Towards metacognitive agents: integrating confidence in sequential decision-making

Baptiste Pesquet

Confidence in natural cognition

Decision-making as a sequential process

- A **decision** is a deliberative process leading to a **choice**.
- Decision-makers need **time** to collect and process informative cues.
- Decision-making is often modeled as an **accumulation-to-threshold** process [Gold and Shadlen, 2007].
- The balance between response time and accuracy (when available) is called the **Speed/Accuracy Trade-off** [Heitz, 2014].

Models of sequential decision-making

For binary choices, a popular model is the **Diffusion Decision Model** [Ratcliff and McKoon, 2008].

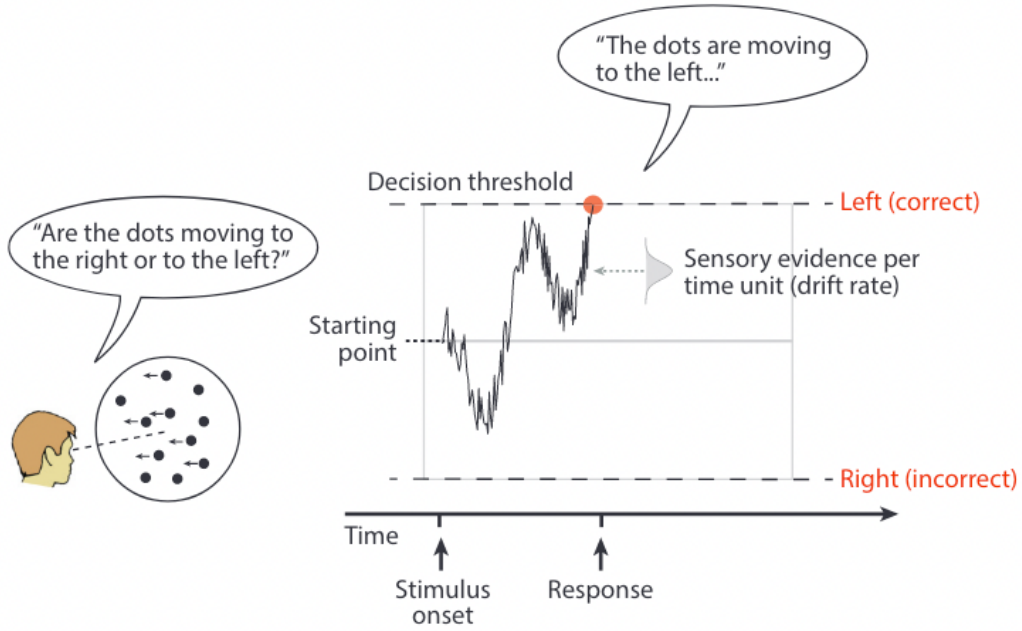


Image credits: [Forstmann et al., 2016]

Multi-alternative decisions are often modeled as a **race** between accumulators for each possible choice.



Image credits: SequentialSamplingModels.jl

Confidence in decision-making

- **Uncertainty** is inherent to all stages of neural computation [Fleming, 2024].
  - It refers to probabilistic representations of information in the brain.
- **Confidence** quantifies the degree of **certainty** associated with a decision.
  - It refers to scalar values derived from those distributions [Meyniel et al., 2015].
- More formally, confidence can be defined as the **probability** that a choice is correct given the evidence [Pouget et al., 2016].

Computing confidence in sequential decision-making models

In **decisional focus models**, confidence is directly indexed by the state of evidence at the time of choice.

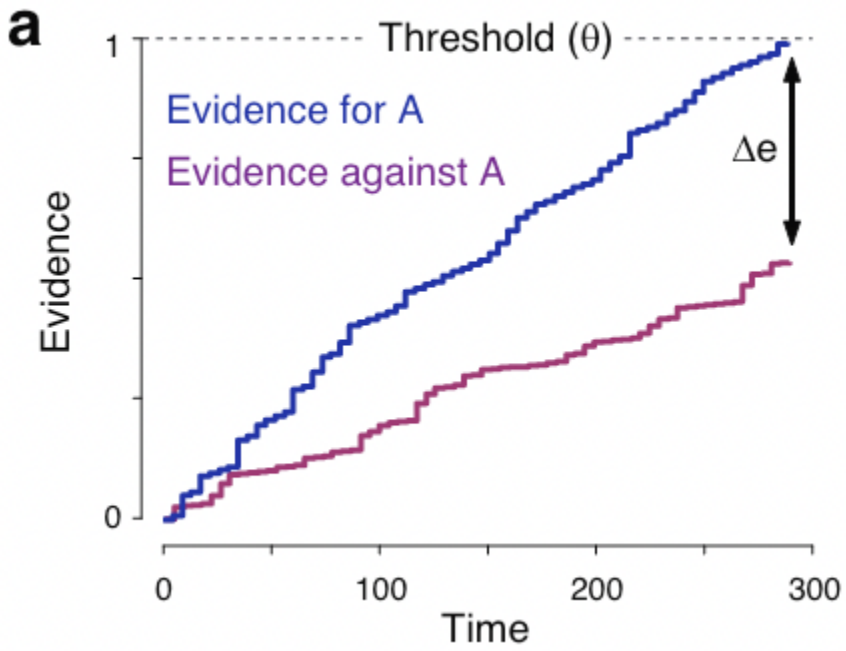


Image credits: [Kepecs et al., 2008]

