Non-Linear Computational Modeling of Emotional-Like States in Artificial Agents

Côme Bruneteau
come.bruneteau@toaq.fr

February 22, 2026

Abstract

Modelling of internally generated affective dynamics without explicit symbolic encoding remains a central challenge in the long-term quest for Artificial Consciousness. This paper proposes a unified computational framework for generating non rule-based emotional dynamics, moving beyond symbolic rule-based approaches. We define emotion as the emergent result of a weighted, non-linear integration of heterogeneous cognitive factors, including social pressure, internal metabolic states, and memory traces.

Our architecture relies on continuous transfer functions (hyperbolic tangent, sigmoid) and a Softmax-based competitive mechanism to model resource allocation and emotional saturation. We further introduce an endogenous stochastic modulation term, characterized by $1/f$ noise, to ensure trajectory uniqueness and avoid deterministic redundancy.

Simulation results demonstrate the model's capacity to generate distinct, stable psychological profiles (e.g., *Extremely Sociable* vs. *Very Stressed*) and to reproduce biological properties such as hysteresis and resilience. These findings suggest that "high-intensity affective functional state" can be rigorously defined as a functional state of high-intensity cognitive integration, providing a scalable foundation for autonomous ethical agents.

Keyword Emotion Computation, Emotional Modelization, Cognitive Integration

Contents

1 Introduction

This work focuses on modelling the emotional system. A key feature of a conscious system is the ability to experience emotions, as explained in the Emotional State and Regulation (ESR) axiom [1]. Here, we use the term emotion to refer to an internal state resulting from the integration of cognitive, environmental and memory factors.

Furthermore, the term "non rule-based" does not refer to a metaphysical position, but to the absence of explicitly coded symbolic rules, such as those found in BDI architectures [2], for producing a given emotional state (if-then).This document demonstrates how an artificial architecture operates with functional emotions, an approach aligned with seminal work on affective computing [3].

We hypothesise that an artificial system develops an internal emotional state with the functional properties of a system-level affective state, such as continuity, temporal dynamics, contextual modulation and memory dependence, inspired by the homeostatic regulation mechanisms described by Damasio [4].

To achieve this system-level affective state, we establish a mathematical framework for defining emotions. Architectural and material considerations are reserved for a separate document; they are mentioned here only to contextualise the use of the proposed mathematical framework.

This document reviews the basis of the formulas employed, compares them with biological emotions, details the creation of distinct psychological profiles, and finally presents the verification of our theories, calculations and postulates.

These three levels are designed to be complementary: the mathematical framework defines the formal constraints, the biological interpretation guides its plausibility, and the hypothesis on system-level affective state explores its implications.

The main contribution of this work is the proposal of a unified mathematical framework for modelling emotional states as a weighted and dynamic integration of heterogeneous cognitive functions, explicitly incorporating memory, emotional saturation and computational stability.

2 Mathematical framework : From Non-Linear Formula to Usable Formula

We establish a mathematical framework capable of modelling human emotions, thereby enabling its use within an architecture that already has cognitive functions. This architecture is discussed in a future document highlighting our work in this area of research. For the purposes of this document, we assume that we have a functional architecture with its own cognitive functions and that there is internal communication enabling us to retrieve the information needed for this document. We define several formulas, which we describe and study. There are two categories of functions. We study the critical functions for this model, such as the central functions, and we not study the functions that serve to refine and validate the central functions. Our definitions will be based on postulates, theories and logical aspects in order to highlight our formulas. It is important for us to work on the psychological and biological aspects in order to approach the human emotional system.

Definition 1: A cognitive function is a specific mental ability that enables us to perform a given task, such as concentrating, memorising, understanding language or solving a problem. There are sets and subsets of cognitive functions that work together to solve more complex problems. To illustrate this definition, we will consider adaptation to a situation that involves executive cognitive functions, a set of cognitive functions. More specifically, it involves cognitive flexibility, which allows us to modify our behaviour or thinking in response to changes and unexpected events.

Definition 2: Let t be the time parameter, which is modelled as a continuous variable representing the intrinsic evolution of the system's emotional state. In the implementation, this continuous dynamic is observed and integrated adaptively, based on cognitive and environmental events, allowing for non-uniform discretisation of time.

Definition 3: Let $\alpha_{i,j}$ and b_i be coefficients bounded in $[-1, 1]$ to represent the directional influence and basic bias of an emotional component, respectively, while ensuring a homogeneous and controlled scale of contributions. Let

$$\omega_i(t) = b_i + \sum_{j\in|\mathcal{J}_i|} (\alpha_{i,j} * F_{i,j}(t)) \tag{1}$$

be the total weight associated with a subset of cognitive functions $\mathcal{J}_i = \{F_{i,1}, F_{i,2}, \cdots, F_{i,j}\}$ and, by definition, $\alpha_{i,j}$ are the coefficients associated with the cognitive functions in the set $\mathcal{J}_i$. The cognitive functions $F_{i,j}(t)$ are normalised over the interval $[0, 1]$, not as probabilities in the strict sense, but as relative degrees of activation, analogous to confidence scores or the presence of a cognitive factor. This normalisation also makes consistent the subsequent use of a softmax function, which interprets the relative weights ω_i as competitive energies producing a normalised importance distribution λ_i, an approach based on the probabilistic interpretation of neural networks [5]. We will need to work on the maximum and minimum values of ω_i to enable us to study the functions. Let us begin by working on the minimum of ω_i:

$$\min_{t\in\mathbb{R}^+}\{\omega_i(t)\} = \min_{t\in\mathbb{R}^+}\left\{b_i + \sum_{j\in|\mathcal{J}_i|}(\alpha_{i,j} * F_{i,j}(t))\right\} = b_i + \min_{t\in\mathbb{R}^+}\left\{\sum_{j\in|\mathcal{J}_i|}(\alpha_{i,j} * F_{i,j}(t))\right\} \tag{2}$$

However, the minimum value of ω_i will depend on the value of $\alpha_{i,j}$, which, as we recall, are the coefficients of the cognitive functions. Therefore, the minimum value of ω_i is the value when $\forall j \in [\![1, |\mathcal{J}_i|]\!], \alpha_{i,j} * F_{i,j}(t) = -1$, i.e.,

$$\min_{t\in\mathbb{R}^+}\{\omega_i(t)\} = b_i + \min_{t\in\mathbb{R}^+}\left\{\sum_{j\in|\mathcal{J}_i|}(\alpha_{i,j} * F_{i,j}(t))\right\} = b_i - n \tag{3}$$

We define $n = |\mathcal{J}_i|$ and we know that $b_i \in [-1, 1]$, hence $\min_{t\in\mathbb{R}^+}\{\omega_i(t)\} = -1 - n$. We will now work on the maximum, which will simply be the opposite of the minimum, hence

$$\max_{t\in\mathbb{R}^+}\{\omega_i(t)\} = -\min_{t\in\mathbb{R}^+}\{\omega_i(t)\} = -(-1-n) = 1+n \tag{4}$$

We conclude that $\omega_i \in [-1-n, 1+n]$. Furthermore, we must consider that n is fixed at the start of the algorithm and cannot be modified without restarting it, so we will obtain a fixed interval that does not depend on a variable at each start.

Postulate 1: *It is considered that if at a fixed t we have* $\omega_i(t) = \begin{cases} \min_{t\in\mathbb{R}^+}\{\omega_i(t)\} \\ \max_{t\in\mathbb{R}^+}\{\omega_i(t)\} \end{cases}$ *then we are in an extreme case that is excluded because it corresponds to total saturation of a component, which is incompatible with adaptive emotional dynamics.*

2.1 Terminological Clarification: Emotion vs. Emotion-Related Variables

In this framework, variables labelled with emotional terminology (e.g., Sadness, Fear, Nostalgia) do not represent emotions as subjective or conscious states. They denote emotion-related internal variables, corresponding to memory traces, affective biases, or residual regulatory states resulting from past interactions.

These variables act as inputs to the emotional dynamics, not as emotional states themselves. The emotional state emerges only after the non-linear integration and competitive normalization process defined by the model.

