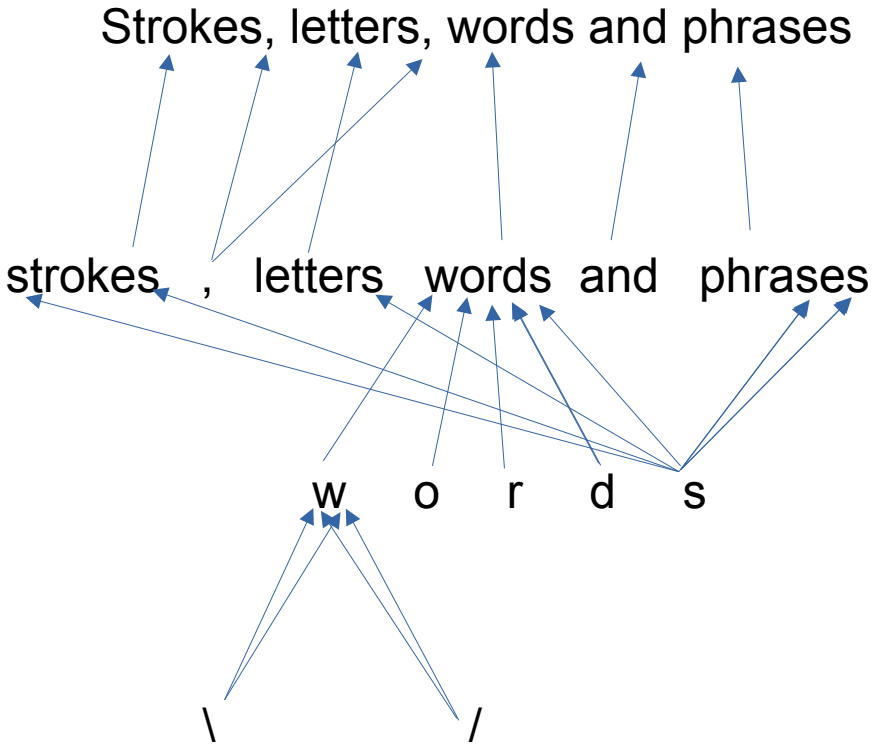
(Hooves AND Tail) AND NOT Red
((White and Black) OR Brown)

=> Horse

https://www.springerprofessional.de/neuro-symbolic-architecture-for-experiential-learning-in-discret/20008336

# Functional equivalence of tensor and graph (symbolic) models

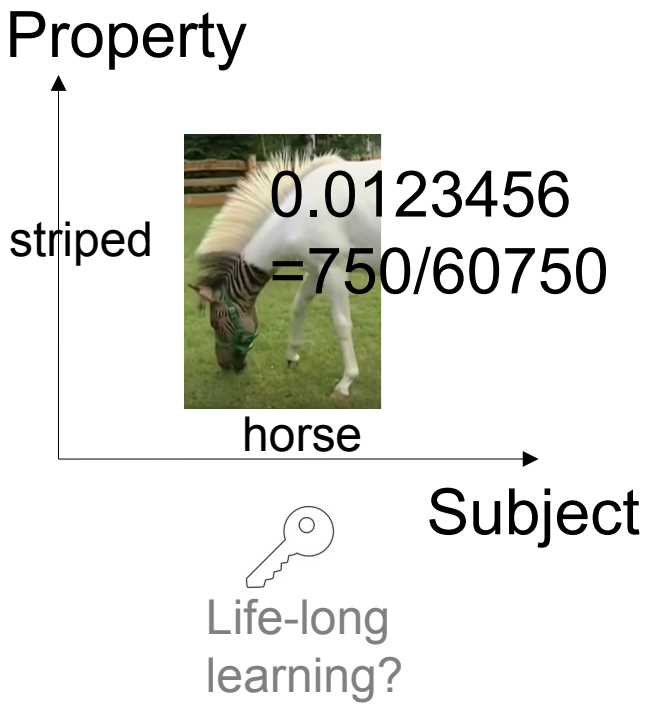


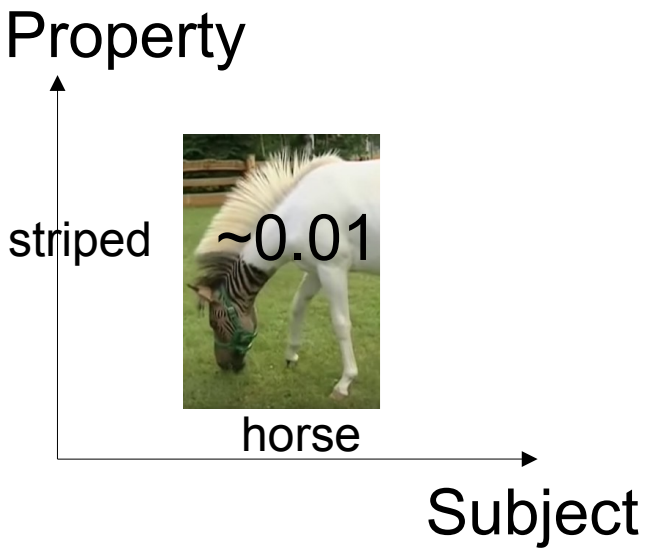

Pedro Domingos, Tensor Logic: The Language of AI https://arxiv.org/pdf/2510.12269

# Functional equivalence of tensor and (symbolic) graph models

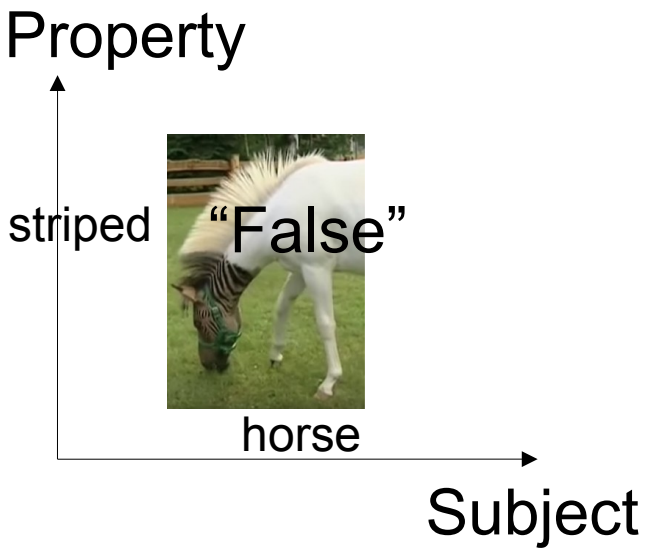
**Truth-Value Tensor**
(NARS/PLN)




