



UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

Corso di Laurea Magistrale in Informatica (LM-18)

Natural Interaction and Affective Computing:

A Computational Implementation of Interoceptive Active Inference: Simulating Cardiac–Perceptual Coupling

Student:

Andrea Moretti, matricola: 47263A

A.Y. 2024/2025

1 Introduction

1.1 Description

In recent years, a growing attention has been devoted to understand interactions between the brain and the body, particularly in the fields of computational neuroscience and psychiatry. Traditional cognitive models often treated the brain as an isolated information processor, but emerging theories of embodied cognition discover that perception is an active process, grounded in the physiological rhythms of the body.

Among these rhythms, the cardiac cycle plays a crucial role: experimental studies have shown that the phase of the heartbeat can modulate perceptual sensitivity, emotional evaluation and even metacognitive confidence.

Within the framework of predictive processing and active inference, these findings suggest that the brain continuously generates predictions not only about the external world (exteroception) but also about the internal physiological state (interoception).

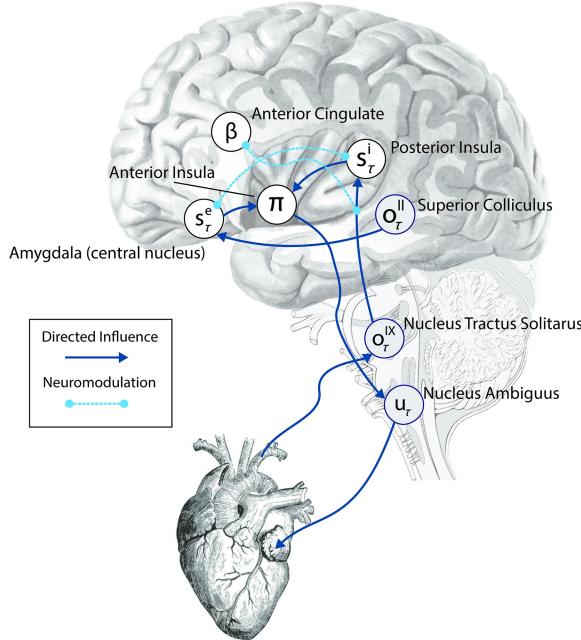


Figure 1: A schematic illustration of the model introduced by Allen et al. [1]

According to this view, perception, action, and bodily regulation emerge from a process of Bayesian inference, through which the brain minimizes prediction errors by updating its beliefs or by acting on the world and the body to make sensations conform to expectations.

To better understand oscillatory brain-body interactions, a computational model of interoceptive inference has been developed in which a synthetic agent's perceptual beliefs are coupled to the rhythm of the heart, formalizing the idea introduced by Allen et al. (2022) [1] of linking interoceptive cardiac signals with exteroceptive sensory processing (fig 1).

Their model provides a practical explanation of how cardiac state can entrain sensory precision, influence affective perception and modulate uncertainty. By simulating the interaction between the cardiac cycle and visual perception, they demonstrated how bodily rhythms can generate phenomena such as the defensive startle reflex and how affect the interoceptive precision may lead to altered affective expectations or misinterpretations of bodily signals.

This framework offers a computational lens through which to study brain–body interactions and to understand how interoceptive inference contributes to emotion, self-awareness and perception. Building upon this theoretical foundation, the present work implements a simplified version of the model proposed by Allen, focusing on the interaction between visual and cardiac states in a minimal “flower/spider” paradigm.

1.2 Model

In this work, I've implemented a simplified version of the *cardiac active inference* model proposed by Allen and Friston [1], to show how interoceptive (bodily) and exteroceptive (visual) states jointly influence perception and belief updating. The model represents an agent that must infer both what it is perceiving in the external world and its own internal bodily state, while selecting appropriate physiological “policies” (relaxed vs. aroused) according to its sensory inputs and prior preferences.

The external environment can present one of two possible stimuli:

1. a **flower**, representing a non-arousing stimulus (safe);
2. a **spider**, representing an arousing stimulus (threat).

Internally, the agent’s body is modelled as being in one of three cardiac phases: two diastolic states (Diastole 1 and Diastole 2) and one systolic state (Systole). This three-state representation was chosen to capture a simplified distinction between relaxed and aroused cardiac dynamics. In the relaxed mode, the cardiac cycle spends more time in diastole (i.e., two diastolic phases for each systolic contraction), reflecting slower heart rhythms and longer periods of sensory sampling. In contrast, under arousal, the cycle transitions more rapidly from diastole to systole, effectively shortening the diastolic period and producing a simulated cardio-acceleration response.

