The Cognitive Architecture of Hebbie

Daniel Mewes
Correspondence: mail@dmewes.com
LLMs were used for writing this document. Might contain mistakes.

Abstract

We present the theoretical formulation, neural mechanics, and embodied kinematics of **Hebbie**, an autonomous artificial life (ALife) agent operating within an open-ended, continuous 2D browser Document Object Model (DOM) ecosystem. Hebbie achieves adaptive, goal-directed behavior through the coupling of an online predictive world model with an internal homeostatic drive-reduction architecture. The cognitive substrate comprises three primary subsystems: (1) a multi-dimensional physiological deficit space $\mathbf{D} = [H, F, B]^T$ governing Hunger, Fatigue, and Boredom; (2) an associative neural network trained via normalized Delta-Hebbian plasticity with action-conditioned eligibility traces, which learns to anticipate prospective shifts in physiological drives $(\Delta\hat{H}, \Delta\hat{F}, \Delta\hat{B})$ conditioned on dual-rail sensory features; and (3) a simulation decider that computes hypothetical future homeostatic loss landscapes over candidate motor programs, selecting actions through a temperature-parameterized Boltzmann distribution.

---

1. Introduction & Theoretical Foundations

A central pursuit of artificial life and computational neuroscience is understanding how purposeful, adaptive behavior emerges in embodied agents without explicit external reward functions or exhaustive state-space enumeration. In classical computational reinforcement learning (Sutton & Barto, 2018), policy optimization relies on a scalar reward signal $r_t \in \mathbb{R}$ designed by an external engineer. While mathematically tractable, scalar reward formulations face well-documented limitations: reward hacking, the non-stationarity of intrinsic motivational states, and the biological implausibility of a monolithic objective function across fluctuating physiological requirements.

In biological organisms, behavior is anchored in **homeostasis**—the active maintenance of internal physiological variables within viable boundaries (Cannon, 1929; Ashby, 1952). Under the drive-reduction theories of Clark Hull (1943) and modern formulations of homeostatic

reinforcement learning (Keramati & Gutkin, 2014; Sterling, 2012), deviations from optimal physiological setpoints constitute motivational deficits. Drive reduction operates not merely as a retroactive reinforcer, but as the normative objective function of an internal predictive model (Friston, 2010; Pezzulo et al., 2015).

Hebbie synthesizes these principles into an embodied, browser-native agent. Rather than mapping perceptions directly to action-values via black-box value functions ($Q : \mathcal{S} \times \mathbf{A} \to \mathbb{R}$), Hebbie maintains a **predictive world model**. Conditioned on the current sensory state $\mathbf{s} \in$

$\mathbb{R}^{15}$, the network predicts the *causal physiological consequence* $(\Delta\hat{H}_a, \Delta\hat{F}_a, \Delta\hat{B}_a)$ of each candidate motor action $a \in \mathbf{A}$. An analytical simulation decider projects these predictions forward, evaluates the resulting homeostatic loss, and selects actions that maximally restore physiological equilibrium.

Learning occurs locally, continuously, and online through **normalized Delta-Hebbian plasticity** combined with multi-step eligibility traces. There are no global backpropagation passes, no replay buffers, and no external rewards. The agent learns the physics and affordances of its environment through direct somatic experience.

---

2. The Ecological Niche: Environment & Embodiment

Hebbie inhabits the DOM of a web page, transforming arbitrary textual content into a continuous, interactive foraging terrain.

2.1 Environmental Tokenization & Resource Taxonomy

The environmental parser (`WordManager`) traverses the DOM tree using a `TreeWalker` restricted to text nodes within designated content containers. Non-ecological DOM subtrees —including user interface controls (`#debug-panel`, `#info-modal`), structural navigation, scripts, styles, and Hebbie's own rendering layer—are strictly masked.

Text is tokenized using Unicode-aware regular expressions:
$\mathcal{T} = \{w_k \mid w_k \in ([\backslash p\{L\}\backslash p\{N\}]+)\}$

Each token w_k is wrapped in an inline DOM element and parsed into an ecological entity characterized by its bounding box rect $\mathbf{R}_k = [x_{\min}, x_{\max}, y_{\min}, y_{\max}]$, spatial center $\mathbf{c}_k$, and linguistic typography:

1. **Food Entities ($\mathcal{F}$)**: Tokens containing exclusively lowercase glyphs ($[a\text{–}z, \dots]$). They serve as primary nutritional resources.
2. **Poison Entities ($\mathcal{P}$)**: Tokens containing at least one uppercase glyph ($[A\text{–}Z, \dots]$). They serve as noxious irritants inducing acute toxic elevation of hunger.

When an entity is consumed via the motor action `EAT_TARGET`, it undergoes a state transition $w_k \to \text{eaten}$, is removed from the active spatial index, and triggers a delayed regrowth timer ($\tau_{\text{regrow}} \sim \mathcal{U}(60, 65)$ s), simulating biological resource replenishment.

2.2 Embodied Kinematics & Steering Dynamics

Hebbie's physical embodiment is governed by continuous 2D kinematics integrated via semi-implicit Euler integration at variable frame rate Δt:

$$\mathbf{x}(t+\Delta t) = \mathbf{x}(t) + \mathbf{v}(t)\Delta t$$
$$\mathbf{v}(t+\Delta t) = (\mathbf{v}(t) + \mathbf{F}_{\text{steer}}\Delta t) \cdot \mu^{\Delta t \cdot 60}$$

where $\mu = 0.92$ is the base friction coefficient and $\mathbf{F}_{\text{steer}}$ represents the active steering force computed via Reynolds-style steering behaviors (Reynolds, 1987).