**Numerical Tensor**
(ANN/Bayesian Logic)

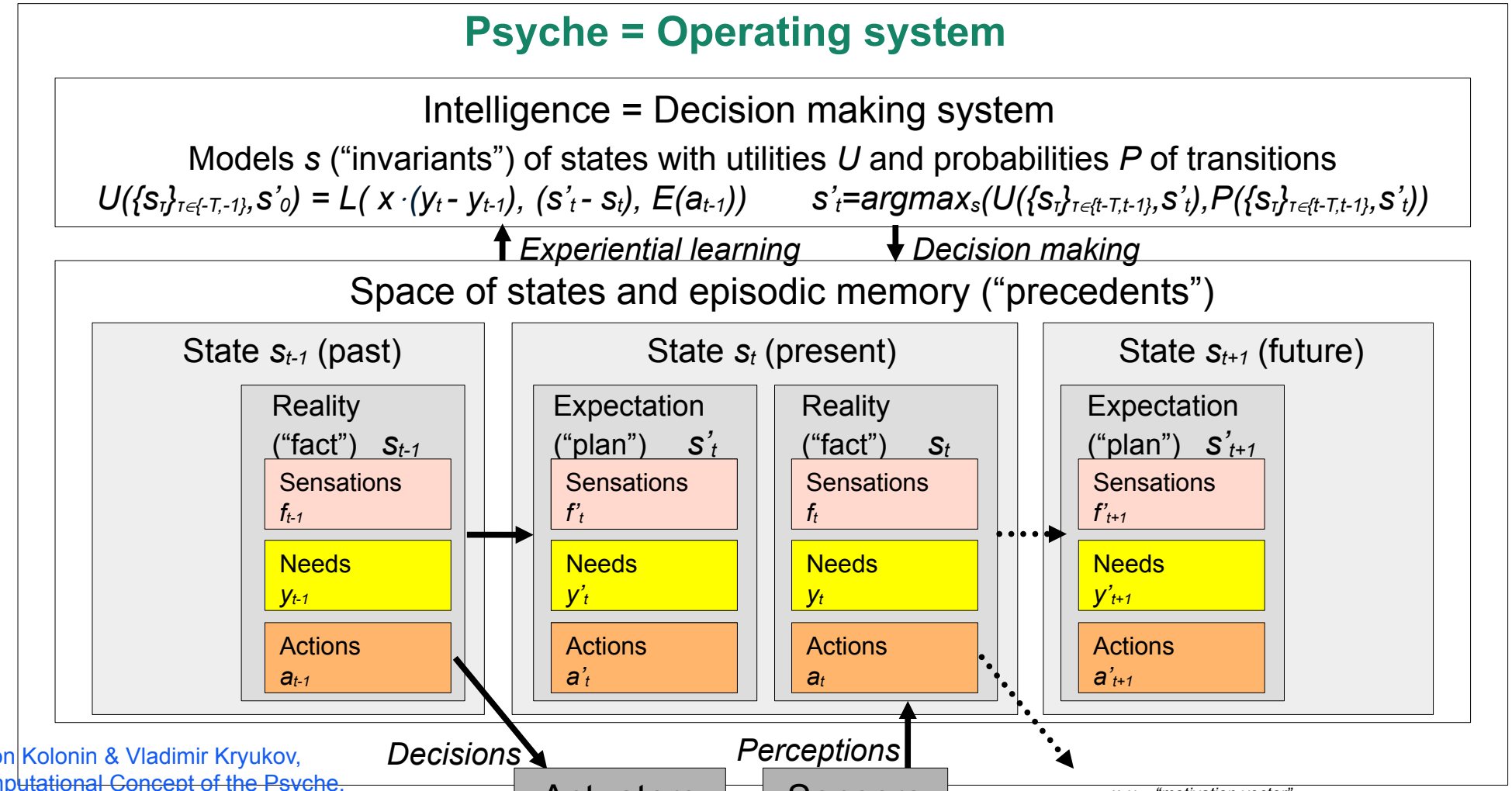



**Boolean Tensor**
(Boolean Logic)




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

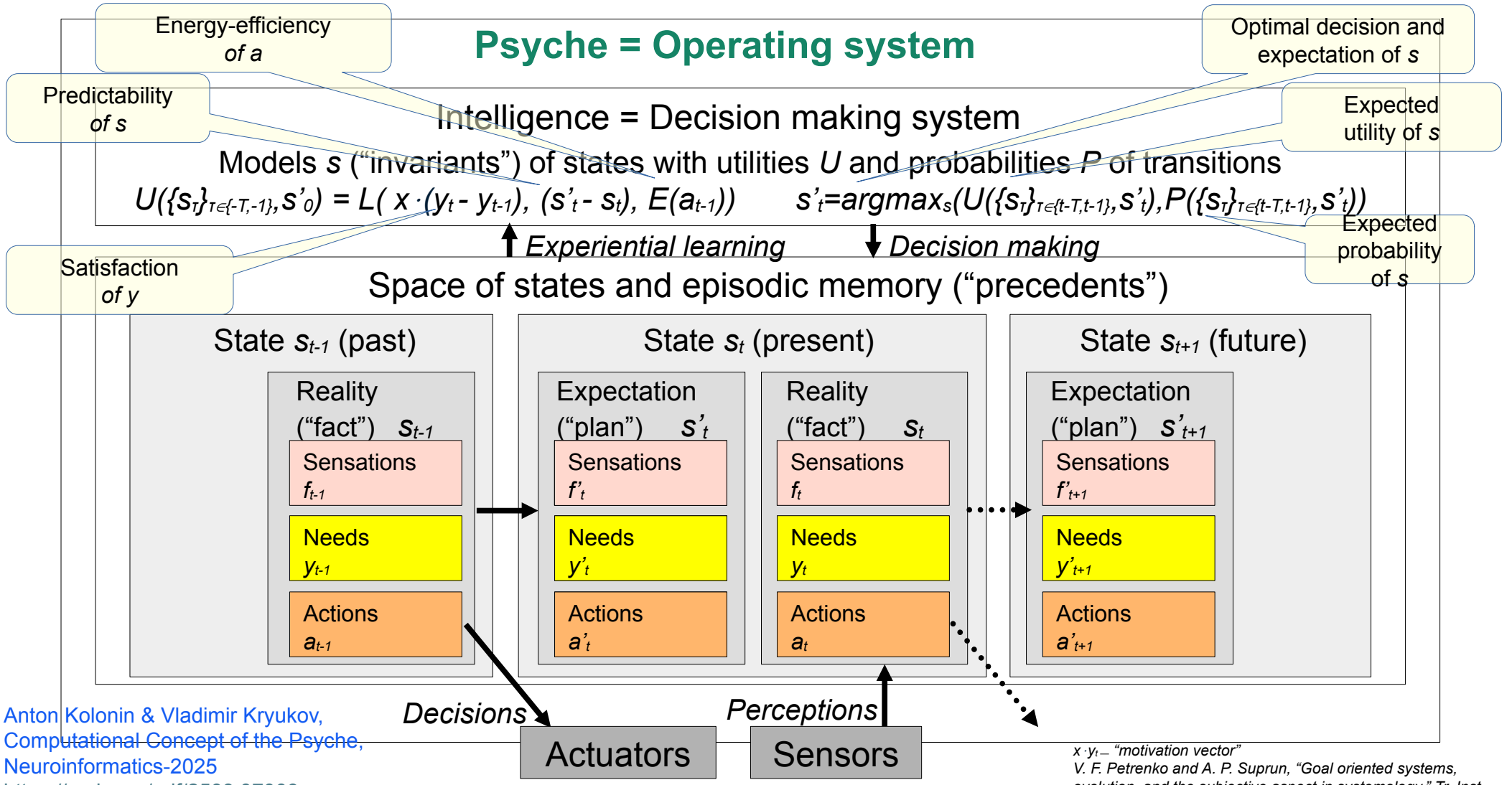
# Computational Concept and Cognitive Architecture of Artificial Psyche




