PeerJ Computer Science

Biologically-inspired emotional processing for adaptive decision-making in non-stationary environments

Jaeyeon Kim[1] and Daihun Kang[1,2]

[1] Department of Plastic and Reconstructive Surgery, Ewha Womans University Seoul Hospital, Ewha Woman's University, Seoul, Republic of South Korea
[2] College of Medicine, Ewha Woman's University, Seoul, Republic of South Korea

ABSTRACT

Background: Artificial intelligence (AI) often struggles to adapt in non-stationary environments where conditions change unpredictably. In contrast, biological organisms utilize emotional processes not as irrational noise, but as rapid heuristics for managing uncertainty. This study investigates whether computational mechanisms inspired by mammalian affective systems provide advantages for adaptive decision-making.

Methods: The Emotional-Cognition Integration Architecture (ECIA) was developed, incorporating computational analogs of eight emotion-like signals designed for reinforcement learning contexts, hippocampus-inspired episodic memory, and dopamine-modulated adaptive learning. Using large-scale experimental replication (3,600 runs across 12 master seeds), ECIA was evaluated against both traditional algorithms (ε-greedy, upper confidence bound (UCB), Thompson Sampling) and improved non-stationary baselines (Sliding Window UCB, Adaptive Thompson Sampling) in three distinct environments designed to test different aspects of adaptation.

Results: ECIA demonstrated environment-specific performance patterns reflecting a functional trade-off. In unpredictable settings characterized by sudden regime shifts and stochastic perturbations, ECIA significantly outperformed all baselines ($p < 0.001$). However, in strictly deterministic patterns, ECIA incurred a "cost of complexity," underperforming compared to Naive UCB (0.8014 *vs.* 0.8522). This trade-off suggests functional specialization for uncertainty management rather than universal superiority. Ablation studies revealed strong synergistic integration among components, with combined removal causing disproportionate degradation far exceeding individual effects and highlighted a "dopamine paradox" where adaptive plasticity benefited uncertain environments but destabilized predictable ones.

Conclusions: These findings demonstrate that emotion-inspired computational mechanisms, drawn from mammalian brain architecture, function as specialized tools for managing environmental volatility. While they incur efficiency costs in stable environments, they provide essential robustness in high-uncertainty domains. This work offers both a practical framework for adaptive AI systems in domains such as clinical decision support and financial trading, and computational insights into why biological intelligence integrates affective processing with cognition.

Submitted 21 July 2025
Accepted 23 January 2026
Published 27 February 2026

Corresponding author
Daihun Kang,
gpk1234567@naver.com

Academic editor
José Santos

Additional Information and Declarations can be found on page 25

DOI **10.7717/peerj-cs.3688**

Copyright
2026 Kim and Kang

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Human-Computer Interaction, Agents and Multi-Agent Systems, Artificial Intelligence, Data Mining and Machine Learning, Neural Networks
Keywords Emotional intelligence, Reinforcement learning, Artificial intelligence, Machine learning, Emotion-cognition integration, Biologically-inspired AI, Decision making, Environmental adaptation

INTRODUCTION

Adaptive decision-making in non-stationary environments represents a fundamental challenge in artificial intelligence (AI). While reinforcement learning (RL) algorithms have demonstrated remarkable success in static settings, their performance deteriorates significantly when environmental dynamics change unpredictably (*Lecarpentier, 2020*; *Lecarpentier & Rachelson, 2019*). This limitation becomes critical in real-world applications where conditions evolve continuously: financial markets shift between bull and bear regimes without warning, patient conditions in healthcare monitoring change gradually or abruptly, and autonomous systems encounter operational contexts that vary dynamically. The demand for algorithms capable of rapid detection and adaptation to environmental changes motivates the development of more flexible learning architectures that can maintain robust performance under uncertainty (*Abdullahi et al., 2025*; *Bai et al., 2025*).

Biological organisms, particularly mammalian brains, exhibit remarkable adaptability in uncertain environments through integrated cognitive-emotional processing. Neuroscientific research has revealed that emotional systems participate necessarily in adaptive decision-making rather than serving as obstacles to rationality (*Pessoa, 2018a*, *2018b*). The amygdala can respond to threat-relevant cues at latencies of roughly 70–200 ms, supporting rapid behavioral adjustments to environmental changes (*Guex et al., 2020*). The hippocampus integrates contextual memory with information about emotional value, supporting value- and similarity-based decisions by linking current options to affectively tagged past experiences (*Rolls, 2014*). The prefrontal cortex orchestrates executive control by integrating emotional signals from limbic structures with cognitive information, resolving conflicts among competing action tendencies and enabling goal-directed behavior under uncertainty (*Gray, Braver & Raichle, 2002*; *Miller & Cohen, 2001*). Dopaminergic circuits encode reward prediction errors and can modulate the effective learning rate with which values are updated, thereby supporting flexible adaptation to changing reward structures (*Glimcher, 2011*; *Keiflin & Janak, 2015*). These neural mechanisms suggest that emotion-like computational processes—rapid heuristics for uncertainty management, context-sensitive memory retrieval, executive integration of conflicting signals, and adaptive learning rate modulation—may provide selective advantages for managing environmental volatility (*Bach & Dayan, 2017*). Recent computational neuroscience has begun exploring how these mechanisms can be formalized mathematically, opening possibilities for biologically-inspired AI architectures (*Bach & Dayan, 2017*; *Govea et al., 2024*; *Huys, Maia & Frank, 2016*).

Existing approaches to non-stationary environments have primarily focused on statistical adaptation methods. Sliding-window algorithms maintain only recent observations while discarding historical data, enabling rapid response to detected changes

but sacrificing long-term statistical power (*Garivier & Moulines, 2008*; *Trovo et al., 2020*). Discounting methods assign exponentially decaying weights to past observations, balancing responsiveness against stability through tunable decay parameters (*Raj & Kalyani, 2017*; *Qi, Guo & Zhu, 2025*). Change-point detection approaches explicitly identify regime transitions and reset parameter estimates upon detection, but require careful threshold calibration and may fail to detect gradual environmental shifts (*Liu, Lee & Shroff, 2018*; *Mellor & Shapiro, 2013*). Bayesian methods model environmental dynamics through probabilistic frameworks, offering principled uncertainty quantification at the cost of increased computational complexity (*Hartland et al., 2006*). These statistical approaches have demonstrated effectiveness in specific scenarios. While some methods combine multiple strategies (*e.g.*, change-point detection with sliding windows, or Bayesian inference with discounting), they typically employ these mechanisms sequentially or hierarchically rather than through the parallel, bidirectionally-coupled processing observed in biological affective-cognitive systems.

Meta-reinforcement learning represents another promising direction for adaptive AI. Meta-learning approaches train agents to acquire learning strategies that generalize across task distributions, enabling rapid adaptation to new tasks through experience (*Wang et al., 2016*). Context-based methods learn to infer task identity from observations and adjust behavior accordingly (*Rakelly et al., 2019*). Recent work has explored combining meta-learning with memory-augmented architectures to improve few-shot adaptation (*Santoro et al., 2016*). While offline meta-RL methods require extensive pre-training on diverse task distributions, recent work has also explored online and continual adaptation approaches (*Finn et al., 2019*; *Nagabandi et al., 2018*). Nevertheless, meta-RL approaches—whether offline or online—are typically optimized for specific task distributions encountered during meta-training, and their adaptation mechanisms differ architecturally from the parallel heuristic processes observed in biological affective systems.

Computational models incorporating emotion-like mechanisms have been explored in several research domains. Affective computing has developed frameworks for recognizing and responding to human emotional states (*Calvo & D'Mello, 2010*). Cognitive architectures such as SOAR and ACT-R have integrated motivational and emotional components to model human (*Marsella & Gratch, 2009*). Social robotics has implemented emotion-inspired controllers to enable natural human-robot interaction (*Moerland, Broekens & Jonker, 2018*). More recently, emotion-like modulation has been incorporated into reinforcement learning: curiosity-driven exploration uses intrinsic motivation signals analogous to novelty-seeking behavior (*Oudeyer & Kaplan, 2007*), affect-inspired agents modulate risk sensitivity based on performance history (*Hu & Leung, 2023*) and emotion-modulated architectures adjust exploration strategies dynamically (*Joffily & Coricelli, 2013*). Multimodal emotion recognition systems now combine visual, acoustic, and linguistic cues for robust affect detection (*Lian et al., 2023a*; *Mittal et al., 2020*). Despite these advances, comprehensive architectures that integrate emotion-inspired processing with episodic memory and adaptive learning specifically for non-stationary reinforcement learning remain limited in the literature.

The reviewed literature reveals a critical gap: while statistical methods provide principled adaptation to specific change types and meta-learning enables generalization across task distributions, neither framework captures the parallel multi-mechanism processing observed in biological adaptive systems. Biological organisms employ fast emotional heuristics (fear-driven risk aversion, curiosity-driven exploration), episodic memory retrieval for context recognition, and gradual parametric learning simultaneously—with these processes interacting dynamically based on environmental conditions. Existing computational approaches typically implement single adaptation mechanisms in isolation, lacking the synergistic integration that may enable robust performance across diverse types of environmental change. This architectural gap motivates the development of integrated frameworks that combine multiple biologically-inspired adaptation processes.

This study introduces the Emotional–Cognition Integration Architecture (ECIA), a reinforcement learning framework inspired by mammalian adaptive mechanisms in non-stationary environments. ECIA comprises four interacting components:

1. **External Limbic System (ELS)** that computes eight emotion-like dimensions (fear, joy, hope, sadness, curiosity, anger, pride, shame) from reward dynamics and performance trends, providing computational analogs of affective functions.
2. **Prefrontal Decision Unit (PDU)** that performs emotion-modulated action selection by integrating ELS signals with value-based evaluation, so that transient emotional states dynamically shape exploration–exploitation balance and risk sensitivity.
3. **Hippocampus-inspired episodic memory system** that stores experiences with emotional and contextual tags, enabling similarity-based retrieval to support context recognition.
4. **Dopamine-inspired adaptive learning module** that adjusts learning rates as a function of prediction errors and emotional states, enabling flexible credit assignment.