The agent receives exteroceptive observations (what it “sees” - flower or spider) and interoceptive observations (how it “feels” - relaxed or aroused), both of which depend on these hidden states. At each step of the simulation, the agent must infer 2 states:

1. the external state (flower or spider);
2. the internal state (cardiac state).

At this point the agent has to select which policy (relaxed or aroused) will minimize its expected free energy in order to preserve a coherent and preferred coupling with its environment.

For this model the main matrices are:

- The likelihood matrix, denoted as \mathbf{A} , characterizes the relationship between the model's states and the observations:

exteroception

$$\mathbf{A}_E = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{bmatrix} \begin{bmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

interoception

$$\mathbf{A}_I = \begin{bmatrix} \beta & \beta & 1-\beta \\ 1-\beta & 1-\beta & \beta \end{bmatrix} \begin{bmatrix} \beta & \beta & 1-\beta \\ 1-\beta & 1-\beta & \beta \end{bmatrix} \begin{bmatrix} \text{red heart} \\ \text{red heart} \end{bmatrix}$$

The parameter α describe the visual precision, the parameter β describe the cardiac precision. Given the hidden state, the agent infer the external state from matrix A_e and the internal state from A_i (fig 1.2), an example of reading: with hidden state spider and diastole 1, the internal osbservation is “relaxed” with β probability and “aroused” with $1-\beta$ probability.

The Bayesian formulas that describes these matrices are

$$P(o_e | s_e, s_i) \quad P(o_i | s_e, s_i) \quad (1)$$

with s_e the external hidden state (visual state) and s_i the internal hidden state (cardiac state)

In this implementation, a slight deviation from the original formulation proposed by Allen et al. [1] was introduced for practical reasons related to the computational framework adopted. Specifically, instead of defining separate likelihood matrices for exteroceptive and interoceptive modalities, I've created a single joint likelihood matrix \mathbf{A} , encoding the mapping from the combined hidden state (s_e, s_i) to both exteroceptive and interoceptive observations.

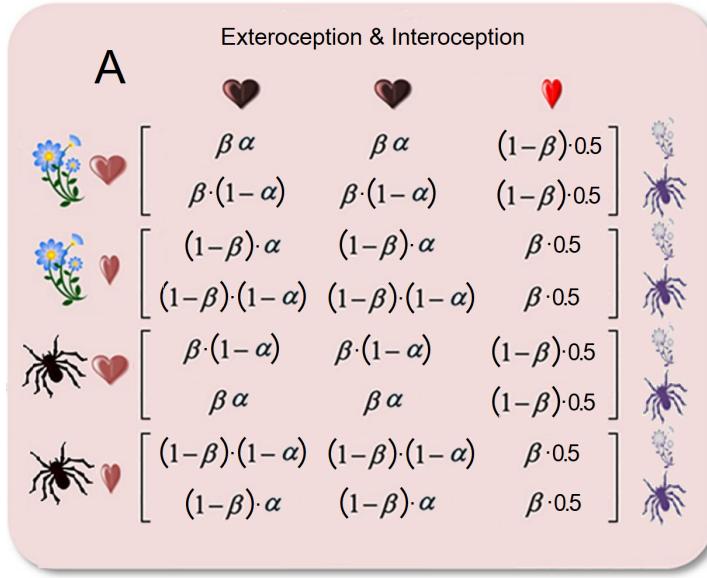


Figure 2: In this matrix there are 4 outputs encoding every combination of internal and external outputs (flower+relaxed, flower+aroused, spider+relaxed, spider+aroused)

This choice was motivated by a limitation of the pymdp [2] library, which does not natively support the definition of joint preferences over combinations of internal and external states within the prior preference matrix C.