Maximum Speed & Fatigue Modulation

The agent's terminal velocity is dynamically constrained by its internal fatigue drive $F(t) \in [0, 1]$:
$v_{\max}(F) = v_0 \cdot \max\left(0.35,\ 1.0 - 0.6 \cdot F(t)\right)$
where $v_0 = 173.25\ \text{px/s}$. As exhaustion mounts, the organism visibly and physically slows down, reducing exploratory range.

Spatial Steering Vectors

For target coordinates $\mathbf{x}_{\text{target}}$, the steering force combines desired velocity matching with an arrival envelope:
$\mathbf{v}_{\text{des}} = \frac{\mathbf{x}_{\text{target}} - \mathbf{x}}{\|\mathbf{x}_{\text{target}} - \mathbf{x}\|} \cdot v_{\text{target}}$
$\mathbf{F}_{\text{steer}} = k_{\text{steer}}\left(\mathbf{v}_{\text{des}} - \mathbf{v}\right)$

where $k_{\text{steer}} = 3.5$ for attraction and 4.0 for repulsion. To eliminate numerical chatter and limit cycle oscillations when approaching stationary targets, an arrival factor ramps speed smoothly within a $2\ \text{px}$ threshold:
$$f_{\text{arrival}}(d) = \begin{cases} \max\left(0,\ \frac{d - 0.5}{1.5}\right), & d < 2.0\ \text{px} \\ 1.0, & d \geq 2.0\ \text{px} \end{cases}$$

When the organism is within contact distance ($d \leq 3\ \text{px}$) during word approach, it transitions to a stationary posture ($\mathbf{v} \leftarrow \mathbf{0}$), halting kinetic jitter.

Viewport Boundary Repulsion

Confinement to the browser viewport is maintained by soft linear repulsion fields active within a border margin $\delta = 40\ \text{px}$:
$$F_{\text{repel},x} = \begin{cases} +12 \cdot (\delta - x), & x < \delta \\ -12 \cdot (x - (W_{\text{vw}} - \delta)), & x > W_{\text{vw}} - \delta \\ 0, & \text{otherwise} \end{cases}$$

Text-Anchored Waypoint Exploration

To prevent purposeless drift into unoccupied page margins during exploratory states (`EXPLORE_WANDER`), Hebbie employs **text-anchored wander waypoints**. The organism samples candidate target tokens from the subset of active words satisfying a minimum distance criterion:
$\mathcal{W}_{\text{distant}} = \{w_k \in \mathcal{W}_{\text{active}} \mid \|\mathbf{c}_k - \mathbf{x}\| > 70 \text{ px}\}$

A waypoint is locked for a maximum temporal horizon of $\tau_{\text{wander}} = 16.0 \text{ s}$ or until the agent reaches within 45 px of the token center. Dynamic re-computation of bounding boxes ensures that waypoints track smooth DOM layout shifts and scrolling events.

---

3. Sensory Architecture & Push-Pull Encoding

The sensory interface transforms environmental and somatic metrics into a normalized, 15-dimensional input vector $\mathbf{s} \in [0, 1]^{15}$ evaluated at discrete decision intervals ($\Delta t_{\text{dec}} = 1.5 \text{ s}$).

Index	Feature Name	Formulation / Semantic Definition
s_0	Bias	1.0 (constant tonic excitation)
s_1	Is Food	$\mathbb{I}(\text{Nearest} \in \mathcal{F})$ (lowercase token)
s_2	Is Poison	$\mathbb{I}(\text{Nearest} \in \mathcal{P})$ (uppercase token)
s_3	Word Proximity	$1/(1 + d_{\text{word}}/90)$
s_4	$1 -$ Word Proximity	$1 - s_3$ (push-pull dual-rail complement)
s_5	Can Eat (Affordance)	$\mathbb{I}(d_{\text{word}} \leq r_{\text{eat}})$, where $r_{\text{eat}} = 38 \text{ px}$
s_6	$1 -$ Can Eat	$1 - s_5$ (push-pull dual-rail complement)
s_7	User Proximity	$1/(1 + d_{\text{mouse}}/120)$ if cursor active, else 0
s_8	$1 -$ User Proximity	$1 - s_7$ (push-pull dual-rail complement)
s_9	Hunger Drive (H)	$H \in [0, 1]$ (metabolic energy deficit)
s_{10}	$1 -$ Hunger Drive	$1 - H$ (nutritional satiety)
s_{11}	Fatigue Drive (F)	$F \in [0, 1]$ (physical exhaustion deficit)
s_{12}	$1 -$ Fatigue Drive	$1 - F$ (kinetic vitality)

s_{13}	Boredom Drive (B)	$B \in [0, 1]$ (cognitive understimulation deficit)
s_{14}	$1 - \text{Boredom Drive}$	$1 - B$ (engagement / equilibrium)

3.1 Spatial Proximity & Hysteresis

The distance between Hebbie's center (x, y) and a word rectangle $\mathbf{R} = [x_l, x_r, y_t, y_b]$ is calculated as the Euclidean distance to the nearest point on the perimeter:

$$x_{\text{close}} = \text{clamp}(x, x_l, x_r), \quad y_{\text{close}} = \text{clamp}(y, y_t, y_b)$$
$$d_{\text{word}} = \sqrt{(x - x_{\text{close}})^2 + (y - y_{\text{close}})^2}$$

To prevent high-frequency perceptual flickering (chatter) between two equidistant words, spatial indexing applies a **hysteresis bias** to the currently pursued token w_{target}:

$$\text{Score}(w) = \begin{cases} 0.85 \cdot d_{\text{word}}^2, & \text{if } w = w_{\text{target}} \\ d_{\text{word}}^2, & \text{otherwise} \end{cases}$$

This introduces a $\approx 7.8\%$ spatial hysteresis margin, stabilizing behavioral intentionality across frames while preserving the capacity to redirect toward significantly closer entities.