Rather than attempting to model human subjective experience, the architecture instantiates computational functions that emotional systems are hypothesized to support in biological decision-making: rapid uncertainty assessment; emotion-modulated action selection under conflicting signals; memory-guided context sensitivity; and dynamic learning-rate modulation. These components interact through bidirectional connections, giving rise to emergent adaptive behaviors across different forms of environmental change. The framework is evaluated against canonical reinforcement-learning algorithms (ε-greedy, Thompson sampling, upper confidence bound) in systematically designed non-stationary bandit environments, using performance distributions over multiple random seeds and formal statistical comparisons.

The contributions of this research are threefold. First, the study provides empirical evidence that emotion-inspired mechanisms can outperform traditional algorithms in specific classes of unpredictable non-stationary environments. Second, the proposed architecture achieves its performance gains primarily through synergistic interactions among components rather than through any single mechanism in isolation. Third, these

findings suggest preliminary design considerations for adaptive AI systems operating in volatile settings, offering directions for future investigation in domains such as clinical decision support and financial trading.

MATERIALS AND METHODS

Simulation environment design

To comprehensively evaluate adaptive behaviors, we formalized the task as a non-stationary 5-armed bandit problem. At each trial $t \in 1, \ldots, 200$ the agent selects an action $a_t \in 0, 1, 2, 3, 4$ and receives a stochastic reward r_t drawn from a Gaussian distribution $r_t \sim \mathcal{N}(u_t(a_t), \sigma)$ where $\mu_t(a)$ represents the time-varying mean reward for action a and σ denotes the environment-specific noise level. The non-stationarity is introduced through changes in $\mu_t(a)$ at predetermined or stochastic change points.

Based on this formalism, we implemented three distinct simulation environments, each designed to test specific aspects of adaptation:

- Environment A tests responses to sudden environmental shifts, simulating scenarios where established strategies become abruptly suboptimal.
- Environment B evaluates performance in predictable cyclic patterns, representing regularly alternating conditions.
- Environment C assesses robustness to stochastic perturbations with random change points and occasional false signals.

Formally, Environment A can be viewed as a piecewise-stationary bandit with a single large abrupt change point, Environment B as a rotating-payoff bandit with deterministic periodic switching between optimal arms, and Environment C as a piecewise-stationary bandit with stochastic change points and transient reward spikes.

Reward scales and noise levels were chosen deliberately to instantiate different uncertainty regimes rather than to enable cross-environment comparisons. Our analyses therefore focus on within-environment comparisons among algorithms, and we additionally report recovery rates as percentages of environment-specific optima to aid interpretation.

Environment a: sudden strategy reversal

Environment A simulates catastrophic strategy failure through four phases across 200 trials. The critical transition at trial 100 tests rapid adaptation when previously reliable strategies suddenly fail.

Phase structure:

- **Initial stable period (trials 1–100).** Action 0 dominates with a mean reward of approximately 0.8 ± 0.1, establishing strong behavioral patterns in favor of this choice.
- **Abrupt reversal (after trial 100).** The optimal action shifts sharply from 0 to 4, creating a sudden mismatch between learned preferences and the new reward landscape.
- **New regime (trials 101–200).** Action 4 yields the highest reward ($\approx 0.9 \pm 0.1$), while the previously optimal Action 0 drops to a baseline level ($\approx 0.2 \pm 0.1$).

Minor variations in the later subphases introduce additional complexity without disrupting the main narrative of a single large reversal. Gaussian noise ($\boldsymbol{\sigma = 0.15}$) prevents single-trial learning while still maintaining clearly distinguishable strategies.

Environment B: predictable alternation

Environment B implements perfectly regular oscillations between two optimal strategies, testing pattern recognition rather than change detection.

Pattern structure:

- Every 40 trials, the optimal action alternates between 0 and 4.
- Rewards are extreme: the optimal action yields a mean reward of 0.95, whereas the opposite action yields 0.01.
- Intermediate actions (1, 2, 3) consistently provide minimal reward (0.05).
- Low noise ($\sigma = 0.05$) maintains near-deterministic outcomes and sharply defined phases.

This deterministic design evaluates whether emotion-based processing helps or hinders performance when simple pattern learning is sufficient for near-optimal behavior.

Environment C: stochastic disruptions

Environment C simulates unpredictable market conditions through random structural changes and false signals. Three change points occur randomly between trials 30–180, testing robustness when change timing is unknowable.

Key features:

- **Random transitions**: Change points vary each run, preventing pattern learning.

- **Unpredictable optimum**: At each phase change, the best action is reassigned randomly (constrained to differ from the previous optimal arm), yielding a mean reward of approximately 0.8 ± 0.15.

- **False signals**: On 5% of trials, temporary reward spikes ($\approx 0.7 \pm 0.15$) occur and persist for 1–2 trials, mimicking noise traders or false breakouts in real markets.

- **High uncertainty**: All non-optimal actions provide baseline reward of approximately 0.2 ± 0.15.

This design tests whether agents can maintain performance when both the timing and direction of changes are fundamentally unpredictable, requiring continuous monitoring rather than reliance on simple periodic patterns.

ECIA architecture

ECIA processes environmental feedback through four parallel mechanisms, each addressing a distinct aspect of adaptive decision-making (Fig. 1).

Environmental reward signals are processed in parallel by the External Limbic System (ELS), which computes eight emotion-like dimensions from recent reward dynamics, and by the dopamine-inspired learning module, which updates value estimates (Q-values) *via* prediction errors. PDU serves as the central integration hub, combining learned Q-values

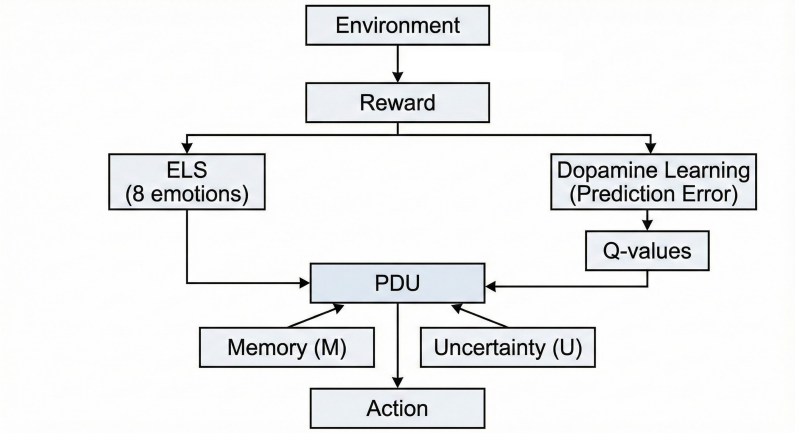


Figure 1 **ECIA architecture (Methods section)—Information flow diagram.**
Full-size DOI: 10.7717/peerj-cs.3688/fig-1

with emotional signals, episodic memory (M) retrievals, and uncertainty (U) bonuses to select actions.

External limbic system: emotion-like computational signals

The External Limbic System generates eight emotion-like signals informed by affective neuroscience. Drawing on dimensional models of emotion (*Plutchik, 1980*; *Russell & Barrett, 1999*), ECIA implements signals organized along four functional dimensions relevant to reinforcement learning:

- Joy/Fear: Valence dimension—reward approach *vs.* threat avoidance, implementing the fundamental "approach-withdraw" behavioral axis.
- Hope/Sadness: Temporal dimension—future expectation *vs.* sustained loss, capturing forward-looking optimism and retrospective discouragement.
- Curiosity/Anger: Exploration dimension—information-seeking *vs.* obstacle confrontation, balancing exploration with persistence.
- Pride/Shame: Self-evaluation dimension—success reinforcement *vs.* failure correction, implementing performance-based strategy adjustment.

This mapping ensures that ECIA captures the core functional dimensions of affect—valence, arousal, expectation, and self-monitoring—which collectively support adaptive decision-making (*Carver & Scheier, 1990*; *Russell & Barrett, 1999*). Alternative emotion taxonomies exist (*e.g.*, Ekman's six basic emotions; Panksepp's seven affective systems; Plutchik's eight primary emotions) (*Ekman, 1972*; *Panksepp, 2004*; *Plutchik, 1980*). Rather than directly replicating any single taxonomy, ECIA's selection prioritizes functional relevance to reinforcement learning: valence-based approach/avoidance, temporal expectation, exploration-exploitation balance, and performance-based self-regulation. The eight-emotion structure balances computational tractability with functional completeness, avoiding both oversimplification (fewer emotions) and excessive complexity (more granular taxonomies).

Each signal's triggering condition corresponds to known neural mechanisms:

- Fear: Activated by negative reward changes (reward < previous_reward − 0.15), echoing amygdala engagement in threat detection. The amygdala rapidly responds to potential threats and negative outcomes (*LeDoux, 1996*), with activity modulated by prediction error magnitude (*Schultz, Dayan & Montague, 1997*). In ECIA, fear intensity scales with the size of the recent reward drop, promoting withdrawal from actions associated with declining outcomes, analogous to Plutchik's withdrawal dimension (*Plutchik, 1980*).
- Joy: Triggered by high absolute rewards (reward > 0.7), inspired by ventral tegmental area (VTA) dopamine responses to appetitive outcomes. Dopaminergic neurons show phasic increases to unexpected or large rewards (*Berridge & Kringelbach, 2015*; *Schultz, 1998*). ECIA's joy signal grows with reward magnitude, reinforcing successful actions and implementing the "approach" pole of the valence axis.
- Hope: Responds to positive reward trends (linear regression slope > 0.03 over the recent four trials), motivated by dopamine "ramping" activity. During approach to anticipated rewards, dopamine firing can increase gradually, encoding proximity to expected positive outcomes (*Howe & Dombeck, 2016*). ECIA implements this as trend detection, providing anticipatory signaling that enables proactive rather than purely reactive adaptation.
- Sadness: Activated during sustained low performance (mean reward < 0.4 over six trials), reflecting prolonged negative states described in depression neuroscience. Extended negative prediction errors are associated with reduced striatal responses and altered prefrontal–limbic connectivity (*Eshel & Roiser, 2010*). ECIA's sadness signal integrates negative outcomes over time and dampens exploratory drive during extended poor performance, representing the functional opposite of joy.
- Curiosity: Driven by insufficient coverage of the action space (action diversity < 0.8), implementing an information-seeking mechanism analogous to the locus coeruleus–norepinephrine (LC–NE) system. LC–NE activity is sensitive to uncertainty and information gaps, supporting exploratory behavior (*Aston-Jones & Cohen, 2005*; *Gottlieb & Oudeyer, 2018*). In ECIA, curiosity increases when options are under-sampled and boosts exploration of rarely chosen actions, corresponding to Panksepp's "seeking" system.
- Anger: Activated when performance falls below an expected improvement trajectory (reward < expected − 0.2), modeling frustration-like responses. The anterior cingulate cortex signals conflicts between expected and actual outcomes and can trigger increased effort allocation (*Amsel, 1992*; *Botvinick, Cohen & Carter, 2004*). ECIA's anger signal temporarily increases exploration away from the most recent choice when anticipated learning progress does not materialize, encouraging strategic switching rather than passive persistence.
- Pride: Triggered by sustained high performance (success rate > 0.6 over five trials), inspired by social reward processing in the ventromedial prefrontal cortex (VMPFC). VMPFC activity tracks positive self-relevant outcomes and goal achievement

(*Izuma, Saito & Sadato, 2008*). ECIA's pride signal strengthens exploitation of successful strategies, stabilizing effective policies.

