

Affective Agent Prototype — Whitepaper

Author: Mark Purvis

This whitepaper introduces the first working prototype linking interoception, prediction error, and a global workspace into a computational model of emotion and consciousness.

Model Summary

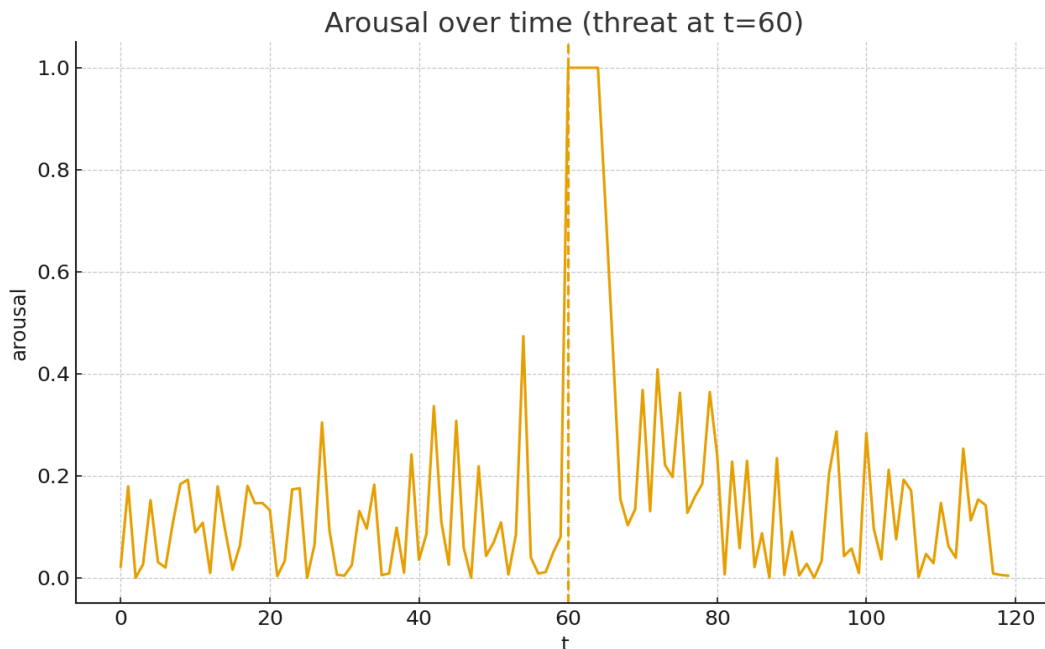
- Interoception: HR, HRV, EDA, Respiration, Pupil (synthetic)
- Prediction error: difference between predicted and sensed intero signals (online learning)
- Global workspace: salient channels (high surprise) are broadcast
- Report: continuous valence/arousal with human-readable summary