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

$x \cdot y_t$ – "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)*

# Computational Concept and Cognitive Architecture of Artificial Psyche




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

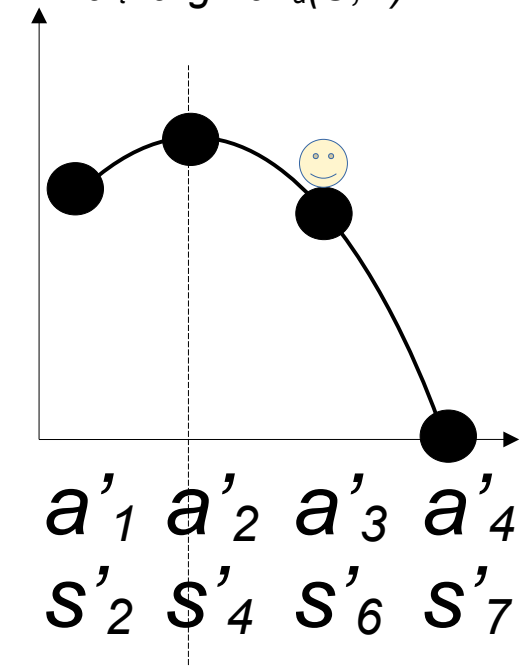
$x \cdot y_t$ – "motivation vector"
V. F. Petrenko and A. P. Suprun, "Goal oriented systems, evolution, and the subjective aspect in systemology," Tr. Inst. Sistem. Analiza RAN 62 (1) (2012)

# 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 | **0.7** |
| $s_1$ | $s'_3$ | $a'_1$ | $y'_2$ | … | 0.4 | 0.5 | |
| $s_1$ | $s'_4$ | $a'_2$ | $y'_3$ | … | 1.0 | 0.8 | **0.8** |
| $s_1$ | $s'_5$ | $a'_2$ | $y'_4$ | … | 0.0 | 0.2 | |
| $s_1$ | $s'_6$ | $a'_3$ | $y'_5$ | … | 0.6 | 1.0 | **0.6** |
| $s_1$ | $s'_7$ | $a'_4$ | $y'_6$ | … | 0.0 | 1.0 | **0.0** |

$s'_t = argmax_s(U,P)$

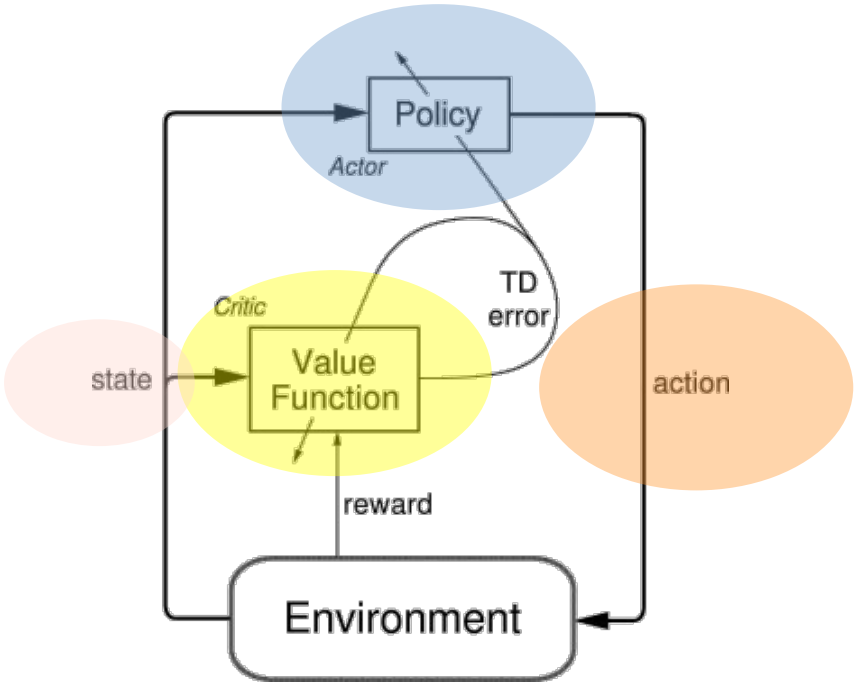
$a'_t = argmax_a(U,P)$




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

# Implementation options

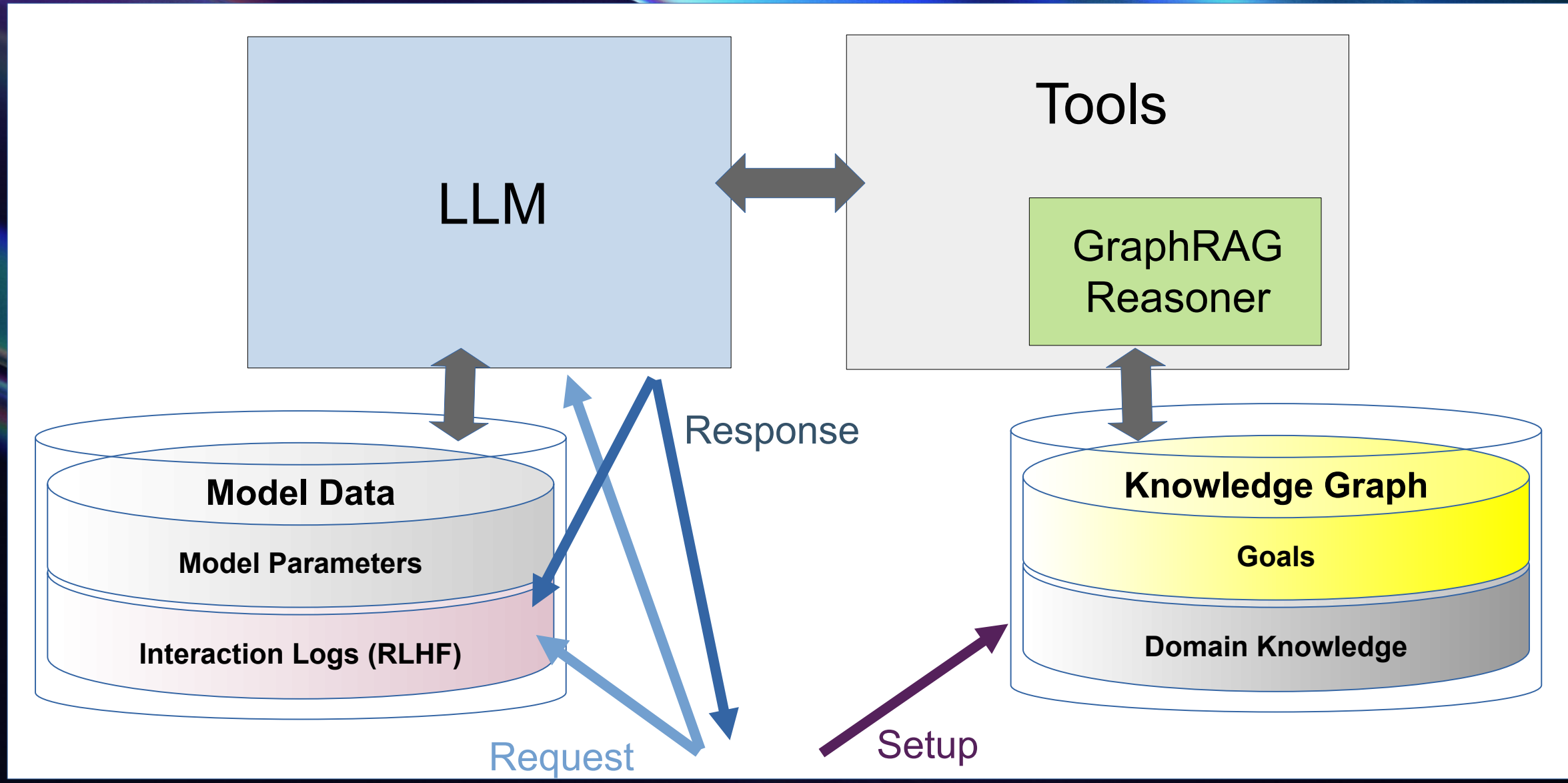
## 1) Reinforcement learning based on “actor-critic” model




