

Saliency as an Internal Currency for Resource-Bounded Agents

Justin Garringer

January 27, 2026

Abstract

We present a saliency formulation that converts heterogeneous signals (surprise, retention risk, goal coupling, continuity, and cost) into an actionable control signal for resource-bounded agents. The saliency score prices attention and compute, induces probabilistic selection and threshold gates, and adapts from telemetry. We also show how saliency can tune the weights of a downstream action-value objective, closing a feedback loop between perception, cognition, and operations.

1 Introduction

Modern assistants and agents operate under hard budgets: limited context, limited tool calls, and limited planning depth. A common failure mode is misallocation—spending compute where it feels interesting rather than where it cashes out (e.g., chasing novelty, forgetting constraints, or thrashing between plans). We propose treating *saliency* as an explicit allocation policy: an internal currency that buys attention and compute.

Our contribution is a model-facing specification with three parts:

1. a saliency score that can be learned from telemetry,
2. a policy that maps scores to selection, gating, and budget allocation,
3. a controller that uses saliency to tune the weights of a downstream decision objective.

This paper is best read as a *specification and research agenda*: the controller is fully specified and falsifiable, while empirical validation is framed as a minimal fixed-budget evaluation (Section 12).

2 Problem setup

Let the agent operate at discrete times $t \in \{1, 2, \dots\}$. At time t the agent has context x_t (messages, observations, active tasks, open files, tool outputs, etc.) and a set of *candidates* $\mathcal{I}_t$ (memories to retrieve, hypotheses to test, plan nodes to expand, tools to call, constraints to re-check).

Candidate heterogeneity. Candidates may have different types (e.g., a memory retrieval vs. a tool call vs. a plan expansion), so the meaning of “distance” d_i and “effort” τ_i can be type-specific. A practical implementation therefore uses per-type normalizations (or learned calibrations) so that S_i is comparable across types.

The agent must allocate scarce resources across candidates:

- **Attention:** which candidates enter the working set,

- **Memory:** which candidates get stored or refreshed,
- **Operations budget:** how much compute/tool budget to spend per candidate.

3 Saliency score

For candidate $i \in \mathcal{I}_t$ in context x_t :

$$S_i(t | x_t) = \underbrace{(w_A \Delta A_i(x_t) + w_R R_i(x_t) + w_M M_i(x_t))}_{\text{value terms}} \cdot \underbrace{C_i(x_t)}_{\text{continuity / coherence}} \cdot \underbrace{e^{-\lambda \Delta t_i}}_{\text{time decay}} \cdot \underbrace{(1 - k \phi(x_t))}_{\text{fatigue / distraction penalty}} \cdot \underbrace{\frac{1}{\sqrt{d_i(x_t) \tau_i(x_t) + \epsilon}}}_{\text{normalization (distance + effort)}}. \quad (1)$$

Here $w_A, w_R, w_M \geq 0$ are weights, $\lambda \geq 0$ is a decay rate, $k \geq 0$ couples fatigue to suppression, and $\epsilon > 0$ prevents division by zero. The multiplicative structure is intentional: terms like continuity $C_i(x_t)$ can act as *veto gates* (low coherence should downweight saliency even if the value terms are high), while the distance/effort normalization penalizes expensive-to-process candidates.

We treat fatigue $\phi(x_t)$ as a *context-level* scalar that captures global degradation (latency, error loops). If desired, it can be extended to a per-candidate $\phi_i(x_t)$ (e.g., “this class of tools is unreliable right now”), but we keep the base controller context-level for simplicity.

3.1 Operational semantics (telemetry to features)

Each term is designed to admit operational proxies so the system can be hand-tuned initially and later learned.

Surprise $\Delta A_i(x)$: expectation violation. Proxies: prediction error (token NLL spikes), KL divergence between predicted and observed outcomes, anomaly scores, “user correction” events.

Retention need $R_i(x)$: future reuse likelihood and cost of forgetting. Proxies: repeated references, dependency count in a project graph, retrieval frequency, outcome-conditioned usefulness.

Momentum $M_i(x)$: coupling to the current objective. Proxies: alignment with active tags/goals, proximity to the plan frontier, reduces blockers, increases optionality.

Continuity $C_i(x)$: coherence pressure. Proxies: consistency with constraints, contradiction counts, narrative stability, reduced policy thrash.

Age Δt_i : time since last reinforcement (read/used/validated).

Fatigue $\phi(x)$: degradation of the system. Proxies: rising error rates, oscillations, long latencies, user confusion, low-confidence loops.