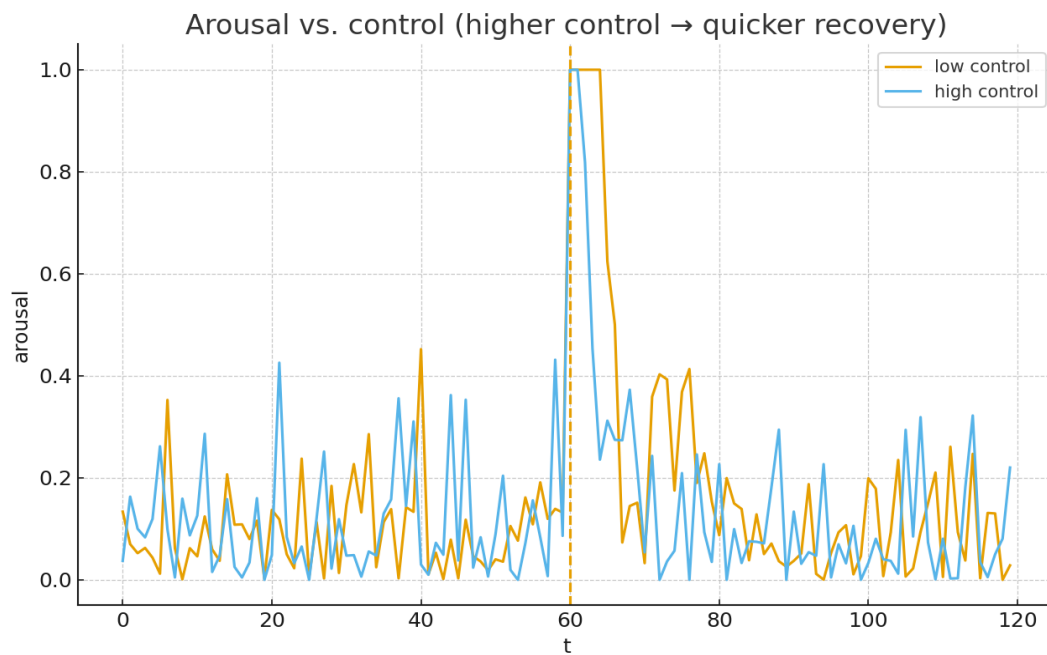
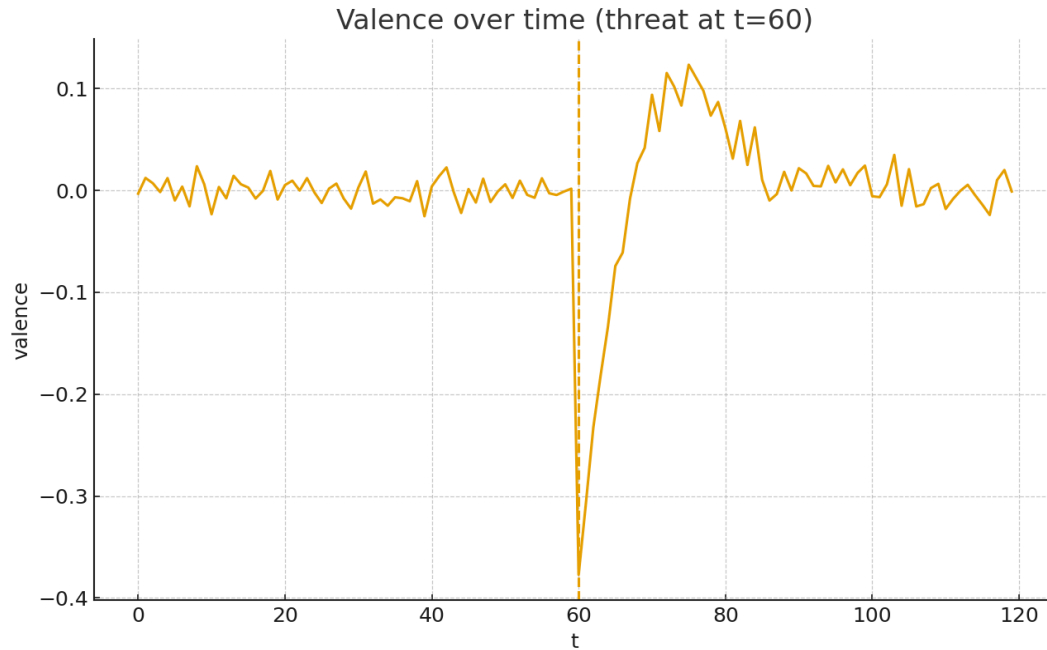
Equations (informal)

- Prediction update: $\text{pred} \leftarrow (1-\alpha) * \text{pred} + \alpha * \text{sensed}$
- Salience: $\text{sal} = |(\text{sensed} - \text{pred}) / \text{scale}|$
- Arousal: weighted positive prediction error across channels
- Valence: goal-congruence proxy from HR(-), HRV(+), EDA(-), Resp(-), Pupil(+)

Results (synthetic)

- Threat at $t=60$ produces a sharp **arousal increase** and **valence drop**, then recovery.
- Higher **control** shortens recovery time (resilience).





Next Steps

- Add voice/face/context features for fine-grained emotions.
- Personalization adapters per individual.
- Validate on real datasets (WESAD/DEAP).