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

# Implementation options

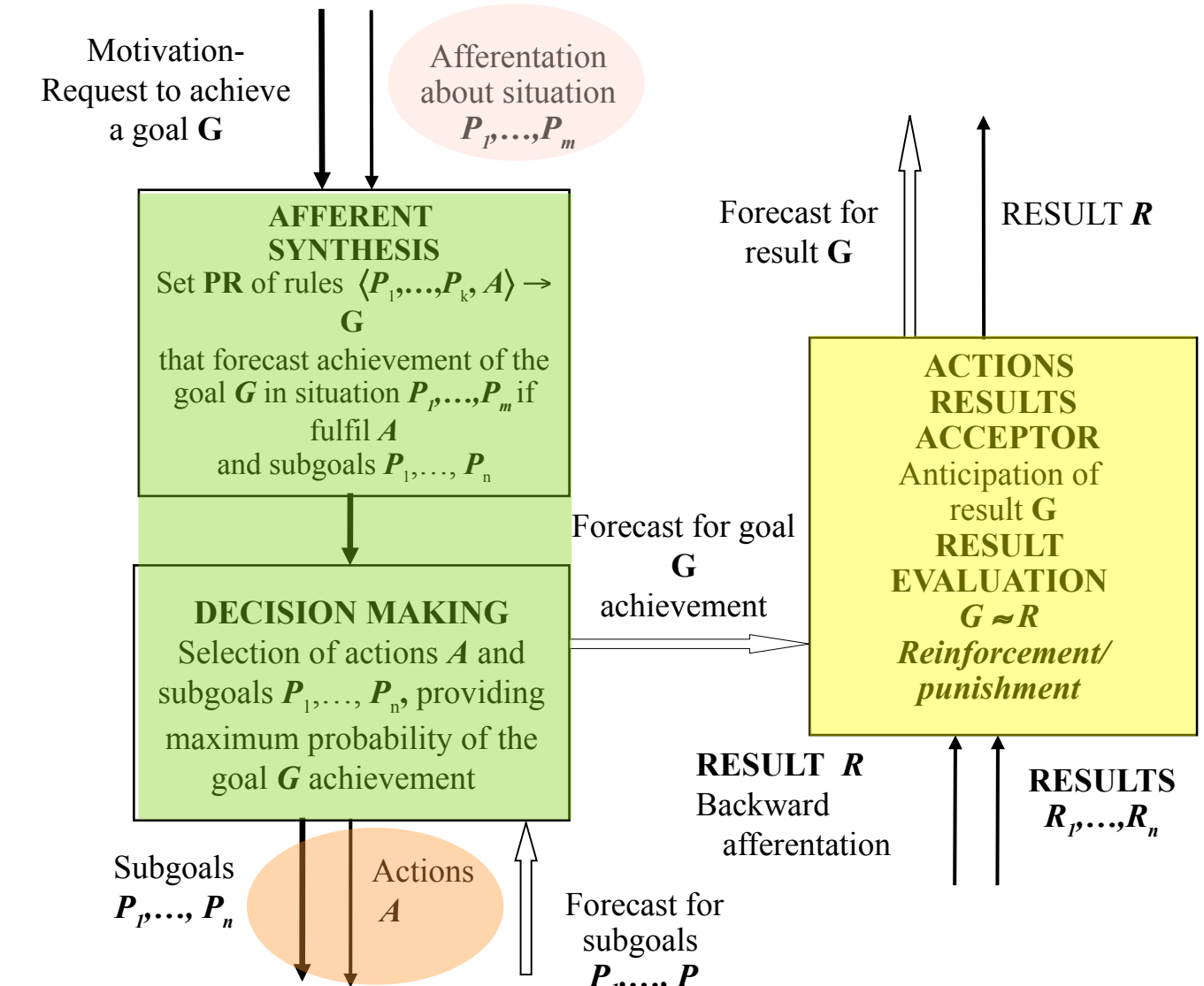
## 3) Cognitive architecture based on LLM and (active) GraphRAG




Leonie Monigatti, "The Evolution from RAG to Agentic RAG to Agent Memory", November 3, 2025
https://www.leoniemonigatti.com/blog/from-rag-to-agent-memory.html

# Implementation options

## 2) Cognitive architecture based on probabilistic logic ("Discovery")




$$Prob(G \mid P_1,\ldots,P_k,P_1,\ldots,P_n,A) = Prob(rule)\cdot Prob(P_1)\cdot\ldots\cdot Prob(P_n)$$

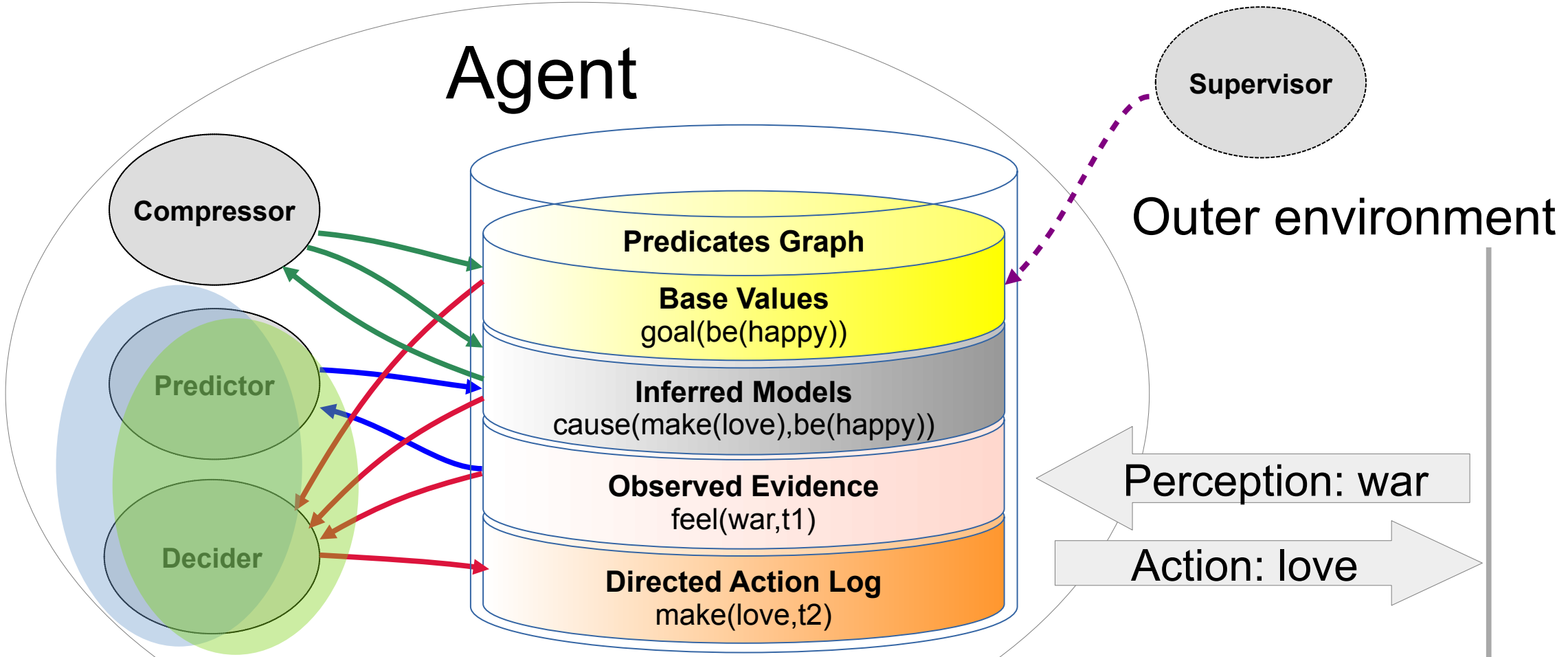
Evgenii Vityaev, Alexander Demin: Adaptive Control of Modular Robots // Conference Paper in Advances in Intelligent Systems and Computing, Conference: First International Early Research Career Enhancement School on Biologically Inspired Cognitive Architectures, Springer, August 2018

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.

