Decision-making via evidence accumulation and metacognitive self-regulation

Baptiste Pesquet

June 23, 2026

Outline

Introduction

Motivation

- Metacognition: key skill to adapt to one's environment and goals.
- Current AI systems still limited in that area.
- Doorway to metacognition: *confidence* (subjective estimate of a decision's quality).

$\Rightarrow$ New, metacognitive models of decision-making and learning inspired by natural cognition.

Related work

- Decision-making as sequential accumulation of evidence until a threshold is crossed [1], [2], [3], [4].
- Joint models of reinforcement learning and evidence accumulation [5], [6].
- Confidence as probability of decision correctness [7], distance from the decision criterion [8] or between accumulators [9].

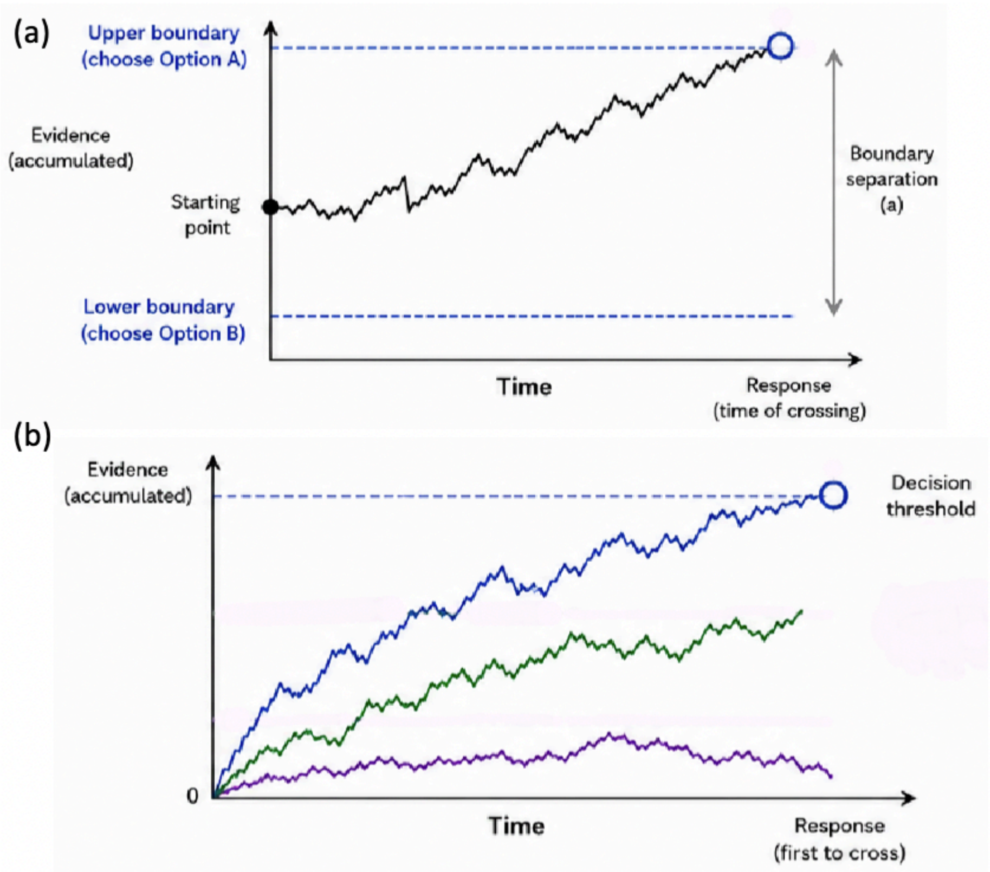

Figure 1: Difference between single- and multi-accumulator models of sequential decision-making.