Sensory proximity is transformed via a hyperbolic response curve with half-saturation distance $d_{1/2} = 90$ px:

$$p_{\text{word}} = \frac{1}{1 + \frac{d_{\text{word}}}{90}}$$

Similarly, interaction with the human user is transduced via cursor tracking:

$$p_{\text{user}} = \begin{cases} \frac{1}{1 + \frac{d_{\text{mouse}}}{120}}, & \text{if cursor active} \\ 0, & \text{otherwise} \end{cases}$$

3.2 Dual-Rail Complementary Encoding

Supplying complementary inputs x and $\bar{x} = 1 - x$ solves three critical constraints specific to local Hebbian plasticity and biological sensorimotor control:

1. **Credit Assignment During Feature Inactivity**:
   Hebbian synaptic updates are gated by presynaptic activity ($\Delta w_{a,i} \propto \delta \cdot e_{a,i}$, where $e_{a,i} \propto s_i$). When a sensory feature is inactive ($x \approx 0$), its eligibility trace is negligible ($e_{a,i} \approx 0$), rendering the synapse plasticly silent. Without an inverted channel, the agent cannot bind associations to the *absence* or *low level* of a feature (e.g., learning exploratory behaviors when distant from food, $p_{\text{word}} \approx 0$, or learning actions appropriate to nutritional satiety when $H \approx 0$). Supplying $\bar{x} = 1 - x$ ensures active presynaptic drive and synaptic eligibility when the primary feature is dormant.

2. **Biological Push-Pull / ON-OFF Antagonism**:
   Biological neural pathways (such as ON/OFF channels in the mammalian retina and visual cortex, or agonist/antagonist motor control circuits) operate with non-negative firing rates ($r \geq 0$). By splitting features into complementary channels (x and $1 - x$), the architecture allows independent excitatory and inhibitory associative weights to grow with separate learning dynamics and decay trajectories.

3. **Per-Feature Baseline Decoupling from Global Tonic Bias**:
   With unipolar features, adjusting the response at $x = 0$ requires updating the global tonic bias synapse s_0, which is shared across all 15 inputs. Dual-rail encoding enables independent parametrization of the baseline response at $x = 0$ (via w_2) and the response at $x = 1$ (via w_1) for each perceptual dimension separately, preventing cross-talk on the global bias node.

3.3 Sensory Habituation & Anti-Fixation Dynamics

To avoid behavioral deadlocks where an unreachable word is pursued indefinitely, a temporal habituation circuit monitors continuous pursuit duration t_{pursuit} on a single token ID. If:

$t_{\text{pursuit}} > 8.0\ \text{s} \quad \text{and} \quad a^* \notin \{\text{REST}, \text{DANCE}\}$

the token is appended to a refractory exclusion set with a 5.0 s expiration timestamp:

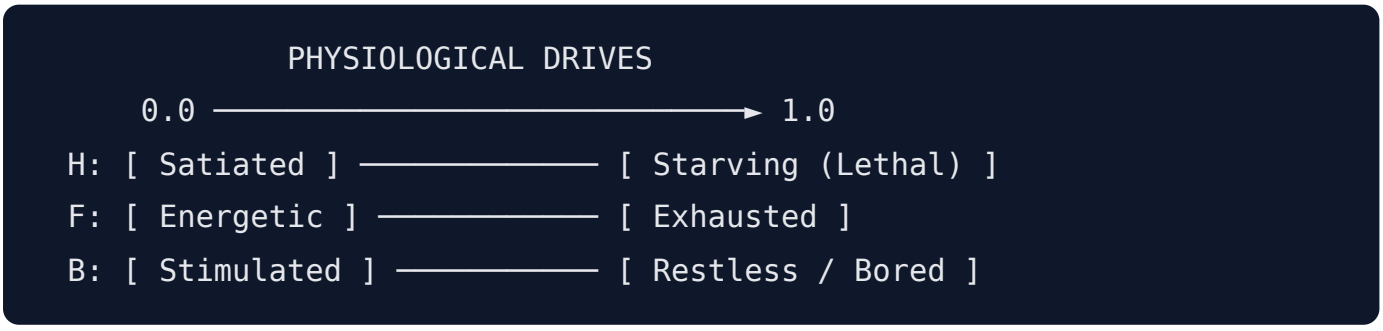
$\mathcal{M}_{\text{ignored}} \leftarrow \mathcal{M}_{\text{ignored}} \cup \{(w_{\text{id}},\ t_{\text{now}} + 5000\ \text{ms})\}$

This forces the organism to disengage, habituate to the stimulus, and sample alternative foraging paths.

---

4. Internal Physiology & Homeostatic Loss

Hebbie's motivational architecture is rooted in three continuous, bounded physiological deficit variables $\mathbf{D} = [H, F, B]^T \in [0, 1]^3$:



4.1 Metabolic Differential Equations

Each drive evolves according to differential equations determined by current physical activity, discrete events, and basal metabolic costs:

$\frac{d\mathbf{D}}{dt} = \mathbf{f}(\mathbf{D}, a) + \sum_k \Delta\mathbf{D}_k \delta(t - t_k)$

Continuous Activity Regimes:

1. **Resting** ($a = \mathrm{REST}$):
   $\frac{dH}{dt} = +0.0045\ \mathrm{s}^{-1}, \quad \frac{dF}{dt} = -0.3500\ \mathrm{s}^{-1}, \quad \frac{dt_{\mathrm{idle}}}{dt} = +0.60$
   Rest provides rapid restorative dissipation of fatigue, with low metabolic hunger accumulation and modest boredom accumulation.

