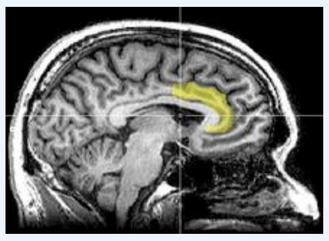
Representations for Cognitive Control in Anterior Cingulate Cortex

Thomas R. Colin (thomas.colin@ugent.be), Clay B. Holroyd

A *self-prediction* role for *anterior cingulate cortex?*

1. Anterior Cingulate Cortex (ACC)



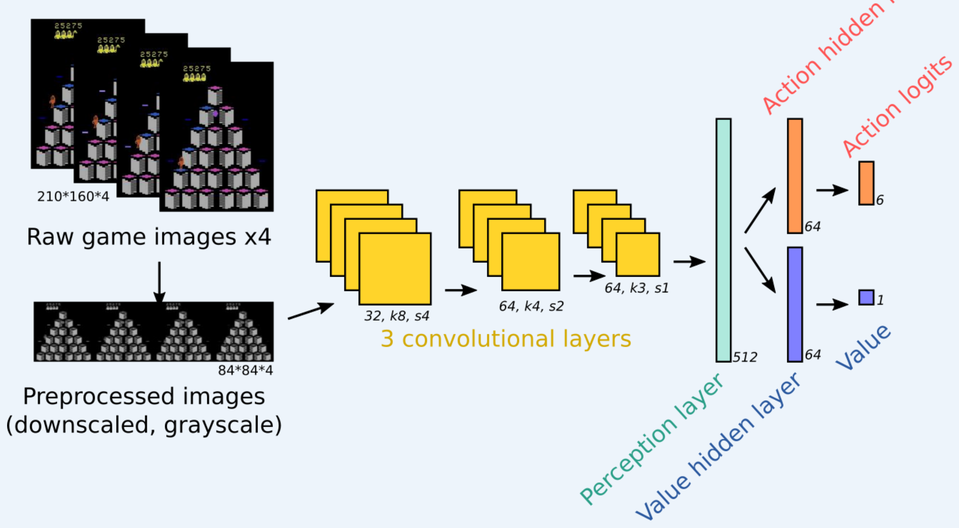
The ACC is involved in "everything" but its exact function remains unclear:

- Cost of control [1]
- World model [2]
- Hierarchical RL [3, 4]

In this ***modeling study*** we investigate ACC function; this poster presents preliminary results..

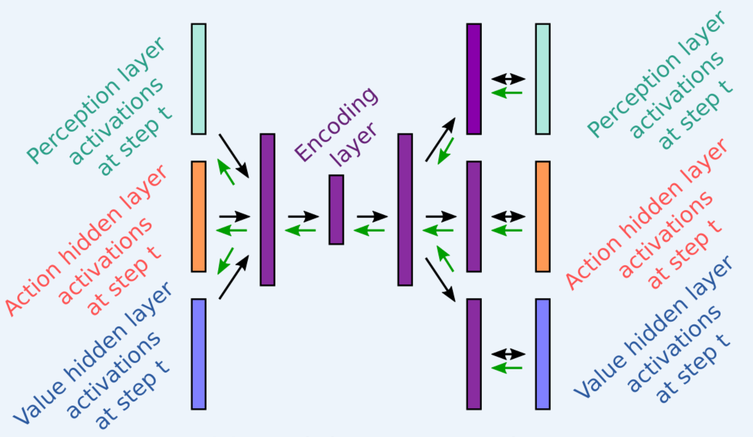
2. Base Model

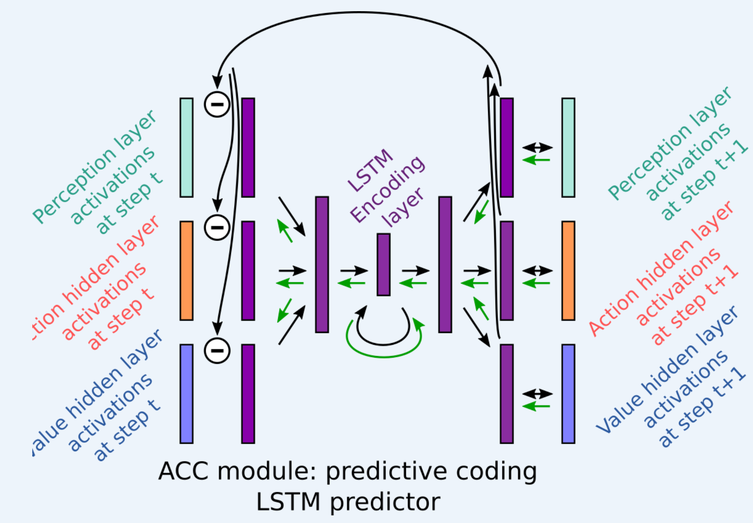
ACC lesions result in only mild impairments; the role ACC is most likely supportive. To account for this, we use a base-model (an actor-critic network trained on the Atari game Q*bert – see box *Q* bert*) to simulate habitual behavior.