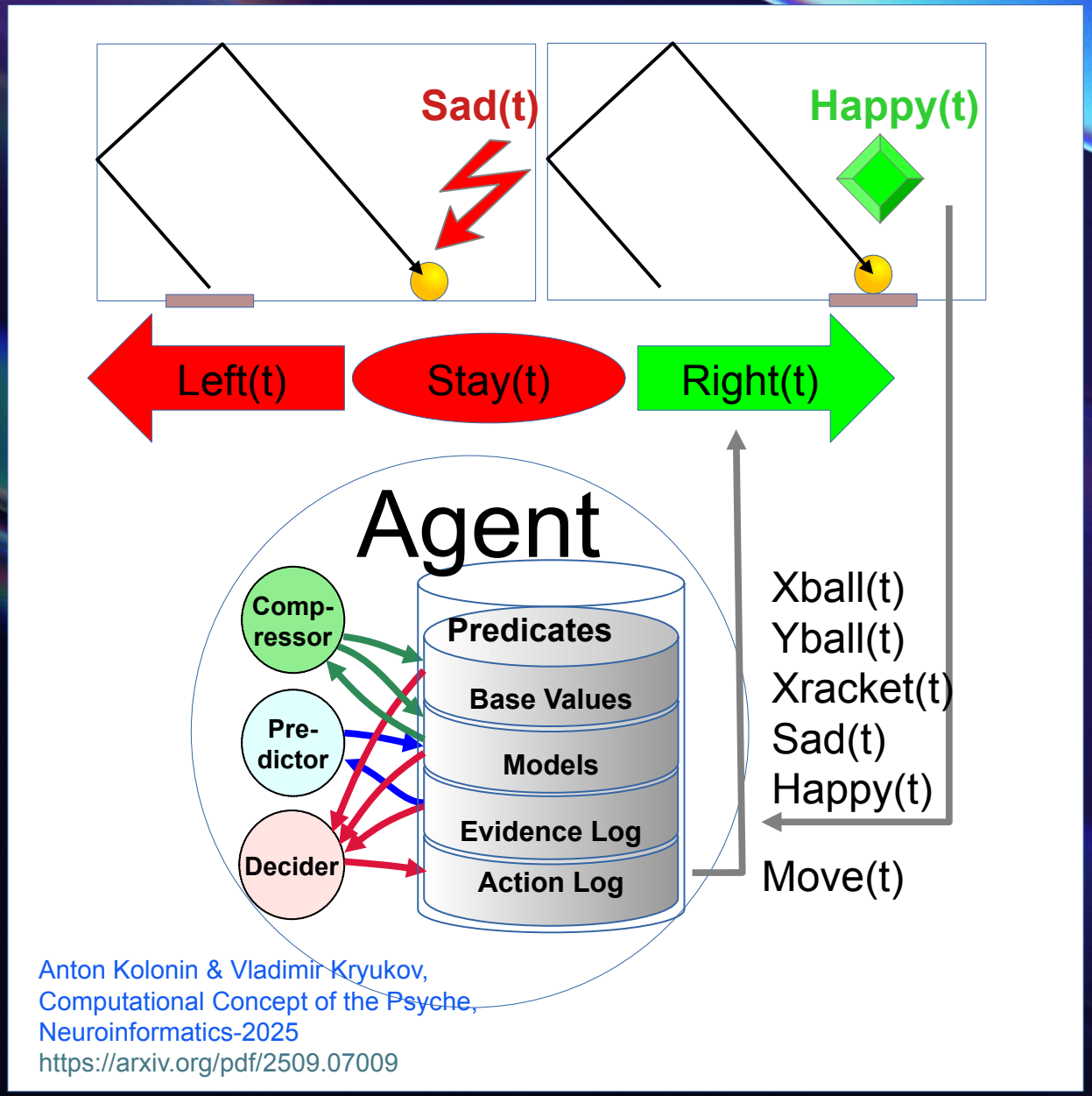
# Implementation options

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

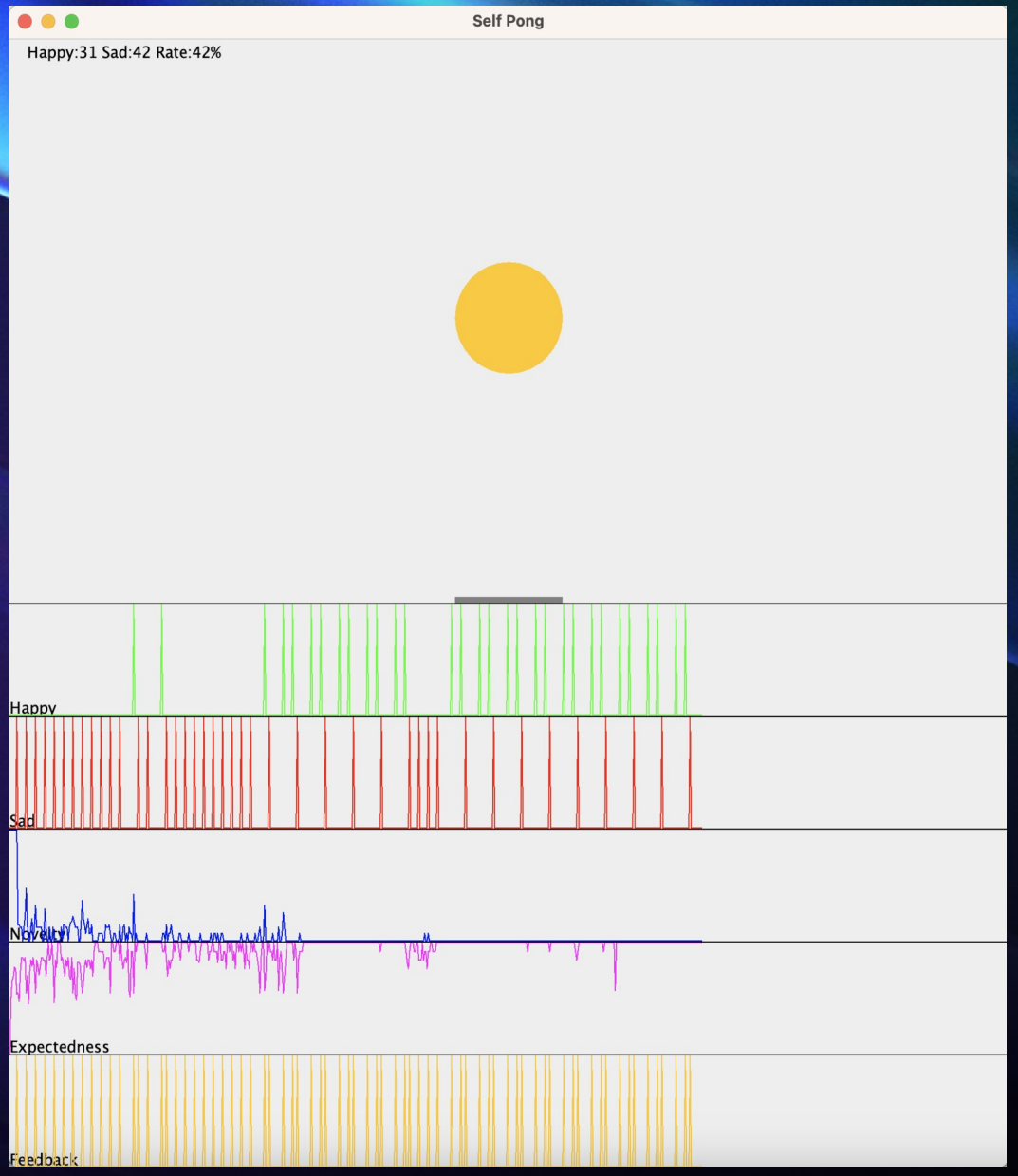



Kolonin, A. (2022). Neuro-Symbolic Architecture for Experiential Learning in Discrete and Functional Environments. In: Goertzel, B., Iklé, M., Potapov, A. (eds) Artificial General Intelligence. AGI 2021. Lecture Notes in Computer Science(), vol 13154. Springer, Cham.
https://doi.org/10.1007/978-3-030-93758-4_12

