REGULATING THE PAST: COMPUTATIONAL AND NEURAL SIGNATURES OF PERSISTENCE CONTROL

Anonymous authors
Paper under double-blind review

ABSTRACT

Continual learning requires preserving useful knowledge while adapting rapidly when that knowledge becomes obsolete. Biological learners solve this problem across a continuous history. Yet adaptive-learning models primarily regulate how strongly new evidence updates a state, while analyses of human learning often reset the learner at block boundaries. Such resets can conceal how accumulated confidence constrains subsequent adaptation. We study a complementary control problem: how strongly should inherited state continue to constrain learning? We call this persistence control, and instantiate it as prior-directed tempering in the Hierarchical Gaussian Filter, a widely used model of human learning under uncertainty. In simulated HGF learners with continuous histories, accumulated confidence restricted the influence of new evidence even while volatility remained represented. Tempering restored responsiveness, and mechanism recovery distinguished persistence control from surprise-modulated volatility control. In non-stationary bandits, relaxing prior influence was beneficial when contexts changed frequently, whereas preserving it was advantageous when contexts endured. Human behavior favored graded persistence control over carrying everything forward or resetting at detected changepoints. Fitted trajectories showed that persistence in reward beliefs rose after an isolated misleading outcome and was released after contradictory evidence accumulated. Successful learners also reconfigured persistence more strongly around genuine reversals rather than maintaining uniformly lower persistence. Feedback-locked frontal-midline theta covaried with this protection following misleading outcomes. Together, these results identify persistence as a distinct target of adaptive control. Continual learning, in machines and people, may depend not only on how strongly new evidence is incorporated but also on when prior beliefs should remain influential and when they should be released.

1 INTRODUCTION

Adaptive behavior depends on the past remaining influential, but not indefinitely. As evidence accumulates, beliefs become more stable and past experience increasingly constrains future behavior. Yet the same persistence becomes a liability when contingencies change. Because isolated errors and genuine environmental changes arrive through the same stream of feedback, responding too readily can discard useful knowledge, whereas responding too slowly can preserve beliefs that no longer apply. Adaptive learning therefore requires deciding when history should continue to constrain behavior and when its influence should be relaxed.

Biological learners face this problem continually: they carry one history across changing contexts, and memory research suggests that their forgetting is regulated rather than purely passive (Richards & Frankland, 2017; Anderson & Hulbert, 2021). The computational models used to study them often lack this property: they either carry everything forward or restart at known boundaries. This motivates our central question: is controlling what persists computationally distinct from controlling how rapidly beliefs update? We formalize this distinction, characterize when persistence control helps or harms an artificial learner, and then ask whether human behavior and EEG exhibit the same selective regulation across a continuous history.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Restarting at boundaries is common in behavioral and neurocomputational studies, which often make changing environments tractable by fitting each block or context as a separate learning problem. Such independent-block modelling remains appropriate when blocks truly instantiate separate tasks. But resetting the learner at a known boundary removes the inherited state whose influence we wish to measure. It can therefore obscure whether confidence accumulated in one context constrains adaptation in the next.

The Hierarchical Gaussian Filter (HGF) provides a useful test bed because it represents beliefs, posterior precision, and higher-order volatility within a single hierarchical learner (Mathys et al., 2011). This lets us ask whether a learner can represent environmental change yet still be constrained by its accumulated past. During prolonged sequential learning, posterior precision can accumulate until new evidence has little leverage over inherited belief, even while the model continues to represent volatility. Increasing volatility or update gain can restore responsiveness, but may also amplify isolated stochastic outcomes. We therefore separate two controls. Update control determines how strongly new evidence changes the current state; persistence control determines how much inherited state is presented to that update. We instantiate the latter through prior-directed tempering before each update.

Continual-learning methods usually treat forgetting as damage, protecting earlier knowledge through parameter regularization, replay, or gradient constraints (Kirkpatrick et al., 2017; van de Ven et al., 2022; Saha et al., 2021). Yet continual learners can face the complementary problem of becoming too resistant to change. This appears in different forms across learning systems: sequential Bayesian penalties can overconstrain later inference, deep networks lose responsiveness during extended nonstationary training (Huszár, 2018; D'Oro et al., 2023; Dohare et al., 2024). Several remedies correspondingly relax learned state toward a reference, including forgetting factors in adaptive filtering (Kulhavý & Zarrop, 1993), Bayesian forgetting (Kurle et al., 2020), shrink-and-perturb (Ash & Adams, 2020), regeneration toward initial weights (Kumar et al., 2023), and neuro-inspired active forgetting (Wang et al., 2023). These approaches motivate a broader question: when should a learner actively reduce the influence of what it has already learned? Our persistence-control formulation makes that decision explicit and tests when such release helps or harms.

Adaptive inference provides a different route to flexibility by changing the gain applied to new evidence. Volatility-sensitive Bayesian learners increase updating when uncertainty or prediction errors signal change; models that jointly estimate stochasticity and volatility help distinguish persistent noise from environmental change (Behrens et al., 2007; Mathys et al., 2011; Nassar et al., 2010; Piray & Daw, 2020, 2021). Bayesian Online Change Detection instead infers abrupt regime changes from a run-length posterior (Adams & MacKay, 2007). Our BOCD comparator accumulates evidence for a changepoint and resets the inherited HGF state when this evidence exceeds a fitted threshold. Persistence control differs from both: it grades how much inherited state survives before the current observation is incorporated. These alternatives allow us to ask whether persistence control contributes beyond faster updating, noise attribution, or explicit boundary detection.

