A Logical Model for Fixation Error Detection: Managing Uncertainty

Gabrielle Porcher, Frédéric Boulanger, Nicolas Sabouret

Laboratoire Méthodes Formelles
Laboratoire Interdisciplinaire des Sciences du Numérique

April 23, 2026

Outline of the Presentation

Introduction: Human Error

ANR IDEFIX project

Logical Model and Related Work

The Problem of Uncertainty

Logical Frameworks with Uncertainty

Connecting the Frameworks!

Conclusion

Critical Systems and Humans

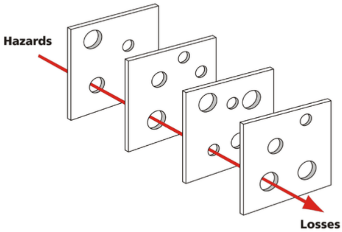


Figure: Swiss Cheese Model

Layers:

- Human actors (experts, managers, advisors)
- Decision support systems
- Staff training
- Dissemination of instructions
- More or less autonomous systems

Humans remain at the center of decision-making

Human Error

Decision-making:

- Result of multiple cognitive processes
- Influenced by many factors
- The flaw: cognitive biases

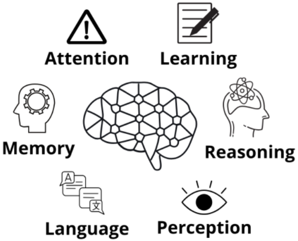


Figure: Cognitive Processes

Medical errors are one of the **main** causes of mortality[1], with 75% linked to **cognitive errors**[2].

[1] Makary and Daniel, 2016

[2] Graber et al., 2024

Fixation bias

Fixation errors are human errors caused by excessive focus on one idea, solution, piece of information, or perspective, to the point of ignoring other possibilities.

Anchoring Bias

Tunnel effect

Confirmation bias

ANR IDEFIX project

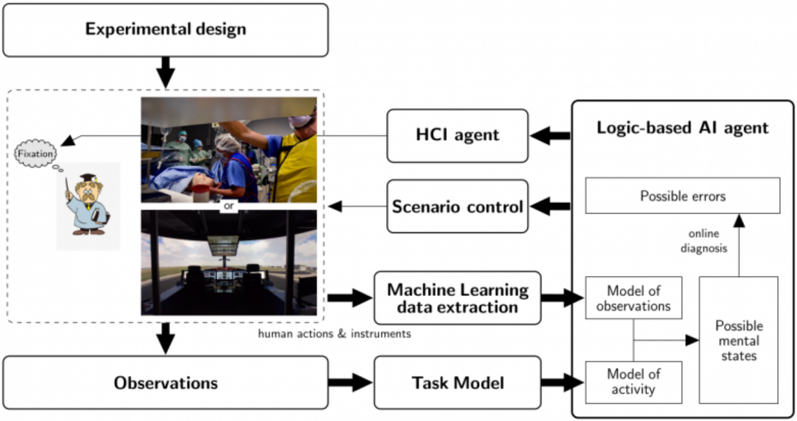
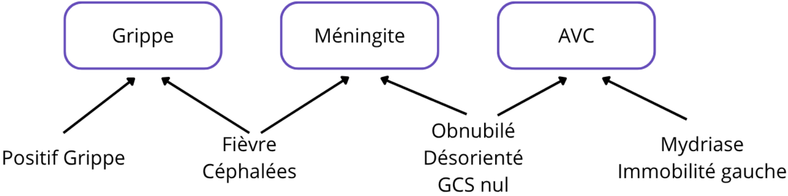


Figure: IDEFIX ANR PROJECT

Exemple de scénario

Meningitis or Flu + Stroke?



Work in Formal Logic

- Transparent: provides explanations for alerts
- BDI logic, epistemic and doxastic logics
- Problem: retrieving what the agent believes from their actions
- The logic of inconsistencies: a formal model for the analysis of human error[a]

If Idefix detects fixation,
then he barks

[a] Valentin Fouillard, 2022

Definition of Fixation

We have access to:

- Events (observations or actions)
- A set of specialized reasoning rules

Example scenario:

- Very high blood pressure, high fever, disorientation...
- Low sat $\rightarrow$ O_2 mask
  meningitis $\rightarrow$ LP
  stroke $\rightarrow$ mydriasis

We define fixation as a gap between the possibility of a lead and how much the operator seems to consider it.