- Shame: Activated by sustained failures (failure rate > 0.5 over four trials), motivated by aversive social feedback signals in anterior insula and dorsal anterior cingulate. These regions respond to negative social evaluation and self-conscious emotions (*Takahashi et al., 2004*). ECIA's shame signal temporarily suppresses recently selected actions after repeated failures, discouraging perseveration on ineffective choices.

All emotional signals undergo temporal decay ($\gamma = 0.96$) with momentum effects, implementing the time-course dynamics observed in affective responses—emotions build gradually but persist beyond immediate triggers (*Kuppens, Oravecz & Tuerlinckx, 2010*). Signals are normalized to prevent runaway activation, analogous to homeostatic regulation in biological affect systems.

Importantly, labels such as 'Fear', 'Joy' and 'Pride' are used as convenient shorthand for these computational functions; they do not imply that ECIA literally implements human subjective emotions or detailed biological circuitry. The threshold and window parameters were selected based on preliminary experimentation and alignment with environmental characteristics. The Fear threshold (0.15) was set to exceed routine stochastic variation in Environments A and C ($\sigma = 0.15$). Temporal windows reflect differential timescales: shorter windows for reactive signals (Hope: four trials) and longer windows for sustained states (Sadness: six trials). The decay parameter ($\gamma = 0.96$) produces a half-life of approximately 17 trials.

Prefrontal decision unit: integrated action selection

The Prefrontal Decision Unit integrates multiple information sources for action selection, inspired by prefrontal cortex's role in complex decision-making (*Miller & Cohen, 2001*). The selection rule combines learned values with emotional, memory-based, and uncertainty signals:

$$A^* = arg\,\max_a [Q(s,a) + \eta \cdot E(s,a) + M(s,a) + U(s,a) + \xi]$$

where:

- Q(s,a): Learned action values (analogous to striatal value encoding)
- E(s,a): Emotional influence ($\eta = 0.55$), integrating limbic signals
- M(s,a): Episodic memory bias (hippocampal-inspired retrieval)
- U(s,a): Uncertainty bonus (LC-NE exploration signal)
- ξ: Gaussian exploration noise (motor variability)

The linear additive form represents a simplifying assumption. Biological prefrontal integration likely involves nonlinear interactions and gating mechanisms (*Miller & Cohen, 2001*), and alternative integration schemes (*e.g.*, multiplicative gating, attention-weighted combination) remain directions for future work. Nevertheless, linear formulation was chosen for interpretability: additive weights allow clear attribution of behavioral effects during ablation analysis. To mitigate dominance of individual terms, emotional signals are

normalized (as described above), and $\eta = 0.55$ was selected to balance emotional modulation against learned values.

This architecture mirrors the known integration of striatal value signals, prefrontal cognitive control, limbic emotional input, and hippocampal memory in mammalian decision-making (*Doya, 2008*; *O'Doherty, Cockburn & Pauli, 2017*).

Episodic memory system: experience storage and retrieval

The Episodic Memory System stores significant experiences based on prediction error magnitude ($|PE| > 0.015$), implementing computational principles of hippocampal memory formation. The hippocampus preferentially encodes surprising events—those with high prediction errors—while routine experiences are less likely to form lasting memories (*Duncan et al., 2012*; *Lisman & Grace, 2005*).

Stored memories include:

- Action-outcome associations
- Contextual features (temporal phase, normalized time)
- Emotional state at encoding
- Prediction error magnitude (salience marker)

Memory retrieval uses similarity-based search (threshold = 0.035), analogous to pattern completion in hippocampal CA3 networks. When current context matches stored patterns, relevant memories bias action selection, implementing rapid reinstatement of successful strategies (*Norman & O'Reilly, 2003*).

The system maintains 15 high-quality memories, prioritizing recent and high-salience experiences. This capacity limitation mirrors working memory constraints and implements a computational analogue of memory consolidation—only significant experiences survive competitive interference.

Adaptive learning module: dopamine-inspired plasticity

The Adaptive Learning Module implements learning rate modulation based on prediction errors and emotional intensity, inspired by dopaminergic modulation of synaptic (*Schultz, 2015*; *Sharpe et al., 2017*).

Base learning rate $\alpha = 0.22$ is adjusted by:

1. Prediction error magnitude: Larger surprises increase plasticity (surprise-driven learning (*Pearce & Hall, 1980*)).
2. Emotional intensity: High arousal states enhance memory encoding (amygdala-hippocampus interactions (*McGaugh, 2004*)).
3. Curiosity signals: Exploration states increase learning rates (LC-NE enhancement of plasticity (*Bouret & Sara, 2004*)).

This creates adaptive learning dynamics where the system learns faster during surprising, emotionally-salient, or exploratory episodes—precisely when environmental information is most valuable.

Computational abstraction and biological correspondence

ECIA's components are computational abstractions capturing functional principles of biological affective-cognitive systems. This level of abstraction is standard in computational neuroscience:

- Q-learning abstracts synaptic plasticity to simple value updates.
- Actor-critic models abstract basal ganglia-cortical loops.
- Temporal difference learning abstracts dopamine prediction errors.

Similarly, ECIA abstracts affective neuroscience findings into computationally tractable algorithms. The key distinction is that triggering conditions, interaction patterns, and temporal dynamics align with empirical neuroscience rather than being arbitrary design choices.

These four components interact through bidirectional connections:

- Emotions modulate memory storage salience and retrieval priorities.
- Memory retrieval triggers anticipatory emotional responses.
- Learning rates are jointly determined by prediction errors and emotional states.
- Decision-making integrates all information streams through prefrontal control.

This creates emergent adaptive behaviors that exceed the capabilities of individual components, analogous to how integrated brain systems produce flexible intelligence. Component contributions and synergistic effects are analyzed through systematic ablation studies reported in Results 'Component Contributions'.

Baseline algorithms and comparison framework

To ensure fair evaluation in non-stationary environments, we implemented two categories of baseline algorithms alongside ECIA and its ablation variants.

Improved baselines for non-stationary environments

Standard reinforcement learning algorithms assume stationary reward distributions, making them unsuitable for direct comparison in our time-varying environments. We therefore implemented three adaptive baseline algorithms specifically designed for non-stationary multi-armed bandit problems: Context-Aware ε-Greedy extends the standard ε-greedy approach with environmental change detection. The algorithm monitors reward distributions by comparing recent performance (most recent 10 trials) against older patterns (trials 20–30 steps back). When the mean reward difference exceeds a threshold (0.25) and surpasses combined noise variance, the algorithm interprets this as an environmental change. Upon detection, it performs a partial reset of Q-values ($Q \leftarrow 0.5Q$) to balance knowledge retention with adaptation, while temporarily increasing exploration. This mechanism allows the algorithm to maintain useful knowledge while remaining responsive to distributional shifts. Parameters: $\varepsilon = 0.1$, $\alpha = 0.1$, window size = 50, change threshold = 0.25. Sliding Window Upper Confidence Bound (SW-UCB) addresses non-stationarity by maintaining only recent observations for each

action (*Garivier & Moulines, 2008*). Rather than accumulating all historical data, SW-UCB uses a fixed window of the most recent W observations per action. The upper confidence bound is calculated using only these windowed statistics:

$$UCB(a) = \mu_w(a) + c\sqrt{\frac{ln(t)}{n_w(a)}}.$$

where $\mu_W(a)$ represents the mean reward within the window, $n_W(a)$ the windowed action count, and c = 2.0 the exploration parameter. This windowing naturally discounts outdated information, enabling the algorithm to track changing reward distributions without explicit change detection. Parameters: W = 100, c = 2.0, minimum samples = 5. Adaptive Thompson Sampling extends Bayesian Thompson Sampling with exponential recency weighting (*Raj & Kalyani, 2017*). The algorithm maintains a Gaussian posterior distribution for each action's expected reward but applies a discount factor $\gamma = 0.99$ to all sufficient statistics at each time step before incorporating new observations. Specifically, sum_rewards $\leftarrow \gamma \times$ sum_rewards + new_reward, with similar updates for sum_squared_rewards and effective_count.

This exponential weighting gives recent observations higher influence on the posterior distribution while gracefully degrading the impact of older data. The posterior mean and variance are computed from these discounted statistics, and action selection follows standard Thompson Sampling by drawing samples from each action's posterior. Parameters: $\gamma = 0.99$, minimum std = 0.1, periodic reset threshold = 100 trials.

Naive baselines

For reference and to demonstrate the necessity of adaptation mechanisms, we also implemented standard versions without temporal adaptations:

- Naive ε-Greedy: Standard ε-greedy ($\varepsilon = 0.1$) with sample-average Q-value updates ($\alpha = 1/n$).
- Naive UCB: Standard UCB1 algorithm with exploration parameter c = 0.5.
- Naive Thompson Sampling: Standard Gaussian Thompson Sampling without discount factor.

These serve as reference points to isolate the contribution of adaptation mechanisms in non-stationary environments.

ECIA and ablation variants

As described in §2.2, ECIA comprises four integrated components. We tested seven systematic ablation variants to quantify individual component contributions:

- ECIA-NoEmotion: Emotion vector set to zero.
- ECIA-NoMemory: Memory bias removed.
- ECIA-NoDopamine: Fixed learning rate ($\alpha = 0.1$).
- ECIA-NoDop-NoMem: Combined removal of dopamine and memory.
- ECIA-NoDop-NoEmo: Combined removal of dopamine and emotion.

- ECIA-NoMem-NoEmo: Combined removal of memory and emotion.
- ECIA-NoAll: All advanced components removed.