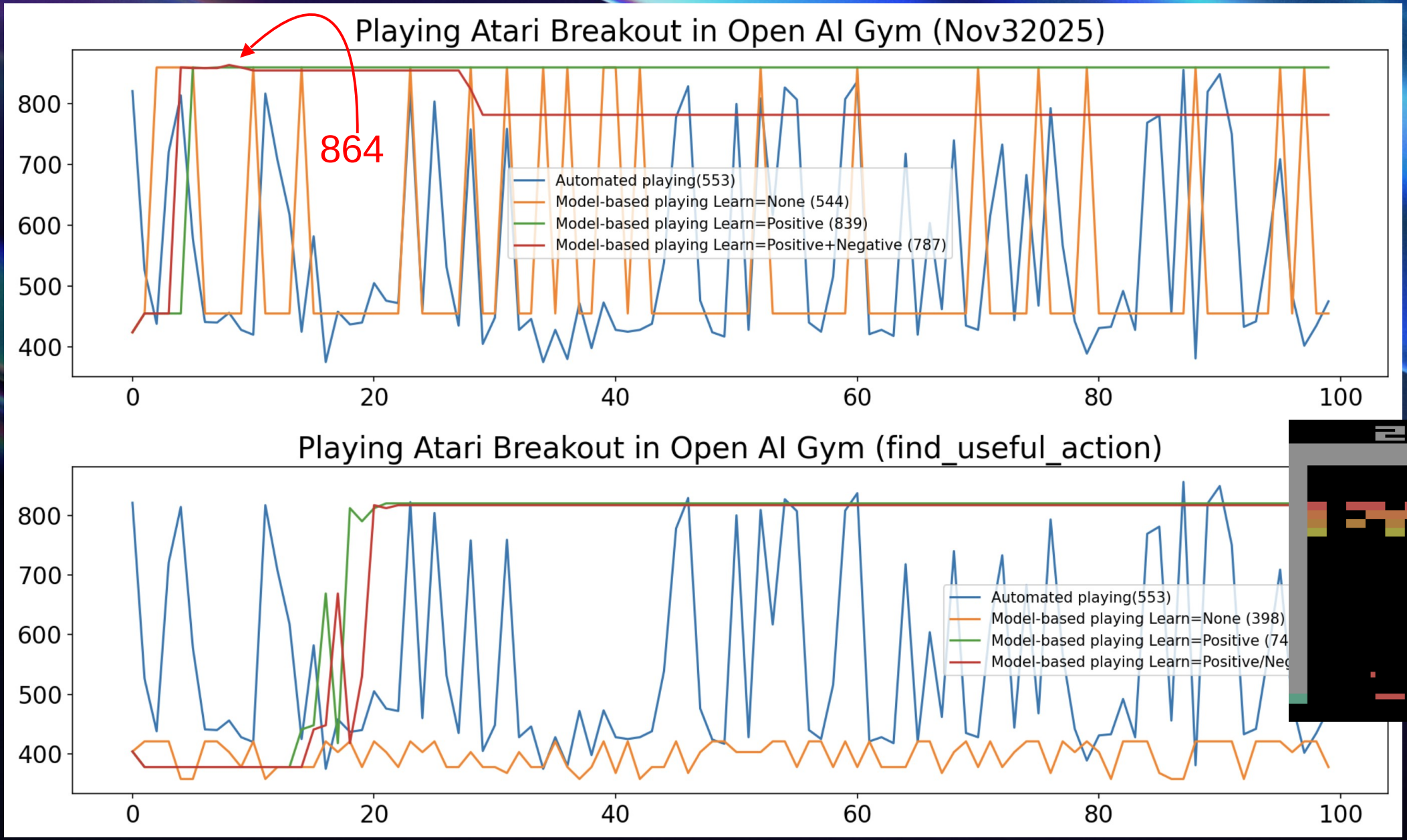
# Cognitive architecture of value-based experiential learning (“Ping-Pong”)

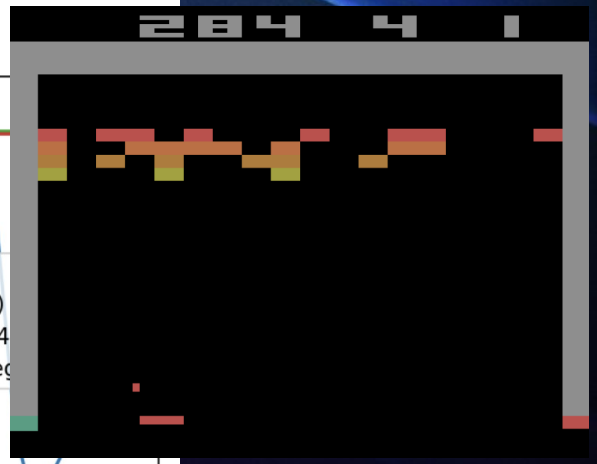



Anton Kolonin & Vladimir Kryukov,
Computational Concept of the Psyche,
Neuroinformatics-2025
https://arxiv.org/pdf/2509.07009