Post-decisional focus models posit that evidence accumulation goes on after decision time to account for confidence.

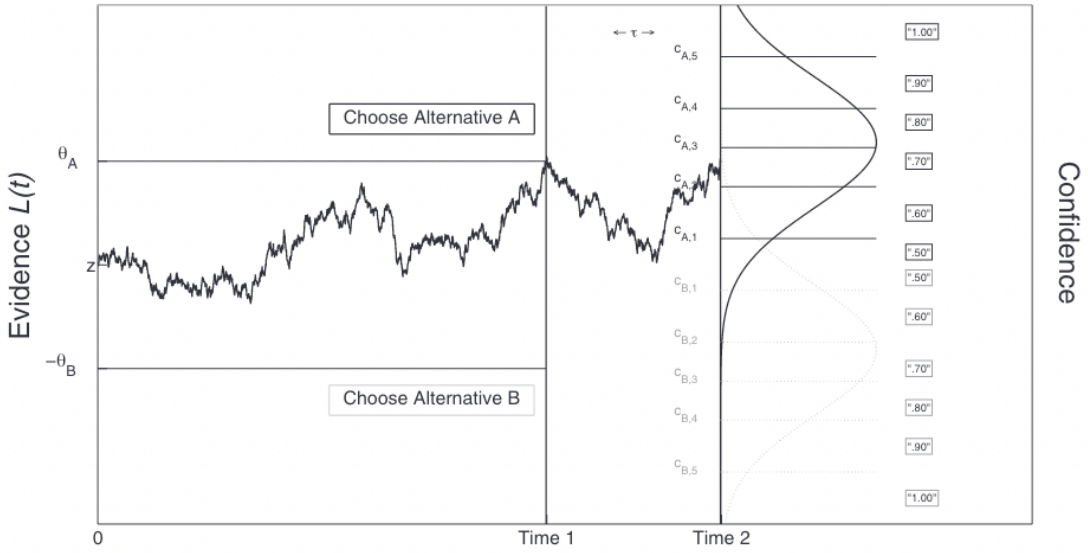


Image credits: [Pleskac and Busemeyer, 2010]

Confidence as a doorway to metacognition

- **Metacognition** is the ability to **monitor** and **regulate** one's cognitive processes [Flavell, 1979].
    - Example: should I study more (or differently) for this exam?
- As part of metacognitive monitoring, confidence judgments may inform the processes of **cognitive control** [Fleming and Lau, 2014].

An emerging field: the neuroscience of confidence

- **Parietal cortex** is related to evidence accumulation during decision-making.
- **Multiple brain areas** seem involved in confidence formation and reporting:
  - antero-medial PFC.
  - anterior PFC and temporal lobe.
  - Brodmann area 46.
  - ventral striatum.

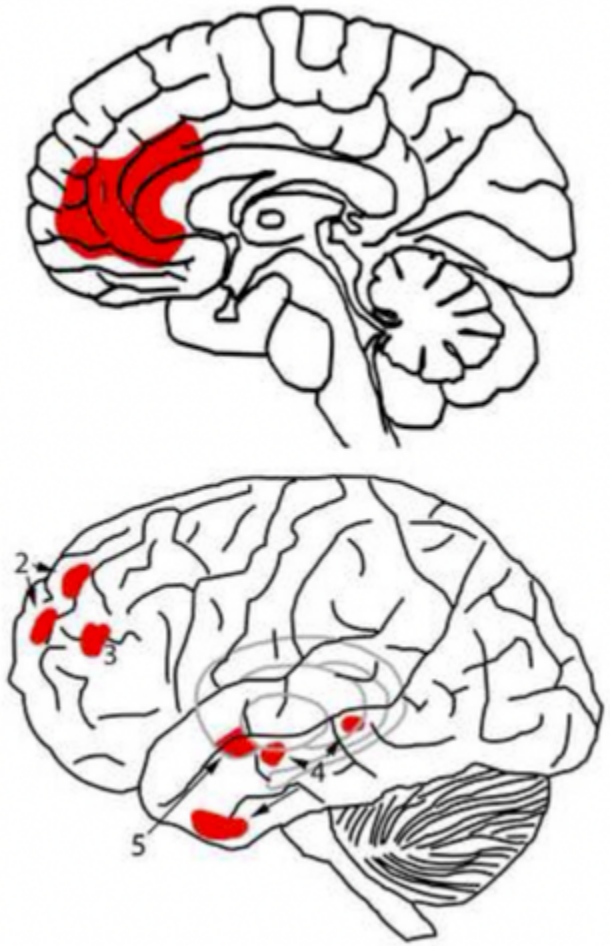


Image credits: [Grimaldi et al., 2015]