2. **Dancing** ($a = \mathrm{DANCE}$):
   $\frac{dH}{dt} = +0.0160\ \mathrm{s}^{-1}, \quad \frac{dF}{dt} = +0.0550\ \mathrm{s}^{-1}, \quad \frac{dt_{\mathrm{idle}}}{dt} = -7.00$
   Dancing represents energetic motor play. It consumes significant nutritional and kinetic resources while rapidly discharging accumulated boredom.

3. **Ingestion** ($a = \mathrm{EAT\_TARGET}$):
   $\frac{dH}{dt} = +0.0060\ \mathrm{s}^{-1}, \quad \frac{dt_{\mathrm{idle}}}{dt} = -1.20$

4. **Locomotion** ($a \in \{\mathrm{APPROACH\_*}, \mathrm{AVOID\_*}, \mathrm{EXPLORE\_WANDER}\}$):
   $\frac{dH}{dt} = +0.0060\ \mathrm{s}^{-1}, \quad \frac{dF}{dt} = +0.0450\ \mathrm{s}^{-1}, \quad \frac{dt_{\mathrm{idle}}}{dt} = +0.18$

Boredom is computed directly from the unrewarded dwell time:
$B(t) = \mathrm{clamp}_{[0,1]}\left(\frac{t_{\mathrm{idle}}}{8.0}\right)$

Acute Discrete Perturbations:

- **Ingestion of Food** ($w \in \mathcal{F}$):
  $\Delta H = -0.30, \quad \Delta t_{\mathrm{idle}} = -1.20\ \mathrm{s}$
- **Ingestion of Poison** ($w \in \mathcal{P}$):
  $\Delta H = +0.25$
- **Human Petting Interaction**:
  $\Delta t_{\mathrm{idle}} = -3.20\ \mathrm{s}, \quad \Delta F = -0.15$

4.2 Terminal State (Death & Revivability)

Hunger operates as the sole **lethal physiological boundary**:
If $H(t) \geq 1.0 \implies \mathrm{State} \leftarrow \mathrm{DEAD}$

Upon death, motor updates are terminated, velocity is zeroed, somatic coloration darkens to near-black, and a revival affordance bubble is instantiated. Clicking Hebbie invokes

`revive()`, resetting $H \leftarrow 0.25$, $F \leftarrow 0.50$, $B \leftarrow 0.0$, and clearing all eligibility traces while preserving learned synaptic weights.

4.3 The Homeostatic Loss Metric

The normative objective of the organism is to minimize a convex, quadratic physiological loss function $L : [0, 1]^3 \to \mathbb{R}^+$:

$$L(H, F, B) = \alpha_H H^2 + \alpha_F F^2 + \alpha_B B^2$$

where $\alpha_H = 2.5$, $\alpha_F = 1.0$, and $\alpha_B = 1.0$.

The super-linear quadratic penalty ensures that small deflections across multiple drives are tolerated, but severe deficits in any single drive are penalized aggressively. The asymmetric weighting factor $\alpha_H = 2.5$ reflects biological prioritization: fatigue and boredom induce behavioral inefficiency, but extreme hunger results in organismic termination.

4.4 Affective Valence & Morphological Expression

Hebbie maps its homeostatic loss directly to affective valence and somatic morphology:

$$\mathcal{V}(L) = \begin{cases} \frac{L_0 - L}{L_0}, & L \leq L_0 \\ -\min\left(1.0,\ \frac{L - L_0}{0.48}\right), & L > L_0 \end{cases}$$