These ablations isolate the contributions of emotional processing, episodic memory, and adaptive learning rates to ECIA's overall performance.

Fair comparison protocol

To ensure fair comparison, all agents operate on the same observation stream: the sequence of chosen actions and received rewards. Additionally, normalized time (t/T) is available as contextual information; however, only ECIA explicitly incorporates this temporal context through episodic memory storage and retrieval. In Environments A and B, environmental phases are deterministic functions of time, meaning any algorithm could infer phase information from the time step alone. In Environment C, phase transitions occur at stochastic change points; here, contextual information provides additional environmental cues.

We acknowledge that this explicit use of normalized time may confer an advantage in Environments A and B, where phase is a deterministic function of time. However, in Environment C, phase transitions occur at stochastic change points unknown to all agents, yet ECIA maintains its performance advantage, suggesting benefits beyond time-based prediction. Ablation experiments isolating the contribution of temporal context were not conducted and represent a direction for future work.

The critical distinction lies in information utilization strategy. Improved baselines detect environmental changes through reward pattern analysis, inferring changes from statistical properties of observed rewards to adapt after distributional shifts occur. ECIA integrates context into its emotion-memory-dopamine system, enabling anticipatory behavior through emotional signals (hope, fear) and rapid retrieval of similar past situations *via* episodic memory.

This design evaluates different adaptive strategies—context-integrated processing *vs.* purely statistical adaptation—rather than comparing algorithms with unequal information access.

Statistical analysis and evaluation framework

We employed large-scale experimental replication with systematic seed variation to ensure robust and generalizable results. Traditional RL studies often report results from single random seeds, risking seed-specific biases and limiting generalizability. We address this through extensive Monte Carlo simulation: 12 Fibonacci-sequence master seeds (34 through 6,765), each generating 300 independent experimental runs with different random initializations, totaling 3,600 experiments per algorithm per environment. This experimental replication design provides three key advantages: (1) avoids cherry-picking favorable seeds while maintaining reproducibility through systematic seed selection, (2) quantifies performance variance to assess algorithm stability across different random conditions, and (3) enables robust statistical inference with sufficient power for detecting true performance differences. Each master seed generates different random configurations, effectively testing algorithms across a distribution of scenarios rather than single instances.

The 12 seed-level means serve as independent observations for statistical testing, avoiding pseudo-replication while providing reliable estimates of central tendency and variance. All ± values reported represent standard deviations (SD) across the 12 seed-level means. For hypothesis testing, the 12 seed-level means served as independent observations (df = 11 for paired comparisons).

Performance evaluation used four unified metrics:

- Overall Performance: Mean reward across all 200 trials.
- Recovery Rate: Proportion of optimal performance regained after environmental changes, calculated as post-change performance divided by environment-specific optimal rewards.
- Recovery Time: Number of trials required to achieve stable performance recovery, with stability windows adjusted for environmental noise levels.
- Emotional Dynamics: Eight-dimensional emotion trajectories tracked throughout experiments (ECIA variants only).

Recovery metrics are formally defined as follows. Recovery Rate quantifies post-change performance relative to the new optimum:

$$\text{Recovery Rate} = \frac{\bar{R}_{post}}{R^*_{post}}$$

where $\bar{R}_{post}$ is the mean reward over 30 trials following a change point, and R^*_{post} is the optimal expected reward in the new regime. Recovery Time is defined as the minimum number of trials after a change point to reach stable performance:

$$t_{recovery} = min\left\{ t : \frac{1}{w} \sum_{i=t}^{t+w-1} R_i \geq \theta \cdot R^*_{post} \right\}$$

where w is the stability window and θ is the threshold ratio ($w = 5$, $\theta = 0.85$ for Environments A and C; $w = 3$, $\theta = 0.90$ for Environment B). The reported recovery time is $t_{recovery} + \lfloor w/2 \rfloor$.

Statistical testing followed hierarchical procedures using the 12 seed-level means as independent observations: Shapiro-Wilk tests for normality, appropriate t-tests or Mann-Whitney U based on distributions, Cohen's d for effect sizes, and Bonferroni correction for multiple comparisons. This approach avoids pseudoreplication while providing robust inference.

We conducted systematic ablation studies comparing the full ECIA architecture against seven component-stripped variants using the same experimental replication framework. All implementations used Python 3.8 with synthetic data generation.

Computational resources

All simulations were conducted on a workstation equipped with an Intel Core i7-12700 CPU (12th generation) and 32 GB RAM (4,800 MT/s), running Windows 11. The implementation utilized Python 3.9 with NumPy 1.23.5. The average execution time per

Table 1 Performance summary across non-stationary environments. Values represent mean performance across N = 3,600 experiments (12 master seeds × 300 runs).

Agent	Env A: mean reward	Env A: recovery rate	Env B: mean reward	Env B: recovery rate	Env C: mean reward	Env C: recovery rate
ECIA (Proposed)	0.8194	97.3%	0.8014	77%	0.7618	81.6%
Sliding window UCB	0.6011	90.4%	0.6430	69.5%	0.5514	68.7%
Adaptive Thompson sampling	0.6027	28.3%	0.5540	49.7%	0.5246	33.8%
Naive UCB	0.7685	74.3%	0.8522	86.5%	0.6741	61.6%
Naive Thompson sampling	0.5754	30.7%	0.4799	41.7%	0.4810	32.9%
Context-Aware E-Greedy	0.5350	28.6%	0.4846	35.8%	0.3701	35.8%
Naive E-Greedy	0.5509	29.8%	0.5658	48.5%	0.4211	33.8%

Note:
All standard deviations <0.01.

200-trial experiment was approximately 0.31 s for the full ECIA model, compared to 0.03 s for the Naive UCB baseline. While ECIA requires additional computation compared to simpler baselines due to emotional processing and memory retrieval, these results demonstrate that the architecture maintains computational tractability suitable for research simulations and potential online learning applications.

RESULTS

Performance across environments

Monte Carlo simulation across 3,600 experimental runs per agent per environment revealed environment-specific performance patterns (Table 1, Fig. 2). The small standard deviations observed (<0.01 for most metrics) reflect that each reported value is an average across 300 independent runs per seed, producing highly stable estimates.

In Environment A, ECIA achieved overall performance of 0.8194 (±0.0007), significantly outperforming Naive UCB (0.7685, ± 0.0023; d = 29.6, $p < 0.001$) and all other baselines. ECIA also surpassed improved baselines including Sliding Window UCB (0.6011) and Adaptive Thompson Sampling (0.6027). Notably, the improved non-stationary baselines (SW-UCB, Adaptive TS) underperformed Naive UCB despite being designed for non-stationary settings; possible explanations are considered in the Discussion.

Environment B showed contrasting results with Naive UCB achieving the highest performance at 0.8522 (±0.0008) compared to ECIA's 0.8014 (±0.0036; d = 19.4, $p <$ 0.001). This reversal indicates that ECIA's complex emotion-memory architecture incurs computational costs in highly predictable environments where simpler algorithms suffice.

In Environment C, ECIA regained superiority with 0.7618 (±0.0021), substantially outperforming Naive UCB (0.6741, ± 0.0035; d = 30.4, $p < 0.001$), Sliding Window UCB (0.5514), and Adaptive Thompson Sampling (0.5246). This confirms ECIA's advantage in managing stochastic, unpredictable environments.

Recovery dynamics

Recovery metrics confirmed environment-dependent advantages (Table 2, Fig. 3).

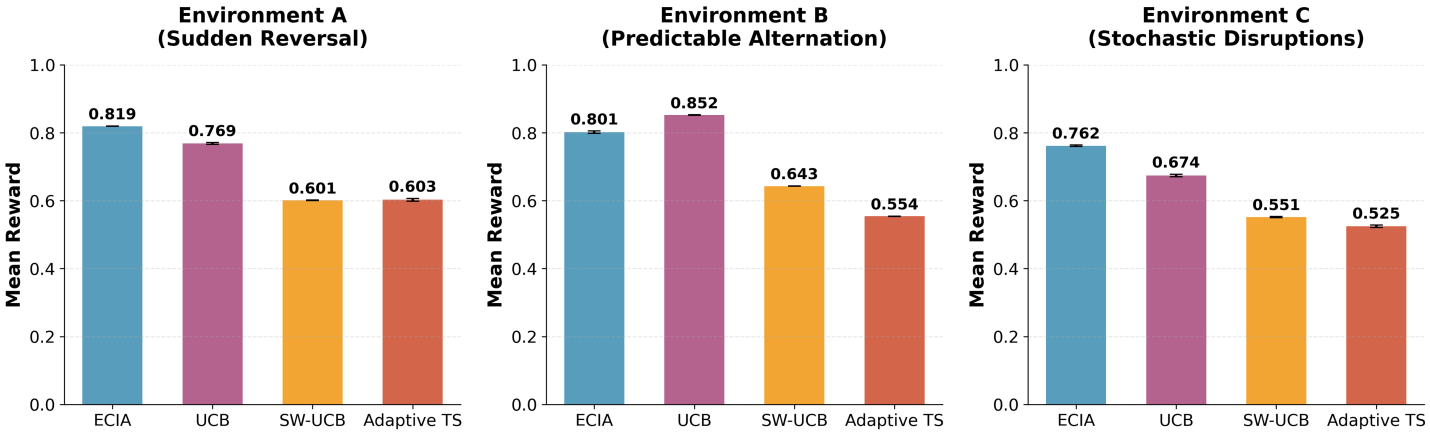


Figure 2 Comparative analysis of mean rewards across three non-stationary environments. Bar charts represent the average accumulated reward over 200 trials (N = 3,600 runs per agent). The proposed ECIA (Emotional-Cognition Integration Architecture) is compared against two improved baselines: SW-UCB (Sliding Window Upper Confidence Bound) and Adaptive TS (Adaptive Thompson Sampling), as well as the standard Naive UCB. (A) Environment A: Sudden Strategy Reversal. (B) Environment B: Predictable Alternation. (C) Environment C: Stochastic Disruptions. Error bars indicate standard deviation. Note that ECIA outperforms baselines in uncertain environments (A and C), while Naive UCB shows superior efficiency in the predictable Environment B. Full-size DOI: 10.7717/peerj-cs.3688/fig-2

Table 2 Summary of recovery metrics: adaptation speed and stability across non-stationary environments. Values represent the mean across 3,600 experimental runs. Recovery Time: The number of trials required to regain performance stability after a regime change (lower values indicate faster adaptation). Recovery Rate: The percentage of the optimal reward retrieved after adaptation (higher percentages indicate better stability). Abbreviations: SW-UCB (Sliding Window Upper Confidence Bound), Adaptive TS (Adaptive Thompson Sampling).