3. ACC Modules

We train multiple ACC modules to monitor the base model's state. LSTM and predictive coding modules (see box *Predictive Coding*) capture temporally extended behavior, accounting for the uniquely long temporal signatures of ACC neurons [5].



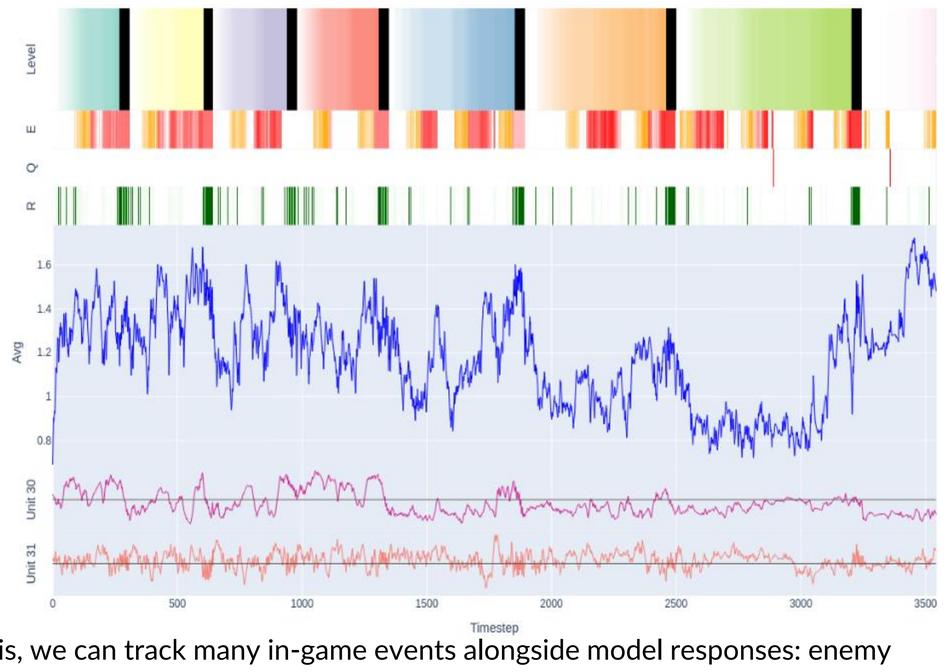


ACC module architectures (cartoon). Two simplified versions of ACC models are shown above; in total we are currently exploring five modules (Auto-encoder, one-step predictor, LSTM predictor, predictive coding LSTM predictor, and a controller module). All modules are abstracting or predicting base model activations. This setup enables exploration of a wide space of possible models.

4. Analysis and preliminary findings

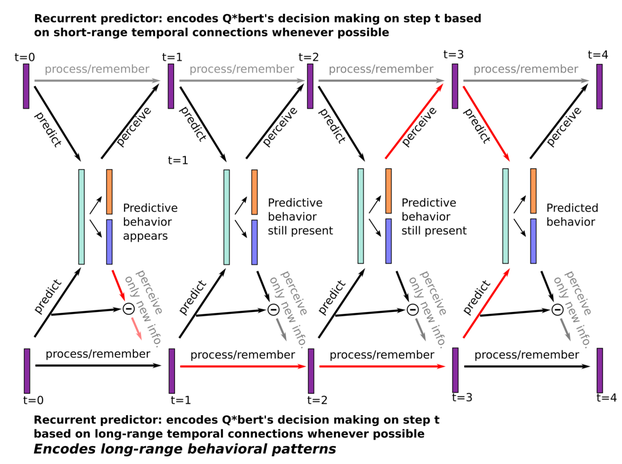
All modules could learn to predict the base model's activations with sufficient accuracy (all achieving an average error < 0.3 when substituting value activation predictions). To investigate the properties of the different models, we conducted RSA (see box *Representational Similarity Analysis*). **Temporally extended structure can be discovered from prediction of a reflex-based agent acting in a Markov problem. This may be a key role of ACC relating to cognitive control.**

Q*bert



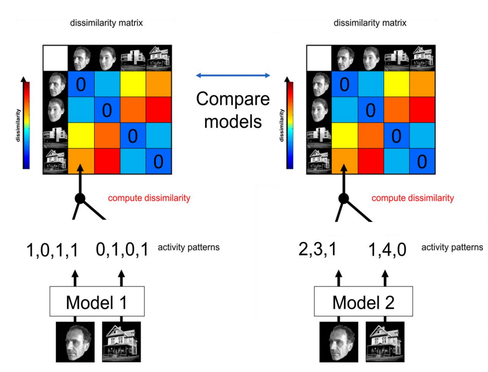
Using simple image analysis, we can track many in-game events alongside model responses: enemy appearance and position, level completion, TD-errors, etc. The plot on the right shows one episode.

