


A Logic-based Model for the Detection of Fixation Bias

Gabrielle Porcher[1,2] Frédéric Boulanger[1] Nicolas Sabouret[2]

[1]Université Paris-Saclay, CNRS, ENS, CentraleSupélec, LMF, Gif-sur-Yvette, France

[2]Université Paris-Saclay, CNRS, LISN, Gif-sur-Yvette, France

September 26, 2025

Context

- Decision-making[a]
- Human error in sensitive fields
- Cognitive biases[b]

[a] Croskerry, 2009
[b] Saposnik et al., 2016

Medical errors are today one of the **leading** causes of mortality[a], of which 75% are estimated to be related to **cognitive errors**[b].

[a] Makary and Daniel, 2016
[b] Graber et al., 2024

Frame

Diagnostiquer les biais cognitifs, 2020.
Valentin Fouillard, Safouan Taha, Nicolas Sabouret, Frédéric Boulanger

- A post-hoc diagnostic model
- The challenge of monitoring: detecting an inconsistent action as early as possible

- IDEFIX ANR Project
- Fixation bias

Methodology

Objective of the model

Represent and monitor an agent in its environment, with a focus on:

- The agent's actions
- The state of the world
- A set of reasoning rules

Methodology: Conceptualization

Data Processing: Symbolic values

Model Expressiveness:

- Time, durations
- Action conditions[1]
- Desires
- Types of action

Bias Detection:
Incoherent action according to reasoning rules and world state

[1] Barot, Lenne and Lourdeaux, 2014

Methodology: Conceptualization

Data Processing: Symbolic values

Model Expressiveness:

- Time, durations
  - Action durations
  - Observation durations
  - Order of events
- Action conditions[1]
- Desires
- Types of action

Bias Detection:
Incoherent action according to reasoning rules and world state

[1] Barot, Lenne and Lourdeaux, 2014

Methodology: Conceptualization

Data Processing: Symbolic values

Model Expressiveness:

- Time, durations
- Action conditions[1]
  - Nomological
  - Regulatory
  - Contextual
  - Favorable
- Desires
- Types of action

Bias Detection:
Incoherent action according to reasoning rules and world state

[1] Barot, Lenne and Lourdeaux, 2014

Methodology: Conceptualization

Data Processing: Symbolic values

Model Expressiveness:

- Time, durations
- Action conditions[1]
- Desires
  - Example:
    [Patient survival $>$ Patient recovery $>$ Patient relief]
- Types of action