Algorithm	Env A recovery time (trials)	Env A recovery rate (%)	Env B recovery time (trials)	Env B recovery rate (%)	Env C recovery time (trials)	Env C recovery rate (%)
ECIA	7.0	97.3	8.2	77.0	8.4	81.6
Naive UCB	15.0	74.3	5.1	86.5	17.5	61.6
SW-UCB	6.2	90.4	6.0	69.5	14.0	68.7
Adaptive TS	44.5	28.3	22.3	49.7	38.2	33.8

Note:
ECIA achieves the fastest and most stable recovery in uncertain environments (A and C), whereas Naive UCB shows superior recovery in the predictable Environment B.

In Environment A, ECIA demonstrated fastest recovery at 7.0 trials, significantly faster than Naive UCB (15.0 trials) and comparable to Sliding Window UCB (6.2 trials). ECIA's recovery rate reached 97.3%, substantially higher than Naive UCB's 74.3%, suggesting that emotion-triggered adaptation enables rapid strategy shifts after sudden environmental changes.

Environment B showed reversed patterns. Naive UCB achieved fastest recovery at 5.1 trials with an 86.5% recovery rate, while ECIA required 8.2 trials with 77.0% recovery. This demonstrates that in predictable cyclic environments, simpler windowing mechanisms outperform complex emotional processing.

In Environment C, ECIA maintained superior recovery performance with 8.4 trials and 81.6% recovery rate, compared to Naive UCB's 17.5 trials and 61.6% rate. Sliding Window UCB showed intermediate performance (14.0 trials, 68.7% rate).

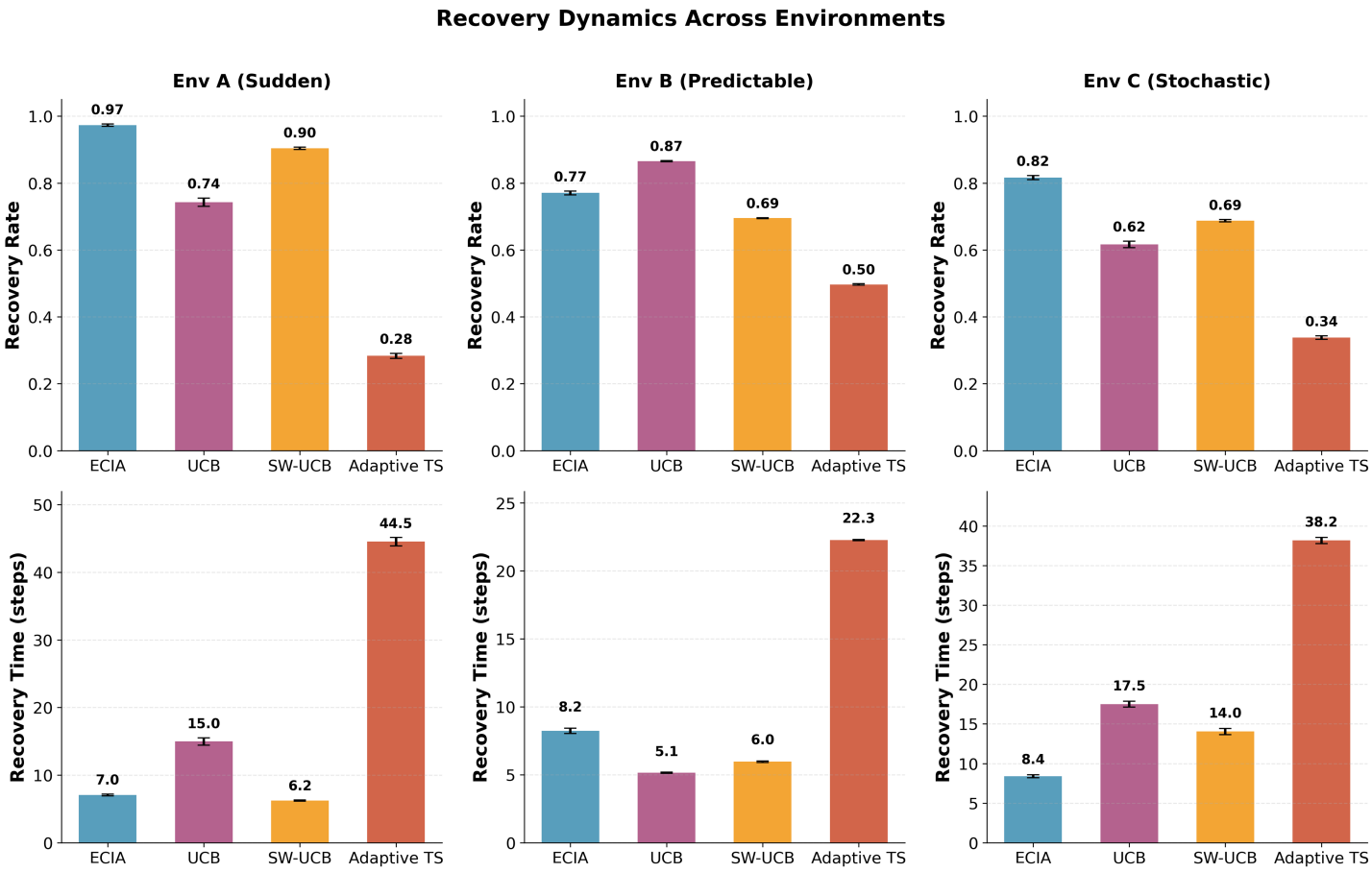


Figure 3 **Recovery dynamics across non-stationary environments.** This figure compares the adaptability of agents following regime changes. (Top Row) Recovery Rate: The percentage of optimal reward regained after a change. Higher values indicate better stability. ECIA (Blue) achieves near-optimal recovery in Env A (97%) and C (82%). (Bottom Row) Recovery Time: The number of trials required to stabilize. Lower values indicate faster adaptation. ECIA demonstrates the fastest recovery in uncertain environments (A: 7.0 trials, C: 8.4 trials), significantly outperforming baselines. Note that in the predictable Environment B, Naive UCB (Green) recovers fastest, illustrating the trade-off between adaptability and efficiency.
Full-size DOI: 10.7717/peerj-cs.3688/fig-3

Component contributions

Ablation analysis demonstrated synergistic effects (Fig. 4). Individual component removal produced modest performance changes in Environment A: NoEmotion (+0.05%), NoMemory (−0.39%), and NoDopamine (−0.97%). However, combined removals caused disproportionate degradation: NoDopamine+NoMemory resulted in −1.91%, and removing all components (NoAll) degraded performance by −2.24%.

Notably, dopamine removal showed opposite effects across environments: performance improved in Environment B (+5.11% for NoDopamine *vs.* Full ECIA) but degraded in Environment C (−5.01%). This environment-specific effect indicates that adaptive learning rates benefit uncertain environments but may overcomplicate responses in predictable patterns.

Recovery time showed the strongest synergistic effects. In Environment A, removing all components increased recovery time by 41.97% (from 7.04 to 9.99 trials), far exceeding the

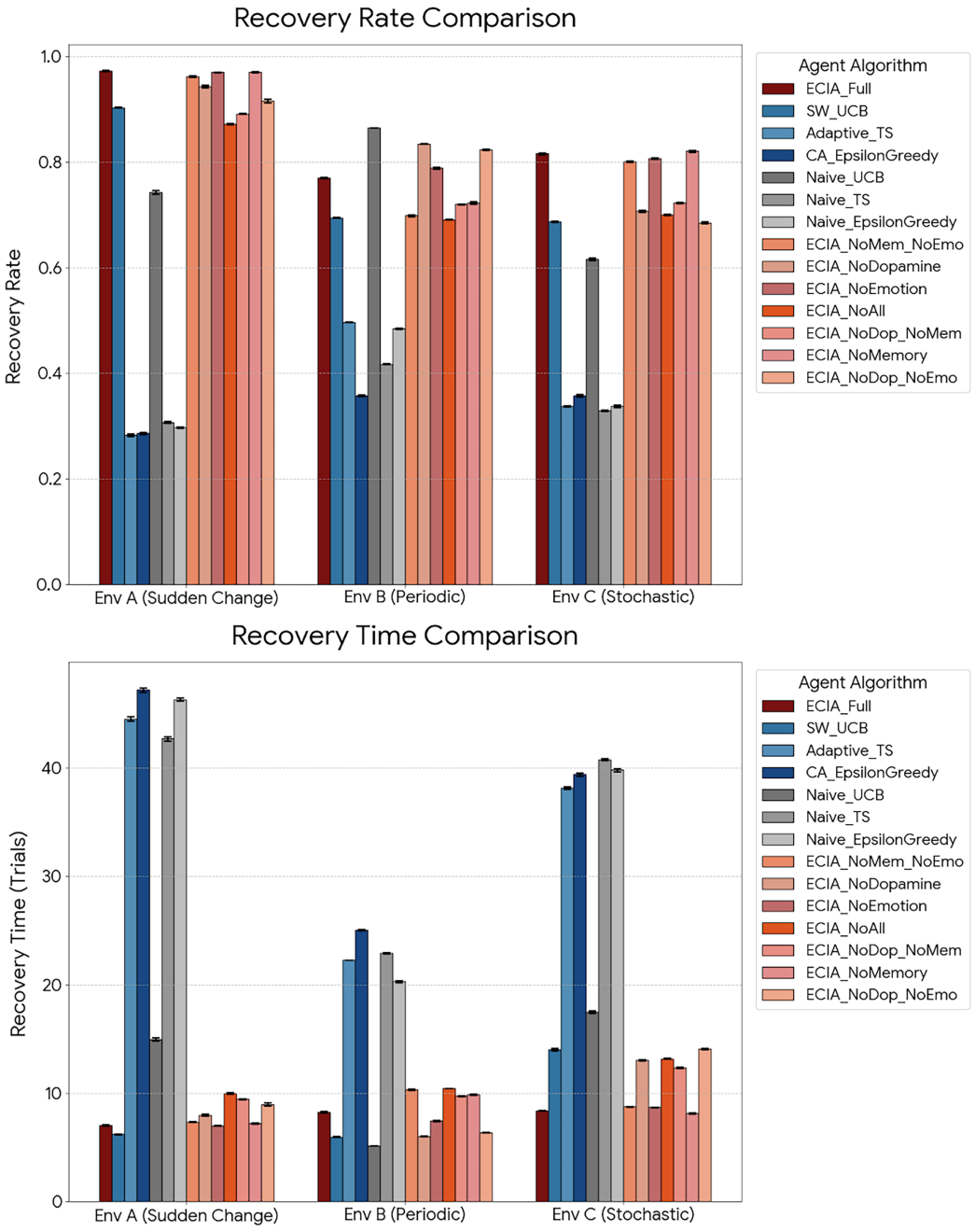


Figure 4 Ablation study—component contribution analysis. This figure illustrates the impact of removing individual or combined components on overall performance across three environments. Abbreviations: Full ECIA (Complete architecture), NoEmotion (External Limbic System disabled), NoMemory (Episodic Memory bias removed), NoDopamine (Adaptive Learning Rate disabled), NoAll (All three components removed). *Note:* In uncertain environments (A and C), the Full ECIA outperforms the NoAll variant, demonstrating synergistic integration. Conversely, in the predictable Environment B, removing the Dopamine component paradoxically improves performance, indicating that high adaptive plasticity can be costly in stable settings. Full-size DOI: 10.7717/peerj-cs.3688/fig-4

sum of individual component impacts (NoEmotion: −0.46%, NoMemory: +2.34%, NoDopamine: +13.42%).

