BORDEAUX INP Ensc

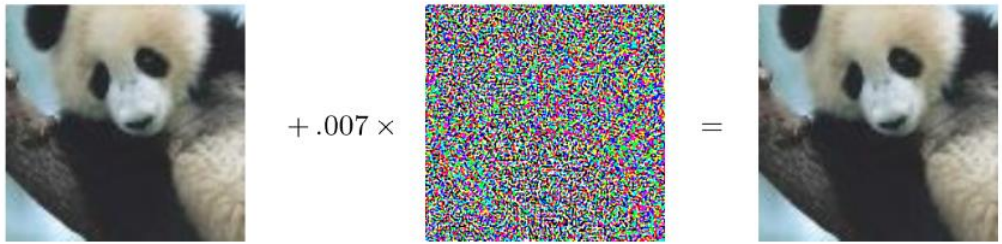
Towards confident AI

Baptiste Pesquet

Inria

WHY CONFIDENCE MATTERS

« A plane parked on the tarmac of an airport »



« Panda »

« Gibbon »

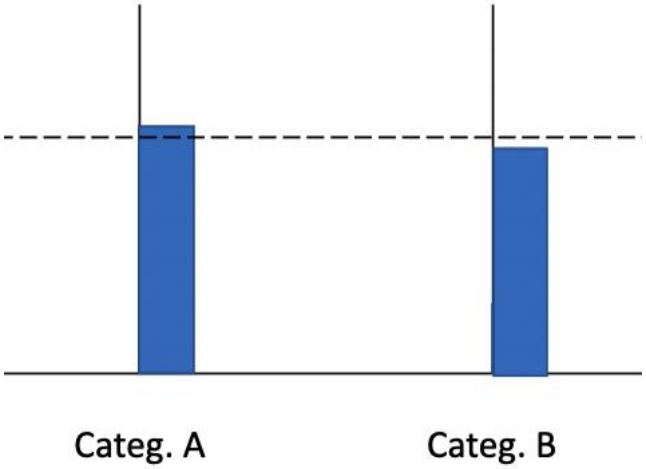
Confiance ai

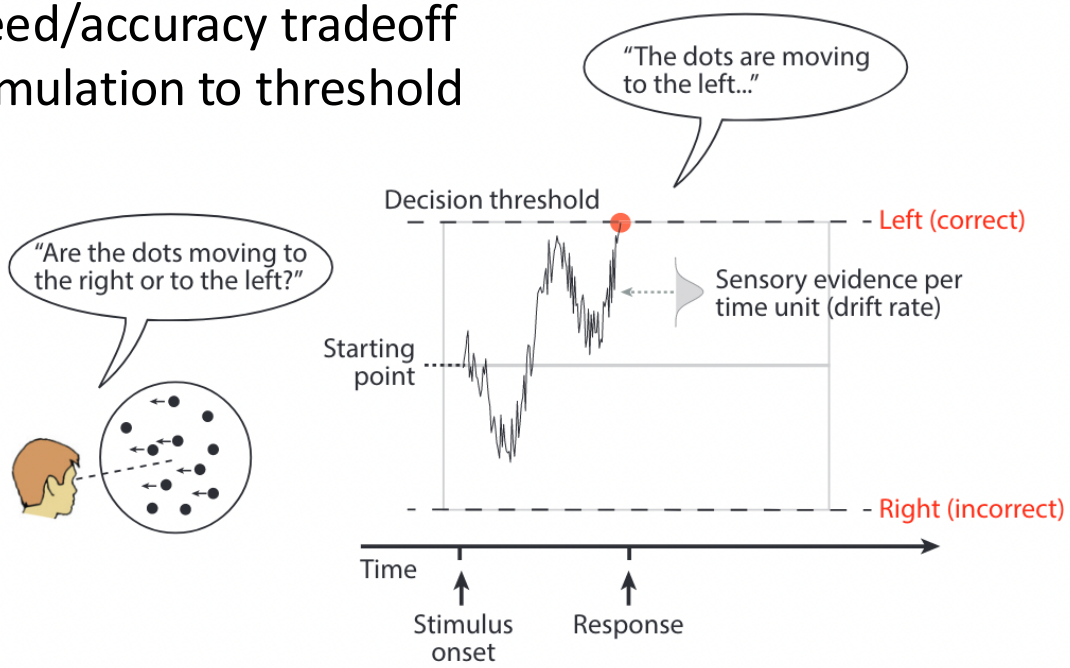
INSPIRATION: HUMAN DECISION-MAKING

DECISION-MAKING

Deliberative process that results in the commitment to a categorical proposition [Gold and Shalden, 2007]