Proposed approach

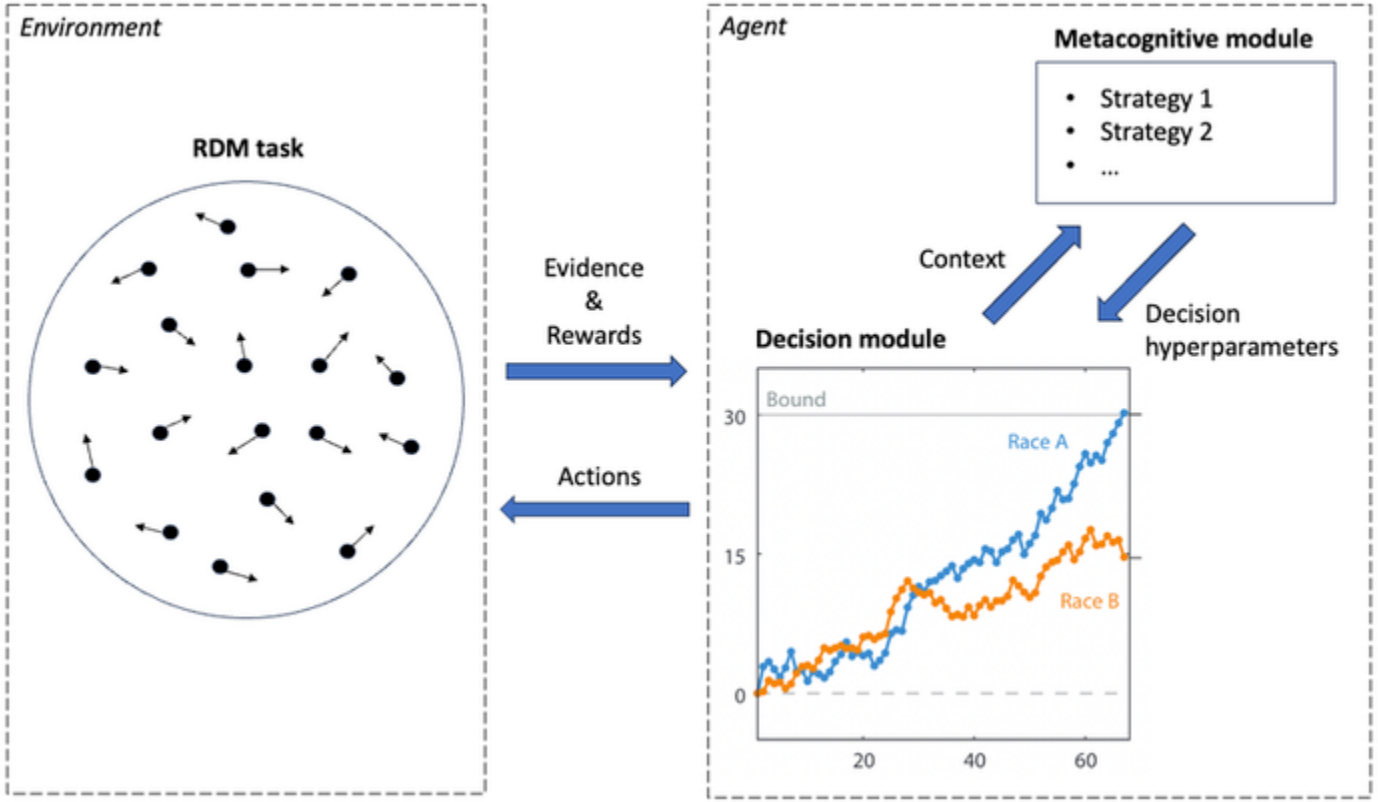
Agent architecture

Model combining:

- a **decision module** based on an evidence accumulation model;
- a **metacognitive module** in which confidence is used to tune the decision hyperparameters: decision threshold and evidence integration rate [Desender and Verguts, 2024].