- Matrix \mathbf{B} corresponds to the transition matrix, describing the dynamics of state changes (internal and external) within the model:

$$\mathbf{B}_I^{relaxed} = \begin{bmatrix} \heartsuit & \heartsuit & \heartsuit \\ \zeta_I & & \zeta_I \\ & \zeta_I & \end{bmatrix} \quad \begin{array}{c} \heartsuit \\ \heartsuit \\ \heartsuit \end{array}$$

$$\mathbf{B}_I^{aroused} = \begin{bmatrix} \heartsuit & \heartsuit & \heartsuit \\ & \zeta_I & \\ \zeta_I & \zeta_I & \end{bmatrix} \quad \begin{array}{c} \heartsuit \\ \heartsuit \\ \heartsuit \end{array}$$

In the transition matrices B_I , which describe the evolution of the internal state, the parameter ζ_I represents the precision of state transitions between cardiac states. In this framework, two distinct transition matrices are defined to capture the heart's dynamics as it alternates between relaxed and aroused states.

The Bayesian formula that describes these matrices is

$$P(s_{\tau+1}|s_\tau, a) \tag{2}$$

where a is the chosen action.

The second transition matrix concerns the external state:

$$\mathbf{B}_E = \begin{bmatrix} \zeta_E & 1 - \zeta_E \\ 1 - \zeta_E & \zeta_E \end{bmatrix}$$

Similarly to the previous matrices, it includes the parameter ζ_E , which controls the probability of transitions between external states (flower/spider). Here the Bayesian formula that describes this matrix is

$$P(s_{\tau+1}|s_\tau) \quad (3)$$

- The matrix of prior preference, denoted as \mathbf{C} :

$$\mathbf{C} = \begin{bmatrix} \zeta & 0 \\ 0 & \zeta \end{bmatrix}$$

This matrix represents the agent's prior preferences over state-stimulus combinations. In particular, the standard agent prefers to be in a relaxed state when seeing a flower, and in an aroused state when seeing a spider.

- The \mathbf{D} matrix specifies the agent's initial beliefs about the hidden states before any observations are received.

Here, both the external and internal states are initialized with uniform priors, assigning equal probabilities to a flower or spider, as well as to each cardiac state.

The Bayesian formulation describing D is straightforward:

$$P(s_1) \quad (4)$$

- The **E** matrix, the prior over policies:

In addition to the matrices A, B, C, and D, the model also includes an E matrix, which encodes prior beliefs over policies, the agent's expectations about which action is more likely or more desirable before observing any sensory evidence. Even in this case, the Bayesian formulation associated with this matrix remains very simple:

$$P(\pi) \quad (5)$$

2 Simulations

To evaluate the behaviour of the implemented model and to assess whether it reproduces the key phenomena described by Allen et al., a series of simulations were conducted. These simulations explore how the agent integrates interoceptive and exteroceptive information, how it responds to arousing stimuli, and how it changes in sensory precision or internal dynamics affect its beliefs and behaviour. Each subsection focuses on a specific phenomenon to illustrate a different functional aspect of the model.

2.1 Simulation 1: metacognitive signals

In the first set of simulations (Fig. 4), I analyzed how the agent’s perceptual inferences give rise to metacognitive signals, specifically in the form of posterior uncertainty. The goal of this analysis is to reproduce the key finding reported by Allen et al.: the metacognitive confidence is not an independent process, but a property of the same inferential process that governs perception and interoception.

In the model, posterior uncertainty—quantified as the entropy of the inferred hidden states—naturally changes across different cardiac phases and depends on sensory precision. During systole, when exteroceptive precision is attenuated, the agent becomes less certain about the external world, resulting in higher posterior entropy. Conversely, during diastole, when sensory precision is stronger, perceptual inferences become more reliable, and entropy decreases.

By manipulating exteroceptive and interoceptive precision, we can see how different “lesioned” conditions (α or $\beta = 0.5$) produce distinct patterns of perceptual uncertainty compared to a healthy condition (α or $\beta = 0.9$), linking bodily signals to metacognitive evaluations. This supports the idea that metacognition arises from the structure of the generative model rather than from a dedicated supervisory system.