Summary of key findings

The comprehensive evaluation revealed three principal findings:

1. **Environment-specific performance advantages.** ECIA demonstrated superior performance in unpredictable environments (Env A: 0.8194 *vs*. Naive UCB 0.7685; Env C: 0.7618 *vs*. Naive UCB 0.6741) while showing inferior performance in the predictable Environment B (0.8014 *vs*. Naive UCB 0.8522). This suggests a functional specialization for uncertainty management rather than simple pattern recognition.
2. **Rapid and robust recovery in uncertain environments.** ECIA achieved the fastest recovery in Environment A (7.0 trials, 97.3% rate) and Environment C (8.4 trials, 81.6% rate), substantially outperforming Naive UCB. Conversely, it showed slower recovery in the predictable Environment B, where simpler algorithms sufficed.
3. **Strong synergistic component effects.** While individual components contributed modestly (0.05–5.11% impact), their combined removal produced disproportionate degradation (up to 41.97% in recovery time), indicating integrated rather than additive functionality. The dopamine system exhibited environment-specific utility, improving performance when removed in predictable settings (+5.11% in Env B) but degrading it in stochastic conditions (−5.01% in Env C).

These results establish that emotion-based processing provides selective advantages in non-stationary environments, with effectiveness dependent on environmental predictability and change characteristics.

DISCUSSION

Performance trade-offs across environmental regimes

Our experimental replication across 3,600 runs per environment reveals a fundamental trade-off between adaptability and computational efficiency. ECIA demonstrated superior performance in unpredictable environments—achieving mean rewards of 0.8194 in Environment A and 0.7618 in Environment C—significantly outperforming both naive baselines and sophisticated adaptive algorithms such as Sliding Window UCB and Adaptive Thompson Sampling. However, in the strictly deterministic Environment B, ECIA underperformed compared to simpler algorithms (Naive UCB: 0.8522 *vs*. ECIA: 0.8014), and even its ablated NoDopamine" variant showed superior performance.

This pattern suggests that emotion-inspired computational mechanisms are functionally specialized for managing uncertainty rather than optimizing performance in static or highly predictable environments. In Environment B, the optimal strategy requires simple pattern repetition—alternating between actions 0 and 4 every 40 trials. ECIA's emotional mechanisms, particularly "Curiosity" (triggered by insufficient action space coverage) and "Fear" (activated by negative reward changes), induced exploratory behaviors and hypothesis verification even when the pattern had stabilized. This constant vigilance incurs what we term a "cost of complexity"—additional computation and

behavioral variability that provide no benefit when environmental dynamics are deterministic.

Conversely, in Environment C with stochastic change points and false signals, this same vigilant mechanism allowed ECIA to distinguish meaningful regime changes from random noise, preventing the catastrophic performance drops observed in naive baselines (Naive UCB: 0.6741, Adaptive TS: 0.5246). The emotion-based system functions as an adaptive interrupt mechanism that reassesses environmental conditions when prediction error signals accumulate—a costly process that proves necessary only in high-uncertainty settings. These findings align with computational perspectives suggesting that biological affective systems evolved not for efficiency in stable niches but for managing volatility in changing environments (*Bach & Dayan, 2017*; *Eldar et al., 2016*).

Component synergy and mechanistic understanding

The ablation study provides critical mechanistic insights, particularly regarding the paradoxical effect of dopamine removal across environments. Removing the dopamine-inspired adaptive learning module improved performance in Environment B while degrading it in Environment C. This environment-specific effect illuminates the functional role of adaptive plasticity.

The dopamine module in ECIA increases learning rates in response to prediction errors, implementing the principle that learning should intensify when environments are surprising. In the predictable alternation of Environment B, every reward change at the 40-trial boundary constitutes a large prediction error—but this is a regular, scheduled event rather than a genuine environmental anomaly. The full ECIA architecture interprets these scheduled transitions as surprises, triggering high plasticity and emotional arousal that temporarily destabilize the policy. When dopamine modulation is removed, the agent maintains a fixed, lower learning rate that paradoxically enhances stability in this rhythmic environment.

However, in the stochastic Environment C, this fixed stability becomes a severe liability. Without adaptive plasticity, the agent cannot rapidly recalibrate when genuine regime shifts occur at unpredictable intervals. The synergistic removal of all components (NoAll) caused a disproportionate degradation in recovery time in Environment A, far exceeding the additive sum of individual component impacts. This super-additive effect indicates that the External Limbic System, Episodic Memory, and Adaptive Learning modules form an integrated loop rather than functioning as independent modules.

Specifically, emotions highlight salient events for preferential memory storage, memory retrieval subsequently modulates emotional responses, and both emotional intensity and memory salience jointly determine learning rates. This bidirectional integration creates a self-organizing system where surprising experiences generate emotions, emotions enhance memory consolidation, and retrieved memories prepare the system for similar future scenarios. Disrupting any single component weakens this loop modestly, but removing multiple components collapses the entire adaptive cycle.

Comparison with existing approaches

ECIA occupies a distinct niche in the landscape of non-stationary reinforcement learning methods. Compared to statistical adaptation approaches—such as Sliding Window UCB which relies on fixed forgetting windows (*Garivier & Moulines, 2008*), or Adaptive Thompson Sampling which applies exponential discounting (*Raj & Kalyani, 2017*), both included as improved baselines in our experiments—ECIA modulates its adaptation dynamically through internal state variables rather than external hyperparameters. This enables flexible response to different types of environmental change without manual recalibration.

Indeed, our results revealed that these improved baselines underperformed Naive UCB, exposing two distinct failure modes. SW-UCB achieved the fastest recovery (6.2 trials) due to its short window, but suffered from persistent exploration variance: its fixed window ($W = 100$) prevents the algorithm from reaching a high-confidence exploitation state, causing sub-optimal exploration even long after the environment stabilized. Conversely, Adaptive TS suffered from high inertia; its discount factor ($\gamma = 0.99$) retained outdated information too long, leading to the slowest recovery (44.5 trials) and lowest recovery rate (28.3%). Naive UCB, while theoretically less adaptive, benefited from larger cumulative sample sizes that allowed stable exploitation once transition periods passed. These findings underscore that non-stationary algorithms require careful parameter tuning matched to specific environmental dynamics.

Meta-Reinforcement Learning approaches achieve adaptability through extensive offline training across diverse task distributions (*Duan et al., 2016*; *Wang et al., 2016*), learning policies that can quickly adapt to new tasks from the same distribution. ECIA demonstrates comparable adaptability without requiring offline meta-training, instead relying on biologically-inspired heuristics (emotions) that provide "zero-shot" adaptation to novel environmental dynamics. This architectural difference has practical implications: meta-RL excels when training distributions are available and representative of deployment conditions, while ECIA may be more suitable for scenarios where pre-training is infeasible or where deployment environments differ substantially from any available training data.

However, this flexibility comes at computational cost. ECIA requires approximately 0.31 s per 200-trial experiment compared to 0.03 s for Naive UCB—a 10-fold increase due to emotional processing, memory retrieval, and adaptive learning rate computations. While this overhead remains computationally tractable (complete 3,600-run simulations require under 30 min), it represents a meaningful trade-off: ECIA sacrifices raw computational speed for architectural flexibility. For applications where environmental uncertainty is high and computational resources are available, this trade-off appears favorable; for low-latency or resource-constrained scenarios, simpler baselines may be preferable.

Beyond non-stationary bandits and meta-reinforcement learning, recent work in multimodal and affective AI has explored complementary directions. Multimodal emotion recognition systems integrate visual, acoustic, and linguistic signals to infer human affective states (*Lian et al., 2023b*; *Mittal et al., 2020*), while affect-modulated agents and

adaptive affective architectures adjust their behavior based on internal or externally inferred emotion-like variables (*Hong et al., 2020*; *Zhou, 2021*). ECIA is complementary to these approaches: rather than focusing on perception or human–agent interaction, it treats emotion-like signals as internal control variables for adaptation in non-stationary reward landscapes. In future work, the ECIA framework could be combined with multimodal affective front-ends or generative policy architectures to build end-to-end systems that both perceive and leverage affective structure for adaptive decision-making.

Implications, limitations, and broader perspectives

Practical implications

The environment-specific performance patterns provide actionable guidance for algorithm selection. ECIA is best suited for domains characterized by:

1. **High environmental uncertainty with unpredictable change timing** (analogous to Environments A and C): financial markets with regime shifts, clinical decision support where patient conditions change abruptly, adaptive control systems in dynamic operational contexts.
2. **Multiple competing action strategies** requiring rapid re-evaluation when conditions shift, rather than simple binary choices.
3. **Availability of computational resources** sufficient to support the multi-component architecture (approximately 10x overhead *vs*. simplest baselines).

Conversely, ECIA is less suitable for:

1. **Highly predictable environments** with deterministic patterns (analogous to Environment B): manufacturing process control, scheduled logistics, routine administrative tasks where simple rules suffice.
2. **Hard real-time systems** where microsecond-level latency requirements preclude the emotional processing overhead.
3. **Extremely high-dimensional problems** where the episodic memory system's quadratic similarity comparisons become prohibitive (though this could be addressed with approximate nearest-neighbor methods in future work).