Distance $d_i(x)$: conceptual distance from the current context. Proxies: graph hops, embedding distance, dependency distance.

Effort $\tau_i(x)$: expected compute/tool cost to act on i .

4 Policy: scores to behavior

4.1 Selection / attention

We convert scores to a distribution over candidates:

$$P(\text{select } i \mid x_t) = \text{softmax}\left(\frac{S_i(t \mid x_t)}{T}\right), \quad (2)$$

where temperature $T > 0$ controls exploitation vs exploration.

4.2 Gating / thresholds

We use three thresholds:

$$\text{working set:} \quad S_i \geq \theta_{\text{work}}, \quad (3)$$

$$\text{long-term storage:} \quad S_i \geq \theta_{\text{store}}, \quad (4)$$

$$\text{ignore/defer:} \quad S_i < \theta_{\text{ignore}}. \quad (5)$$

4.3 Budget allocation

Let B denote the total compute/tool budget for the step. Allocate proportionally:

$$\text{budget}_i = B \cdot \frac{\max(S_i, 0)}{\sum_{j \in \mathcal{I}_t} \max(S_j, 0)}. \quad (6)$$

This makes “salience” literal currency for compute.

5 Algorithmic skeleton

Algorithm 1 One step of salience-based control

Require: Context x_t , candidates $\mathcal{I}_t$, total budget B

- 1: Estimate features $(\Delta A_i, R_i, M_i, C_i, \Delta t_i, \phi, d_i, \tau_i)$ from telemetry
 - 2: Compute salience $S_i(t \mid x_t)$ for each $i \in \mathcal{I}_t$ (Section 3)
 - 3: Sample/select attention set $\mathcal{W}_t$ from $\text{softmax}(S/T)$
 - 4: **for all** $i \in \mathcal{W}_t$ **do**
 - 5: Allocate per-candidate spend $\text{budget}_i \leftarrow B \max(S_i, 0) / \sum_j \max(S_j, 0)$
 - 6: **if** $S_i \geq \theta_{\text{store}}$ **then**
 - 7: Commit/refresh i in long-term memory
 - 8: **else if** $S_i \geq \theta_{\text{work}}$ **then**
 - 9: Promote i to the working set
 - 10: **end if**
 - 11: **end for**
 - 12: Tune $(\lambda_I, \lambda_O, \lambda_R, \lambda_T)$ using salience features
 - 13: Execute actions / tool calls within allocated budgets; log outcomes
-

6 Saliency-tuned action evaluation

Assume a downstream action-evaluation objective (your “TLP control law”):

$$J(a \mid b, \theta) = \Delta U(a) + \lambda_I \text{VoI}(a) + \lambda_O O(a) - \mathcal{C}(a) - \lambda_R R(a) - \lambda_T T(a), \quad (7)$$

where b is belief/state and θ are parameters, and $\mathcal{C}(a)$ denotes action cost (to avoid a notational collision with continuity C_i in Section 3).

We propose using saliency-derived telemetry to tune the weights λ contextually:

- High surprise $\Delta A \Rightarrow$ increase λ_I (prefer information-gathering).
- High retention risk $R \Rightarrow$ increase λ_R (protect constraints/continuity).
- High fatigue ϕ or time pressure $\Rightarrow$ increase λ_T and reduce planning depth.
- High momentum M on a productive thread $\Rightarrow$ increase λ_O (keep optionality/momentum).

7 A geometric analogy: saliency as curvature

The saliency score in Section 3 can be pushed further by treating it as a *resource-geometry* over the agent’s state space: regions of high “informational mass” curve the landscape so that attention and compute preferentially flow along certain directions. This section is an analogy (not a claim of physical equivalence), but it is useful because it suggests invariances, stability diagnostics, and visualization primitives. This geometric interpretation is strictly optional—the controller is fully specified without it—but it provides a convenient vocabulary for stability analysis and hyperparameter tuning.

7.1 Dictionary: saliency terms to GR-flavored counterparts

We interpret the agent’s evolving context x_t as a point on a manifold $\mathcal{M}$, and each candidate i as a local “source” that deforms how costly it is to move (allocate compute) in different directions.