Methodology

Architecture

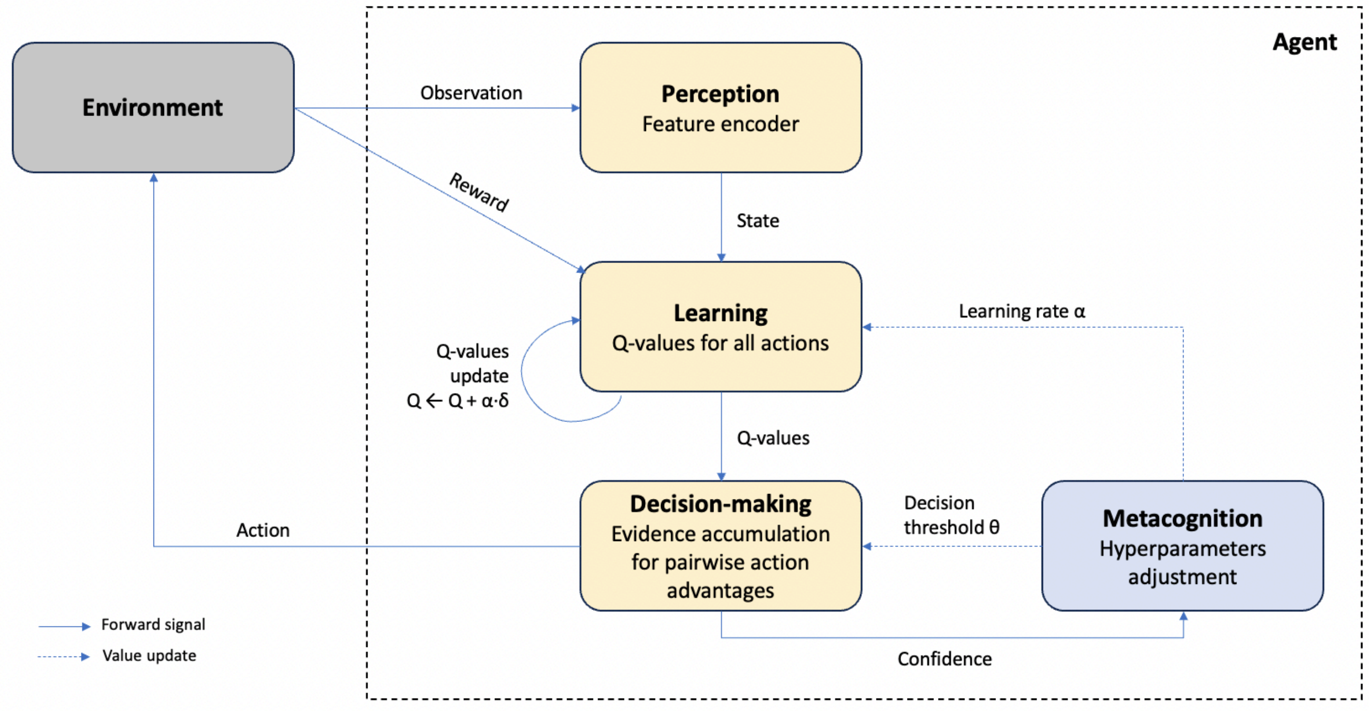

Figure 2: Proposed architecture for a metacognitive agent interacting with its environment.

Learning

Action values estimations updated via Q-Learning (Equation 1) using the reward prediction error δ (Equation 2).

$$Q(S, a) \leftarrow Q(S, a) + \alpha \cdot \delta \tag{1}$$

$$\delta = r + \gamma \cdot \max_a Q(S', a) - Q(S, a) \tag{2}$$

Architecture

Decision-making

- Race between pairwise accumulators $X_{i,j}$ for each action i over actions $j \neq i$.
- Drift rates $v_{i,j}$ (speed of evidence accumulation) defined as a weighted sums between the *advantage* (difference) and *sum* terms for action values (Equation 4).

$$\mathrm{d}X_{i,j}(t) = v_{i,j}\mathrm{d}t + \sigma\mathrm{d}W_{i,j}(t) \tag{3}$$

$$v_{i,j} = w_d\big(Q(S, a_i) - Q\big(S, a_j\big)\big) + w_s\big(Q(S, a_i) + Q\big(S, a_j\big)\big) + v_0 \tag{4}$$