Bias Detection:
Incoherent action according to reasoning rules and world state

[1] Barot, Lenne and Lourdeaux, 2014

Methodology: Conceptualization

Data Processing: Symbolic values

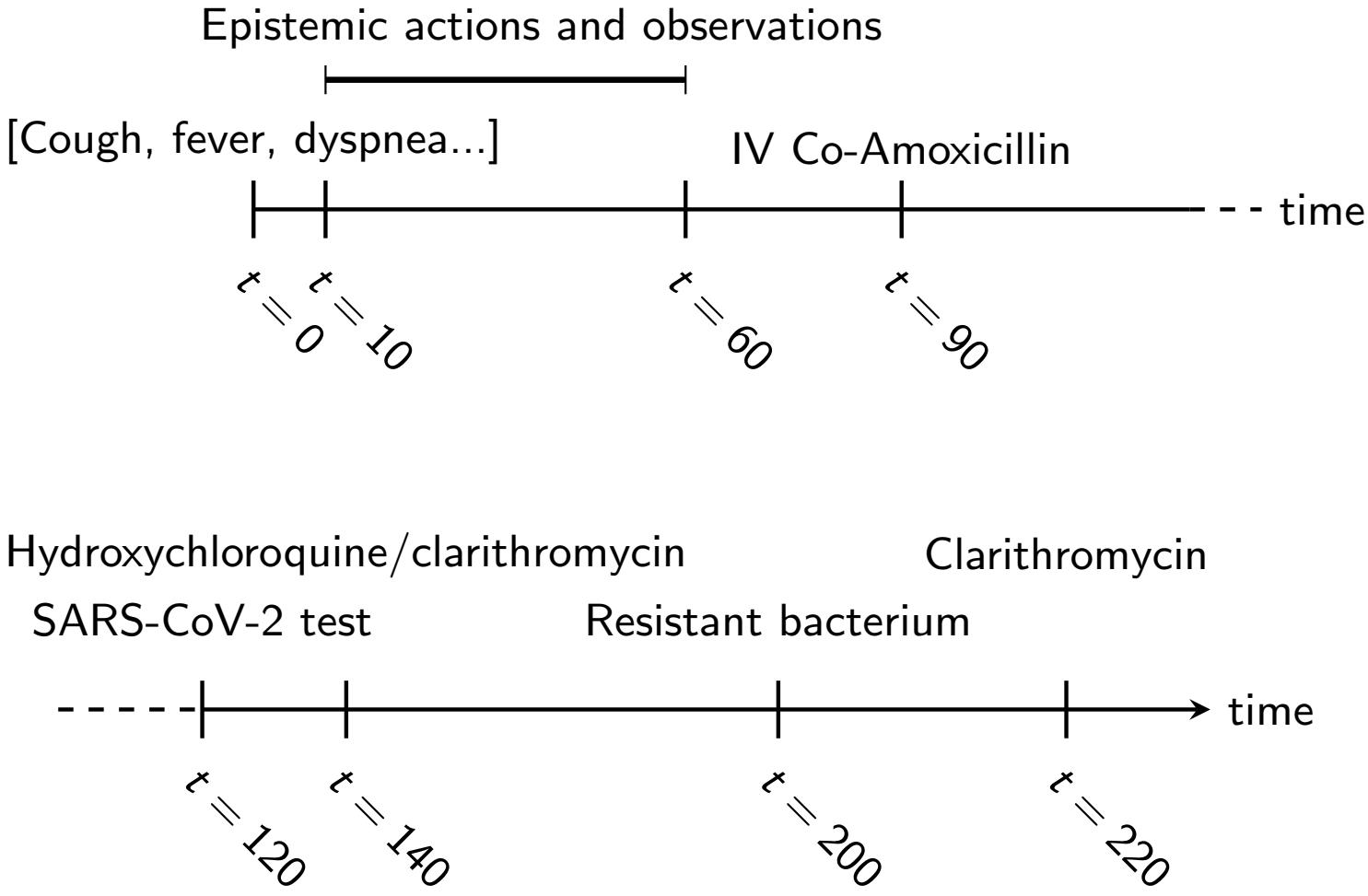
Model Expressiveness:

- Time, durations
- Action conditions[1]
- Desires
- Types of action
  - Epistemic actions
  - Pragmatic actions

Bias Detection:
Incoherent action according to reasoning rules and world state

[1] Barot, Lenne and Lourdeaux, 2014

Methodology: Test Scenario [2]



[2] Bertaux, Alameda, Tataw and Kenfak, 2020

Solution

Structure with two vectors evolving in parallel:

- A vector ϕ of physical possibilities
- A vector ψ of psychological belief distributions

Observations modify the weights associated with the diagnostics ϕ, actions modify the weights associated with the diagnostics ψ.