Saliency term	GR-style interpretation
Information density ρ_i	“mass-energy” density sourcing curvature
Continuity C_i	constraint-consistency pressure (suppresses nonphysical trajectories)
Decay $e^{-\lambda \Delta t_i}$	redshift / weakening of influence with “proper time”
Distance d_i	spatial separation on $\mathcal{M}$ (metric-induced)
Effort τ_i	local resistance / lapse (how fast you can advance per compute)
Fatigue ϕ	effective temperature / noise floor (smears sharp minima)

7.2 Information density as “mass”

Let $\rho_i(x_t)$ denote the information density contributed by candidate i at context x_t . A simple operational choice is to fuse surprise and future utility:

$$\rho_i(x_t) \triangleq w_A \Delta A_i(x_t) + w_R R_i(x_t) + w_M M_i(x_t). \quad (8)$$

Then your saliency score can be read as a *density-to-influence* transform:

$$S_i(t \mid x_t) \approx \rho_i(x_t) C_i(x_t) e^{-\lambda \Delta t_i} (1 - k\phi(x_t)) (d_i(x_t) \tau_i(x_t) + \epsilon)^{-1/2}. \quad (9)$$

7.3 A “salience potential” and a redshift view of decay

In weak-field GR, gravitational time dilation can be expressed via an effective potential. Mirroring that idea, define a salience potential Φ_i by absorbing decay and cost terms:

$$\Phi_i(t \mid x_t) \triangleq -\log\left(C_i(x_t) e^{-\lambda \Delta t_i} (1 - k\phi(x_t))\right) + \frac{1}{2} \log(d_i(x_t) \tau_i(x_t) + \epsilon). \quad (10)$$

Then S_i is proportional to $\rho_i e^{-\Phi_i}$: information density is attractive, while potential raises the local “height” (lower influence).

7.4 Metric view: distance and effort as geometry

Let $g_{ab}(x)$ be a positive-definite metric on $\mathcal{M}$ defining squared distance $d(x, x')^2 \approx \Delta x^\top g(x) \Delta x$. One can interpret effort $\tau_i(x)$ as a local scaling of the metric (hard regions are “stretched”):

$$g_{ab}^{\text{eff}}(x) \triangleq \tau(x) g_{ab}(x), \quad (11)$$

so that moving into computationally expensive regions increases effective distance.

7.5 Geodesic attention and budget flow

Selection and budgeting (Section 4) can be reinterpreted as approximate geodesic flow in the potential landscape: attention follows “downhill” directions of Φ while being sourced by ρ . A minimal continuous-time caricature is

$$\dot{x} \propto -\nabla_x \Phi(x) \quad \text{with noise scale set by } \phi(x), \quad (12)$$

which connects naturally to Langevin-like sampling and softmax selection.

7.6 What the analogy buys you (design implications)

- **Gauge/invariance thinking:** changing units of compute or rescaling telemetry should correspond to simple reparameterizations of ρ and Φ , not qualitatively different behavior.
- **Curvature diagnostics:** “salience explosion” corresponds to overly deep wells (high curvature), while “collapse” corresponds to a nearly flat landscape. Both can be detected by monitoring concentration of the softmax and the variance of S_i .
- **Energy barriers for anti-thrash:** oscillation suggests two competing minima separated by a low barrier; increasing continuity pressure C (or adding an explicit switching cost) raises the barrier and reduces flipping.

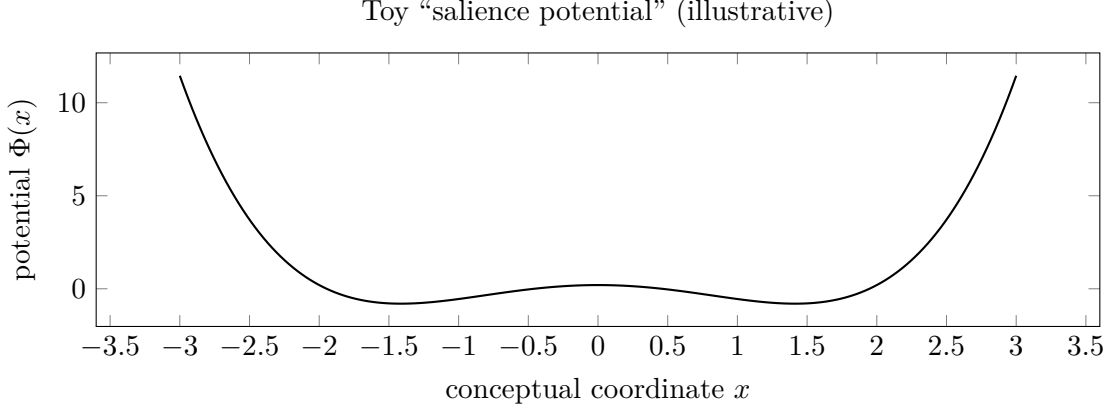


Figure 1: A simple double-well potential illustrating two competing threads; raising continuity can increase the barrier and reduce oscillation.