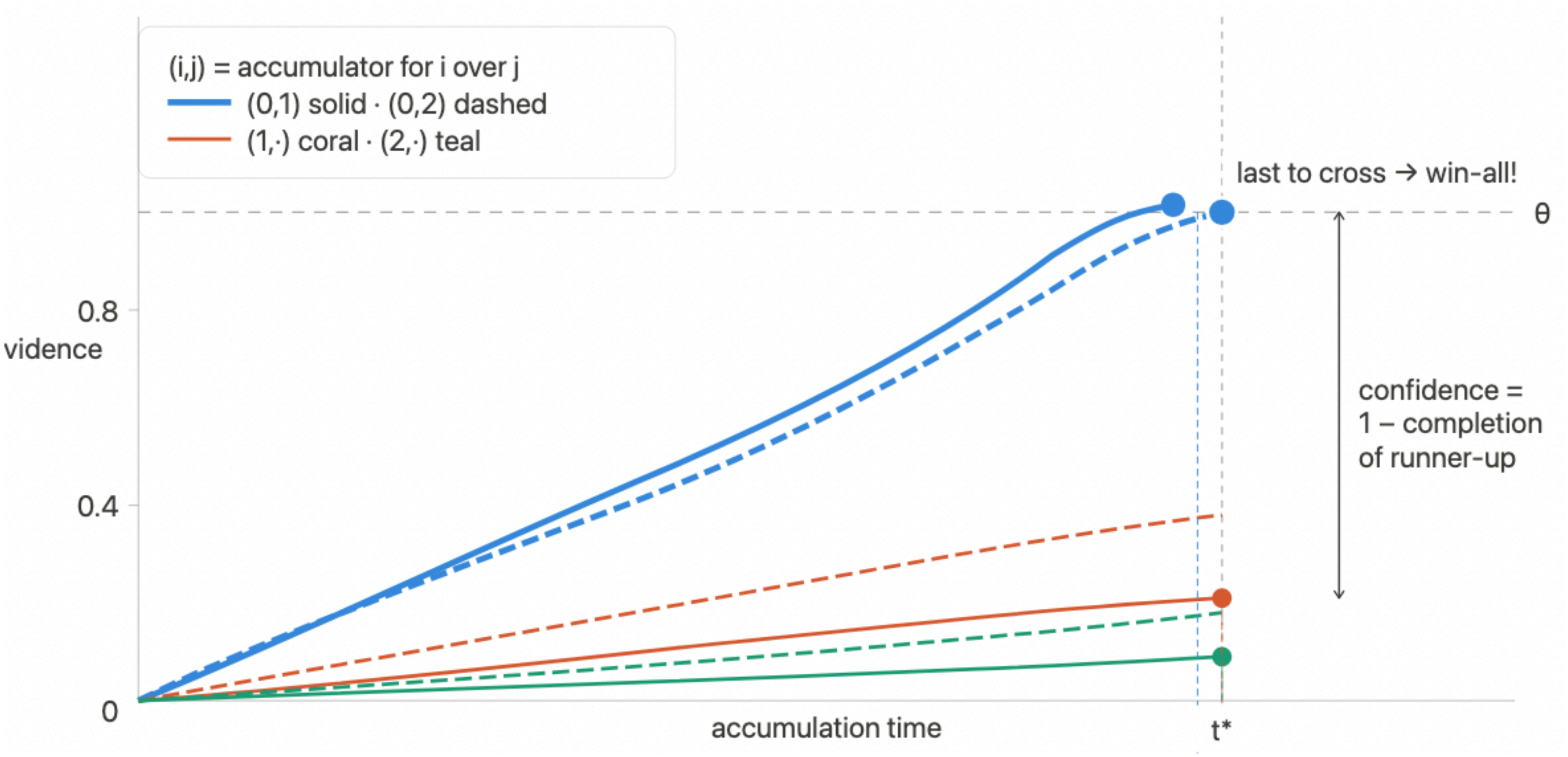

Figure 3: Illustration of the evidence accumulation dynamics for decision-making. Each of the three possible actions is associated with two accumulators representing the pairwise advantages over other actions. The confidence is computed as the normalised distance between winner's and runner-up's slowest accumulators at decision time.