Association dictionaries

- Between action/observation, diagnosis, and weight

Fixation bias detection

- Deviation δ_D between the vectors over a duration δ_T

Results

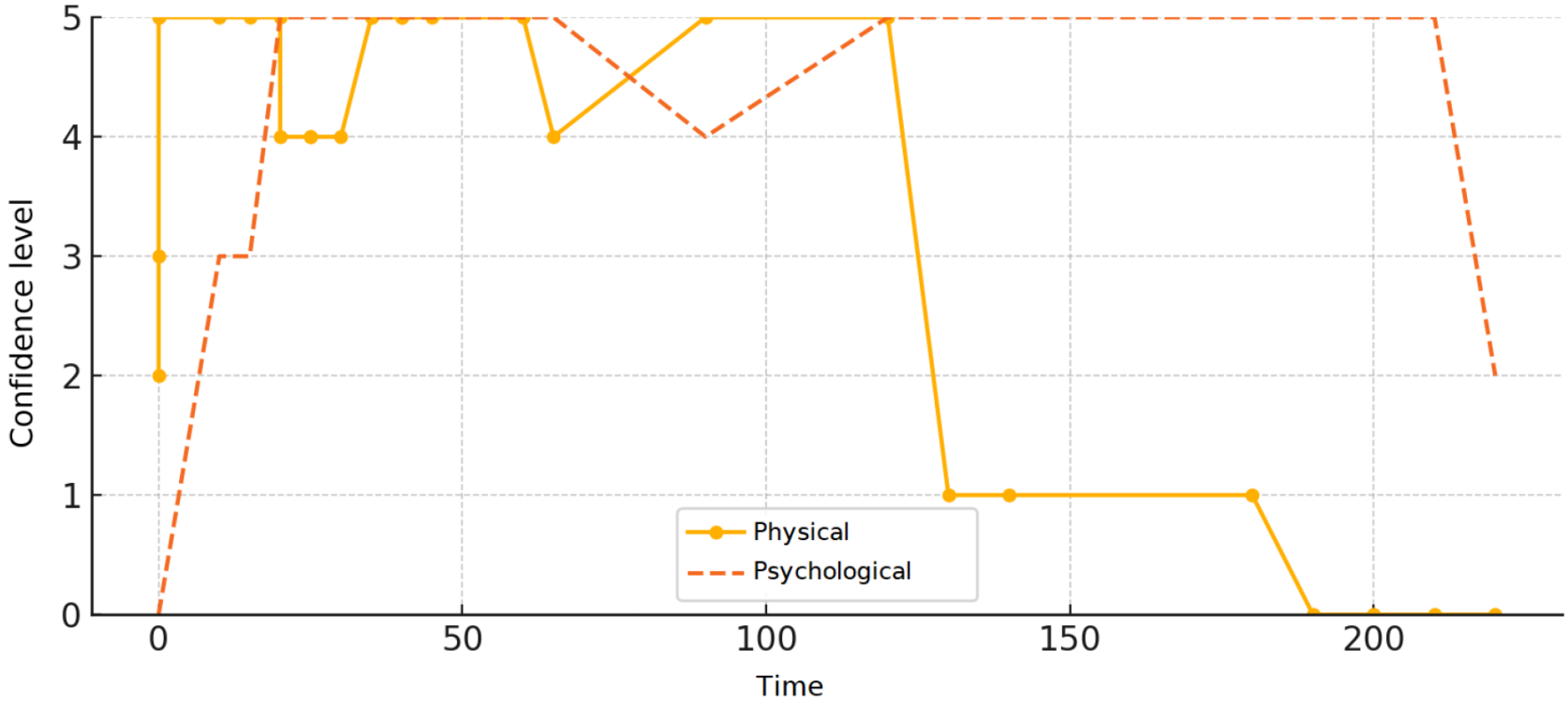


Figure: Evolution of the weights associated with the “covid” diagnosis in the physical and psychological vectors