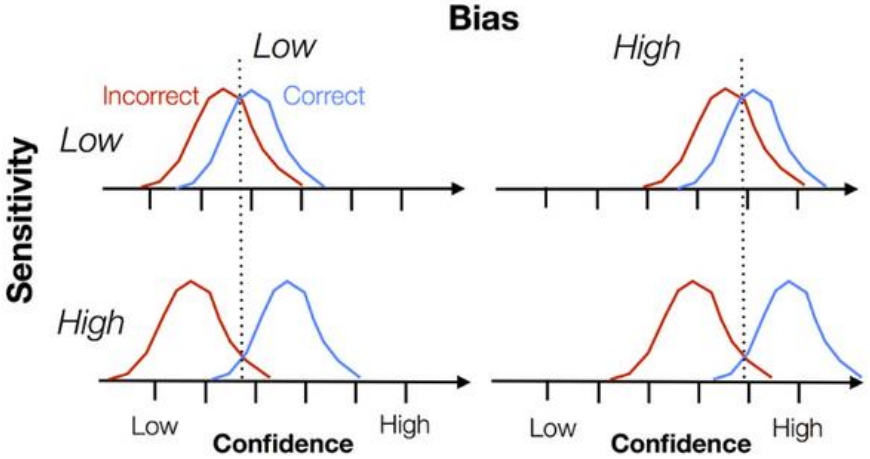
- Sequential nature, speed/accuracy tradeoff
- Model: evidence accumulation to threshold





CONFIDENCE

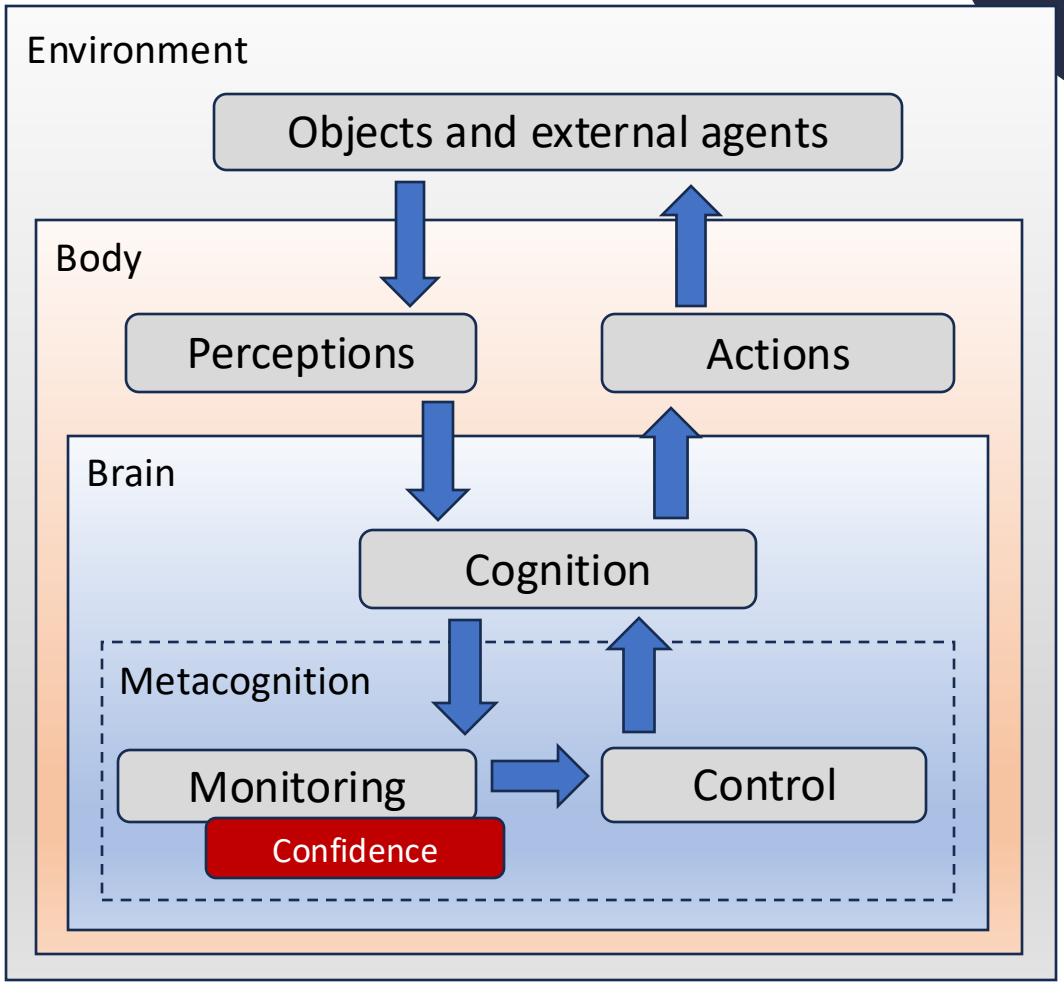
- Quantification of the degree of certainty associated to a decision
  - **Sensibility**: capacity to distinguish correct from incorrect decisions
  - **Bias**: difference in confidence despite constant task performance
  - **Efficiency**: level of sensitivity given a certain level of task performance
- Metacognitive activity