Architecture

Metacognition

Exploits the confidence c computed at decision time to dynamically adjust the decision threshold θ (Equation 5) and the Q-value learning rate α (Equation 6).

$$\theta \leftarrow \theta - \eta_{\theta}\left(c - c_{\text{target}}\right) \tag{5}$$

$$\alpha \leftarrow \alpha - \eta_{\alpha}\left(c - c_{\text{target}}\right) \tag{6}$$

Test environment

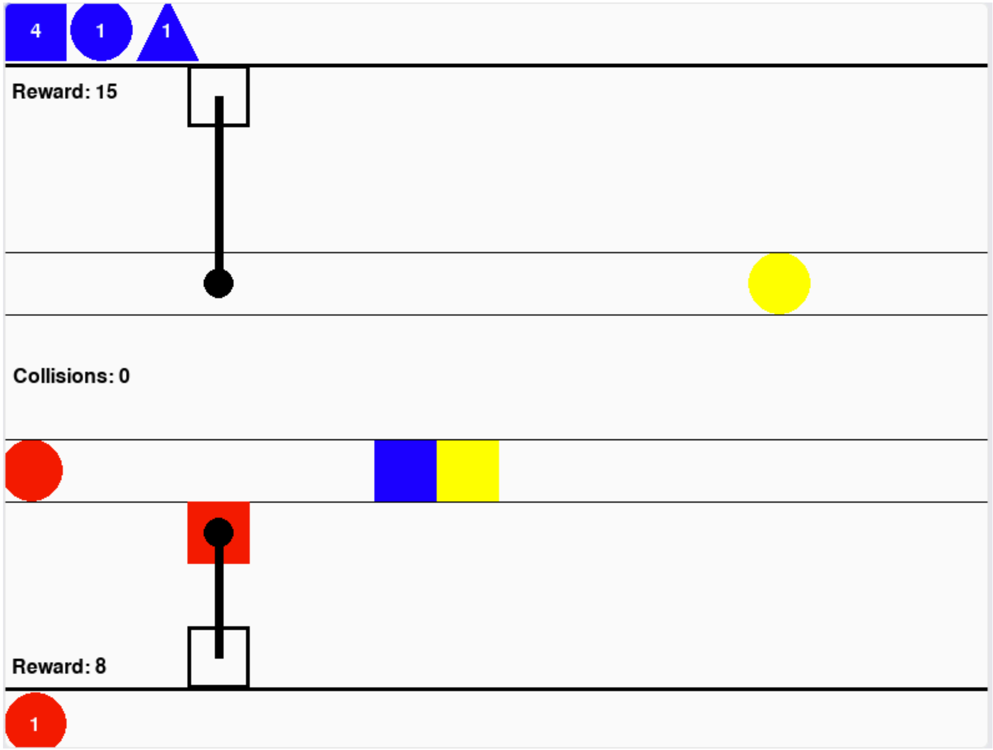

Figure 4: The collaborative object sorting task between a deterministic (top) and a cognitive (bottom) agent.

Preliminary results

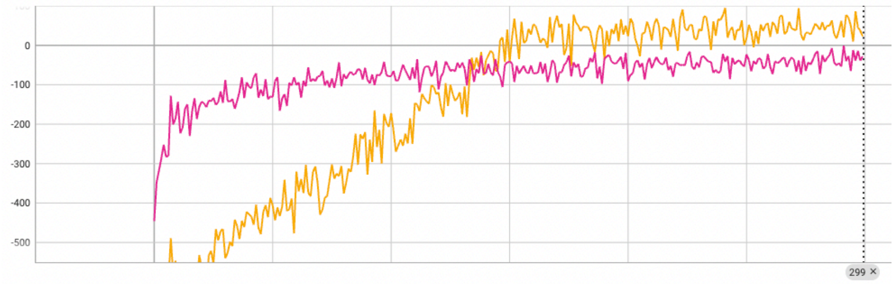

Figure 5: Cumulated rewards at the end of training for the metacognitive (magenta) and Q-Learning agents.