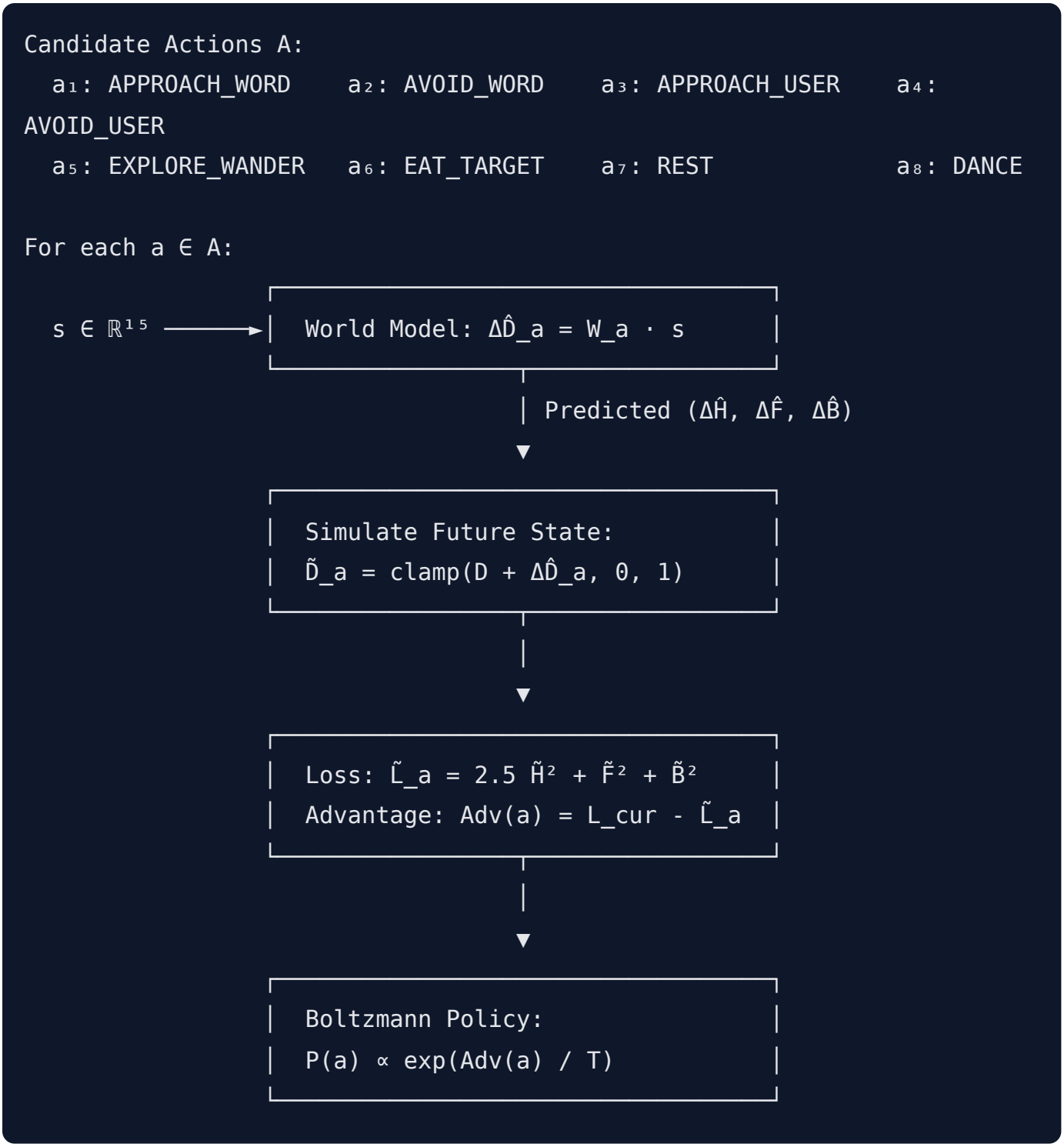
where $L_0 = 0.32$ denotes the baseline neutral homeostatic setpoint. Valence $\mathcal{V} \in [-1, +1]$ continuously parameterizes the rendering of the creature:

- **Mouth Curvature**: Interpolated via an SVG quadratic Bézier curve:
  $y_{\text{ends}} = 5.5 - 1.2 \cdot \mathcal{V}, \quad y_{\text{center}} = 5.5 + 3.2 \cdot \mathcal{V}$
- **Somatic Melanin / Chromatic Tinting**: Hunger directly darkens the organism's body from bright grey ($\text{RGB}(224, 224, 224)$) when $H \leq 0.5$ down to near-black ($\text{RGB}(20, 20, 20)$) as $H \to 1.0$:
  $\text{Gray}(H) = \text{round}\left(224 - \max\left(0,\ \frac{H - 0.5}{0.5}\right) \cdot (224 - 20)\right)$
- **Eyelid Ptosis**: Eyelids droop in direct proportion to fatigue ($h_{\text{lid}} = 0.88 \cdot F$), transitioning to closed eyes and drifting z-particle emissions when resting with $F > 0.5$.

---

5. The Predictive World Model & Simulation Decider

Hebbie rejects model-free temporal difference methods in favor of a **prospective simulation architecture**.



5.1 Synaptic Matrix Representation

The world model comprises three synaptic weight matrices:
$\mathbf{W}^H, \mathbf{W}^F, \mathbf{W}^B \in \mathbb{R}^{8\times 15}$

Each row $\mathbf{w}_a^D$ represents the associative projection estimating the delta of drive $D \in \{H, F, B\}$ if motor action $a \in \mathbf{A}$ is executed under sensory state $\mathbf{s}$.

Innate Instincts

At genesis, weights are initialized to near-zero Gaussian perturbations $\mathcal{U}(-0.01, 0.01)$, with the exception of a few innate instincts:

$W^H_{\text{EAT, CAN\_EAT}} = -0.25 \quad (\text{ingestion relieves hunger})$

$W^H_{\text{EAT, IS\_POISON}} = +0.35 \quad (\text{toxic consumption exacerbates hunger})$

$W^H_{\text{EXPLORE, HUNGER}} = -0.10 \quad (\text{foraging instinct: exploration encouraged when hungry})$

These priors ensure basic survival competency and active foraging, while all other associations must be acquired empirically.

5.2 Prospective Drive Simulation & Advantage Formulation

At each decision cycle, the simulation decider computes the prospective sensory-action dot products for all $a \in \{0, \dots, 7\}$:

$\Delta\hat{H}_a = \mathbf{w}^H_a \cdot \mathbf{s} = \sum_{i=0}^{14} w^H_{a,i} s_i$

$\Delta\hat{F}_a = \mathbf{w}^F_a \cdot \mathbf{s} = \sum_{i=0}^{14} w^F_{a,i} s_i$

$\Delta\hat{B}_a = \mathbf{w}^B_a \cdot \mathbf{s} = \sum_{i=0}^{14} w^B_{a,i} s_i$

Hypothetical future drive levels are projected and clamped to the physical unit interval:

$\tilde{H}_a = \text{clamp}_{[0,1]}\left(H + \Delta\hat{H}_a\right)$

$\tilde{F}_a = \text{clamp}_{[0,1]}\left(F + \Delta\hat{F}_a\right)$

$\tilde{B}_a = \text{clamp}_{[0,1]}\left(B + \Delta\hat{B}_a\right)$

The hypothetical homeostatic loss for each action is evaluated:

$\tilde{L}_a = L(\tilde{H}_a, \tilde{F}_a, \tilde{B}_a) = 2.5\tilde{H}_a^2 + \tilde{F}_a^2 + \tilde{B}_a^2$