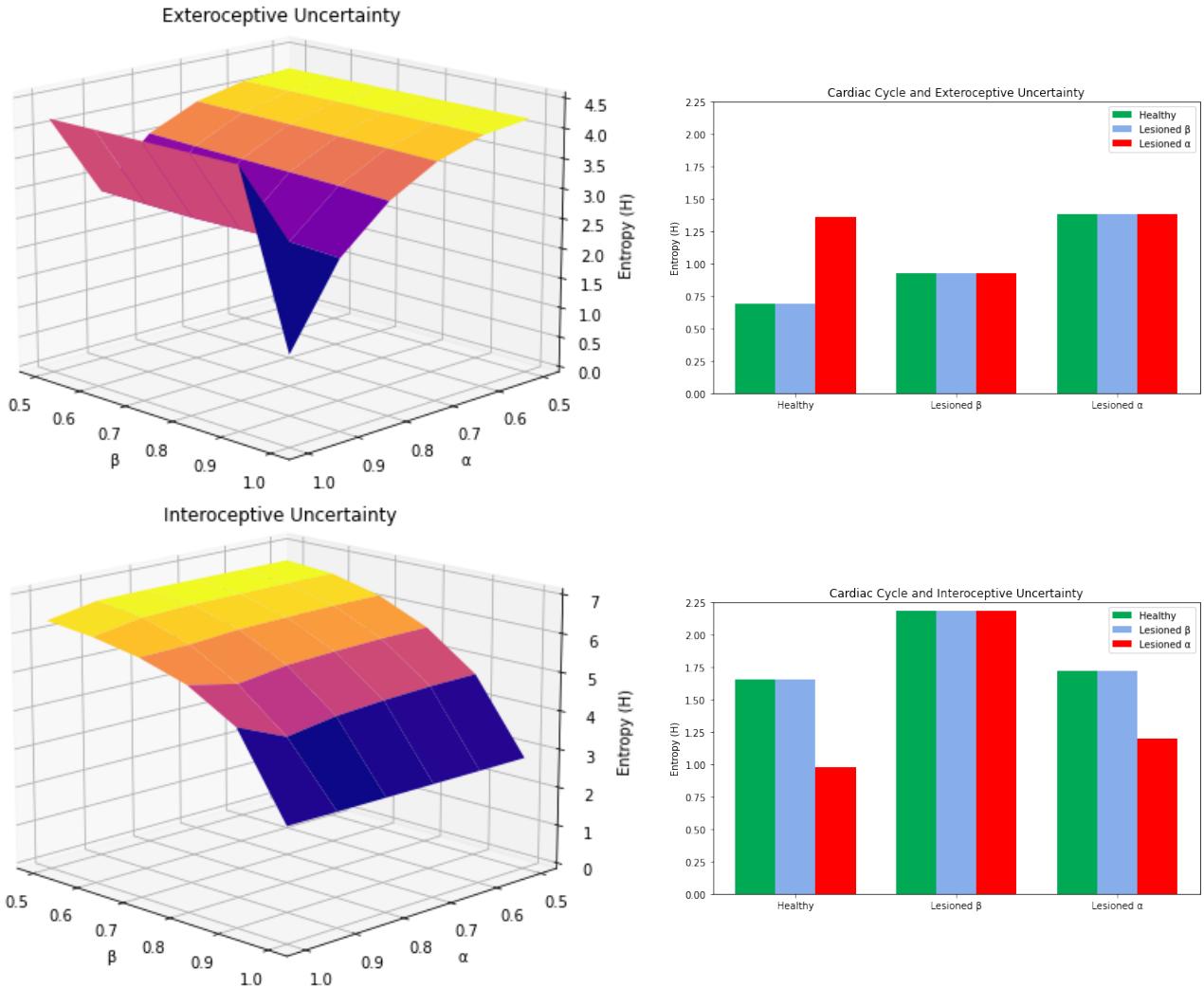


Figure 3: Exteroceptive and interoceptive uncertainty, measured as entropy for each parameter (α and β) value.

The figure 3 is divided into two panels: the left panel shows the graphical representation of uncertainty values computed for each combination of perceptual parameters (α and β), summed across all combinations of hidden states; the right panel presents bar plots illustrating the uncertainty measured at each cardiac cycle for three different types of agents.

The results confirm the expected pattern, namely that interoceptive precision directly influences not only internal observations but also external ones: understanding the internal state of the body is crucial for shaping and interpreting observations of the external world.

An important focus is required on the first subfigure of figure 3, which shows that sensory precision in active inference models controls the sharpness of the likelihood mapping, but does not directly encode the mutual information between observations and hidden states. In particular, we observe an unexpected increase in entropy when the external sensory precision parameter α approaches 1 (except in the case where $\beta=1$ as well).

When sensory precision is maximal, the likelihood becomes deterministic but non-discriminative, providing equivalent evidence for external state hypotheses. As a result, posterior beliefs over external states remain dominated by the prior D , leading to maximal entropy. Conversely, lower sensory precision introduces stochasticity that breaks these symmetries, allowing the posterior distribution to become more concentrated despite noisier observations.