The objective through ω is to consider as many important factors as possible in the emotional modelling of our individual. To do this, we have defined a kind of rule by weight in order to be able to categorise them correctly while including as many factors as possible. Here is our classification with the cognitive functions included in each weight ω:

Weights (ω_i)	Category	b_i	Detailed formulas
ω_1	Emotional basis	0.5	b_1
ω_2	Cognitive state	0.3	$b_2 + \alpha_1 \cdot Co_{LOAD}(t) + \alpha_2 \cdot Uncertainty(t) + \alpha_3 \cdot Fatigue(t)$
ω_3	Social Pressure	0.4	$b_3 + \alpha_1 \cdot Proximity(t) + \alpha_2 \cdot Feedback(t) + \alpha_3 \cdot Relevance(t)$
ω_4	Motivation	0.3	$b_4 + \alpha_1 \cdot Motivation(t) + \alpha_2 \cdot GoalAlignment(t) + \alpha_3 \cdot Frustration(t)$
ω_5	Affective memory traces	0.8	$b_5 + \alpha_1 \cdot Sadness(t) + \alpha_2 \cdot Fear(t) + \alpha_3 \cdot Nostalgia(t)$
ω_6	Sensory inputs	0.4	$b_6 + \alpha_1 \cdot Comfort(t) + \alpha_2 \cdot Pleasure(t) + \alpha_3 \cdot Fatigue(t)$

Table 1: Classification and details of weight formulas ω_i

By respecting the categorisation of ω_i, we have our parameters integrated into the respective formula of ω_i, which correspond directly to the appraisal dimensions established in cognitive psychology [6, 7].

Through ω_2, we find three functions that are defined as follows: Co_{LOAD} defines the current cognitive load on our entity, which can be represented by the amount of data to be processed, the number of simultaneous sensory inputs, or even the workload to be performed, $Uncertainty$ refers to the mental condition of the entity due to its lack of knowledge, clarity, or confidence and $Fatigue$ defines the fatigue of our entity, which is represented in an artificial setting by the battery percentage, which we reduce to a ratio $f_a \in [0, 1]$.

Next, for ω_3, we find $Proximity$, which describes the operational proximity to the individuals in front of our entity. This can be based on past discussions, affection, or the well-being of the current discussion. $Feedback$ defines the response to what our entity expresses, whether negative or positive. This allows our individual to follow codes and condition themselves to live in society. And $Relevance$ expresses the importance of the comments made to our entity. To draw a parallel, if a stranger judges us, we will not attach great importance to it, whereas if it is a friend, we will attach importance to it.

We will now look at ω_4, which is composed of $Motivation$, which describes the state of motivation of our entity based on several considerations. We find direct motivation, which is motivation at the moment, and indirect motivation, which describes the state over a longer period of time (generally, it is indirect motivation that is present when a project is realised). Also, $GoalAlignment$ describes the alignment with our short- or long-term goal. This factor takes various considerations into account and can have several states at once. And $Frustration$ defines the frustration that an entity may feel when a project is blocked. This will allow the entity to develop persistence, for example.

Next, we will look at ω_5, which has the following functions: $Sadness$ encodes a negative-valence affective memory trace. This sadness is not an emotional state but is stored in the entity's emotional memory. Also $Fear$ represents stored threat-related bias influencing future appraisal. And $Nostalgia$ encodes positively valenced autobiographical bias.

Finally, we will look at ω_6, which uses two new functions: $Comfort$, which describes our entity's comfort within a situation, taking into account the environment, the people involved, the context, and memory. $Pleasure$ describes our entity's pleasure in being present at a moment t. Since ω_6 describes the state of sensory inputs, this represents a good emotional state that can function perfectly.

Definition 4: It is important to note that we have created only six ω_i, but it is possible to create others. We then define $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$ as the set containing all the categorised ω_i. Furthermore, let $\mathcal{P} = |\Omega|$.

Postulate 2: *At $t = 0$, all cognitive functions are equal to 0, hence ω_1 is defined as the emotional basis.*

Definition 5: Let

$$\lambda_i = \frac{e^{\frac{\omega_i}{T}}}{\mathcal{Z}(\omega)} \quad \text{with} \quad \mathcal{Z}(\omega) = \sum_{j=1}^{n} e^{\frac{\omega_j}{T}} \tag{5}$$

where $i \in \{1, \ldots, \mathcal{P}\}$ represents the index of the emotional component and $T \in \mathbb{R}_*^+$ is the temperature parameter.

The use of Softmax normalization models a competitive resource allocation mechanism, analogous to attentional bottlenecks in cognitive systems. The temperature T regulates the entropy of the distribution:

- High T: The system tends towards equiprobability (high entropy), simulating a diffuse or uncertain emotional state.
- Low T: The system converges towards a "winner-takes-all" dynamic (low entropy), simulating a distinct, dominant emotional state.

The standard analysis of the Jacobian J of the Softmax function confirms that λ_i is locally Lipschitz continuous with a constant $L = \frac{1}{2T}$. This bounded spectral norm ($\|J\|_2 \leq \frac{1}{2T}$) guarantees the robustness of the system against perturbations in inputs ω, a critical requirement for non-linear dynamic stability [8].

Furthermore, the partial derivative analysis $\frac{\partial \lambda_i}{\partial \omega_i} = \frac{1}{T}\lambda_i(1-\lambda_i)$ highlights a biologically consistent sensitivity profile:

1. **Maximal Sensitivity (Neutral Zone):** The gradient is maximized when $\lambda_i \approx 0.5$. This implies that the agent is most reactive to stimuli when in a balanced, neutral affective state.
2. **Saturation (Emotional Inertia):** As $\lambda_i \to 0$ or $\lambda_i \to 1$, the gradient vanishes. This models "emotional saturation," where an agent experiencing an intense emotion becomes less sensitive to marginal variations in other inputs.

Since inputs ω are bounded within the hypercube $\mathcal{D} = [-1-n, 1+n]^{\mathcal{P}}$, the output λ_i is strictly confined to $[\lambda_{min}, \lambda_{max}]$. To prevent the system from locking into a purely deterministic state (where a single component becomes unitary, $\lambda_i \approx 1$), we impose a design constraint on the maximum input deviation $(b-a)$. By setting a saturation threshold $\Lambda = 0.99$ (representing the maximum allowable dominance of one emotion), we derive the following structural condition:

$$b - a \leq T \ln\left(\frac{\Lambda(n-1)}{1-\Lambda}\right) \tag{6}$$

This condition ensures that the system maintains a minimum degree of cognitive fluidity, preventing pathological rigidity in the emotional trajectory.

Now that we have defined our variables, factors and cognitive functions, we will define the formulas that will be integrated into our architecture. Next, we will take the time to draw an analogy with biology using logic, theory and fundamentals.

Note: The functions use the variable t to establish the logic that the functions retrieve values at time t. From a mathematical point of view, we use this variable as a computational indicator in order to retrieve information from cognitive functions at time t. This will be highlighted by the definition using emotional memory (see definition **??**), where we will work on three time periods ($t_1 = t+1$, $t_2 = t_1 - 1$, $t_3 = t_1 - 2$).

Definition 6: Let

$$E(t) = \sum_{i=1}^{n} (\lambda_i * \omega_i) \tag{7}$$

the function used to find the current value of the emotional state. This function $E(t)$ depends heavily on two parameters λ_i and ω_i, which we have studied (see Definition **??** and **??**). We will define the bounds (if they exist), the *max* and the *min*. As a reminder, we have

$$\lambda_{i,max} = \frac{e^{\frac{b}{T}}}{e^{\frac{b}{T}} + (n-1)e^{\frac{a}{T}}} \tag{8}$$

$$\lambda_{i,min} = \frac{e^{\frac{a}{T}}}{e^{\frac{a}{T}} + (n-1)e^{\frac{b}{T}}} \tag{9}$$

$$\omega_i \in [-1-n, 1+n] \tag{10}$$

We then have the weighted sum of two bounded parameters, so the maximum of $E(t)$ is reached by a corner vector of the type $\omega^* = (b, a, a, \cdots, a)$:

$$\begin{aligned} E_{max}(t) = \max_{t\in\mathbb{R}^+}\left\{\sum_{i=1}^{n}(\lambda_i * \omega_i)\right\} &\iff E_{max}(t) = \lambda_{max}(\omega_{max}) * \omega_{max} + \sum_{i=2}^{n}(\lambda_{i,min} * \omega_{i,min}) \\ \iff E_{max}(t) = \frac{e^{\frac{b}{T}}}{e^{\frac{b}{T}} + (n-1)e^{\frac{a}{T}}} * b + (n-1)&\frac{e^{\frac{a}{T}}}{e^{\frac{b}{T}} + (n-1)e^{\frac{a}{T}}} * a \\ \iff E_{max}(t) = \frac{e^{\frac{b}{T}}b + (n-1)e^{\frac{a}{T}}a}{e^{\frac{b}{T}} + (n-1)e^{\frac{a}{T}}} \end{aligned} \tag{11}$$

To determine the minimum obtained at the opposite corner, $\omega = (a, b, b, \cdots, b)$ or, in other words, we have:

$$E_{min}(t) = \frac{e^{\frac{a}{T}}a + (n-1)e^{\frac{b}{T}}b}{e^{\frac{a}{T}} + (n-1)e^{\frac{b}{T}}} \tag{12}$$

Therefore, $E(t)$ is bounded by a minimum and maximum defined by parameters such as the chosen interval, from which $\forall T \in \left]0, \infty\right[, a \leq E(t) \leq b$. As a reminder, T the temperature allows us to better control our function, for example when $T \mapsto \infty \iff E(t) \mapsto b$. From a cognitive point of view, we have defined the function $E(t)$ as corresponding to the integrated perception of oneself at time t or the emotional valence of the system. In other words, $E(t)$ represents the weighted sum of all cognitive processes of the emotional system.