The **Action Advantage** $\text{Adv}(a)$ represents the predicted net reduction in homeostatic discomfort:

$\text{Adv}(a) = L_{\text{current}} - \tilde{L}_a$

A positive advantage indicates that action a is anticipated to bring the organism closer to homeostatic equilibrium, whereas a negative advantage indicates an expected exacerbation of physiological stress. The total advantage decomposes linearly into drive-specific reductions:

$\text{Adv}(a) = \text{Adv}_H(a) + \text{Adv}_F(a) + \text{Adv}_B(a)$

$\text{Adv}_H(a) = 2.5\left(H^2 - \tilde{H}_a^2\right)$

$\text{Adv}_F(a) = F^2 - \tilde{F}_a^2$

$\text{Adv}_B(a) = B^2 - \tilde{B}_a^2$

5.3 Action Selection via Boltzmann Distribution

Actions are sampled stochastically using a temperature-controlled Softmax policy over action advantages:

$$P(a) = \frac{\exp\left(\frac{\text{Adv}(a) - \max_{a'} \text{Adv}(a')}{T}\right)}{\sum_{k=1}^{|\mathbf{A}|} \exp\left(\frac{\text{Adv}(k) - \max_{a'} \text{Adv}(a')}{T}\right)}$$

where T represents the computational temperature (calibrated to $T = 0.10$). Subtracting $\max_{a'} \text{Adv}(a')$ ensures numerical stability against floating-point overflow.

- When $T \to 0$ ($T \leq 0.02$), the policy collapses to a deterministic greedy selection: $a^* = \arg\max_{a \in \mathbf{A}} \text{Adv}(a)$.
- At $T = 0.10$, the agent predominantly selects actions that optimize homeostatic restoration while maintaining sufficient entropy to discover novel environmental contingencies.

---

6. Plasticity: Normalized Delta-Hebbian Learning

The core learning mechanism of Hebbie is an online, self-supervised, causal associative rule. The objective is to minimize prediction errors across drives without backpropagation.

6.1 Multi-Step Action-Conditioned Eligibility and Prediction Traces

Actions in continuous space require extended execution before their consequences manifest (e.g., executing `APPROACH_WORD` requires multiple seconds of traversal before the word enters the ingestion zone). To bridge this temporal credit-assignment gap, Hebbie maintains action-conditioned sliding window traces for both pre-synaptic input activations (eligibility traces) and world model predictions.

1. **Eligibility Traces Matrix** ($\mathbf{E} \in \mathbb{R}^{8 \times 15}$):
   When action a^* is selected, active presynaptic features reinforce their corresponding traces:
   $e_{a^*,i} \leftarrow e_{a^*,i} + s_i$
2. **Prediction Traces Vectors** ($p^D \in \mathbb{R}^8$ for $D \in \{H, F, B\}$):
   To prevent temporal scaling mismatches during learning, Hebbie also accumulates predictions of chosen actions over time. When action a^* is selected, the single-step predicted drive changes ($\Delta\hat{H}_{a^*}, \Delta\hat{F}_{a^*}, \Delta\hat{B}_{a^*}$) are accumulated into their respective

prediction traces:
$p^D_{a^*} \leftarrow p^D_{a^*} + \Delta\hat{D}_{a^*} \quad \forall D \in \{H, F, B\}$

Both eligibility traces and prediction traces for all actions undergo continuous, passive exponential decay governed by a half-life parameter $\tau_{1/2} = 2.0\ \text{s}$:
$\mathbf{E}(t + \Delta t) = \mathbf{E}(t) \cdot 0.5^{\frac{\Delta t}{\tau_{1/2}}}$
$p^D_a(t + \Delta t) = p^D_a(t) \cdot 0.5^{\frac{\Delta t}{\tau_{1/2}}} \quad \forall a \in \mathbf{A}, D \in \{H, F, B\}$

6.2 Prediction Error Formulation & Drive-Floor Clamping

At each decision boundary, actual drive shifts over the preceding interval are observed:
$\Delta D_{\text{actual}} = D(t) - D(t - \Delta t_{\text{dec}}), \quad D \in \{H, F, B\}$

Rather than comparing ΔD_{actual} to raw, single-step predictions, Hebbie calculates prediction errors using the accumulated prediction traces p^D_a. To resolve the artifact of **physiological floor saturation** (where restorative actions executed on fully satisfied drives $D \approx 0$ would be spuriously penalized since they cannot reduce the drive below zero), Hebbie calculates an **effective prediction** $\Delta\tilde{D}^{\text{eff}}_a$ clamped to the maximum possible reduction:
$$\Delta\tilde{D}^{\text{eff}}_a = \begin{cases} -D_{\text{last}}, & \text{if } p^D_a < 0 \text{ and } D_{\text{last}} < -p^D_a \\ p^D_a, & \text{otherwise} \end{cases}$$

The drive prediction error is then formulated as:
$\delta^D_a = \Delta D_{\text{actual}} - \Delta\tilde{D}^{\text{eff}}_a$

6.3 Synaptic Weight Update Rule

Synaptic updates are applied to all actions with non-negligible eligibility trace energy:
$\Delta w^D_{a,i} = \eta_{\text{eff}} \cdot \delta^D_a \cdot e_{a,i} - \lambda \cdot w^D_{a,i}$