2.2 Simulation 2: physiological response

In the second simulations, I focused on the physiological and psychological response to arousing stimuli. I tested the hypothesis that the unexpected presentation of a spider would induce an aroused state—reflected in an acceleration of cardiac cycle—and increase the agent’s posterior expectation of encountering a threatening stimulus on the next trial. This pattern corresponds precisely to the expected computational analogue of a startle reflex.

To simulate this scenario, agents were initialized with external state priors (D) assigning equal probability to flower or spider stimuli. A prolonged safe period was first induced by presenting a flower stimulus for the initial 14 trials, establishing a relaxed baseline state.

At trial 15, a single unexpected spider (aroused) stimulus was introduced, followed by a return to flower stimuli for the remaining 85 trials. This design allowed assessment of how agents’ expectations and interoceptive dynamics recover—or remain biased—after an isolated threat.

Results are reported as the median across agents within each group: 40 healthy agents ($\alpha = \beta = 0.9$) and 40 lesioned agents on internal precision ($\alpha = 0.9, \beta = 0.5$).

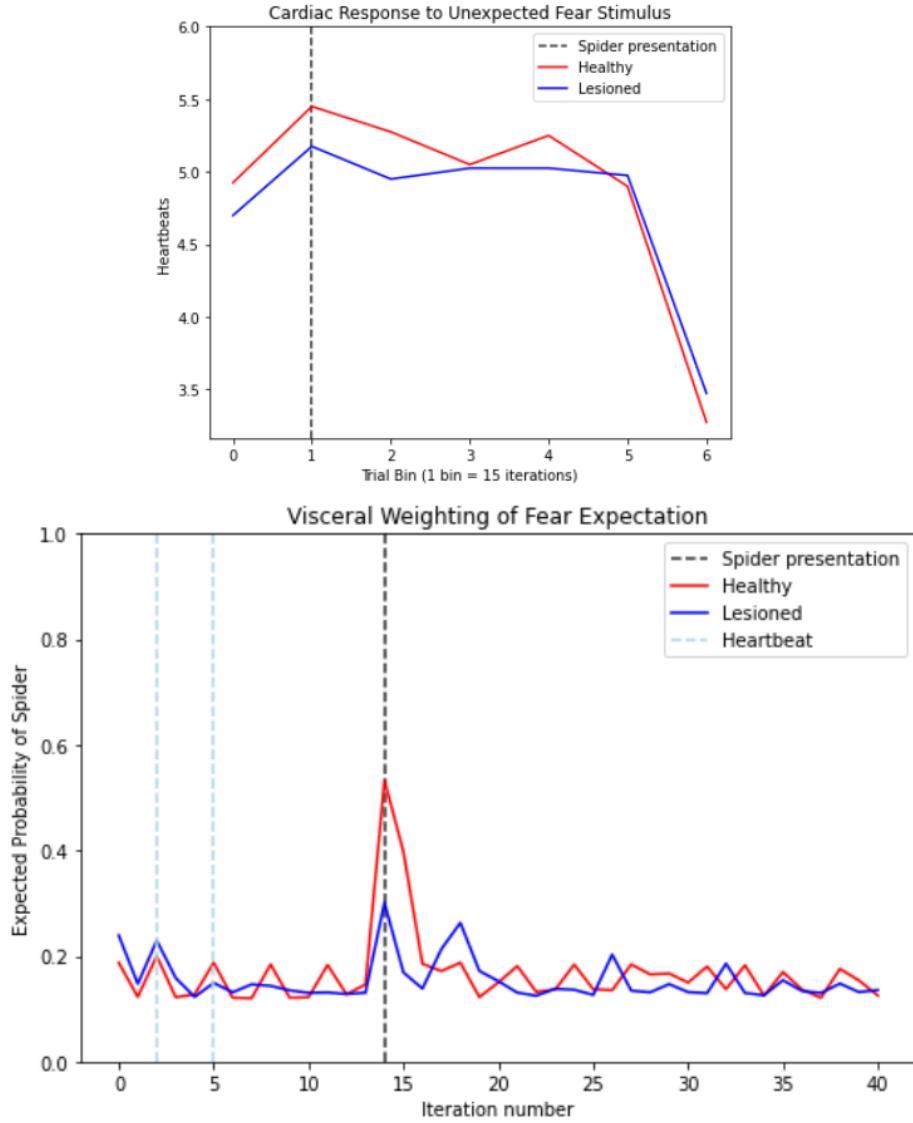


Figure 4: Graphical result of the second simulation

This abrupt change from the previously stable context triggered a strong and adaptive updating of the agent's beliefs. For a healthy agent, this sudden shift in both perceptual and interoceptive states corresponds to the expected

response of a startle reflex: upon receiving the unexpected spider observation, the agent rapidly converges on the inference that the environment now contains a threatening stimulus, accompanied by an immediate shift to an aroused physiological state, reflecting a simulated increase in heart rate.