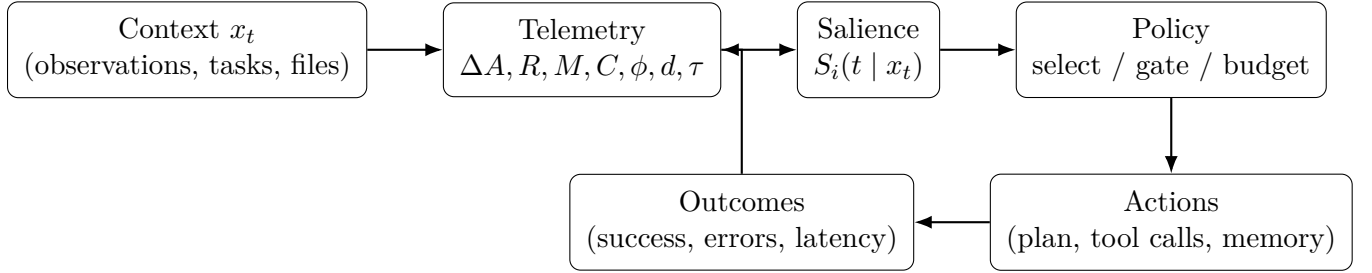


Figure 2: Salience as a closed-loop controller: telemetry feeds feature estimates, which determine salience, which allocates resources, producing outcomes that update telemetry.

7.7 Visualization: a toy salience well

8 Diagrams

8.1 Salience control loop

8.2 Budget market intuition

8.3 Thresholding as a gate

8.4 Example: scores induce budgets

8.5 Temperature controls exploration

8.6 Telemetry-to-feature estimation pipeline

9 Failure modes and telemetry signatures

We monitor the following signatures:

- **Salience collapse:** S_i nearly equal for all i , softmax becomes noise.
- **Salience explosion:** one thread monopolizes budget and misses reality shifts.

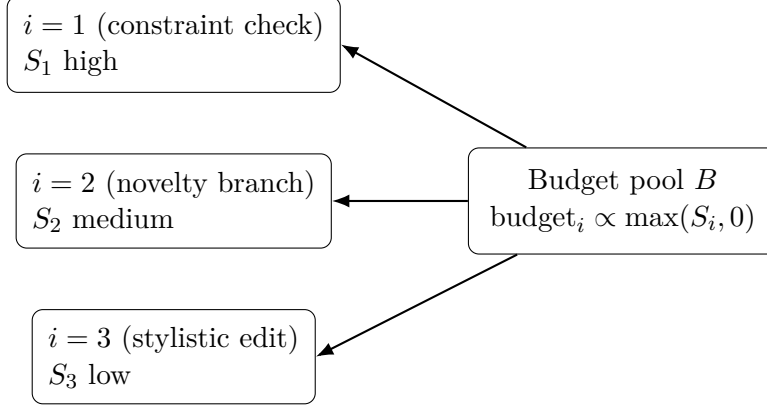


Figure 3: Saliency turns into spend: higher-score candidates receive more compute/tool budget.

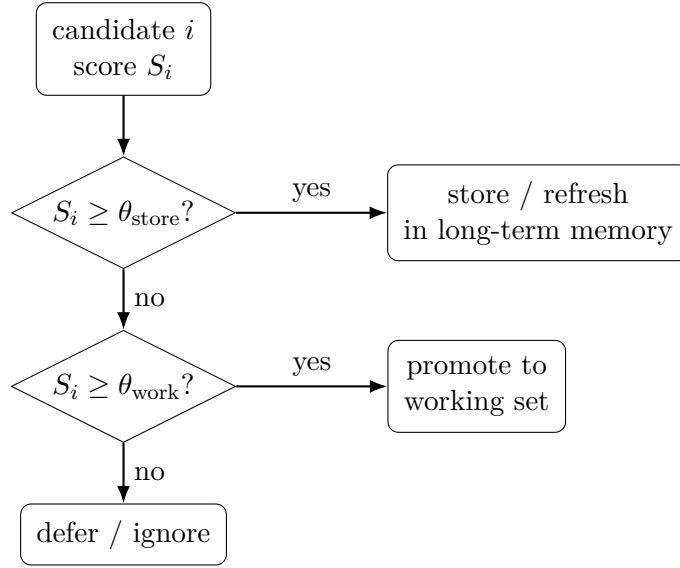


Figure 4: A simple gating view of saliency thresholds.