METACOGNITION

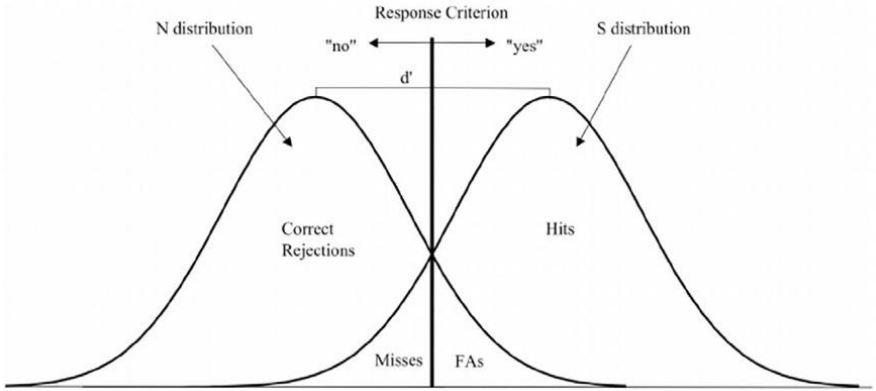
Ability to consider, understand and regulate one's cognitive processes [Fleming, 2024]

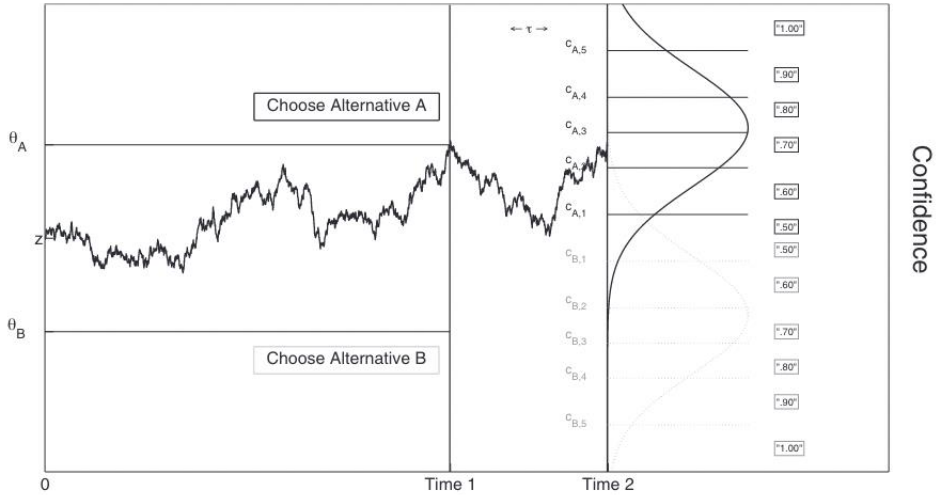
Key skill to adapt to complex problems and changing environments



COMPUTING CONFIDENCE

- Inspiration from Signal Detection Theory framework: meta-d', M-ratio, AUROC2 [Fleming and Lau, 2014]
- Sequential dimension: Balance of Evidence, 2DSD [Pleskac and Bussemeyer, 2010]





WHAT'S NEXT?

- Long-term goal: augment artificial agents with the ability to evaluate and use their confidence for performance, explainability and acceptability

- Next steps:
  - Finish review of confidence for decision-making
  - Define a cognitive architecture based on studied principles
  - Choose an application task for experimentations

THANKS FOR LISTENING!