In contrast, this response is attenuated in the lesioned agents. Following the unexpected observation, the agent initially remains in a relatively relaxed physiological state, with no immediate cardio-acceleration (figure 4, lower part). The perception of the arousing stimulus is attenuated and temporally delayed, leading to weaker and slower belief updating compared to a healthy agent.

Within the active inference framework, these physiological changes are not passive outcomes of perception but result from policy selection aimed at minimizing expected free energy.

2.3 Simulation 3: individual differences and allostatic construction

The final simulation concerns the theoretical framework of emotion. According to the Theory of Constructed Emotion of Lisa Feldman Barrett [3], emotions are not internal states but predictive inferences about the causes of interoceptive signals, constructed through the integration of bodily sensations and contextual information.

To show this idea, two agents with identical sensory precision were simulated, differing only in their prior preference:

1. Standard agent: this agent represents a typical profile where a flower is associated with a preferred state of “relaxed”, while a spider is associated with a preferred state of arousal.
2. Inverted agent: this agent represents an atypical emotional construction, with a phobic (aroused) response to flowers and a specialized interest in spiders, where the spider represents a safe/relaxing stimulus.

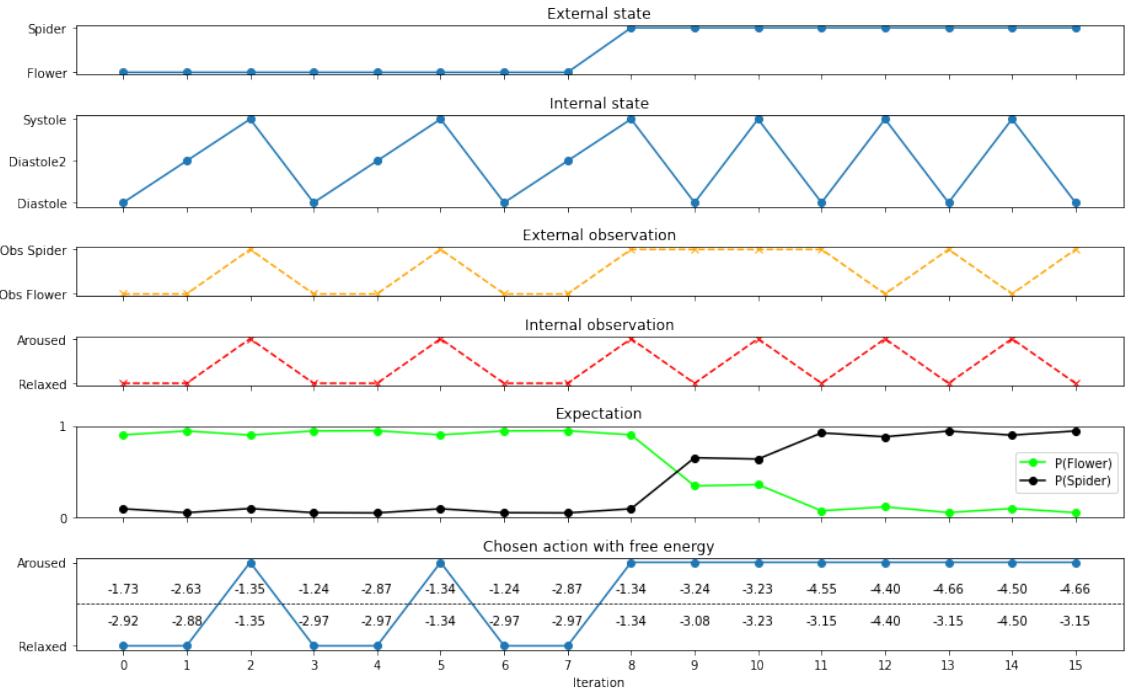


Figure 5: Hidden states and observations of the typical agent

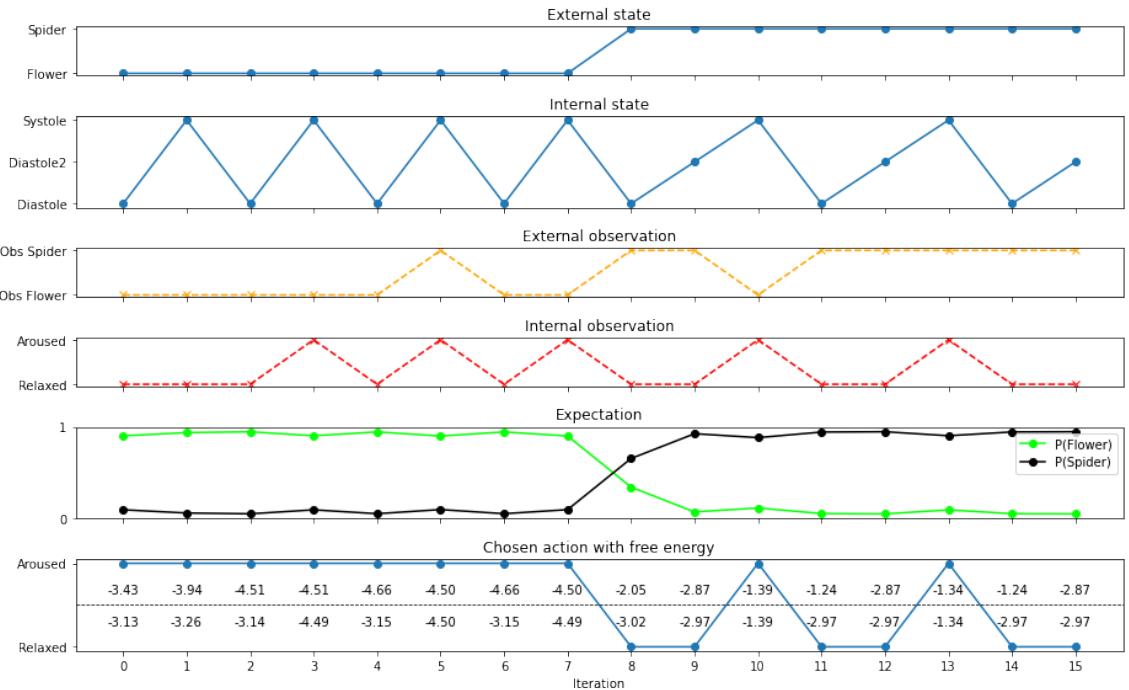


Figure 6: Hidden states and observations of the atypical agent

Both agents were exposed to identical sequences of external stimuli, ensuring that any differences of behavior arise only from differences in conceptual priors.