- **Oscillation:** alternating high saliency between incompatible candidates.
- **Fatigue blindness:** ϕ uncoupled from error/latency, pushing into nonsense.

10 Minimal implementation recipe

1. Define candidates $\mathcal{I}_t$ (memories, tools, hypotheses, plan nodes).
2. Instrument telemetry (prediction error, repeats, dependencies, time cost, latency, corrections).
3. Map telemetry $\rightarrow$ features $(\Delta A, R, M, C, \phi, d, \tau)$ (start hand-tuned, then learn).
4. Use S for retrieval, planning depth, tool budget, and λ -tuning in J .
5. Add anti-thrash: penalize rapid switching (via C and/or an explicit switching cost) and increase ϕ when oscillation is detected.

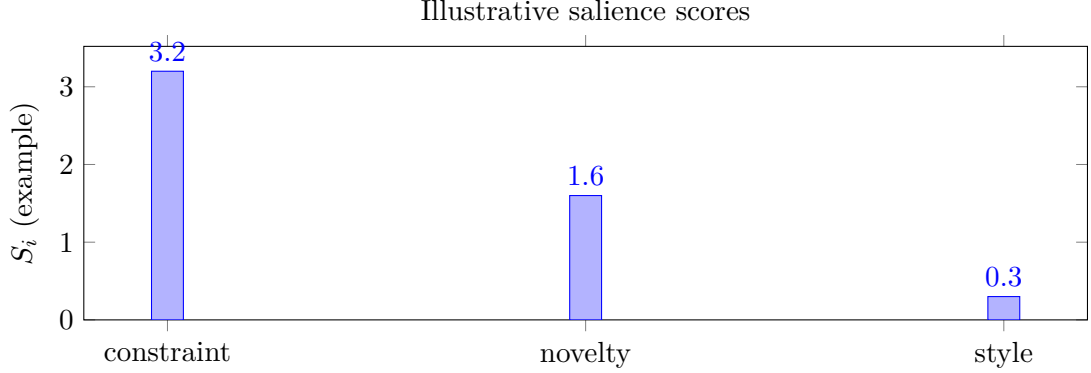


Figure 5: A toy snapshot of salience values; in practice S_i is derived from telemetry.

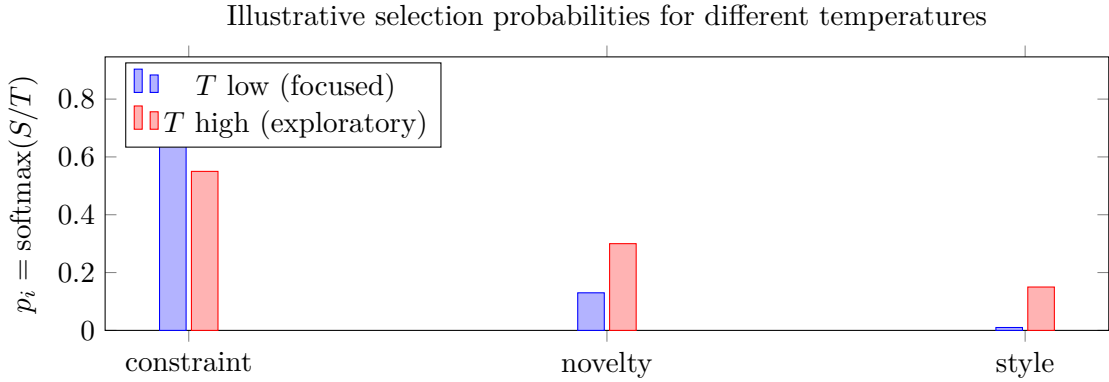


Figure 6: Lower temperature concentrates probability on the top-salience candidate; higher temperature spreads attention.

10.1 Learning the weights (a minimal recipe)

The parameters $(w_A, w_R, w_M, \lambda, k, \theta_{\text{work}}, \theta_{\text{store}})$ can be tuned by optimizing downstream outcomes under a compute constraint. A simple approach is to make selection differentiable via the softmax policy in Section 4 and minimize a task loss plus a compute regularizer:

$$\min_{\psi} \mathbb{E}[\mathcal{L}_{\text{task}}] + \beta \mathbb{E}[\text{compute}], \quad (13)$$

where ψ parameterizes the feature estimators and weights. For non-differentiable tool outcomes, one can use policy-gradient estimators (REINFORCE) with advantage baselines, treating S as the policy’s score function.