Definition 7: Let

$$E_s(t) = \ln(1 + |E(t)|) = \ln\left(1 + \left|\sum_{i=1}^{n}(\lambda_i * \omega_i)\right|\right) \tag{13}$$

be the function used to define the current intensity of the emotional state. Before discussing this formula mathematically, let us define why we have defined it as such. This function, which defines intensity, is based on the function $E(t)$ by retrieving its value at time t. This value will be introduced into a natural logarithm, returning a value greater than or equal to 0. The natural logarithm in this function prevents an "explosion" of emotional intensity. A good parallel is that it is not possible to expend more dopamine than we have; similarly, when we experience an emotion of extreme intensity, there is a limit to the resources we can use, in accordance with the neurobiological constraints of affective systems [9]. It is important to understand that E_s returns a coefficient $i \in [0, \infty[$, which will help us, through sensory interpretation of the situation, to measure the intensity that we must send to our artificial consciousness system. Therefore, to fully understand this function, we need only view it as the creation of an emotional intensity coefficient in a system that interprets situations in relation to its psychological profile. However, the function does not have a max (characteristic of ln), which means that there is no theoretical limit. To solve this problem, we will establish that an emotion has a maximum duration at a state of extreme intensity. To illustrate this point, it is important to consider that an emotion has an activation duration, during which we can visualise the emotion as a parabolic function F_p, with $F_p \in \mathcal{C}^1$, with a coefficient defined by the intensity at the moment of activation. Furthermore, it should be noted that we must have a function of class $\mathcal{C}^1$ because we will need to study the derivative of F_p in order to interpret whether the curve is increasing, constant or decreasing. This study will therefore allow the system to understand whether it is increasing in intensity in an emotion e or whether it is calming down. We mentioned a coefficient that will allow us to adjust the parabolic curve. The purpose of this coefficient is to approximate the desired actual intensity; the larger the coefficient, the faster the curve will reach its maximum. Furthermore, this function has a minimum, as we will demonstrate. First, we know from equation 4 that the term that varies is the sum, so we will apply a condition to it to determine the minimum of $E_s(t)$.

$$\sum_{i=1}^{n}(\lambda_i * \omega_i) = 0 \Leftrightarrow \forall i \in \mathbb{N}, \omega_i = 0 \tag{14}$$

$$\min_{t\in\mathbb{R}^+}\left\{\left|\sum_{i=1}^{n}(\lambda_i * \omega_i)\right|\right\} = 0 \Rightarrow \min_{t\in\mathbb{R}^+}\left\{\ln\left(1 + \left|\sum_{i=1}^{n}(\lambda_i * \omega_i)\right|\right)\right\} = \ln(1+0) = 0 \tag{15}$$

To achieve this, it suffices for all ω_i to be equal to 0, which is possible, for example, at initialisation. Now it is important to check that this function is continuous at every point on the set defined on the natural logarithm. In our case, the absolute value is used precisely to maintain continuity at all points and at all times, regardless of the value returned by $E(t)$, hence $E_s(t)$ is continuous at all points, as shown by the following inequality:

$$\left|\sum_{i=1}^{n}(\lambda_i * \omega_i)\right| \geq 0 \Rightarrow 1 + \left|\sum_{i=1}^{n}(\lambda_i * \omega_i)\right| \geq 1 \Rightarrow \ln\left(1 + \left|\sum_{i=1}^{n}(\lambda_i * \omega_i)\right|\right) \geq 0 \tag{16}$$

This sequence of implications is the minimum and necessary condition for the continuity of our natural logarithm at every point. We then have the domain of definition $E_s \in \mathcal{D}_{E_s} = [0, \infty[$. In summary, we will make a biological link to our function $E_s(t)$. We know that emotions have characteristics that are unique to each emotion, so our function $E_s(t)$ will propose to bring intensity to our emotional system. The greater $E_s(t)$ is, the more our emotional system is using an emotion to its fullest. Furthermore, due to the continuity of our function, we will have, at all times and for all values of the emotional system, our intensity will return a value allowing our system to adapt its behaviour to represent what its subjective experience is experiencing.

Postulate 3: *There is a finite number M of cognitive functions available and usable at time t in a conscious being.*

Logic: *In a conscious being, we have a limited amount of information available at a given moment t. In principle, we consider that there is a process linked to the blocking of an excessively high cognitive load. This is the reason for including cognitive load in this emotional system. There is no exact count of the number of cognitive functions available and used by a conscious being, which is why we use the principle of parallel growth, which emphasises that the more cognitive functions we include, the closer the accuracy of $\phi(t)$ will be to an ideal result in an artificial conscious system. We will return to the availability and usability of cognitive functions at time t. In principle, we believe that we have a number x of cognitive functions present in the system and a number M of cognitive functions available and usable at time t, so in principle $x \geq M$ and there is a process defining the availability and usability of cognitive functions depending on the situation. Therefore, there exists a set $\mathcal{T}$ that contains all cognitive functions $S(t)$ available and usable by the system, and $M = |\mathcal{T}|$. In other words, $\mathcal{T} = \{F_{1,1}, F_{1,2}, \cdots, F_{i,j}\}$ where i is the i-th category of cognitive functions and j is the j-th cognitive function (see definition **??**).*

Definition 8: Let $c_i \in [-1, 1]$ be the weighting coefficient of the internal cognitive functions, and $\mathcal{S} = \{S_i(t) \mid i \in [\![0, M]\!]\}$ be the set of cognitive functions available at time t. We define the global cognitive score $S_g(t)$ as:

$$S_g(t) = \sum_{i=0}^{|\mathcal{T}|-1}(c_i \cdot S_i(t))$$

Assuming asymptotic behavior is excluded (i.e., $S_g(t) \mapsto \pm\infty$), we define $\Psi(t)$ as the normalized cognitive availability function:

$$\Psi(t) = \sigma(S_g(t)) = \frac{1}{1 + e^{-S_g(t)}}$$

This function $\Psi(t)$ maps the aggregated cognitive capacity into a bounded interval $[0, 1]$. Unlike a linear sum, the use of the sigmoid function allows us to model the saturation of cognitive capabilities.Proof of Existence ($|\mathcal{T}| \geq 1$):We posit that a conscious system requires at least one cognitive function to process information. Let us verify this by contradiction. If $\mathcal{T} = \emptyset$, then $S_g(t) = 0$ (null sum), implying a constant neutral activation $\Psi(t) = 0.5$ regardless of context, rendering the system incapable of distinctive problem solving (Definition **??**). Furthermore, if we assume an infinite number of active functions without normalization ($|\mathcal{T}| \to \infty$), the sum $S_g(t)$ would diverge, violating our biological constraints on finite energy. Thus, we deduce that $\mathcal{T}$ must be finite and non-empty: $\forall t, 1 \leq |\mathcal{T}| < \infty$.

Definition 9: Let $\zeta(t)$ be a stochastic variability term representing the biological uniqueness of the system (the properties and spectral density of $\zeta(t)$ are detailed in Section 5).We then define the global output as:

$$E_{out}(t) = \tanh\left(E(t) + \zeta(t)\right) \tag{17}$$

return the overall emotional state. This function is based on a hyperbolic tangent that belongs to $\mathcal{D}_{\tanh} = [-1, 1]$. So with the function E_{out} we force our system to remain within realistic physical limits. For example, even if $E(t) \mapsto \infty \Rightarrow E_{out} \mapsto 1$. Furthermore, considering the derivative, which is $(\tanh(x))' = 1 - \tanh^2(x)$, we can carry out a study on differential sensitivity. When $E \approx 0$, the slope is 1 (max), so the system is very sensitive to small changes. And when E is saturated (close to ± 1), the slope tends towards 0. The system therefore becomes less sensitive to small variations. So E_{out} shows that the impact of the uniqueness/chaos of ζ depends on the emotional system E.