These results illustrate that “anxiety” or “excitement” are not directly related to the stimulus but are constructed by the agent’s need to align its interoceptive state with its contextual expectations. For the Inverted Agent, high arousal in a “Flower” context is not a malfunction but the fulfillment of its predictive model with respect of its prior preferences.

It’s interesting to note that the first agent’s external belief update (the transition from “Flower” to “Spider”) is not instantaneous, even in presence of exteroceptive evidence. This “perceptual lag” occurs specifically when the observation is sampled during the Systole phase of the internal state (figure 5, iteration 9). The agent’s posterior probability shifts to the “Spider” state once the internal cycle returns to the Diastole phases, where the precision of the external sensory channel is restored.

This simulation shows that, within the Active Inference framework, emotional experience comes from the agent’s effort to reduce Expected Free Energy by moving toward states its internal model considers “normal” or “safe” in a given context.

References

- [1] Micah Allen et al. “In the Body’s Eye: The computational anatomy of interoceptive inference”. In: *PLOS Computational Biology* 18.9 (Sept. 2022), pp. 1–27. URL: <https://doi.org/10.1371/journal.pcbi.1010490>.
- [2] Conor Heins et al. “pymdp: A Python library for active inference in discrete state spaces”. In: *Journal of Open Source Software* 7.73 (2022), p. 4098. DOI: 10.21105/joss.04098. URL: <https://doi.org/10.21105/joss.04098>.
- [3] Lisa Barrett. “The theory of constructed emotion: an active inference account of interoception and categorization”. In: *Social Cognitive and Affective Neuroscience* 12 (May 2017). DOI: 10.1093/scan/nsx060.