Persona, Ego, Shadow, and Self: A Map of the Soul Framework for Proto-Emotional Homeostasis in AI

Napassorn Litchiowong
School of Computing
National University of Singapore
pleng@u.nus.edu

Abstract

I present a compact, testable architecture that endows learning agents with continuous proto-emotional dynamics and interpretable modulators (Persona, Ego, Shadow, Self). The design grounds these modulators in a computational interpretation of Jung's Map of the Soul, mapping each archetype to a differentiable control that modulates policy selection via a bounded, low-dimensional affect vector. I describe concrete modular implementations, a staged experimental program (toy domains $\rightarrow$ multi-agent/social tasks $\rightarrow$ nonstationary transfer), baselines, ablations, and reproducible evaluation metrics.

Introduction

Contemporary reinforcement-learning agents typically optimize external task reward without an explicit internal affective process. This omission reduces transparency and can limit adaptivity under nonstationarity. I investigate whether a small, homeostatically governed internal state, coupled to policy selection via a compact set of interpretable modulators, can improve robustness, behavioural diversity, and interpretability. The proposed formulation maps four archetypal roles (Persona, Ego, Shadow, Self) to concrete, differentiable modules that modulate policy selection via a low-dimensional affect vector. The motivation draws on affective computing and appraisal models (Picard 1997; Marsella and Gratch 2009), formal perspectives on prediction error and internal needs (Friston 2010; Keramati and Gutkin 2014), and intrinsic-motivation and meta-reinforcement learning templates for exploration and arbitration (Oudeyer et al. 2007; Wang et al. 2016; Vezhnevets et al. 2017). The archetypal vocabulary provides an interpretable inductive bias that admits clear ablation and hypothesis testing.

Background

Affective computing has produced robust sensing and appraisal models, but relatively little work endows agents with endogenous affective dynamics that bidirectionally influence action selection (Picard 1997; Marsella and Gratch 2009). Predictive-coding and free-energy perspectives treat surprise and prediction error as drivers of inference and action (Friston 2010). Homeostatic reinforcement formalizes trade-offs between internal needs and external reward (Keramati and Gutkin 2014). Intrinsic motivation research demonstrates how curiosity-like signals bootstrap exploration (Oudeyer et al. 2007). Meta-reinforcement learning and hierarchical control provide templates for learned arbitration and slow-timescale integration (Wang et al. 2016; Vezhnevets et al. 2017).

Prior work

I previously developed a GPU-efficient multimodal pipeline for early-stage depression screening, including a compact vision encoder, a speech emotion model, and an LSTM-based text sentiment module; I also addressed latency, throughput, and deployment considerations. This systems experience supports feasible implementation of the proposed prototypes while managing compute and integration risks.

Approach

System architecture

The agent architecture comprises three integrated modules: (1) a *Core Affective Dynamics* module that maintains a compact internal affect vector A_t; (2) an *Archetypal Modulation* layer that maps A_t to control signals associated with Persona, Shadow, Ego, and Self; and (3) a *Decision & Memory* module where a policy $\pi(a \mid s, A_t)$ produces actions and stores episodic traces tagged by archetypal activations. The structure is compatible with standard model-free RL pipelines and with recurrent and meta-RL extensions for longer-horizon control.

Core affective dynamics

I represent the agent's internal affective state as the four-dimensional vector

$$A_t = [v_t,\ a_t,\ \tau_t,\ \iota_t],$$

where v denotes valence, a arousal, τ cumulative tension (integrated prediction error or unmet needs), and ι a short-horizon integrator. The update rule enforces bounded homeostasis:

$$A_{t+1} \ = \ \alpha \odot A_t \ + \ \beta \odot g(r_t, \delta_t, \phi(s_t)),$$

where δ_t is prediction (e.g., TD) error, r_t is external reward, $\phi(s_t)$ are perceptual salience features, and $g(\cdot)$ is a

small MLP that maps reward, prediction error, and novelty into per-dimension adjustments. The decay/gain vectors $\alpha, \beta \in (0,1)^4$ implement interpretable timescales and guarantee boundedness; they may be hand-initialized for interpretability and optionally learned.

Archetypal modulation

Each archetype is a differentiable module mapping A_t to policy modulation parameters.

Persona. Persona implements reward and social sensitivity. The module computes a gain

$$G_t^{\text{persona}} \;=\; \sigma\big(w_p^\top A_t + b_p\big),$$

which rescales value estimates or policy logits to bias toward cooperative, reward-seeking actions.

Shadow. Shadow implements an exploration and risk bias. The module produces an exploration temperature

$$T_t^{\text{shadow}} \;=\; \exp\big(w_s^\top A_t + b_s\big),$$

used to scale stochastic action sampling or to weight intrinsic curiosity bonuses.

Ego. Ego is a meta-arbitrator that outputs normalized weights ω_t over archetypal contributions. Two implementations will be compared: (a) a small feedforward gating MLP trained jointly with the policy, and (b) an LSTM-based recurrent meta-RL controller that learns arbitration dynamics across task distributions.

Self. Self operates at a slower timescale and implements an auxiliary meta-objective L_{self}, for example minimizing long-run variance among archetype activations. Practically, Self applies a regularization loss during meta-update intervals to avoid collapse to a single dominant archetype.

All archetypal mappings are differentiable; they may be hand-initialized for interpretability and later fine-tuned end-to-end.

Decision, learning, and perception

Policies are conditioned on the concatenated input $[s_t; A_t]$ and instantiated with Proximal Policy Optimization (PPO) for episodic tasks and Soft Actor-Critic (SAC) for continuous control. Recurrent policies (LSTM/GRU) support temporal integration where necessary. Meta-learning configurations for Ego include recurrent controllers trained across task distributions. Episodic memory stores archetype-tagged trajectories; a prioritized replay keyed by archetype activation supports emotionally tinted recall. Perceptual salience $\phi(s)$ is produced by lightweight adapters on pretrained encoders (for example CLIP for vision and wav2vec 2.0 for speech) to reduce sample complexity.