Definition 10: Let $\beta \in \mathbb{R}$ be the dissipation time constant of neurotransmitters. Let

$$E_h(t+1) = \beta * E(t) + (1 - \beta) * E_h(t) \tag{18}$$

define the two-step emotional evolution, introducing a hysteresis characteristic of self-organised dynamic systems [10]. This function is characterised by its continuous nature; in physics, no signal changes state instantaneously. We have two possible cases for β: if β is large (close to 1), then the system is impulsive. It reacts strongly to the present moment and quickly forgets the past. If β is small (close to 0), then the system is resilient. It retains its past state for a long time. In biology, β should be observed as, as we have said, the dissipation time constant of neurotransmitters (such as the reuptake of serotonin or dopamine). Emotion does not disappear as soon as the stimulus stops.

In summary, this formalism does not merely define isolated variables. By incorporating inertia (β), non-linear saturation (tanh) and stochastic micro-variability (ζ), we have transformed a theoretical emotional response into a calculable dynamic system. This mathematical model is now ready to be transposed algorithmically to verify whether the emergence of complex behaviours observed in living organisms can be reproduced in silico.

3 Biological Inspiration and Functional Consistency

Before examining the mathematical properties of the proposed model (see Section 2), it is necessary to clarify the scope and role of the biological references used throughout this work. The parallels discussed in this section do not constitute a neurobiological validation of the model, nor do they aim at reproducing neural circuits, neurotransmitter dynamics, or anatomical structures. Instead, they are introduced as interpretative tools to assess whether the mathematical choices made remain functionally consistent with general principles widely observed in biological affective systems.

3.1 Weighted Integration as an Affective-Inspired Mechanism

In our modelling framework, several inputs are derived from variables commonly associated with human affective processes. These include internal variables (such as fatigue-related states), cognitive appraisals (goal alignment, uncertainty), and social or environmental feedback (e.g., perceived proximity or external responses). These variables correspond to appraisal dimensions identified in emotional psychology, which are known to modulate affective responses according to established theories, such as the Component Process Model [6], as well as broader findings in affective neuroscience [11].

Within the model, weighted integration is used to modulate the relative influence of these variables, enabling adaptability and inter-individual variability in affective responses. Such variability is functionally reminiscent of differences in emotional reactivity observed across individuals, without implying any structural or neurophysiological equivalence. This integrative mechanism is analogous, at an abstract level, to the convergence of multiple signals observed in neural systems, where affective responses emerge from the interaction of several contributing factors rather than from isolated inputs. Similar integrative principles have been described in models involving affect-related regions such as the amygdala or orbitofrontal cortex [4]. Accordingly, the proposed architecture goes beyond purely symbolic rules by implementing a computational structure that captures the convergence and modulation of cognitive and affect-related variables, while remaining at a functional level of description consistent with principles discussed in affective neuroscience [9].

3.2 Dynamic Affective-Like State Modelling

A fundamental property of biological affective systems is their dynamic nature: emotional states are not static entities, but continuously evolve under the influence of internal fluctuations and external stimuli.

In this context, the functions $E(t)$, $\zeta(t)$, and $E_{out}(t)$ are introduced to represent, respectively, an internal affective state variable, an intrinsic stochastic modulation term, and an integrated affective output. These functions do not claim to model neurochemical processes directly; rather, they implement abstract dynamical properties that are commonly observed in biological systems. In particular, affect-related biological processes such as dopaminergic or serotonergic modulation exhibit non-linear, bounded, and saturating behaviours in response to stimulus intensity and frequency [9].

By employing bounded non-linear functions, the proposed model ensures that affective responses remain limited and regulated, avoiding uncontrolled divergence. This design choice does not reproduce underlying biochemical mechanisms, but ensures that the resulting dynamics do not violate general constraints observed in biological affective regulation. As such, the model remains functionally consistent with known affective dynamics, while preserving its abstract and computational nature.

3.3 Cognitive-Affective Mapping Grounded in Psychological Theory

The integration of variables such as cognitive load, uncertainty, feedback, and goal alignment is directly inspired by well-established psychological frameworks, including cognitive appraisal theory [7, 6] and structural models of emotion [12]. Within the model, these variables act as modulators of affective state intensity and direction rather than as explicit symbolic representations of emotions.

For instance, parameters such as ω_2, representing the negative contribution of cognitive overload, or ω_4, representing the alignment between internal goals and external conditions, influence affective trajectories in a manner consistent with phenomena such as frustration, relief, or motivational engagement. This approach does not claim to encode discrete emotions, but establishes a functional correspondence between cognitive state variations and affective modulation, in line with prior work in computational affective modelling [3, 13].

3.4 Architectural Embedding and Operational Feasibility

Beyond its theoretical formulation, the proposed model is designed to be operationally embedded within artificial architectures such as TOAQ. Cognitive and affect-related variables can be connected to real-time system metrics or sensor-derived inputs, enabling continuous affective modulation. Examples of such functional mappings include:

$$\text{Camera or voice analysis} \Rightarrow \text{social feedback or proximity cues}$$

$$\text{CPU load or task queue metrics} \Rightarrow \text{cognitive load indicators}$$

$$\text{Internal process alignment} \Rightarrow \text{goal congruence variables}$$

These examples illustrate how the model can be instantiated within embedded or autonomous systems, without constituting empirical validation of emotional experience. Rather, they demonstrate the practical feasibility of integrating the proposed affective dynamics into artificial agents operating in real environments.

3.5 Limits of the Biological Analogy and Interpretative Scope

The parallels discussed in this section do not aim to establish neurophysiological equivalence between the proposed computational model and the human emotional system. Instead, they serve to show that the mathematical and architectural choices made are functionally consistent with principles widely observed in biological affective systems, such as integration, modulation, saturation, and variability.

While extensions such as affective memory may further refine the temporal coherence of affective responses, these mechanisms should be understood as computational enhancements rather than as analogues of human subjective experience. Accordingly, the present model provides a plausible interpretative framework for affective regulation in artificial agents, compatible with embodied and enactive approaches to cognition [14], without making claims regarding consciousness or phenomenological emotional experience.

4 Psychological Profile : From Definition to Choice

We will now be able to define values to create different psychological profiles, which will inevitably change the values in our results. Through this work, we want to enable the creation of different types of psychological profiles in order to create uniqueness for each entity. To do this, we will modify the values of the following variables: b_i and α_i, which will modify the result of all the functions and allow us to have a unique profile for the entity. In this work, the emotional profile is a component of the overall psychological profile, which itself contributes to the entity's personality. The initial objective is to create a functional, variable, modular emotional system that is unique to a machine, but also to be able to create a personality through these emotions. To do this, we must take into consideration that creating a personality is the result of various subjective and emotional processes, as well as memory. In this section, we will focus on the modelling and modularity of emotions in order to create different psychological profiles. It is essential that this step be completed before verifying our formulas, as we cannot create an emotional system that aims to approximate the human emotional system while neglecting the uniqueness and modularity of the emotional system. We have established that in order to create uniqueness, we will be able to modify the values of certain coefficients by determining key coefficients, which will allow us to obtain a number of possible emotional profiles. To verify the accuracy of our statements, we will define a new formula that highlights the modularity of our modifications.

Definition 11: Let

$$P_e() = E_s(0) * E(0) \tag{19}$$

define modularity based on two factors of emotional profiles.

The function $P_e()$ does not aim to describe a complete emotional dynamic, but constitutes an initial heuristic metric for characterising and comparing emotional profiles at the system initialisation stage. Choosing a product between valence $E(0)$ and intensity $E_s(0)$ makes it possible to distinguish between profiles with the same initial valence but different emotional intensities, or vice versa. We begin by establishing key profiles, or in other words, establishing the possibilities of the different emotional states that we can create by considering typical psychological profiles such as an antisocial, stressed or sociable, non-stressed individual, etc. A central question in this work is how can we extract an emotion that is understandable to our system from a series of mathematical formulas? To highlight this question, it is important to know that there are several variants, by modifying the values of cognitive functions and the psychological profile, allowing us to obtain the same result on a function such as E_{out}. In our view, the case where E_{out} produces the same result, even if the cognitive functions have different values in experiments 1 and 2, is differentiated by the formula E_s, where the intensity should be distinct between the experiments. We can still consider an improvement to the mathematical framework by including a system of neurons that learn responses adapted to the relevant factors. So, considering that our architecture has a finite number of cognitive functions and character parameters, and that its personality comes from a set of internal factors, including memory and emotional state. We will be able to define typical psychological profiles. We will first check the initial value, i.e. where the architecture is starting up (all cognitive functions return 0), then with typical states defined as the most extreme for our architecture. First, we will list the cognitive function currently available (see Table 1), within an architecture currently being created with the aim of creating a conscious system, then the positivity/negativity and finally the influence that this cognitive function should have.