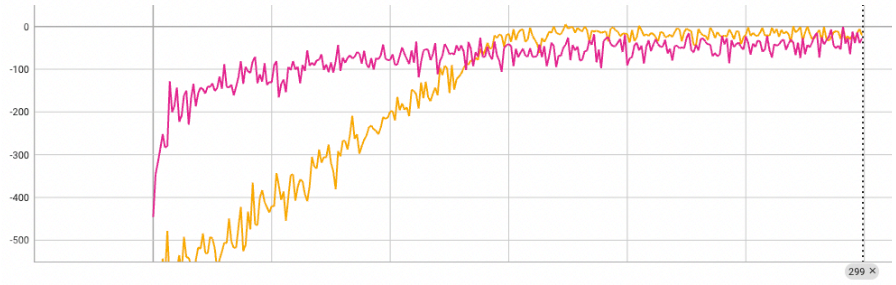

Figure 6: Cumulated rewards at the end of training for the metacognitive (magenta) and DQN agents.

Conclusion

Discussion

- Work at the crossroads of AI and cognitive computational neuroscience.
- Model for a metacognitive artificial agent using its own confidence levels to dynamically adjust its decision- making and learning functions.
- Dedicated testbed: collaborative object sorting task.
- Much work needed to refine architecture and test environment.

References

[1] R. Ratcliff and G. McKoon, "The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks," *Neural Computation*, vol. 20, no. 4, pp. 873–922, Apr. 2008, doi: 10.1162/neco.2008.12-06-420.

[2] M. Usher and J. McClelland, "The Time Course of Perceptual Choice: The Leaky, Competing Accumulator Model," 2001.

[3] S. D. Brown and A. Heathcote, "The Simplest Complete Model of Choice Response Time: Linear Ballistic Accumulation," *Cognitive Psychology*, vol. 57, no. 3, pp. 153–178, Nov. 2008, doi: 10.1016/j.cogpsych.2007.12.002.

[4] D. Van Ravenzwaaij, S. D. Brown, A. A. J. Marley, and A. Heathcote, "Accumulating Advantages: A New Conceptualization of Rapid Multiple Choice.," *Psychological Review*, vol. 127, no. 2, pp. 186–215, Mar. 2020, doi: 10.1037/rev0000166.

[5] L. Fontanesi, S. Gluth, M. S. Spektor, and J. Rieskamp, "A Reinforcement Learning Diffusion Decision Model for Value-Based Decisions," *Psychonomic Bulletin & Review*, vol. 26, no. 4, pp. 1099–1121, Aug. 2019, doi: 10.3758/s13423-018-1554-2.

[6] S. Miletić, R. J. Boag, A. C. Trutti, N. Stevenson, B. U. Forstmann, and A. Heathcote, "A New Model of Decision Processing in Instrumental Learning Tasks," *eLife*, vol. 10, p. e63055, Jan. 2021, doi: 10.7554/eLife.63055.

[7] A. Pouget, J. Drugowitsch, and A. Kepecs, "Confidence and Certainty: Distinct Probabilistic Quantities for Different Goals," *Nature Neuroscience*, vol. 19, no. 3, pp. 366–374, Mar. 2016, doi: 10.1038/nn.4240.

[8] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics.* in Signal Detection Theory and Psychophysics. Oxford, England: John Wiley, 1966, p. xi, 455.

[9] A. Kepecs and Z. F. Mainen, "A Computational Framework for the Study of Confidence in Humans and Animals," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1594, pp. 1322–1337, May 2012, doi: 10.1098/rstb.2012.0037.