The Model

First step:
Propose a diagnosis

- Follow the symptoms
- Link them to diagnoses
- Aggregate them

Second step:
Alert errors

- Track actions
- Link them to diagnoses and symptoms
- Check that the operator follows consistent leads

The Problem of Uncertainty: Logical Inferences

Deduction: Stroke $\rightarrow$ mydriasis

Abduction: Stroke $\rightarrow$ mydriasis

Induction: Stroke $\rightarrow$ mydriasis

The Problem of Uncertainty: Deduction

Stroke $\rightarrow$ mydriasis... But is this always true?

- Depends on size, location, presence of intracranial hypertension... Many and complex nuances!

Logical example:

Bird(x) $\rightarrow$ Flies(x)

Bird(Tweety)

Penguin(Tweety)

Penguin(x) $\rightarrow$ Bird(x)

Several ways to handle: exceptions, rule hierarchy, non-monotonic logics, probabilistic approaches...

Stroke $\xrightarrow{0.9}$ mydriasis

The Problem of Uncertainty: Abduction

It was already complicated... But we also seek to retrieve diseases/failures **from** symptoms!

$$\begin{tikzcd} \text{Stroke} \arrow[r, "0.9"] & \text{mydriasis} \arrow[l, bend right, "?"'] \end{tikzcd}$$

The Problem of Uncertainty: Human Reasoning

- Reason under uncertainty
- Based on degrees of belief [3]
- Based on experiences we do not have access to

[3] Daniel Kahneman, 2011

Logical Frameworks with Uncertainty: Probabilistic Logic

Bayesian models widely used for diagnosis[4] and representation of human reasoning[5]

Problems:

- Require defining precise probabilities for each hypothesis
- Direct interpretation of probabilities
- Human reasoning is opaque, biased, and probabilities are "intuitive"

[4] Heckerman, Horvitz & Nathwani (1992)

[5] Griffiths, Kemp & Tenenbaum (2008 / 2010)

Logical Frameworks with Uncertainty: Possibilistic Logic

- Based on possibility theory[6]
- Uses degrees of necessity N and possibility Π on **facts** and **rules**, example: $\overset{0.8}{\text{low sat}} \xrightarrow{0.9} \text{O2 mask}$
- Operators:
  - Conjunction: min
  - Disjunction: max
  - Negation: 1-
- Testing the descriptive validity of possibility theory in human judgments of uncertainty (Raufaste et al., 2003)

[6] Zadeh (1970), Dubois & Prade (1980)

Logical Frameworks with Uncertainty: Possibilistic Logic

In the context of abduction in the presence of sometimes contradictory information:

- Two hypotheses: pneumonia and COVID
- Two symptoms: nausea (weakly associated with hypotheses) and scan result only associated with pneumonia

If $H = Pneumonia$:

$\pi(nausea, scan\_p | Pneumonia) = \min(0.1, 0.9) = 0.1$

If $H = Covid$:

$\pi(nausea, scan\_p | Covid) = \min(0.1, 0.1) = 0.1$

Belief Function Theory

- Developed by Dempster (1960), formalized by Shafer (1976)

Belief Function Theory

- Developed by Dempster (1960), formalized by Shafer (1976)
- Allows merging uncertain information

Belief Function Theory

- Developed by Dempster (1960), formalized by Shafer (1976)
- Allows merging uncertain information
- Fusion of sensor data or expert opinions

Belief Function Theory

- Developed by Dempster (1960), formalized by Shafer (1976)
- Allows merging uncertain information
- Fusion of sensor data or expert opinions
- Definition of a frame of discernment representing all mutually exclusive and exhaustive hypotheses.
  Example: Meningitis, Stroke, Flu+Stroke