Results

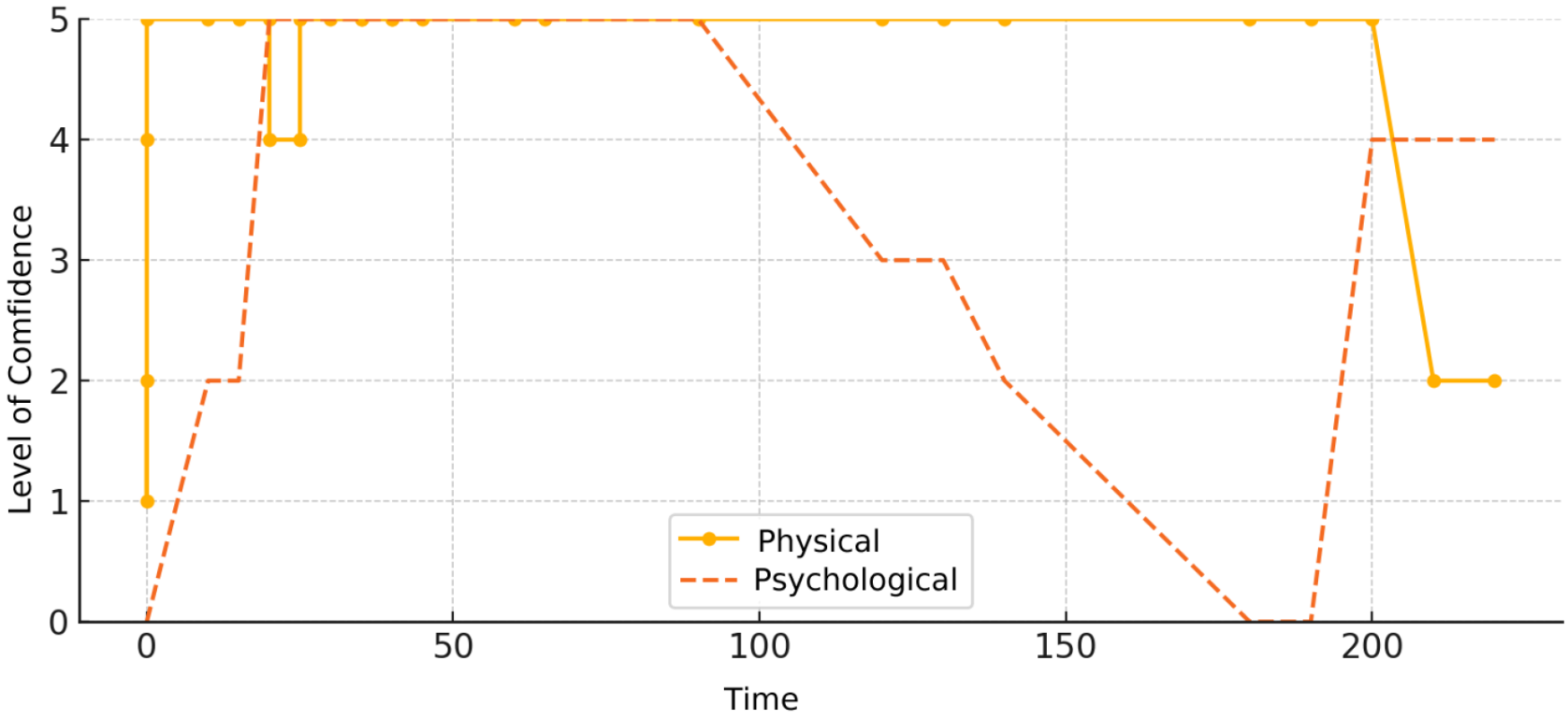


Figure: Evolution of the weights associated with the "Pneumonia 1" diagnosis in the physical and psychological vectors