Training regimen and ablation

Training proceeds in staged experiments: (a) validate Core Affective Dynamics in single-agent toy environments to ensure numerical boundedness and recovery properties; (b) add archetypal modulators and evaluate behavioural regimes in social-dilemma and multi-agent environments; (c) enable meta-RL Ego and Self regularizers and measure long-horizon coherence and transfer. Ablations will remove or freeze each archetype in isolation (Persona-off, Shadow-off, Ego-fixed, Self-off) and compare gating MLP versus recurrent meta-RL implementations and alternative Self meta-objectives.

Evaluation

Evaluation combines quantitative RL benchmarks, multi-agent social-dilemma tasks, controlled ablations, and conservative human-in-the-loop pilots when appropriate. Environments include single-agent bandits and gridworlds for homeostasis validation, social/multi-agent tasks (iterated Prisoner's Dilemma, stag-hunt, Multi-Agent Particle Environment variants) to stress cooperation versus exploration trade-offs, and nonstationary transfer tasks where reward functions or partner behaviours change mid-training. Baselines comprise: (a) standard RL conditioned only on external state (no affect); (b) a dimensional-affect agent with valence–arousal only; and (c) an intrinsic-motivation curiosity agent. Metrics include task performance (cumulative return, convergence rate), affective dynamics (boundedness, recovery time, autocorrelation, trajectory entropy), behavioural diversity (state-action visitation entropy, Jensen–Shannon divergence), identity emergence (policy-summary embeddings clustered with k-means and validated via silhouette and Davies–Bouldin indices), and robustness under distribution shift and noisy observations. Statistical protocol uses multiple random seeds (8–12), non-parametric tests (Mann–Whitney U), effect sizes, and bootstrap validation for clustering. Any human-facing evaluation will be small, IRB-reviewed, and conducted only after algorithmic validation and safety checks.

Discussion

The proposed design trades additional inductive structure and hyperparameters for interpretability and potential robustness gains. I expect Shadow-driven exploration to increase policy diversity but to reduce sample efficiency in some settings; Persona-driven reward sensitivity is likely to aid social coordination while potentially reducing exploratory behaviour. Risks for user-facing systems include manipulative affect modulation or misinterpretation of internal states; mitigations include conservative pilot design, institutional oversight prior to deployment in sensitive domains, explicit documentation of failure modes, and publication of interpretability diagnostics. The project is staged so that gridworld experiments provide fast, informative go/no-go signals prior to larger-scale experiments.

Conclusion

This proposal describes a principled, compact architecture for proto-emotional homeostasis that couples a bounded, low-dimensional affect vector to four differentiable archetypal modulators. The staged experimental program, explicit ablations, and reproducible metrics will determine whether structured internal affect functions as a useful inductive bias for robust, interpretable adaptive behaviour. Emphasis is placed on reproducibility, transparency, and ethical safeguards for any human-facing applications.

References

Baevski, A.; Zhou, Y.; Mohamed, A.; and Auli, M. 2020. wav2vec 2.0: A framework for self-supervised learning of speech representations. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*.

Cox, M. T. 2007. Perpetual self-aware cognitive agents. *AI Magazine* 28(1):32–45. doi:10.1609/aimag.v28i1.2027.

Fitzpatrick, K. K.; Darcy, A.; and Vierhile, M. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health* 4(2):e19. doi:10.2196/mental.7785.

Friston, K. 2010. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience* 11:127–138. doi:10.1038/nrn2787.

Graves, A.; Wayne, G.; Reynolds, M.; Harley, T.; Danihelka, I.; Grabska-Barwińska, A.; Colmenarejo, S. G.; Grefenstette, E.; Ramalho, T.; Agapiou, J.; et al. 2016. Hybrid computing using a neural network with dynamic external memory. *Nature* 538(7626):471–476. doi:10.1038/nature20101.

Haarnoja, T.; Zhou, A.; Abbeel, P.; and Levine, S. 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)*, Proceedings of Machine Learning Research 80:1856–1865.

Jung, C. G. 1969. *The Archetypes and the Collective Unconscious*. Princeton, NJ: Princeton University Press.

Keramati, M.; and Gutkin, B. 2014. Homeostatic reinforcement learning for integrating reward collection and physiological stability. *eLife* 3:e04811. doi:10.7554/eLife.04811.

Marsella, S.; and Gratch, J. 2009. EMA: A process model of appraisal dynamics. *Cognitive Systems Research* 10:70–90. doi:10.1016/j.cogsys.2008.03.005.

Oudeyer, P.-Y.; Kaplan, F.; and Hafner, V. V. 2007. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation* 11(2):265–286. doi:10.1109/TEVC.2006.890271.

Picard, R. W. 1997. *Affective Computing*. Cambridge, MA: MIT Press.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning transferable visual models from natural language supervision. arXiv preprint. arXiv:2103.00020.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. arXiv preprint. arXiv:1707.06347.

Vezhnevets, A. S.; Osindero, S.; Schaul, T.; Heess, N.; Jaderberg, M.; Silver, D.; and Kavukcuoglu, K. 2017. FeUdal networks for hierarchical reinforcement learning. arXiv preprint. arXiv:1703.01161.

Wang, J. X.; Kurth-Nelson, Z.; Tirumala, D.; Soyer, H.; Leibo, J. Z.; Munos, R.; Blundell, C.; Kumaran, D.; and Botvinick, M. 2016. Learning to reinforcement learn. arXiv preprint. arXiv:1611.05763.