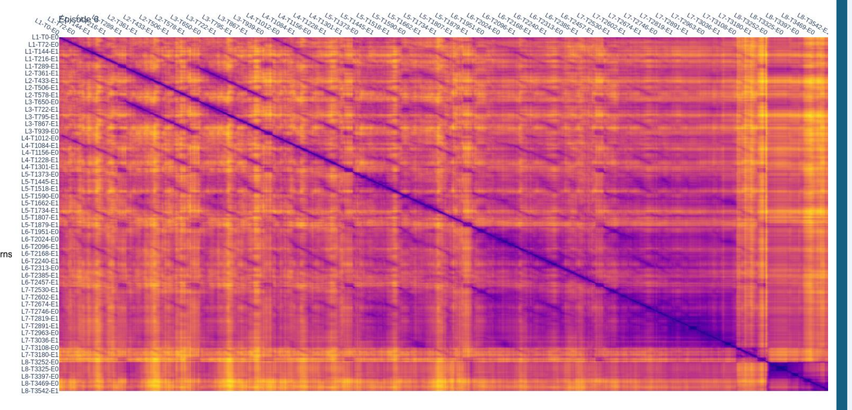
Predictive Coding



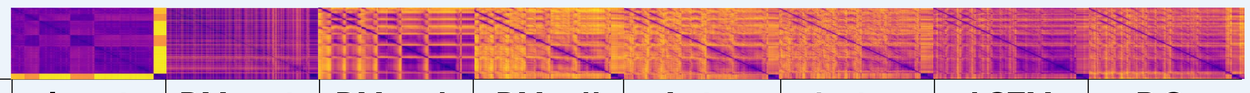
Predictive coding (left cartoon) is used to force the model to encode long-range temporal dependencies in its recurrent weights. It's not yet clear whether this worked (right figure).

Representational Similarity Analysis





Left: cartoon explanation of RSA, adapted from [6]. Right, representational dissimilarity matrix for one ACC module (predictive coding) over one episode (~3500 steps).



	Input image	BM: act. h	BM: val. h	BM: all	Auto-encoder	1-step pred.	LSTM pred.	P.C. LSTM
Level	0.44	0.04	0.30	0.34	0.32	0.32	0.31	0.29
Level %	0.18	0.04	0.18	0.22	0.25	0.18	0.19	0.24
Reward	0.00	0.16	0.21	0.24	0.19	0.09	0.16	0.14
TD-error	0.10	0.29	0.23	0.29	0.24	0.12	0.16	0.22
Value	0.03	0.09	0.67	0.54	0.44	0.34	0.37	0.29
Action	0.01	0.04	0.01	0.02	0.02	0.02	0.01	0.01
Egg?	0.01	0.00	-0.03	-0.02	-0.02	0.01	0.00	0.03
Snake?	0.01	0.00	0.10	0.11	0.08	0.08	0.10	0.10
Snake loc	0.16	0.04	0.07	0.09	0.03	0.08	0.04	0.12
Q*bert loc	0.11	-0.03	0.05	0.13	0.19	0.25	0.18	0.28

How do these models behave? Spearman correlation coefficients between model RDMs (columns) (BM stands for base model) and RDMs of various events and quantities (rows). Each RDMs is visualized next to the column and row headings.

References

[1] Shenhav, A., Cohen, J. & Botvinick, M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nat Neurosci* **19**, 1286–1291. https://doi.org/10.1038/nn.4384
[2] Akam, T., Rodrigues-Vaz, I., Marcelo, I., Zhang, X., Pereira, M., Oliveira, R. F., Dayan, P., & Costa, R. M. (2021). The Anterior Cingulate Cortex Predicts Future States to Mediate Model-Based Action Selection. *Neuron*, *109*(1), 149–163.e7. https://doi.org/10.1016/j.neuron.2020.10.013
[3] Holroyd, C. B., & Verguts, T. (2021). The best laid plans: computational principles of anterior cingulate cortex. *Trends in Cognitive Sciences*, *25*(4), 316-329, https://doi.org/10.1016/j.tics.2021.01.008.
[4] Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in cognitive sciences*, *16*(2), 122-128, https://doi.org/10.1016/j.tics.2011.12.008 .
[5] Emmanuel Procyk, Vincent Fontanier, Matthieu Sarazin, Bruno Delord, Clément Goussi, Charles R.E. Wilson (2021), The midcingulate cortex and temporal integration, *International Review of Neurobiology*, 158, 395-419, https://doi.org/10.1016/bs.irn.2020.12.004.
[6] Nikolaus Kriegeskorte, Marieke Mur, and Peter Bandettini (2008), Representational similarity analysis – connecting the branches of systems neuroscience, *Frontiers in Systems Neuroscience*, 2, https://doi.org/10.3389/neuro.06.004.2008
[7] Colin, T. R., Ikink, I., & Holroyd, C. B. (2025). Distributed representations for cognitive control in frontal medial cortex. *Journal of Cognitive Neuroscience*, *37*(5), 941-969, https://doi.org/10.1162/jocn_a_02285