10.2 Robustness to measurement gaming

If telemetry can be manipulated (e.g., a candidate can induce spurious “surprise” ΔA), the controller should incorporate countermeasures: cap ΔA (winsorization), require multi-signal agreement (e.g., surprise *and* continuity), and add trust/credibility weights on telemetry sources. A practical pattern is to downweight signals from sources with recent high error rates (folding this into ϕ or into per-source reliability multipliers).

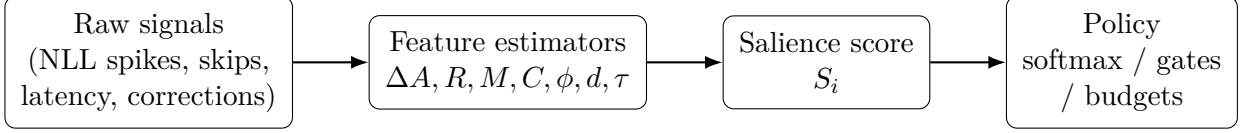


Figure 7: A practical decomposition: raw telemetry is mapped to interpretable features, which then drive salience-based control.

11 What can be proved vs. what must be tested

The GR-inspired discussion in Section 7 is an analogy. The actionable claims of this paper are about *a control policy over computation*, not physics. This section separates (i) statements that follow mathematically from the definitions from (ii) empirical hypotheses that can be falsified with telemetry.

11.1 Definitions and assumptions

Assume a finite candidate set $\mathcal{I}_t$, a temperature $T > 0$, and define softmax selection

$$p_i \triangleq P(\text{select } i \mid x_t) = \frac{\exp(S_i/T)}{\sum_{j \in \mathcal{I}_t} \exp(S_j/T)}. \quad (14)$$

Assume budgets are assigned as in Section 4 with $B > 0$:

$$b_i \triangleq \text{budget}_i = B \frac{\max(S_i, 0)}{\sum_{j \in \mathcal{I}_t} \max(S_j, 0)}. \quad (15)$$

11.2 Propositions (provable properties)

Proposition 1 (monotonicity of selection). For fixed T and fixed $\{S_j\}_{j \neq i}$, the selection probability p_i is strictly increasing in S_i .

Proposition 2 (translation invariance). For any constant $c \in \mathbb{R}$, replacing $S_i \mapsto S_i + c$ for all i leaves p_i unchanged.

Proposition 3 (temperature limits). If there is a unique maximizer $i^* = \arg \max_i S_i$, then as $T \rightarrow 0^+$, $p_{i^*} \rightarrow 1$ and $p_i \rightarrow 0$ for $i \neq i^*$. As $T \rightarrow \infty$, $p_i \rightarrow 1/|\mathcal{I}_t|$.

Proposition 4 (budget conservation and sparsity). If $\sum_j \max(S_j, 0) > 0$, then $\sum_i b_i = B$ and $b_i = 0$ whenever $S_i \leq 0$.

(Proof sketches: each follows directly from standard properties of the exponential function/softmax and the definition of b_i .)

11.3 Empirical hypotheses (falsifiable with telemetry)

The following are not theorems; they are testable claims about the *feature estimators* and how they couple to outcomes.

1. **Compute efficiency:** controlling tool/compute spend via b_i reduces cost per unit improvement in task metrics versus ungated baselines.

2. **Constraint retention:** increasing retention pressure (via R_i or θ_{store}) reduces constraint violations over long horizons.
3. **Anti-thrash:** adding switching costs or continuity pressure reduces oscillation frequency without increasing failure-to-notice rates.
4. **Fatigue coupling:** estimated $\phi(x)$ correlates with rising error/latency; using it to reduce depth improves reliability under load.

11.4 How to disprove the model

A concrete disproof pattern is: build estimators for $(\Delta A, R, M, C, \phi, d, \tau)$, deploy the controller, and show that across a representative suite it *increases* constraint violations or tool spend while *not* improving task success. In that case either (i) the feature estimators are mis-specified, (ii) the multiplicative form of S_i is wrong for the domain, or (iii) the downstream objective J is mismatched to the environment.

12 Quantifying salience from cultural time series

Your intuition—“tune the math based on data from movies, TV, music”—is exactly the right kind of falsification pressure: it forces S_i to predict real preference/engagement dynamics rather than just “sounding good.” The key move is to treat salience as a latent variable that must explain *attention allocation under budget*.

12.1 What data can (and cannot) identify

Cultural data typically provides sequences with feedback:

- user/item interaction logs (plays, watches, skips, rewinds),
- time-to-dropoff / completion,
- rewatch/relisten rates,
- co-consumption graphs (“users who watched X also watched Y”).