Results

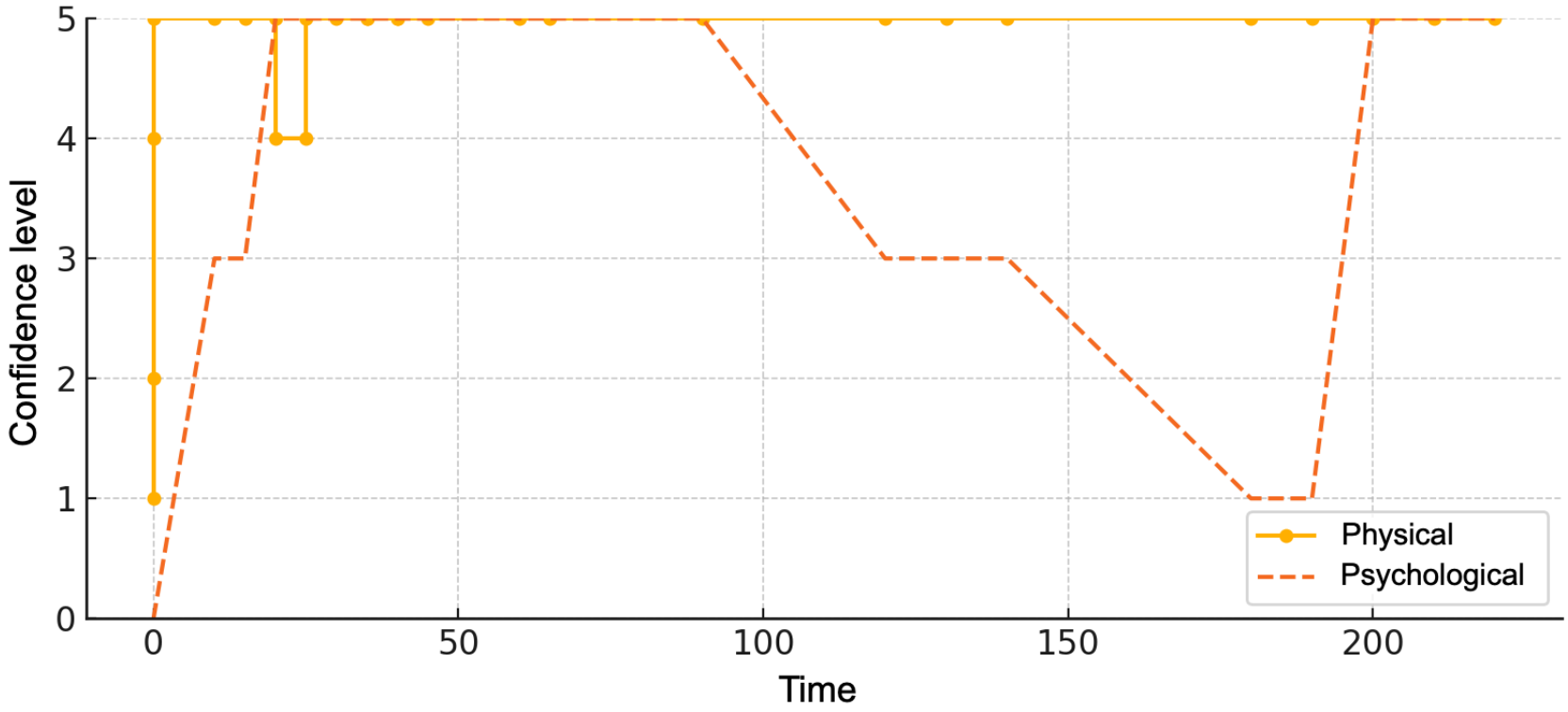


Figure: Evolution of the weights associated with the "Pneumonia 2" diagnosis in the physical and psychological vectors

Critical Analysis

- When to raise an alert?
- Parameters: time window and weight difference
- Risk of error and desensitization
- Next steps:
  - Implementation of action types and durations, conditions...

Thank you for your attention.
Any questions ?

Références I

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine*, 84(8), 1022–1028.

Saposnik, G., Redelmeier, D., Ruff, C., & Tobler, P. (2016). Cognitive biases associated with medical decisions: A systematic review. *BMC Medical Informatics and Decision Making*, 16(1), 138.

Makary, M. A., & Daniel, M. (2016). Medical error—the third leading cause of death in the US. *BMJ*, 353, i2139.

Graber, M. L., et al. (2024). Cognitive errors in diagnosis: missing the forest for the trees. *Diagnosis*.

Reason, J. (2000). Human error: models and management. *BMJ*, 320(7237), 768–770.

Hartigan, I., et al. (2020). Avoiding cognitive biases in clinical decision-making. *Clinical Nursing Research*, 29(6), 370–377.

Al-Khafaji, A., et al. (2022). Cognitive aids in emergency medicine. *Emergency Medicine Journal*, 39(5), 345–350.

Sutton, R. T., et al. (2020). An overview of clinical decision support systems. *Journal of Biomedical Informatics*, 117, 103678.

Rao, A. S., & Georgeff, M. P. (1995). BDI agents: From theory to practice. *Proceedings of the First International Conference on Multi-Agent Systems*, 312–319.

Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50(2), 510–530.

Références II


Voinson, M., Billiard, S., & Alvergne, A. (2015). Beyond rationality: modeling individual decision making based on cognitive biases. *Proceedings of the Royal Society B*, 282(1813), 20141497.


Fouillard, V., Taha, S., Sabouret, N., & Boulanger, F. (2020). Diagnostiquer les biais cognitifs. *Technical report*, CentraleSupélec.


Barot, C., Lenne, D., & Lourdeaux, D. (2014). Une ontologie des conditions d'action en environnement virtuel. *ICST Transactions on Ambient Systems*, 14(6), 1–15.


Projet ANR IDEFIX, (2024).