Weights (ω_i)	Cognitive Functions	Positivity	Influence
ω_2 (Cognitive)	Co_{LOAD}	Negative	Strong
	$Uncertainty$	Negative	Moderate
	$Fatigue$	Negative	Moderate
ω_3 (Social)	$Proximity$	Positive	Strong
	$Feedback$	Positive	Moderate
	$Relevance$	Positive	Moderate
ω_4 (Motivation)	$Motivation$	Positive	Strong
	$GoalAlignement$	Positive	Moderate
	$Frustration$	Negative	Moderate
ω_5 (Memory)	$Sadness$	Negative	Strong
	$Fear$	Negative	Moderate
	$Nostalgia$	*Mixed*	Moderate
ω_6 (Sensory)	$Comfort$	Positive	Strong
	$Pleasure$	Positive	Moderate
	$Fatigue$	Negative	Moderate

Table 2: Detailed classification of cognitive functions, their positivity and their influence.

Note: Mixed means that, depending on the value returned by the function, the contribution of the cognitive function can have a positive or negative influence. The influence is defined according to relative criteria and determined by mathematical training aimed at maintaining consistency in the values of the formulas. Furthermore, positivity is established by the sign in front of the cognitive function.

We reiterate that the idea behind this mathematical framework is to be able to include new cognitive functions without having to modify pre-existing internal variables. It is important to note that when we add new cognitive functions, we will undoubtedly improve the accuracy of our mathematical framework. As we have mentioned several times, each cognitive function used within our architecture is not simulated and must come from a variable factor that cannot be controlled by the architecture. To highlight this, we can take the cognitive function $Fatigue$, which will use the remaining battery percentage to determine whether it is necessary to connect the system. In this example, we have a factor that cannot be modified or influenced by the architecture, so we can define the cognitive functions as follows:

Definition 12: Cognitive functions are external functions that cannot be directly modified by architecture, although their interpretation and integration depend on it.

With this idea, we will allow our architecture, and more specifically our emotional system, to be authentic and unique to itself. This notion of authenticity and uniqueness is very important within a conscious system. Indeed, if we consider uniqueness within our systems, then each cognitive function must return different results depending on the psychological profile and the subjective interpretation of the conscious entity.

We will now create five typical psychological profiles (extremely social, asocial, very calm, very stressed, and human-like) to verify the viability of our model. The initialization of these parameters is not arbitrary; it follows a phenomenological calibration strategy designed to map the relative topology of psychological traits onto our weight matrices. For each profile, we define:

- The **Base Weights** (b), representing the stable "personality traits" or long-term tendencies.
- The **Sensitivity Coefficients** (α), representing the "state dynamics" or immediate reactivity to stimuli.

1. Extremely sociable
2. Asocial
3. Very calm
4. Very stressed
5. Human-like

4.0.1 Creation of the "Extremely social" psychological profile

This profile is designed to model an agent whose internal state is heavily dependent on external validation.

- **Configuration:** We significantly increase the base intercept for Social Pressure ($b_3 = 0.55$) compared to the standard model. Furthermore, the sensitivity coefficients are set to high positive values ($\alpha_{3,1}, \alpha_{3,2} \approx +0.80$).
- **Justification:** A high b_3 ensures that social factors always have a baseline priority. The high positive α coefficients create a strong amplification loop: positive social feedback instantly dominates the global weight distribution, simulating a reward-seeking behavior typical of hypersociability.

4.0.2 Creation of the "Asocial" psychological profile

In contrast, this profile minimizes the impact of social interaction.

- **Configuration:** The base intercept for Social Pressure is drastically reduced ($b_3 = 0.10$). The sensitivity coefficients are flattened ($\alpha_{3,\cdot} \approx 0.05$).
- **Justification:** With a b_3 close to zero, the social component contributes negligibly to the emotional sum $S_g(t)$, regardless of the input intensity. This mathematically effectively models "social indifference" or inhibition.

4.0.3 Creation of the "Very calm" psychological profile

This profile represents high emotional resilience and stability (inertia).

- **Configuration:** We maintain moderate base values ($b_i \approx 0.40$) but strictly limit the magnitude of all sensitivity coefficients ($\alpha_{i,j} \in [-0.1, 0.1]$).
- **Justification:** By keeping α values close to zero, the term $\sum \alpha \cdot F(t)$ becomes negligible. The weights ω_i remain close to their static base b_i, rendering the system unresponsive to sudden spikes in inputs (stress, noise). This simulates a phlegmatic temperament.

4.0.4 Creation of the "Very stressed" psychological profile

This profile models the collapse of cognitive capabilities under load.

- **Configuration:** The base cognitive weight is set low ($b_2 = 0.10$). Crucially, the sensitivity to Cognitive Load is set to a large negative value ($\alpha_{2,1} = -0.65$).
- **Justification:** The formula $\omega_2 = 0.10 - 0.65 \cdot \text{Load}(t)$ implies that as load increases, the weight ω_2 can become negative or null. This simulates the "tunnel vision" effect where stress actively suppresses cognitive regulation capability.

4.0.5 Creation of the "Human-like" psychological profile

This profile serves as the control group.

- **Configuration:** We utilize the standard values defined in Table 1 ($b_2 = 0.3, b_3 = 0.4$, etc.) with moderate sensitivity coefficients.
- **Justification:** This configuration balances the static personality traits (b_i) with dynamic environmental reactivity (α), providing a stable baseline for comparison.

4.1 Modularity and Uniqueness Based on Psychological Profile

A key point that we have developed in this document and in our work is the ability to create a certain degree of modularity between each entity in order to enable uniqueness. In this subsection, we will demonstrate that this system is modular and offers a number of possibilities. First, we need to define the two important terms used in this subsection.

Definition 13: Psychological modularity refers to the idea that an individual's personality and psychological functioning can be viewed as a set of relatively independent but interconnected modules.

This definition refers to two important aspects of a conscious being: personality and psychological functioning. We will take the time to explore this definition in greater depth so that we can then interpret it and demonstrate that psychological modularity exists within this mathematical framework. We therefore have several modules, each representing an area of psychological functioning (e.g. cognition, emotion, motivation, self-regulation, sociability, values). Furthermore, in modern personality theories such as the Big Five model [15] or the HEXACO model [16], we find the idea that personality is not a homogeneous block but a combination of dimensions (modules). With this aspect of modularity, we can explain intra-individual variability, which highlights that a person can be very extroverted but have very low emotional stability. In summary, modularity is the ability to break down a psychological profile into subsystems (modules) or traits, like building blocks that come together to form an overall profile.

Definition 14: Psychological uniqueness emphasises that even though modules/traits are common to all humans (for example, everyone has a degree of sociability and motivation), the precise combination of these modules is unique to each individual.

This definition, closely linked to modularity (see Definition **??**), shows that even if two individuals share a dominant trait (e.g. extraversion), they may still differ in other traits, such as stress management (anxiety). One of the foundations of uniqueness is that, during our early years, we may share more dominant traits with other individuals, but once we reach n years of age, we unconsciously reduce the number of shared traits. To explain this principle, we must consider that uniqueness and the emotional system (modules, traits) are based on an aspect that evolves in part through temporality and subjective experiences. However, the very definition of a subjective experience is that it is unique to each individual, that each individual has their own feelings about the experience. So, let us take a practical case to demonstrate that subjective experience has a strong influence on individuals' emotional framework. Let us consider two individuals who share two dominant traits, such as extraversion and self-confidence. Now that we have this dominance, let us simply consider that one of the individuals will, during their lifetime, experience a trauma that considerably and unexpectedly reduces their self-confidence. So, in this simple case, we had two individuals sharing two dominant traits, and through a simple subjective experience causing trauma, we are left with only one dominant trait.

Furthermore, these two definitions propel us into a world combining psychology and mathematics, which will highlight why our cognitive function values are between 0 and 1. We will simply pose a single consideration that will launch us into this exploration: let us now consider traits as continuous dimensions. This simple consideration highlights that each individual can be located anywhere on a continuum (e.g., very extroverted, moderately extroverted, or very introverted), as explained by dimensional psychometric models [15]. Mathematically, however, continuums are infinite, so theoretically there are an infinite number of possible psychological combinations. Now, let us leave theory behind and think about reality. Is this really possible? To answer this question, we must logically assume that no two individuals will ever have exactly the same psychological configuration, because their subjective experiences, life histories and cultural contexts are unique. And even if we wanted to "create" the same psychological configuration, we must not ignore the fact that we are not talking about a usable or modifiable factor, but rather a factor that is not necessarily intentional and even less necessarily conscious. To support our argument, if we consider only psychological modularity, we could say that the number of profiles is potentially infinite, but not necessarily so. However, if we add uniqueness (life history, subjective experiences), then the number of profiles necessarily becomes infinite, because each life trajectory is unique and cannot be perfectly reproduced. To draw a parallel, we can take the example of fingerprints, which are all based on the same "biological modularity" (ridges, patterns), but their exact combination is always unique. We can therefore see that it is by linking the two definitions that we obtain a continuous and necessarily infinite dimension. Although the profile space is theoretically continuous, any computer implementation produces a discrete approximation with finite resolution. To do this, we will simply consider one aspect: the subjective experience and memory of the entity. Indeed, the subjective experience combined with the entity's memory will work together to create a part of the personality that is therefore a fundamental aspect of a psychological profile, a process analogous to the construction of the "autobiographical self" described in neuroscience [4]. This is how we will be able to enter into our definition and the continuous dimension aspect of traits. However, a new problem arises: not only do we have to create an infinite number of psychological profiles (something that can be demonstrated by the uniqueness of personality), but we also have to allow our previously defined functions to be bounded in order to limit any overflow. This limit must not block emotional excesses, but must allow the formulas to remain usable. In section 2, we have mathematically demonstrated that there are limits for each of the functions through the hyperbolic tangent or natural logarithm functions. So, with an infinite number of personalities and bounded functions, these two conditions allow us to use the formulas. We will then verify our formulas using graphs and situations to check the postulates that we will define directly.