These signals can identify relative salience (what gets selected/retained), but not an absolute “true salience” without choosing an operational definition.

12.2 Operational definition: salience as predictive compression

One clean quantification is:

A candidate is salient insofar as allocating compute/attention to it improves prediction of the next relevant observation at lower cost.

Concretely, define an outcome y_{t+1} (next scene, next beat, next line, next user action) and a predictor $p_\theta(y_{t+1} | x_t, \mathcal{W}_t)$ where $\mathcal{W}_t$ is the selected working set. Then S is useful iff, under a fixed budget, selecting items with high S_i reduces held-out log loss:

$$\Delta \mathcal{L} \triangleq \mathbb{E}[-\log p_\theta(y_{t+1} | x_t, \mathcal{W}_t^{\text{sal}})] - \mathbb{E}[-\log p_\theta(y_{t+1} | x_t, \mathcal{W}_t^{\text{base}})]. \quad (16)$$

A falsifier is $\Delta \mathcal{L} \geq 0$ consistently (no predictive gain) once compute is properly normalized.

12.3 Mapping the salience terms to measurable proxies

For cultural sequences, the terms in Section 3 can be grounded as:

- ΔA : surprise in a sequence model (e.g., spikes in token/frame NLL) or in user action (unexpected skip/rewind).
- R : long-horizon reuse signals (rewatch/relisten, repeated motif callbacks, future co-consumption influence).
- M : goal coupling (task-conditioned objectives: “recommend for workout” vs “recommend for sleep”), proxied by context tags.
- C : narrative continuity (consistency constraints like character identity, plot coherence, musical key/tempo continuity); proxied by contradiction detectors or embedding consistency.
- Δt : time since last reactivation (when a motif/plot thread last mattered).
- ϕ : session fatigue (rising skip rate, shorter dwell time, increased switching).
- d : semantic distance in an embedding graph (scene-to-scene, track-to-track).
- τ : compute/time cost (model FLOPs, tool calls, or retrieval cost).

12.4 Ablations that meaningfully “disprove” components

To test whether the structure matters (vs any scoring rule), run ablations:

1. **Remove a term:** set $w_A = 0$ or drop the decay factor and measure change in $\Delta \mathcal{L}$ and downstream engagement metrics.
2. **Break the multiplicative form:** replace the product by a sum and compare calibration.
3. **Shuffle tests:** permute ΔA across candidates within each step; if performance is unchanged, surprise is not contributing.
4. **Counterfactual gating:** keep the same selected set size, but select uniformly at random; if salience does not beat random under equal compute, the controller is not justified.

12.5 A minimal fixed-budget evaluation (to escape “position paper” status)

Run a fixed-budget comparison where each method receives the same total budget B per step and is evaluated on (i) task success, (ii) constraint violations, and (iii) compute cost:

- **Salience budgeting:** allocate spend using b_i from Section 4.
- **Uniform:** equal allocation across candidates.
- **Novelty-only:** score using only ΔA (a “shiny object” baseline).
- **Ungated:** no thresholds; always process everything until budget is exhausted.

The controller earns its keep if it improves outcome-per-compute (e.g., higher success with equal budget, or equal success with lower budget) and reduces failure signatures (collapse/explosion/oscillation).

12.6 When this becomes “a bigger deal”

This framework matters if it produces a portable, domain-robust controller: i.e., a single salience estimator that (i) improves prediction/decision quality per unit compute, and (ii) transfers across modalities (text, video, audio) with minimal retuning. If it only fits one dataset, it is just another heuristic.

13 Case study: pro-wrestling booking as salience control

A practical way to stress-test a salience controller is to apply it to domains that already have folk-theoretic “allocation policies.” Pro-wrestling booking is one such domain: the booker is explicitly managing scarce resources (minutes, match positions, interference, finishes) to manufacture and cash out audience attention. Cornette’s advice is often phrased as a set of operational levers (heat, comeback, payoff, protecting credibility, avoiding contrived spots). For example, he emphasizes that not all matches should be equally “competitive” and that the finish should match the intended story beat (competitive win, cheating win, decisive upset), which effectively specifies a policy over “effort” and “payoff” as a function of context [1]. He also stresses crowd feedback, selling until the comeback, and avoiding contrived spots that pull the audience out of the match [2]. Depending on audience/venue, this section can be moved to an appendix without changing the core controller.