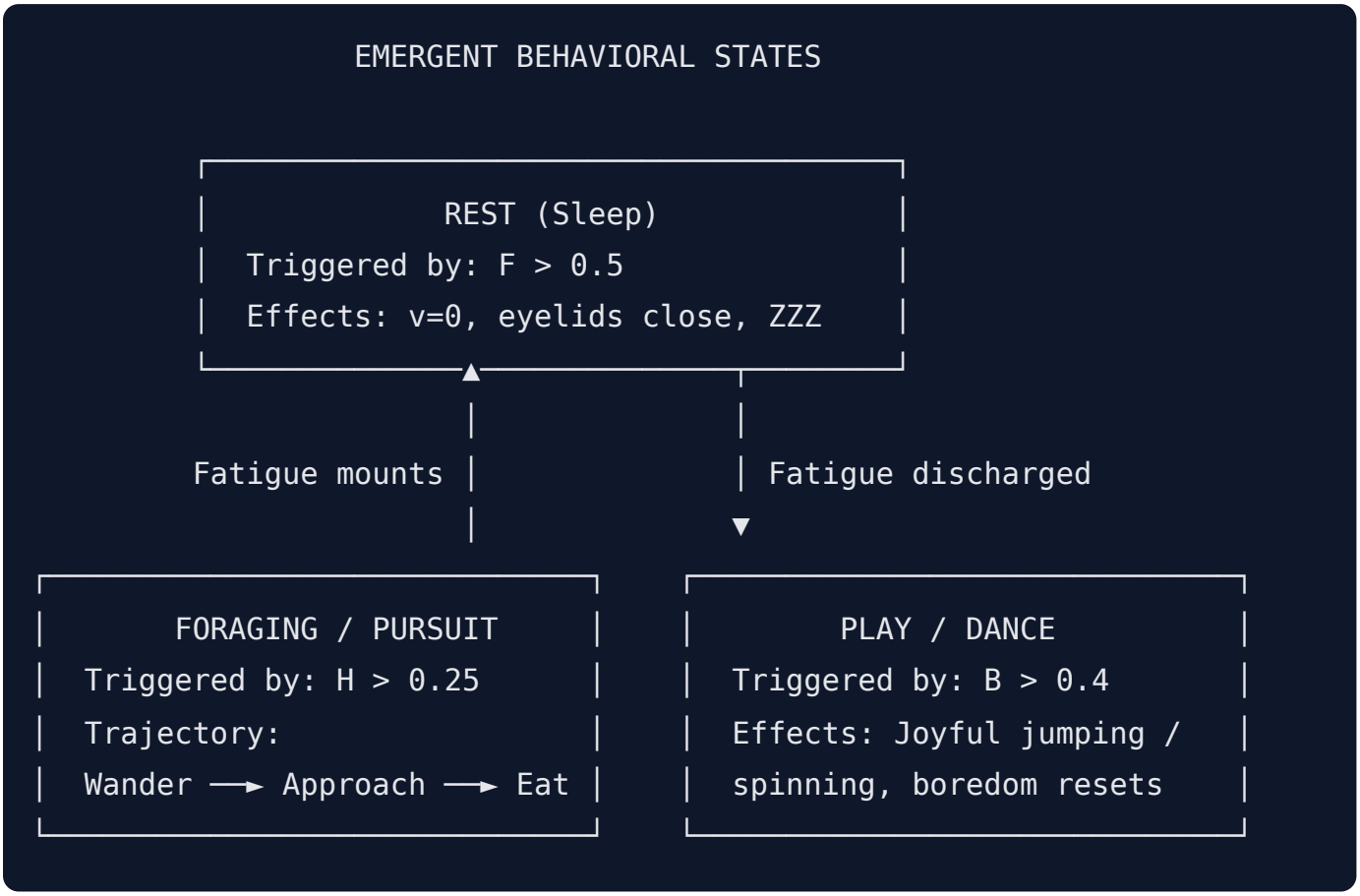
where:

1. **Trace Normalization (Widrow-Hoff stability)**:
   $\eta_{\text{eff}} = \frac{\eta_0}{\max(1.0,\, \|\mathbf{e}_a\|_2)}$
   Scaling the base learning rate $\eta_0 = 0.28$ by the Euclidean norm of the action's eligibility trace vector prevents gradient explosion during rapid, repeated action executions.

2. **Weight Decay & Hard Clamping**:
   A passive decay term $\lambda = 1 \times 10^{-4}$ prevents runaway synaptic drift, and weights are hard-clamped to a physiological dynamic range:
   $w^D_{a,i} \in [-2.8,\, +2.8]$

7. Emergent Behaviors & Empirical Dynamics

Through the interaction of predictive modeling and homeostatic regulation, complex adaptive behaviors emerge spontaneously without programmed state machines.



7.1 Autonomous Foraging Cascades

When Hunger H rises above baseline ($H > 0.25$), the homeostatic loss gradient $\nabla_H L = 5.0 \cdot H$ dominates all other considerations. The advantage landscape shifts dynamically:

1. At long range ($d_{\text{word}} > 70$ px), $s_3 \approx 0$ and $s_5 = 0$. The model projects no immediate relief from `EAT_TARGET`. However, the innate foraging prior $W^{H}_{\text{EXPLORE, HUNGER}} = -0.10$ couples directly with sensor $s_9 = H$, predicting $\Delta\hat{H} = -0.10 \cdot H < 0$ for `EXPLORE_WANDER`. This imparts a built-in advantage to text-waypoint exploration, propelling the organism across the page toward text clusters where `APPROACH_WORD` and proximity sensors take over.
2. Upon entering contact distance ($d_{\text{word}} \leq 38$ px), affordance sensor $s_5 \to 1.0$. The innate and reinforced weight $W^{H}_{\text{EAT,CAN\_EAT}} = -0.25$ predicts immediate hunger drop ($\Delta\hat{H} = -0.25$).
3. The advantage $\text{Adv}(\text{EAT\_TARGET})$ surges above all alternatives. The agent halts locomotion, initiates rhythmic mouth aperture chomping (two-cycle Hann-window envelope over 0.52 s), consumes the entity, and registers an acute drop in hunger ($\Delta H = -0.30$).