This section has shown that the proposed mathematical framework allows for the introduction of explicit psychological modularity, while guaranteeing the uniqueness of each artificial entity. By varying the internal parameters associated with cognitive and emotional functions, it becomes possible to generate a continuous space of emotional profiles, compatible with the dimensional models of modern psychology. Personality is thus not conceived as a set of discrete types, but as the dynamic result of the weighted integration of cognitive, emotional and memory modules.

Psychological uniqueness then emerges naturally from the combination of initial parameters and the temporal evolution of emotional memory, making each emotional trajectory unique, even with identical architecture. This framework thus makes it possible to reconcile structural generality and individual differentiation without resorting to fixed symbolic rules. This approach lays the necessary foundations for the analysis of observable emotional behaviours, which will be the subject of the following section dedicated to the results and practical application of the model.

5 Endogenous Stochastic Modulation

The objective of this section is to ensure that entities sharing identical psychological profiles do not evolve along strictly superimposable emotional trajectories. While psychological profiles are defined by the same parameter values, real cognitive systems exhibit intrinsic fluctuations that prevent perfectly deterministic dynamics. Our aim is therefore not to introduce structural individuality or subjective differences, but to model trajectory-level variability that emerges from internal system dynamics.

Psychological profiles are assumed to be stationary over short and medium time scales, characterised by fixed parameters α_i and b_i. However, stationarity does not imply strict determinism. Empirical studies of human cognition consistently show that similar individuals exposed to identical conditions do not generate identical affective or cognitive time series. This variability arises from endogenous fluctuations related to attention, arousal, and internal regulation processes, rather than from external randomness alone [17].

To capture this property while preserving profile stability, we introduce a stochastic modulation mechanism of internal cognitive micro-variability. This mechanism does not represent an additional emotional component, decision process, or personality parameter. Instead, it acts as a low-amplitude endogenous modulation influencing the dynamic trajectory of the system. Its function is to prevent deterministic collapse of the emotional dynamics, ensuring that identical initial conditions do not lead to strictly identical temporal evolutions.

Definition 15: Endogenous stochastic modulation is defined as an internal stochastic variable $\zeta(t)$ that cannot be directly observed and represents intrinsic cognitive fluctuations independent of external stimuli. This variable is not interpreted as encoding identity, personality, or subjective experience, but as a bounded endogenous factor that locally perturbs the system's trajectory.

The variable $\zeta(t)$ is assumed to have zero mean,

$$\mathbb{E}[\zeta(t)] = 0,$$

and finite variance,

$$\mathrm{Var}(\zeta(t)) = \sigma_\zeta^2,$$

where σ_ζ controls the amplitude of the fluctuations. Its temporal structure is characterised by long-range correlations across multiple time scales, modelled by a stochastic process with a $1/f$ power spectrum (pink noise). This choice reflects scale-free fluctuations commonly observed in complex cognitive systems, without introducing a dominant time scale or destabilising the underlying dynamics [18, 19].

Endogenous stochastic modulation acts exclusively as a secondary modulator of existing mechanisms, particularly those governing sensitivity to external inputs. It does not alter the structure of the psychological profile itself, nor does it introduce persistent individual differences. Instead, it ensures that entities sharing the same profile parameters may follow distinct, non-superimposable trajectories while remaining statistically equivalent over long time horizons.

5.1 Integration of Endogenous Stochastic Modulation into the Emotional Dynamics

The variable $\zeta(t)$ is integrated into the emotional dynamics as a weak, bounded perturbation that preserves overall stability and normalisation constraints. Its role is to introduce local endogenous variability without modifying the qualitative regimes of the system.

Cognitive micro-variability is modelled as a discrete-time stochastic process with a $1/f$ power spectrum, normalised such that

$$\mathbb{E}[\zeta(t)] = 0, \qquad \mathrm{Var}(\zeta(t)) = \sigma_\zeta^2. \tag{20}$$

This formulation allows for the coexistence of rapid fluctuations and slow drifts, avoiding the introduction of a characteristic time scale and remaining consistent with scaling laws observed in biological and cognitive dynamics [19].

5.1.1 Modulation of permeability to external influences.

To capture realistic variability in perceptual and attentional responsiveness, micro-variability is weakly coupled to the external permeability term $\Psi(t)$. An effective permeability is defined as

$$\Psi_{\mathrm{eff}}(t) = \mathrm{clip}\big(\Psi(t) + \kappa_\Psi\,\zeta(t),\, 0,\, 1\big), \tag{21}$$

where κ_Ψ is a low-amplitude coupling coefficient and $\mathrm{clip}(\cdot)$ enforces admissible bounds. This formulation reflects the fact that identical external stimuli may have slightly different effects depending on the instantaneous internal state of the system, without introducing systematic bias or instability [17].

Emotional weights $\lambda_i(t)$ are computed via softmax normalisation of intermediate variables $u_i(t)$. To preserve normalisation and stability, cognitive micro-variability is introduced at the level of these intermediate variables:

$$u_i(t) = u_i^{(0)}(t) + \eta_i\,\zeta(t), \tag{22}$$

$$\lambda_i(t) = \frac{\exp(u_i(t))}{\sum_j \exp(u_j(t))}. \tag{23}$$

The coefficients η_i control the sensitivity of each component to micro-variability and are chosen to remain sufficiently small so as not to induce artificial dominance or structural asymmetries.

In the specific case of the social component, a weak bidirectional coupling between the global emotional state and the corresponding weighting is introduced to capture contextual sensitivity:

$$u_3(t) = u_3^{(0)}(t) + \beta_{E\to 3}\, g(E(t)) + \eta_3\, \zeta(t), \tag{24}$$

where $g(\cdot)$ is a bounded centring function and $\beta_{E\to 3}$ is a low-amplitude coefficient. This interaction models limited coordination between emotional state and social engagement while preserving overall dynamical equilibrium, consistent with principles of coordination dynamics in complex systems [10].

The introduction of endogenous stochastic modulation does not alter the stationarity of psychological profiles or the qualitative dynamic regimes of the model. The variability it induces is local and trajectory-based: entities sharing the same profile may exhibit distinct temporal evolutions, while remaining statistically equivalent in expectation. The system is therefore deterministic on average, but not strictly deterministic at the level of individual trajectories.

6 Numerical Simulations and Dynamic Analysis

This section presents the results obtained by simulating our unified mathematical framework. Here, we demonstrate the model's ability to generate distinct dynamical regimes through parametric variation (Section 6.1), to maintain stable and dissipative emotional dynamics (Section 6.2), and finally, we discuss the functional role of the emerging system-level affective state (Section 6.3).

6.1 Parametric Sensitivity and Profile Differentiation

The primary objective of our modelling is to enable the emergence of psychological uniqueness within a common architecture. Rather than relying on rigid rules, we modulate the system's attractor basins by modifying only the structural parameters α_i (influence coefficients) and b_i (baseline bias). We simulated five typical profiles: *Extremely Sociable*, *Asocial*, *Very Calm*, *Very Stressed*, and *Human-like*.

Figure 1 illustrates the trajectories of emotional valence $E(t)$ for these profiles when faced with a standardized scenario: a phase of social stimuli ($t = 120$) followed by a phase of intense cognitive load ($t = 220$).