Experimental validation

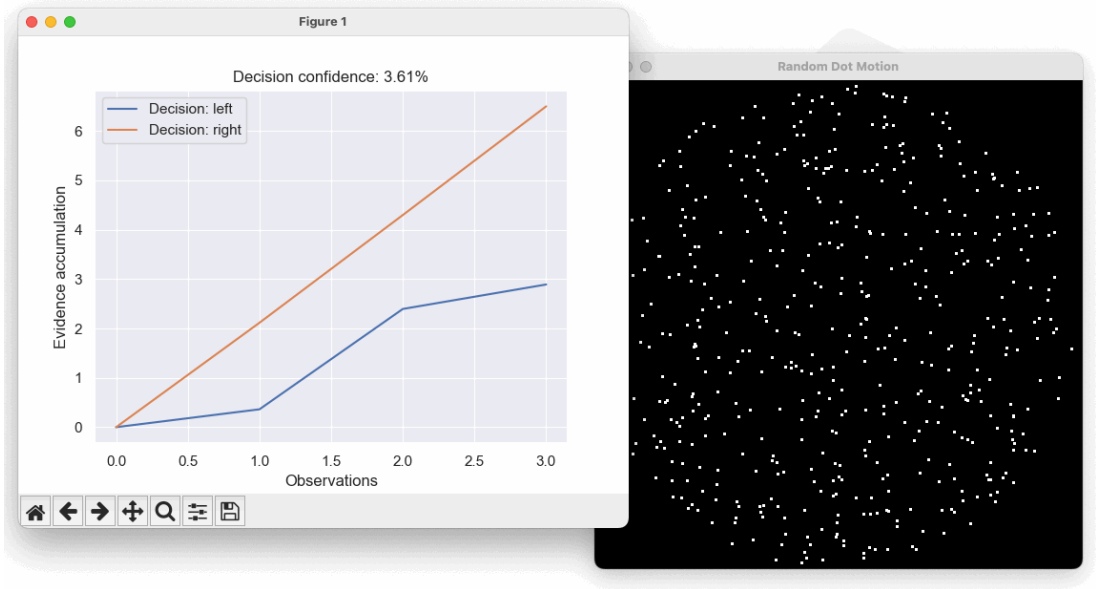
Model was assessed on a classic **perceptual task**: Random Dot Motion discrimination.

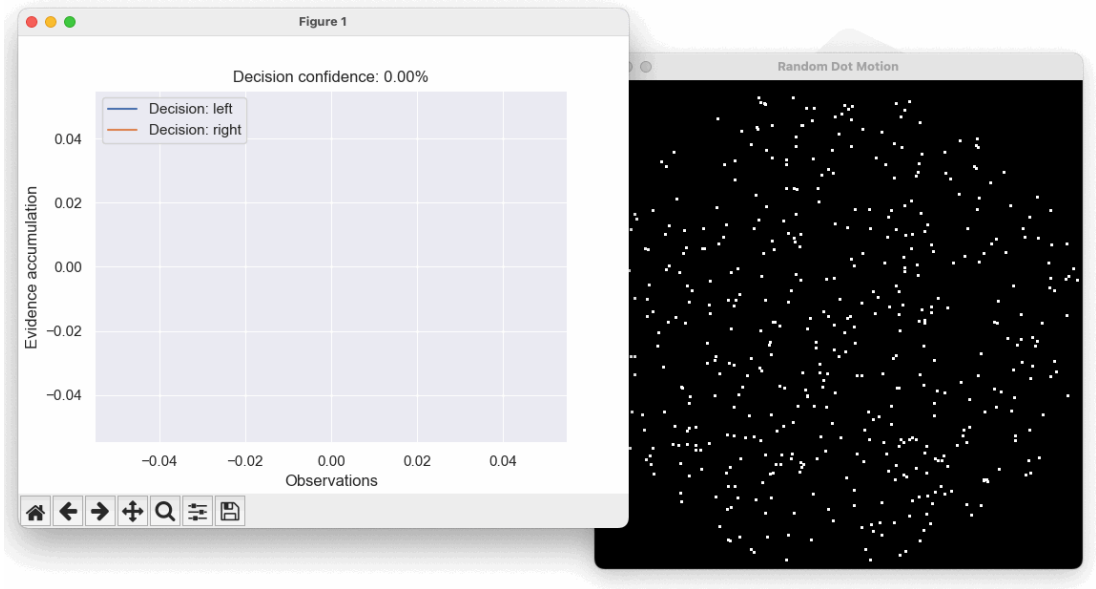


Included image credits: [Mamassian, 2016]

Preliminary results

- Confidence is **correlated** with dot motion coherence, as is (oppositely) decision time.
- Model is able to implement the **SAT**.





Future work

- Refine the hyperparameter tuning process.
- Add different metacognitive strategies.
- Implement model on a human/robot collaborative task.
- ...

Thanks for your attention!

Any questions?