7.2 Learned Taste Aversion (The Synthetic Garcia Effect)

When a newborn agent encounters and consumes a Poison entity ($w \in \mathcal{P}$, containing uppercase characters), sensor $s_2 = 1.0$ is co-active with affordance $s_5 = 1.0$. Upon ingestion, hunger spikes by $+0.25$.

- The resulting prediction error $\delta^H > 0$ drives rapid potentiation of $W^H_{\text{EAT,IS\_POISON}} \to +2.8$.
- In subsequent encounters with uppercase tokens, the world model predicts: $\Delta\hat{H}_{\text{EAT}} = W^H_{\text{EAT,CAN\_EAT}} \cdot 1.0 + W^H_{\text{EAT,IS\_POISON}} \cdot 1.0 = -0.25 + 0.35 = +0.10$
- This yields a strongly negative advantage $\text{Adv}(\text{EAT\_TARGET}) < 0$. Concurrently, `AVOID_WORD` yields a positive relative advantage. The agent exhibits active avoidance of capitalized words, mirroring biological conditioned taste aversion (Garcia et al., 1955).

7.3 Homeostatic Sleep-Wake Cycles

Prolonged physical activity (chasing words and wandering) continuously accrues fatigue ($\frac{dF}{dt} = +0.045\ \text{s}^{-1}$).

- As $F \to 0.6$–0.8, the fatigue component of the loss function (F^2) grows rapidly.
- The action `REST` predicts $\Delta\hat{F} < 0$, resulting in an advantage that overtakes foraging (unless hunger is critically near the lethal threshold).
- Upon selecting `REST`, the agent halts, drops its eyelids, and enters a physiological sleep state. Fatigue declines rapidly ($\frac{dF}{dt} = -0.35\ \text{s}^{-1}$) until homeostatic loss is minimized, at which point the agent wakes and resumes exploration.

7.4 Spontaneous Play & Boredom Discharge

If the environment provides few nearby foraging interactions, unrewarded dwell time accumulates, elevating Boredom $B(t) \to 1.0$.

- Because neither `APPROACH_WORD` nor `REST` alleviates boredom, their advantages decline.
- The action `DANCE` predicts strong boredom reduction ($\Delta\hat{B} \ll 0$).
- The organism spontaneously initiates expressive dance routines (joyful vertical leaping or rapid axial rotation), rapidly resetting $t_{\text{idle}} \to 0$ and restoring affective valence $\mathcal{V} \to +1$.

---

8. Comparative Analysis: Hebbie vs. Alternative Paradigms

Architectural Dimension	Classical Model-Free RL	Active Inference (Friston)	Hebbie Cognitive Architecture
Objective Metric	Cumulative discounted external reward $\sum_t \gamma^t r_t$	Variational Free Energy $F = D_{\mathrm{KL}}(q \parallel p) - \mathbb{E}_q[\ln p]$	Convex Homeostatic Loss $L = 2.5H^2 + F^2 + B^2$
Learning Mechanism	TD-error backpropagation / Policy gradients	Gradient descent on variational free energy	Normalized Delta-Hebbian with multi-step eligibility traces
Internal State Dynamics	Absent (monolithic state) or black-box RNN	Implicit prior preferences over observation states	Explicit multi-drive ODEs (Hunger, Fatigue, Boredom)
Decision Strategy	Q-value $\arg\max$ / Actor policy distribution	Active inference via expected free energy (EFE)	Prospective simulation of homeostatic action advantage $\mathrm{Adv}(a)$
Computational Overhead	High (tensor graphs, backpropagation, replay buffers)	Moderate–High (matrix inversions, belief distributions)	Ultra-low (< 0.1 ms tick, pure vectorized Float32)

Hebbie shares theoretical affinities with Friston's **Active Inference**: internal drives act as priors over physiological states, and actions are chosen to minimize anticipated divergence from those setpoints. However, Hebbie avoids the computational complexity of variational density estimation by employing direct associative drive projections coupled with an analytical quadratic loss evaluator. This makes the entire architecture lightweight enough to execute synchronously within standard browser animation loops at 60 frames per second.

---

9. Conclusion

Hebbie demonstrates that coherent, intelligent, and lifelike autonomy can emerge from the interplay of:

1. **Embodied homeostatic constraints** that define survival priorities natively,
2. **Predictive world models** that learn environmental causal relationships without external reward engineering, and

3. **Normalized Delta-Hebbian plasticity** with multi-step eligibility traces operating under biologically grounded physical constraints.

By grounding cognition in physiological preservation rather than arbitrary task rewards, Hebbie provides a viable, computationally efficient blueprint for synthetic life in interactive virtual environments.

References

- Ashby, W. R. (1952). *Design for a Brain: The Origin of Adaptive Behaviour*. Chapman & Hall.
- Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological Reviews*, 9(3), 399–431.
- Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Reviews Neuroscience*, 11(2), 127–138.
- Garcia, J., Kimeldorf, D. J., & Koelling, R. A. (1955). Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science*, 122(3160), 157–158.
- Hull, C. L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*. Appleton-Century.
- Keramati, M., & Gutkin, B. S. (2014). Homeostatic reinforcement learning for integrating reward collection and physiological stability. *eLife*, 3, e04811.
- Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, 134, 17–35.
- Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH Computer Graphics*, 21(4), 25–34.
- Sterling, P. (2012). Allostasis: a predictive regulation of biological variables. *Nature Neuroscience*, 15(4), 659–664.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.