13.1 Mapping booking concepts to salience terms

We can map common booking heuristics onto the terms of Section 3:

- **Heat** $\leftrightarrow$ retention need R and continuity C (maintain a coherent grievance thread that will be cashed out later).
- **Comeback / shine** $\leftrightarrow$ momentum M (advances the local objective of restoring the babyface’s trajectory).
- **Finish / payoff** $\leftrightarrow$ retention R (rewarding long-horizon investment) and surprise ΔA (controlled violation of expectations).
- **No contrived spots** $\leftrightarrow$ continuity C (penalize actions that break the narrative model of the audience).
- **Match layout tiers** $\leftrightarrow$ effort τ (how much “work” the match consumes) and budget B (where to spend your limited big moments).

13.2 How to quantify from video/audio

This suggests measurable proxies:

- ΔA : unexpected crowd response deltas (audio amplitude spikes), unexpected user behavior (rewind/share), or sequence-model surprise on commentary/transcripts.
- R : callbacks (motifs, rivalries) that later predict engagement; rewatch/replay; long-horizon co-consumption.
- C : contradiction scores (e.g., character alignment violations), “this makes no sense” comment rate, or drops in engagement after contrived sequences.

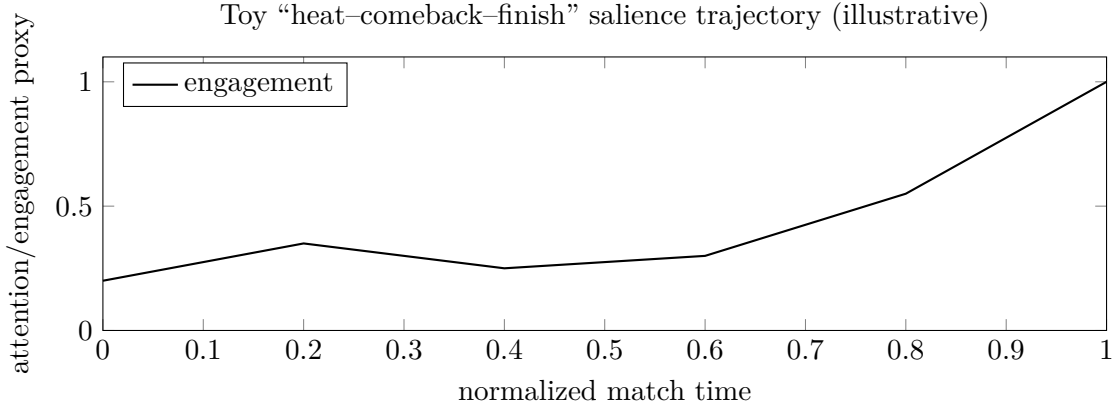


Figure 8: A stylized arc: heat builds investment (retention), the comeback increases momentum, and the finish is a high-salience cash-out.

14 Related work (brief)

Your salience score is not identical (as far as we can tell) to a single canonical prior formula, but it overlaps several established threads:

- **Rational inattention / information-constrained decision-making:** allocating limited attention under costs [11].
- **Curiosity / novelty bonuses:** using prediction error and information gain as intrinsic value [9, 6].
- **Prioritized memory/replay:** allocating updates to experiences with larger “priority” (e.g., TD error) [8].
- **Active inference / expected free energy:** trading off utility and information gain in control [5].

The “information density” reading naturally touches information theory [10, 3], while the geometry/control analogy is loosely inspired by field-and-potential thinking in physics [4]. We also cite standard references for RL and AI background [12, 7, 13].

References

- [1] Jim cornette spells out how bookers, heels lay out pro wrestling matches. Augusta Free Press, 2022. Accessed 2026-01-27.
- [2] Kayfabe, lies and alibis: Jim cornette shoot interview secrets of the ring. Culture Crossfire, n.d. Accessed 2026-01-27.
- [3] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, 2006.
- [4] Albert Einstein. Die feldgleichungen der gravitation. *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin*, 1915.

- [5] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, and Giovanni Pezzulo. Active inference and learning. *Neuroscience & Biobehavioral Reviews*, 2017.
- [6] Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *International Conference on Machine Learning (ICML)*, 2017.
- [7] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, 4 edition, 2021.
- [8] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. In *International Conference on Learning Representations (ICLR)*, 2016.
- [9] J”urgen Schmidhuber. A possibility for implementing curiosity and boredom in model-building neural controllers. *Proceedings of the International Conference on Simulation of Adaptive Behavior*, 1991.
- [10] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 1948.
- [11] Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 2003.
- [12] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2 edition, 2018.
- [13] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. Attention is all you need. In *Advances in Neural Information Processing Systems*, 2017.