Belief Function Theory

- Developed by Dempster (1960), formalized by Shafer (1976)
- Allows merging uncertain information
- Fusion of sensor data or expert opinions
- Definition of a frame of discernment representing all mutually exclusive and exhaustive hypotheses.
  Example: Meningitis, Stroke, Flu+Stroke
- Definition of mass functions: a mass function associates, for each observation, a mass with each subset of the frame of discernment. Example:

Belief Function Theory

- Developed by Dempster (1960), formalized by Shafer (1976)
- Allows merging uncertain information
- Fusion of sensor data or expert opinions
- Definition of a frame of discernment representing all mutually exclusive and exhaustive hypotheses.
  Example: Meningitis, Stroke, Flu+Stroke
- Definition of mass functions: a mass function associates, for each observation, a mass with each subset of the frame of discernment. Example:

"Obnubilation" $\rightarrow$

m(Meningitis, Stroke, Flu+Stroke)=0.7, m(total)=0.3

Belief Function Theory: Mass Fusion

We seek to calculate: obnubilation $\oplus$ low_saturation:

"Obnubilation" $\rightarrow$
m(Meningitis, Stroke, Flu+Stroke)=0.7, m(Total)=0.3
"Low_saturation" $\rightarrow$
m(Flu, Flu+Stroke)=0.5, m(Total)= 0.5

Belief Function Theory: Mass Fusion

We seek to calculate: obnubilation $\oplus$ low_saturation:

"Obnubilation" $\rightarrow$
m(Meningitis, Stroke, Flu+Stroke)=0.7, m(Total)=0.3
"Low_saturation" $\rightarrow$
m(Flu, Flu+Stroke)=0.5, m(Total)= 0.5

1) Calculate conflict K: sum of products of subsets with empty intersection

Belief Function Theory: Mass Fusion

We seek to calculate: obnubilation $\oplus$ low_saturation:

"Obnubilation" $\rightarrow$
m(Meningitis, Stroke, Flu+Stroke)=0.7, m(Total)=0.3
"Low_saturation" $\rightarrow$
m(Flu, Flu+Stroke)=0.5, m(Total)= 0.5

1) Calculate conflict K: sum of products of subsets with empty intersection

2) Dempster's Rule:
m(Flu+Stroke)= $0.7 \times 0.5 = 0.35$
m(Meningitis, Stroke, Flu+Stroke)= $0.7 \times 0.5 = 0.35$
m(Flu, Flu+Stroke)= $0.5 \times 0.3 = 0.15$
m(Total)= $0.3 \times 0.5 = 0.15$

Belief Function Theory: Interpretation

- Plausibility: sum of masses of all sets that intersect it

- Belief: sum of masses of all its subsets

Belief Function Theory: Interpretation

- Plausibility: sum of masses of all sets that intersect it
  Pl(Flu+Stroke) $= 0.35 + 0.15 + 0.35 + 0.15 = 1$
  Pl(Meningitis) $= 0.15 + 0.15 = 0.3$
  Pl(Total) $= 1$
- Belief: sum of masses of all its subsets Bel(Flu+Stroke) $=$ 0.35
  Bel(Meningitis) $= 0$
  Bel(Flu, Flu+Stroke) $= 0.35 + 0.35 = 0.70$
  Bel(Total) $= 1$

Belief Functions & Possibility Theory

$$\text{Plausibility Pl} \longleftrightarrow \text{Possibility } \Pi$$
$$\text{Belief Bel} \longleftrightarrow \text{Necessity } N$$

Belief Functions & Possibility Theory

Plausibility Pl $\longleftrightarrow$ Possibility Π
Belief Bel $\longleftrightarrow$ Necessity N

Action rules with diseases as predicates:
- if a disease is possible, one may
- if a disease is possible, one must
- if a disease is certain, one may
- if a disease is certain, one must

...and their opposites!

Alerts and Conclusion

Model capable of detecting and alerting:

- Errors directly related to diseases/failures or observations
- Diagnoses ignored over time despite being possible
- A privileged diagnosis even though others are possible

What's next?
Urgency levels, simulations to suggest actions, defining masses with professionals

Thank you for your attention!
Any questions?