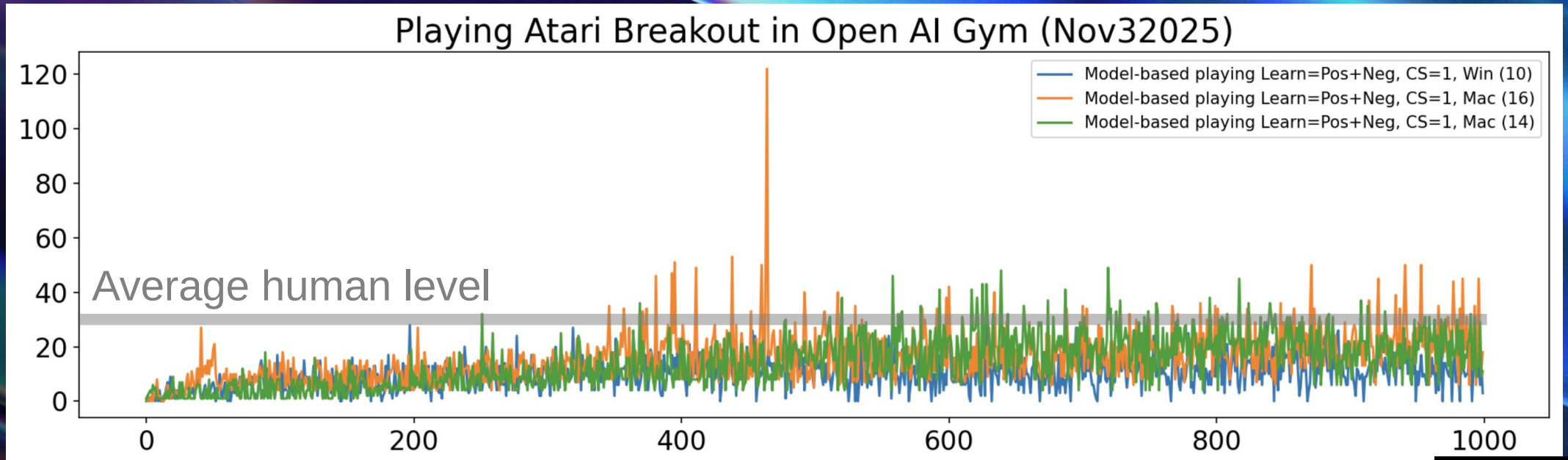


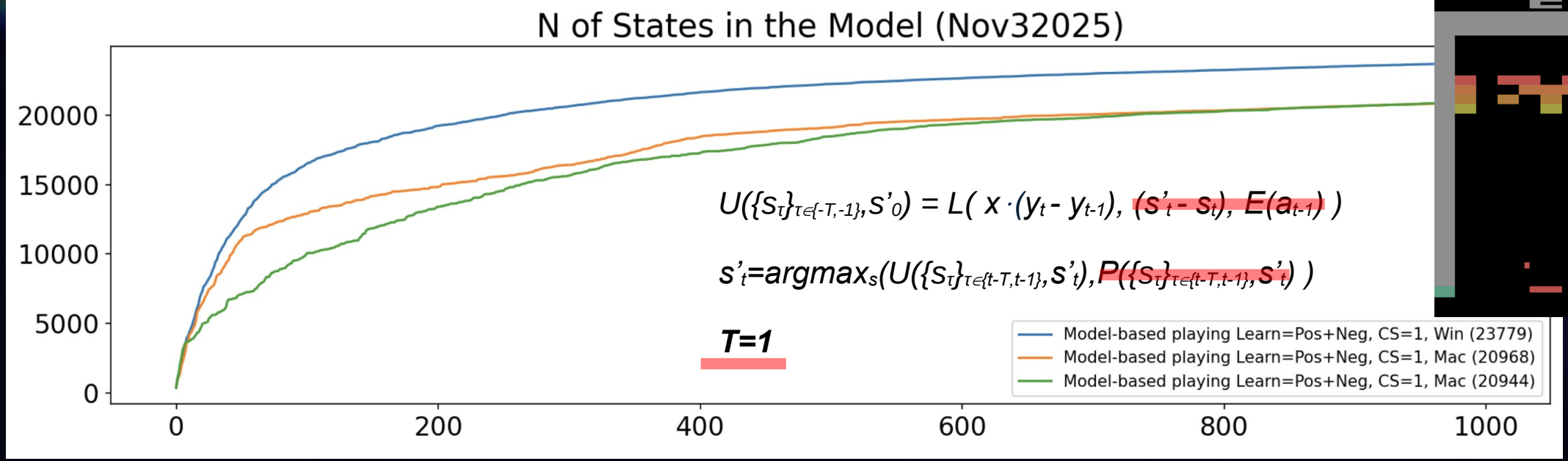
# Imitation learning – decision making on “pre-trained” model (“ Breakout”)

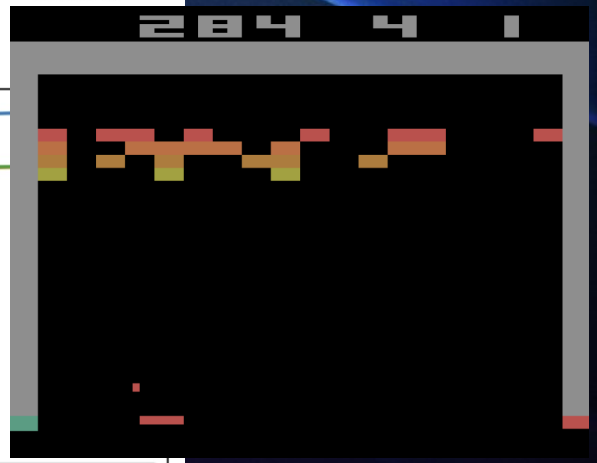


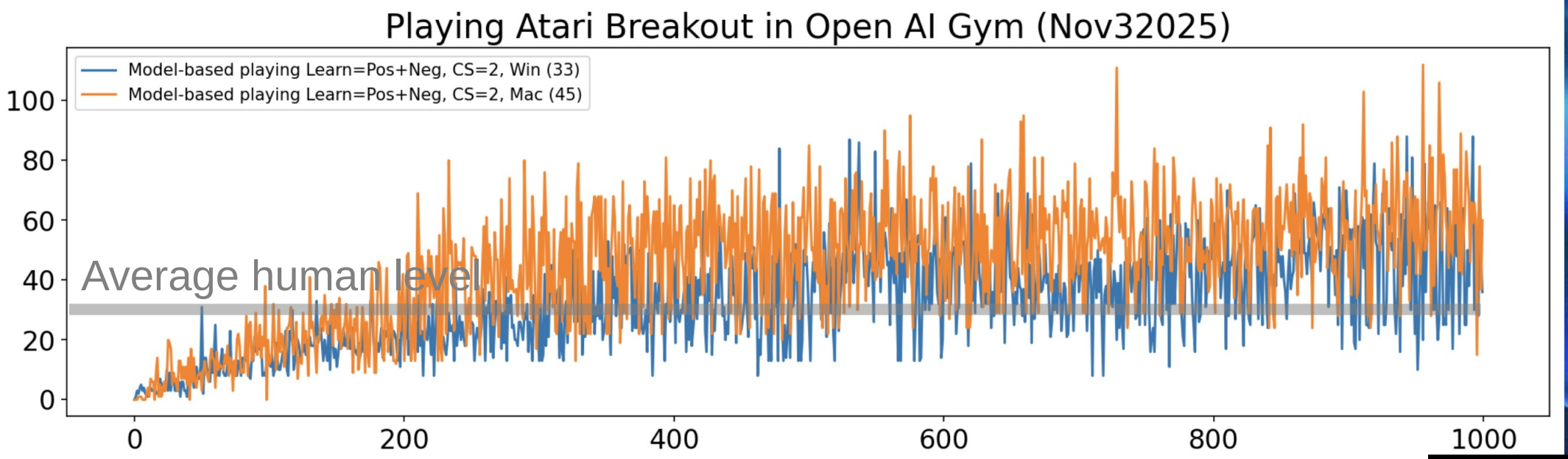
# Reinforcement learning – experiential learning & decision making (“Breakout”)