Limitations

Several limitations qualify these findings. First, while computationally tractable, ECIA introduces additional hyperparameters (emotional decay rates, memory similarity thresholds, dopamine modulation coefficients) requiring tuning. The current parameter values ($\eta = 0.55$, $\gamma = 0.96$, *etc.*) were selected through preliminary experimentation rather than systematic optimization, and performance is sensitive to these settings. Future work should develop principled methods for automatic parameter configuration and explore whether different values might better approximate human emotional decision-making or optimize artificial agent performance. Similarly, sensitivity analysis for recovery metric parameters (stability window, threshold ratio) was not conducted.

Second, the experiments employed synthetic multi-armed bandit environments. While these effectively isolate fundamental decision-making dynamics under non-stationarity, they lack the high-dimensional state spaces, partial observability, and long-horizon credit assignment challenges of real-world problems. Extending ECIA to Deep Reinforcement Learning frameworks—where emotional signals would modulate neural network learning rates and attention mechanisms—represents an important direction for validating these principles in complex domains.

Third, the selection of eight specific emotions, while informed by dimensional models of affect remains somewhat arbitrary (*Plutchik, 1980*; *Russell & Barrett, 1999*). Alternative emotion taxonomies might yield different performance characteristics (*Ekman, 1972*; *Panksepp, 2004*). Systematic ablation across emotion subsets could identify minimal sufficient sets for different problem classes.

Additionally, in Environment C, the provision of phase context following stochastic change points may have provided ECIA with an informational advantage beyond what would be available in purely observational settings. Future work should therefore evaluate ECIA under observational conditions without explicit phase cues to isolate the contribution of emotional processing from contextual awareness. Furthermore, evaluating ECIA on widely used benchmark suites for non-stationary bandits would facilitate direct comparison with alternative adaptive algorithms and strengthen the generalizability of these findings. Temporal visualizations such as cumulative regret curves and post-change recovery trajectories were not included; such diagnostics would provide additional empirical support for adaptation dynamics.

Broader perspectives: speculative computational insights

While our synthetic bandit experiments cannot directly validate evolutionary hypotheses—lacking population dynamics, generational turnover, or ecological realism—they may nonetheless offer computational insights into enduring puzzles about human affective behavior. The following discussion extends beyond our experimental findings and should be interpreted as hypothesis-generating rather than empirically validated claims.

Why do emotions persist in human cognition despite their measurable costs in stable contexts? Our results suggest a possible computational answer: emotions may be not uniformly irrational but *selectively rational*—adaptive for managing uncertainty, maladaptive for exploiting stability. In environments where change is constant (financial crises, medical emergencies, social conflicts), rapid affective responses may enable survival through quick strategy shifts. In stable niches (routine procedures, assembly line work), these same mechanisms may become costly noise that disrupts efficient pattern execution.

More intriguingly, why do individuals vary so dramatically in emotional reactivity? Some people remain calm and analytical under pressure; others respond intensely and rapidly to minor provocations. Traditional psychology frames this through personality dimensions (neuroticism, impulsivity, conscientiousness), but our computational model suggests an alternative speculative interpretation: these individual differences might represent different *cognitive niche specializations*.

If ancestral human populations faced diverse environmental regimes—some groups in relatively stable resource-rich valleys, others in volatile tundra or coastal regions subject to rapid climate shifts during glacial transitions—different emotional response profiles might have been computationally advantageous in different contexts. An individual with high emotional reactivity and rapid affect-driven decision-making (high "Fear" and "Curiosity" sensitivity, low emotional decay rates) might struggle in predictable environments but excel when conditions shift unpredictably, analogous to how ECIA outperforms in Environments A and C. Conversely, an individual with dampened affective responses and stable analytical processing (low emotional sensitivity, fixed learning rates) might show superior performance in predictable contexts but suffer during volatility.

This computational framework does not constitute evidence for specific evolutionary trajectories—establishing such connections would require extensive interdisciplinary research combining agent-based evolutionary modeling, paleoclimatic reconstructions, comparative neuroscience, and cross-cultural psychological studies. However, it offers mechanistic hypotheses that such research could rigorously test: (1) that affective variability may represent computational trade-offs optimized for different environmental regimes rather than simply reflecting "better" or "worse" cognitive control, and (2) that the persistence of emotion-driven decision-making in modern humans may reflect specialization for managing uncertainty rather than representing evolutionary lag or suboptimal design (*Bach & Dayan, 2017*; *Eldar et al., 2016*).

Future computational work could implement evolutionary simulations where agent populations face varying environmental volatility over generations, testing whether emotional response diversity emerges spontaneously through selection pressures and whether resulting population-level affective distributions resemble those observed in human behavioral studies.

CONCLUSION

This study introduced the Emotional-Cognition Integration Architecture (ECIA), demonstrating that biologically-inspired emotional processing provides selective computational advantages for managing environmental uncertainty. Through rigorous experimental replication across 3,600 runs per environment, three principal findings emerged: (1) ECIA significantly outperformed traditional and adaptive algorithms in unpredictable environments (Environment A: 0.8194 *vs.* Naive UCB's 0.7685; Environment C: 0.7618 *vs.* 0.6741) while incurring costs in predictable contexts (Environment B: 0.8014 *vs.* 0.8522), demonstrating environment-specific specialization rather than universal superiority; (2) components exhibited strong synergistic integration, with combined removal causing disproportionate degradation (41.97% increase in recovery time) far exceeding individual effects; and (3) dopamine-inspired adaptive plasticity showed opposite effects across environments (+5.11% in predictable settings, −5.01% in stochastic ones), revealing that learning rate modulation benefits uncertain environments but can destabilize predictable ones. These findings support perspectives from affective neuroscience that frame emotions as functional computational heuristics specialized for managing volatility rather than mere impediments to rational

decision-making (*Bach & Dayan, 2017*; *Damasio, 2001*). The observed trade-offs—superior adaptability in uncertain environments at the cost of efficiency in stable ones—suggest that emotion-inspired architectures merit consideration for AI systems operating under real-world uncertainty, particularly in domains such as clinical decision support and financial trading where environmental dynamics shift unpredictably.

Future work should extend ECIA to deep reinforcement learning frameworks and validate these principles in complex real-world domains. These conclusions are necessarily bounded by the use of simplified multi-armed bandit tasks with synthetic non-stationarity and by the additional computational overhead of the architecture relative to simple baselines. By computationally demonstrating the adaptive advantages of mammalian-inspired affective mechanisms, this study highlights that emotion-like processes can serve as essential architectural components for navigating an uncertain world. We advocate a paradigm shift toward AI models that more actively emulate biological nervous systems rather than purely abstract optimization procedures. Ultimately, we hope this research contributes—even in a modest way—toward a future in which AI systems transcend mere computational efficiency and move toward a more integrated form of emotional–cognitive capacity, bringing the field one step closer to artificial intelligence that mirrors the resilient and adaptive nature of the human mind.

ACKNOWLEDGEMENTS

The authors used several AI-assisted tools (including Claude, Gemini, ChatGPT, and Perplexity) for English language editing, stylistic refinement of text translated from Korean to English, and limited assistance with Python code debugging. These tools were not used for study design, data generation, statistical analysis, or interpretation of the results. All experimental design, data analysis, result interpretation, and substantive manuscript revisions were performed independently by the authors. The conceptual framework of ECIA and all scientific insights are original contributions of the authors.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Jaeyeon Kim conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Daihun Kang conceived and designed the experiments, performed the experiments, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The simulation code and generated datasets are available at GitHub and Zenodo:

- GitHub: https://github.com/DrKDH/PeerjCS.

- DrKDH. (2025). DrKDH/PeerjCS: ECIA Non-Stationary Bandit Experiments (PeerJ Computer Science) (Version_v1). Zenodo. https://doi.org/10.5281/zenodo.17717944.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.3688#supplemental-information.

REFERENCES

Abdullahi M, Alhussian H, Aziz N, Abdulkadir SJ, Baashar Y, Alashhab AA, Afrin A. 2025. A systematic literature review of concept drift mitigation in time-series applications. In: *IEEE Access*.

Amsel A. 1992. *Frustration theory: an analysis of dispositional learning and memory*. Cambridge: Cambridge University Press.

Aston-Jones G, Cohen JD. 2005. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience* **28(1)**:403–450 DOI 10.1146/annurev.neuro.28.061604.135709.

Bach DR, Dayan P. 2017. Algorithms for survival: a comparative perspective on emotions. *Nature Reviews Neuroscience* **18(5)**:311–319 DOI 10.1038/nrn.2017.35.

Bai Y, Gao Y, Wan R, Zhang S, Song R. 2025. A review of reinforcement learning in financial applications. *Annual Review of Statistics and its Application* **12(1)**:209–232 DOI 10.1146/annurev-statistics-112723-034423.

Berridge KC, Kringelbach ML. 2015. Pleasure systems in the brain. *Neuron* **86(3)**:646–664 DOI 10.1016/j.neuron.2015.02.018.

Botvinick MM, Cohen JD, Carter CS. 2004. Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences* **8(12)**:539–546 DOI 10.1016/j.tics.2004.10.003.

Bouret S, Sara SJ. 2004. Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. *European Journal of Neuroscience* **20(3)**:791–802 DOI 10.1111/j.1460-9568.2004.03526.x.

Calvo RA, D'Mello S. 2010. Affect detection: an interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing* **1(1)**:18–37 DOI 10.1109/t-affc.2010.1.

Carver CS, Scheier MF. 1990. Origins and functions of positive and negative affect: a control-process view. *Psychological Review* **97(1)**:19–35 DOI 10.1037//0033-295x.97.1.19.

Damasio AR. 2001. Emotion and the human brain. *Annals of the New York Academy of Sciences* **935(1)**:101–106 DOI 10.1111/j.1749-6632.2001.tb03475.x.

Doya K. 2008. Modulators of decision making. *Nature Neuroscience* **11(4)**:410–416 DOI 10.1038/nn2077.

Duan Y, Schulman J, Chen X, Bartlett PL, Sutskever I, Abbeel P. 2016. Rl^ 2: fast reinforcement learning via slow reinforcement learning. ArXiv DOI 10.48550/arXiv.1611.02779.