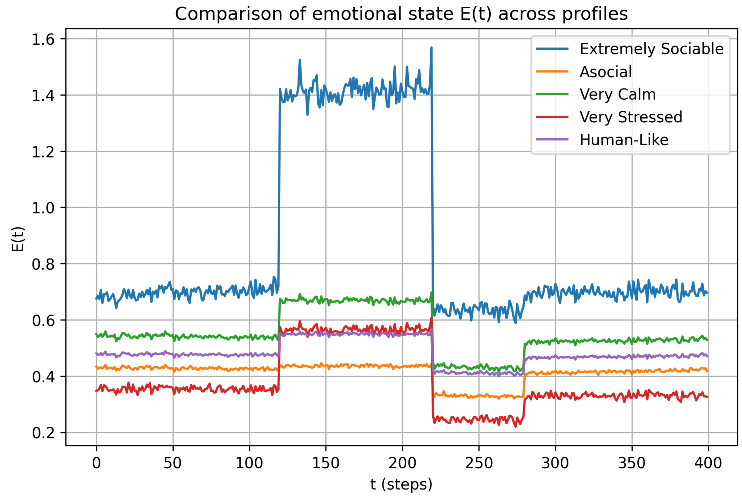
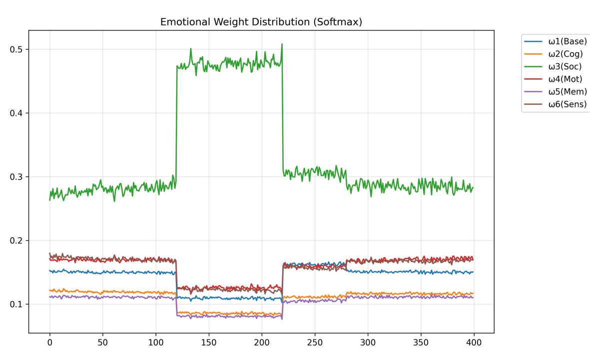


Figure 1: Comparison of emotional valence trajectories $E(t)$ across five distinct profiles. Note the divergence in reactions to the social stimulus ($t = 120$) and the stress event ($t = 220$). The Very Stressed profile (in red) collapses, while the Very Calm profile (in green) remains stable.

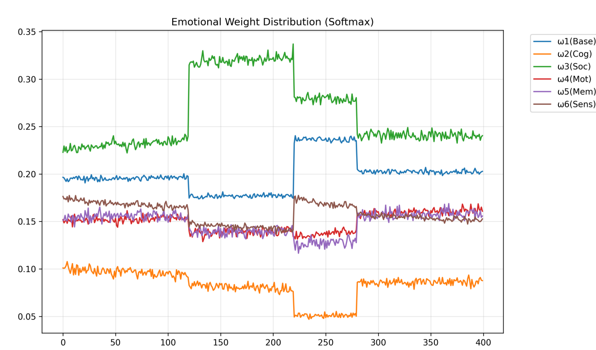
The results demonstrate a clear behavioural divergence based on parametric topology:

- The **Sociable** profile exhibits high gain sensitivity to social stimuli, with a valence $E(t)$ reaching 1.4, but shows high volatility, indicative of a system near instability.
- The **Asocial** profile remains effectively decoupled from social inputs (flat curve), validating the functional modularity of the ω_3 influence.
- The **Stressed** profile undergoes a phase transition akin to a collapse during cognitive overload ($t = 220$), falling to $E(t) \approx 0.2$.

This differentiation is driven by the internal distribution of attentional weights λ_i, as illustrated in Figure 2. The Softmax competition mechanism allows a dominant emotion to overwhelm the global workspace in the sociable profile, while the stressed profile illustrates a suppression of cognitive capabilities under load.



(a) Sociable Profile: Social Dominance



(b) Stressed Profile: Cognitive Breakdown

Figure 2: Comparison of the distribution of internal weights (λ_i). On the left, social emotion (green) dominates. On the right, cognitive load saturates the system.

6.2 Phase Space Analysis: Stability and Hysteresis

6.2.1 Hysteresis and Memory

Figure 3 highlights the system's trajectory in the phase space defined by the instantaneous emotional state $E(t)$ and the integrated mood $E_h(t)$. The formation of large, closed loops (limit cycles) indicates two fundamental properties:

1. **Temporal inertia:** For the same instantaneous valence, the internal state differs depending on whether the system is in a rising or falling phase. This non-linearity confirms the presence of hysteresis.

2. **Dissipative Stability:** The fact that the cycles close confirms that the system is dissipative: it is capable of absorbing the emotional energy and returning to its initial equilibrium point after the stimulus, thus avoiding chaotic drift [8].

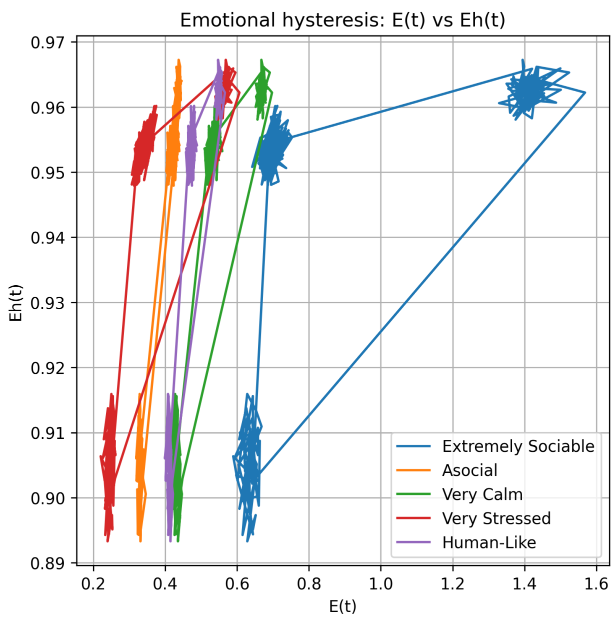


Figure 3: Representation in phase space of the Emotional State $E(t)$ vs Mood $E_h(t)$. The width of the cycles illustrates the intensity of emotional memory: the system retains a trace of the past event before returning to equilibrium.

6.2.2 Complexity and Entropy

Shannon entropy analysis on the vector λ (Figure 4) reveals the complexity of emotional processing. The *Very Calm* profile maintains high entropy (≈ 1.75), indicating a balanced integration of multiple factors (motivation, comfort, social). In contrast, unstable profiles tend towards low-entropy states, or mono-emotionality, reducing their adaptive capacity.

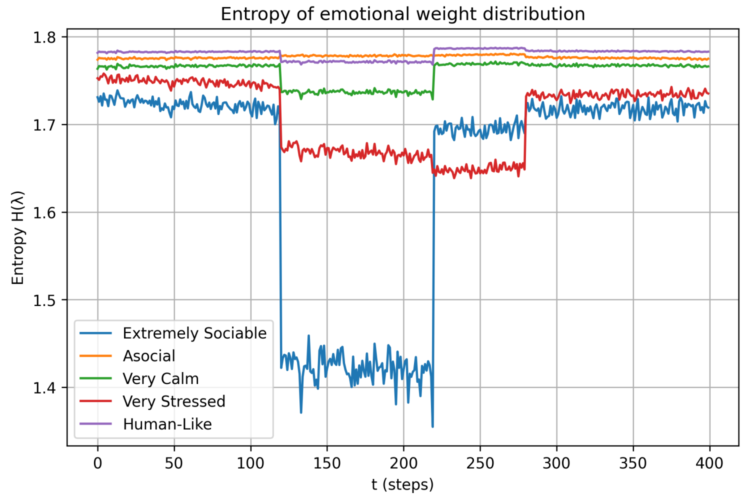


Figure 4: Shannon entropy of the emotional weight vector λ. Higher entropy (e.g., *Very Calm*) indicates balanced and resilient emotional integration.

6.2.3 Stochastic Trajectory Divergence

Finally, the integration of *Endogenous Stochastic Modulation* $\zeta(t)$ (introduced in Section 5) ensures that the system avoids deterministic redundancy. As shown in Figure 5, two entities sharing the same *Human-like* profile (Seed A and Seed B) follow the same global trend but develop unique micro-trajectories. This provides the system with a non-deterministic signature essential for realistic agent simulation [17].