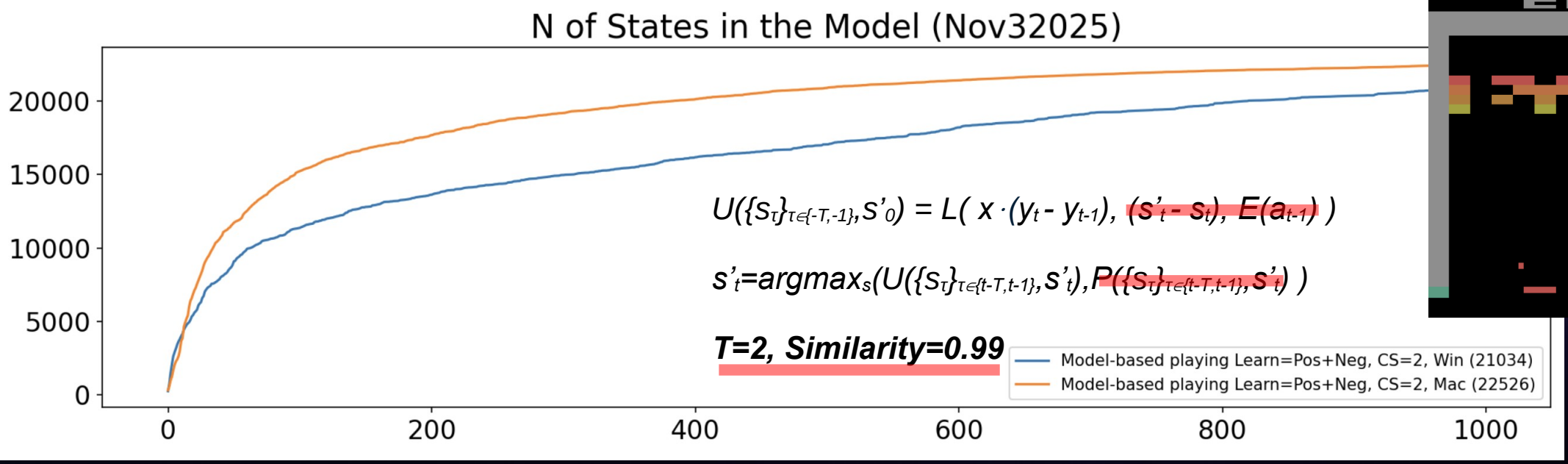


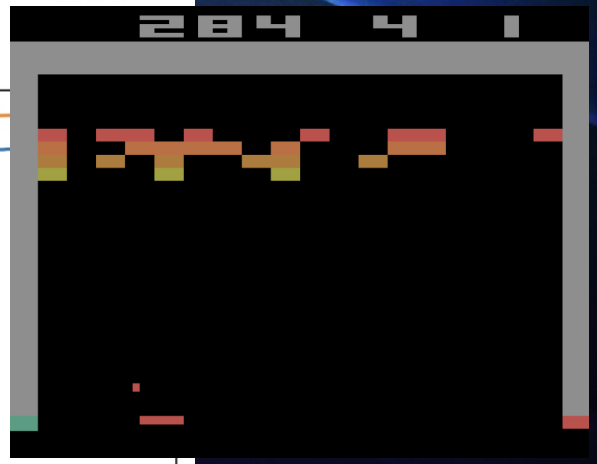



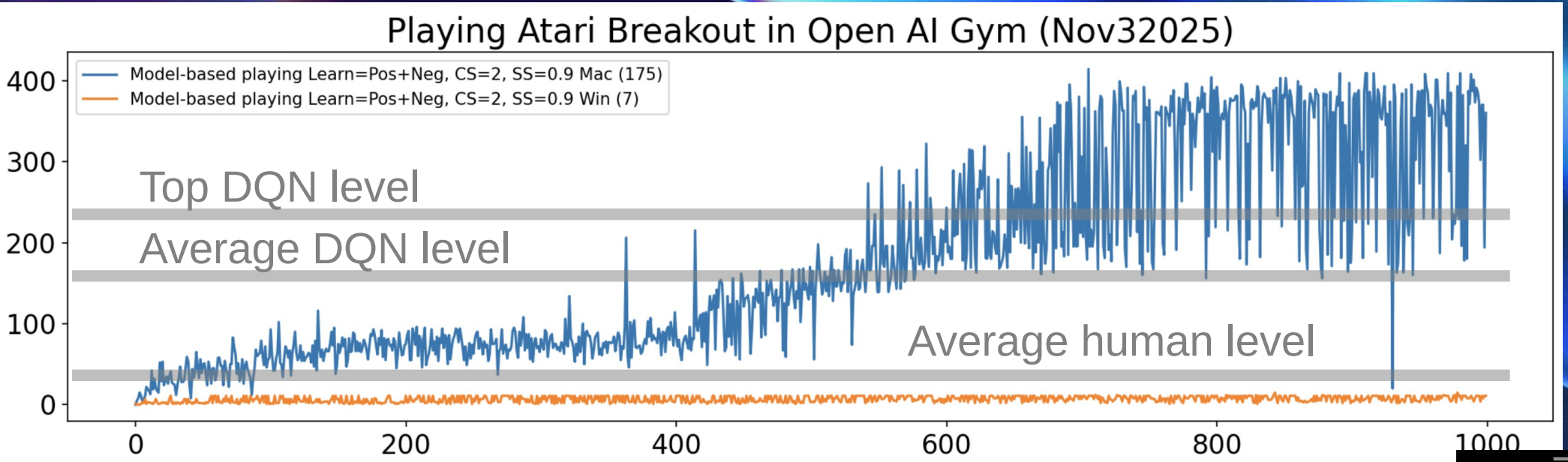
# Reinforcement learning – experiential learning & decision making (“Breakout”)

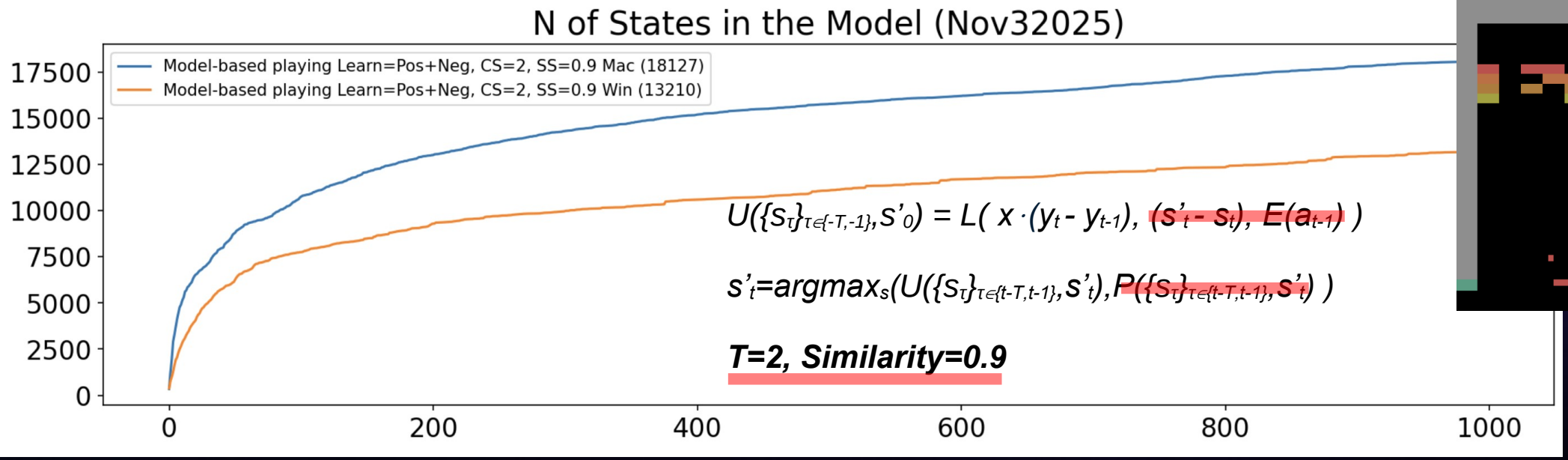


# Reinforcement learning – experiential learning & decision making (“Breakout”)




$U(\{s_\tau\}_{\tau\in\{-T,-1\}}, s'_0) = L(\ x \cdot (y_t - y_{t-1}),\ (s'_t - s_t),\ E(a_{t-1})\ )$

$s'_t = \mathrm{argmax}_s(U(\{s_\tau\}_{\tau\in\{t-T,t-1\}}, s'_t), P(\{s_\tau\}_{\tau\in\{t-T,t-1\}}, s'_t)\ )$

***T=2, Similarity=0.9***

# 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