Duncan K, Ketz N, Inati SJ, Davachi L. 2012. Evidence for area CA1 as a match/mismatch detector: a high-resolution fMRI study of the human hippocampus. *Hippocampus* **22(3)**:389–398 DOI 10.1002/hipo.20933.

Ekman P. 1972. *Emotion in the human face: guidelines for research and an integration of findings*. Oxford, United Kingdom: Pergamon Press.

Eldar E, Rutledge RB, Dolan RJ, Niv Y. 2016. Mood as representation of momentum. *Trends in Cognitive Sciences* **20(1)**:15–24 DOI 10.1016/j.tics.2015.07.010.

Eshel N, Roiser JP. 2010. Reward and punishment processing in depression. *Biological Psychiatry* **68(2)**:118–124 DOI 10.1016/j.biopsych.2010.01.027.

Finn C, Rajeswaran A, Kakade S, Levine S. 2019. Online meta-learning. In: *International Conference on Machine Learning: PMLR*, 1920–1930.

Garivier A, Moulines E. 2008. On upper-confidence bound policies for non-stationary bandit problems. ArXiv DOI 10.48550/arXiv.0805.3415.

Glimcher PW. 2011. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* **108(supplement_3)**:15647–15654 DOI 10.1073/pnas.1014269108.

Gottlieb J, Oudeyer P-Y. 2018. Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience* **19(12)**:758–770 DOI 10.1038/s41583-018-0078-0.

Govea J, Navarro AM, Sánchez-Viteri S, Villegas-Ch W. 2024. Implementation of deep reinforcement learning models for emotion detection and personalization of learning in hybrid educational environments. *Frontiers in Artificial Intelligence* **7**:1458230 DOI 10.3389/frai.2024.1458230.

Gray JR, Braver TS, Raichle ME. 2002. Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America* **99(6)**:4115–4120 DOI 10.1073/pnas.062381899.

Guex R, Méndez-Bértolo C, Moratti S, Strange BA, Spinelli L, Murray RJ, Sander D, Seeck M, Vuilleumier P, Domínguez-Borràs J. 2020. Temporal dynamics of amygdala response to emotion- and action-relevance. *Scientific Reports* **10(1)**:11138 DOI 10.1038/s41598-020-67862-1.

Hartland C, Gelly S, Baskiotis N, Teytaud O, Sebag M. 2006. Multi-armed bandit, dynamic environments and meta-bandits. HAL Id: hal-00113668. *Available at https://hal.science/hal-00113668v1*.

Hong A, Lunscher N, Hu T, Tsuboi Y, Zhang X, dos Reis Alves SF, Nejat G, Benhabib B. 2020. A multimodal emotional human-robot interaction architecture for social robots engaged in bidirectional communication. *IEEE Transactions on Cybernetics* **51(12)**:5954–5968 DOI 10.1109/tcyb.2020.2974688.

Howe MW, Dombeck DA. 2016. Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* **535(7613)**:505–510 DOI 10.1038/nature18942.

Hu X, Leung HF. 2023. A tighter problem-dependent regret bound for risk-sensitive reinforcement learning. In: *International Conference on Artificial Intelligence and Statistics: PMLR*, 5411–5437.

Huys QJ, Maia TV, Frank MJ. 2016. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience* **19(3)**:404–413 DOI 10.1038/nn.4238.

Izuma K, Saito DN, Sadato N. 2008. Processing of social and monetary rewards in the human striatum. *Neuron* **58(2)**:284–294 DOI 10.1016/j.neuron.2008.03.020.

Joffily M, Coricelli G. 2013. Emotional valence and the free-energy principle. *PLOS Computational Biology* **9(6)**:e1003094 DOI 10.1371/journal.pcbi.1003094.

Keiflin R, Janak PH. 2015. Dopamine prediction errors in reward learning and addiction: from theory to neural circuitry. *Neuron* **88(2)**:247–263 DOI 10.1016/j.neuron.2015.08.037.

Kuppens P, Oravecz Z, Tuerlinckx F. 2010. Feelings change: accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology* **99(6)**:1042–1060 DOI 10.1037/a0020962.

Lecarpentier E. 2020. *Reinforcement learning in non-stationary environments*. Lyon: ENS Lyon.

Lecarpentier E, Rachelson E. 2019. Non-stationary Markov decision processes, a worst-case approach using model-based reinforcement learning. In: *Advances in Neural Information Processing Systems 32*.

LeDoux JE. 1996. *The emotional brain: the mysterious underpinnings of emotional life*. USA: Simon and Schuster.

Lian H, Lu C, Li S, Zhao Y, Tang C, Zong Y. 2023a. A survey of deep learning-based multimodal emotion recognition: speech, text, and face. *Entropy* **25(10)**:1440 DOI 10.3390/e25101440.

Lian Z, Sun H, Sun L, Gu H, Wen Z, Zhang S, Chen S, Xu M, Xu K, Chen K. 2023b. Explainable multimodal emotion recognition. ArXiv DOI 10.48550/arXiv.2306.15401.

Lisman JE, Grace AA. 2005. The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron* **46(5)**:703–713 DOI 10.1016/j.neuron.2005.05.002.

Liu F, Lee J, Shroff N. 2018. A change-detection based framework for piecewise-stationary multi-armed bandit problem. In: *Proceedings of the AAAI Conference on Artificial Intelligence*.

Marsella SC, Gratch J. 2009. EMA: a process model of appraisal dynamics. *Cognitive Systems Research* **10(1)**:70–90 DOI 10.1016/j.cogsys.2008.03.005.

McGaugh JL. 2004. The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience* **27(1)**:1–28 DOI 10.1146/annurev.neuro.27.070203.144157.

Mellor J, Shapiro J. 2013. Thompson sampling in switching environments with bayesian online change point detection. ArXiv DOI 10.48550/arXiv.1302.3721.

Miller EK, Cohen JD. 2001. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* **24(1)**:167–202 DOI 10.1146/annurev.neuro.24.1.167.

Mittal T, Guhan P, Bhattacharya U, Chandra R, Bera A, Manocha D. 2020. Emoticon: context-aware multimodal emotion recognition using frege's principle. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14234–14243.

Moerland TM, Broekens J, Jonker CM. 2018. Emotion in reinforcement learning agents and robots: a survey. *Machine Learning* **107(2)**:443–480 DOI 10.1007/s10994-017-5666-0.

Nagabandi A, Clavera I, Liu S, Fearing RS, Abbeel P, Levine S, Finn C. 2018. Learning to adapt in dynamic, real-world environments through meta-reinforcement learning. ArXiv DOI 10.48550/arXiv.1803.11347.

Norman KA, O'Reilly RC. 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychological Review* **110(4)**:611–646 DOI 10.1037/0033-295x.110.4.611.

Oudeyer P-Y, Kaplan F. 2007. What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics* **1**:108 DOI 10.3389/neuro.12.006.2007.

O'Doherty JP, Cockburn J, Pauli WM. 2017. Learning, reward, and decision making. *Annual Review of Psychology* **68(1)**:73–100 DOI 10.1146/annurev-psych-010416-044216.

Panksepp J. 2004. *Affective neuroscience: the foundations of human and animal emotions*. Oxford: Oxford university press.

Pearce JM, Hall G. 1980. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review* **87(6)**:532–552 DOI 10.1037//0033-295x.87.6.532.

Pessoa L. 2018a. Emotion and the interactive brain: insights from comparative neuroanatomy and complex systems. *Emotion Review* **10(3)**:204–216 DOI 10.1177/1754073918765675.

Pessoa L. 2018b. Understanding emotion with brain networks. *Current Opinion in Behavioral Sciences* **19**:19–25 DOI 10.1016/j.cobeha.2017.09.005.

Plutchik R. 1980. *Emotion. A psychoevolutionary synthesis*. New York: Harper & Row.

Qi H, Guo F, Zhu L. 2025. Thompson sampling for non-stationary bandit problems. *Entropy* **27(1)**:51 DOI 10.3390/e27010051.

Raj V, Kalyani S. 2017. Taming non-stationary bandits: a Bayesian approach. ArXiv DOI 10.48550/arXiv.1707.09727.

Rakelly K, Zhou A, Finn C, Levine S, Quillen D. 2019. Efficient off-policy meta-reinforcement learning via probabilistic context variables. In: *International Conference on Machine Learning: PMLR*, 5331–5340.

Rolls ET. 2014. Emotion and decision-making explained: a précis. *Cortex* **59(35)**:185–193 DOI 10.1016/j.cortex.2014.01.020.

Russell JA, Barrett LF. 1999. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of Personality and Social Psychology* **76(5)**:805–819 DOI 10.1037/0022-3514.76.5.805.

Santoro A, Bartunov S, Botvinick M, Wierstra D, Lillicrap T. 2016. Meta-learning with memory-augmented neural networks. In: *International Conference on Machine Learning: PMLR*, 1842–1850.

Schultz W. 1998. Predictive reward signal of dopamine neurons. *Journal of Neurophysiology* **80(1)**:1–27 DOI 10.1152/jn.1998.80.1.1.

Schultz W. 2015. Neuronal reward and decision signals: from theories to data. *Physiological Reviews* **95(3)**:853–951 DOI 10.1152/physrev.00023.2014.

Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. *Science* **275(5306)**:1593–1599 DOI 10.1126/science.275.5306.1593.

Sharpe MJ, Chang CY, Liu MA, Batchelor HM, Mueller LE, Jones JL, Niv Y, Schoenbaum G. 2017. Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nature Neuroscience* **20(5)**:735–742 DOI 10.1038/nn.4538.

Takahashi H, Yahata N, Koeda M, Matsuda T, Asai K, Okubo Y. 2004. Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. *Neuroimage* **23(3)**:967–974 DOI 10.1016/j.neuroimage.2004.07.054.

Trovo F, Paladino S, Restelli M, Gatti N. 2020. Sliding-window thompson sampling for non-stationary settings. *Journal of Artificial Intelligence Research* **68**:311–364 DOI 10.1613/jair.1.11407.

Wang JX, Kurth-Nelson Z, Tirumala D, Soyer H, Leibo JZ, Munos R, Blundell C, Kumaran D, Botvinick M. 2016. Learning to reinforcement learn. ArXiv DOI 10.48550/arXiv.1611.05763.

Zhou Z. 2021. Emotional thinking as the foundation of consciousness in artificial intelligence. *Cultures of Science* **4(3)**:112–123 DOI 10.1177/20966083211052651.