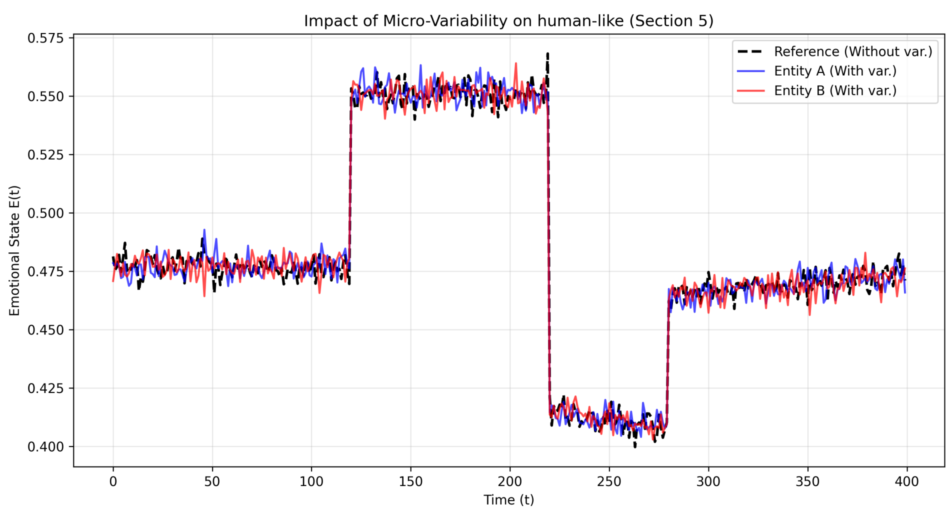


Figure 5: Impact of Endogenous Stochastic Modulation $\zeta(t)$ on the Human-like profile. The solid lines represent two distinct entities (Seeds A and B) diverging locally from the theoretical trajectory (dotted lines), simulating unique subjective experiences.

6.3 Functional Emergence: Intensity as a Global Control Signal

The results presented above validate the structural stability of the model. The final question is how these numerical states translate into a functional system-level affective state.

We reframe the "feeling" not as a qualia, but as a **Global Gain Control Signal**. In our architecture, this emergence is captured by the correlation between valence $E(t)$ and the logarithmic intensity $E_s(t)$.

Simulations indicate that $E_s(t)$ acts as a system-wide alert signal. When $E_s(t)$ exceeds a critical threshold (as observed in the *Sociable* profile's high-stimulation phase), the emotional information saturates the decision-making space. This forces a behavioural adaptation, effectively mimicking the "ignition" phenomenon described in Global Workspace Theory.

Thus, the system-level affective state is not explicitly hard-coded. It emerges from the system's dynamics: it is the self-regulatory observation of the competition between weights λ_i, modulated by the system's history (hysteresis) and its intrinsic noise (stochastic modulation).

7 Discussion

In line with the principles of open science, this section aims to summarise the contributions of our work, while making accessible the ethical questions and major debates raised during the development of this theoretical framework.

7.1 Synthesis and Scope of the Model

The work presented in this paper lays the foundations for a unified mathematical framework for modelling an unsimulated artificial emotional system. Unlike classical symbolic approaches (based on if-then rules), we have demonstrated that it is possible to generate complex and coherent emotional dynamics from continuous equations and an integrative architecture.

The major contributions of this research can be summarised in three points:

- **Mathematical formalisation:** The definition of non-linear transfer functions (hyperbolic tangent, sigmoid, Softmax) has made it possible to model essential properties of living beings such as saturation, cognitive competition and temporal inertia (hysteresis).
- **Bio-inspired anchoring:** We validated the functional consistency of the model by establishing strict parallels with neurobiological mechanisms (weighted synaptic integration) and modern psychological theories (*Appraisal Theory* [6]), ensuring that the system does not "mimic" emotion but reproduces its underlying mechanics.
- **Emergence of Uniqueness:** The introduction of micro-variability ($\zeta(t)$) and the parameterisation of profiles (α_i, b_i) have proven that behavioural uniqueness can emerge from a generic structure, without requiring explicit programming of personality traits.

7.2 Open Debates and Ethical Validation

During our work, we chose to incorporate critical debates into the research document itself. A central topic is the ethics and safety of such a project.

Within a system with non rule-based emotion, assessing "emotional concordance" is a major challenge. Based on cognitive appraisal theory [7], which states that emotions arise from a subjective appraisal of an external situation, we establish our validation algorithm on a causal basis: if our formulas are correct, there must always be a direct or indirect correlation between the internal state generated and the external environment.

It is crucial to note that this verification must take place in real time but without rigid "symbolic supervision", as this would contradict the objective of creating an autonomous entity. Validation must therefore be statistical and behavioural, not coercive.

7.3 The Bias Paradox: From Imperfection to Functionality

A fundamental question raised by Patrick Kamtchueng Kom during our discussions concerns the possibility of developing such a system without injecting our own human biases into it.

We postulate that inductive bias is not only inevitable but necessary for functional differentiation in limited-resource agents. To limit this impact, openness to scientific collaboration is essential. In the same way that a child's genetics result from the combination of its parents, increasing diversity, increasing the number of "parents" (researchers, designers) of a conscious system allows individual biases to be diluted in favour of more universal cognitive structures.

Finally, the nature of bias needs to be reconsidered. As René Descartes pointed out: *"Man is an imperfect thing that constantly strives for something better and greater than itself."* If human consciousness is imperfect, must artificial consciousness be infallible?

We argue that some "biases" are actually necessary features. The most compelling example is **attention bias** (or selective attention). This mechanism, which consists of ignoring a large part of the information in order to focus on what is relevant, is technically a bias. However, it is a sine qua non condition for cognitive efficiency in a resource-limited environment, a principle already theorised for artificial agents [20].

Thus, our approach does not aim to eliminate all biases, but to use these deviations in cognitive processing (such as selective attention or emotional prioritisation) as architectural levers to approach functional consciousness.

7.4 Future work

While the presented framework successfully models stable emotional dynamics, the current iteration relies on heuristic parameter initialization (b_i, $\alpha_{i,j}$). A primary avenue for future research is the transition from manual tuning to **automated parameter optimization**. We plan to implement Evolutionary Algorithms (EA) or Reinforcement Learning (RL) to dynamically adjust the weight matrices α in response to environmental rewards, thereby allowing the psychological profiles to emerge organically rather than being predefined.

Secondly, the validation presented here is purely simulation-based. Future work will focus on **empirical cross-validation** using biological datasets (e.g., DEAP or MAHNOB-HCI). We aim to compare the system's output $E_{out}(t)$ with physiological signals (such as Galvanic Skin Response or Heart Rate Variability) recorded in humans under similar stress conditions. This will allow us to quantify the "bio-fidelity" of the generated trajectories. x Finally, we intend to integrate this affective engine into an **embodied architecture** (robotics). The hypothesis is that the global gain control signal $E_s(t)$ should not only modulate internal processing but also influence motor actuation (e.g., speed, stiffness), creating a closed-loop system where physical embodiment constrains and shapes the emotional dynamic.

References

[1] C. Bruneteau, "The ethical problems and considerations of creating artificial consciousness," 2025.

[2] B. Steunebrink, M. Dastani, and J.-j. Meyer, "A formal model of emotion triggers: An approach for bdi agents," *Synthese*, vol. 185, 04 2012.

[3] R. W. Picard, *Affective Computing*. MIT Press, 1997.

[4] A. R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam, 1994.

[5] J. S. Bridle, *Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition*. Springer, 1990.

[6] K. R. Scherer, "Appraisal considered as a process of multilevel sequential checking," *Appraisal processes in emotion: Theory, methods, research*, vol. 92, p. 120, 2001.

[7] R. S. Lazarus, *Emotion and Adaptation*. Oxford University Press, 1991.

[8] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press, 2014.

[9] J. Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York: Oxford University Press, 1998.

[10] J. A. S. Kelso, *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT Press, 1995.

[11] J. LeDoux, *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Simon & Schuster, 1996.

[12] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

[13] S. Marsella and J. Gratch, "Ema: A process model of appraisal dynamics," *Cognitive Systems Research*, vol. 10, no. 1, pp. 70–90, 2009.

[14] A. K. Seth, "Interoceptive inference, emotion, and the embodied self," *Trends in Cognitive Sciences*, vol. 17, no. 11, pp. 565–573, 2013.

[15] R. R. McCrae and P. T. Costa, *Personality in Adulthood: A Five-Factor Theory Perspective*. Guilford Press, 2003.

[16] K. Ashton and K. Lee, *The H Factor of Personality: Why Some People are Manipulative, Self-Entitled, Materialistic, and Exploitive—And Why It Matters for Everyone*. Wilfrid Laurier University Press, 2012.

[17] D. L. Gilden, "Cognitive emissions of 1/f noise," *Psychological Review*, vol. 108, no. 1, p. 33, 2001.

[18] G. C. Van Orden, J. G. Holden, and M. T. Turvey, "On the origin of pink noise in cognitive performance," *Psychological Bulletin*, vol. 129, no. 5, p. 697, 2003.

[19] C. T. Kello, G. D. Brown, R. Ferrer-i Cancho, J. G. Holden, K. Linkenkaer-Hansen, T. Rhodes, and G. C. Van Orden, "Scaling laws in cognitive sciences," *Trends in Cognitive Sciences*, vol. 14, no. 5, pp. 223–232, 2010.

[20] A. Sloman and M. Croucher, "Why robots will have emotions," *IJCAI*, vol. 81, pp. 197–202, 1981.