At the cognitive level, functional accounts of memory transience and controlled forgetting motivate active regulation of accessibility but do not predict the sign or timing of the fitted gate (Richards & Frankland, 2017; Anderson & Hulbert, 2021). Feedback-locked frontal-midline theta (FMT) provides an independent physiological measure because it is associated with unexpected outcomes and control demands (Cavanagh & Frank, 2014; Sambrook & Goslin, 2015). We therefore proceed algorithm-first: we define persistence control and establish its computational boundary before asking whether human behavior and EEG exhibit the predicted selectivity. We treat FMT as convergent evidence, not as the neural implementation of the gate.

Figure 1 previews the task, persistence-control operation, and its behavioral and neural signatures; Supplementary Tables S1–S2 summarize the model comparisons. We address five questions. First, can a learner that represents volatility nevertheless become constrained by its accumulated past? Second, is persistence control computationally distinguishable from adapting volatility or update gain? Third, under what environmental conditions does releasing inherited state help rather than harm learning? Fourth, do humans selectively protect persistence after isolated misleading evidence and release it as evidence for change accumulates? Finally, does feedback-locked frontal-midline theta covary with this selective protection? Together, these analyses move from defining the computational operation and its boundary conditions to testing whether corresponding selectivity is expressed in human behavior and neural activity.

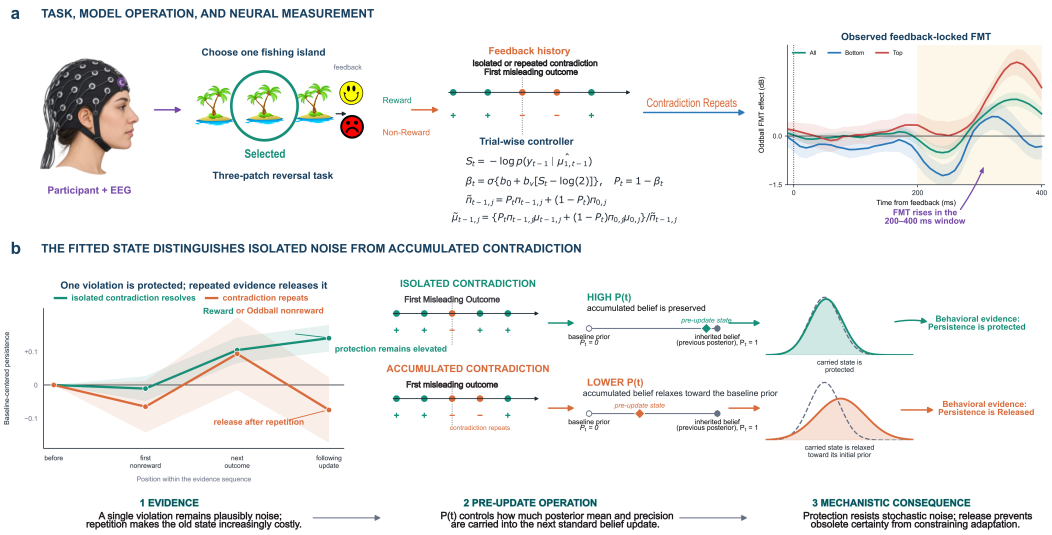

Figure 1: A pre-update gate lets a learner protect inherited beliefs after isolated contradiction and release them after accumulated contradiction. (a) In the three-patch reversal task, feedback history sets a trial-wise gate over inherited belief and precision while EEG records feedback-locked frontal-midline theta (FMT). (b) Fitted persistence rises after one misleading outcome and falls once contradiction repeats: high P_t preserves the inherited belief, whereas lower P_t relaxes it toward the initial (baseline) prior. Panel a writes S_{t-1} as S_t and uses $P_t = 1 - \beta_t$.

2 Persistence Control

2.1 Persistence Control as a Pre-Update Operation

Persistence control inserts one operation into an ordinary learning step. Let z_{t-1} denote the state inherited from the previous trial, z_0 its initial prior, y_t the new observation, and θ the ordinary update parameters. The inserted transformation precedes the update:

$$\tilde{z}_{t-1} = \mathcal{T}(z_{t-1}, z_0; \beta_t), \qquad z_t = \mathcal{U}(\tilde{z}_{t-1}, y_t; \theta). \tag{1}$$

The transformation $\mathcal{T}$ controls how strongly the previous state constrains the next trial; $\mathcal{U}$ controls how new evidence changes the state that survived. The decomposition is representation-agnostic: it specifies where control acts without requiring the retained object to be identical across Bayesian inference, reinforcement learning, and human model fits.

We instantiate the operation in the HGF. Standard HGF starts each trial from the previous posterior, whereas AF-HGF (active-forgetting HGF) first tempers the Gaussian sufficient statistics at levels $j \in \{2, 3\}$ toward their initial priors:

$$\tilde{\pi}_{t-1,j} = (1 - \beta_t)\, \pi_{t-1,j} + \beta_t\, \pi_{0,j}, \tag{2}$$

$$\tilde{\mu}_{t-1,j} = \left[(1 - \beta_t)\, \pi_{t-1,j}\, \mu_{t-1,j} + \beta_t\, \pi_{0,j}\, \mu_{0,j}\right] / \tilde{\pi}_{t-1,j}. \tag{3}$$

The unchanged HGF prediction and update equations then operate on $(\tilde{\mu}_{t-1,j}, \tilde{\pi}_{t-1,j})$. At $\beta_t = 0$, AF-HGF reduces to Standard HGF; at $\beta_t = 1$, the state returns to its initial prior before updating. We use persistence control for the general decision, active forgetting for the prior-directed release operation, β_t for release, and $P_t = 1 - \beta_t$ for persistence, the retained carryover. High P_t therefore means stronger preservation, not stronger forgetting (Supplementary Figure S1; Supplementary Section B).

Static AF-HGF uses a participant-specific constant β. Trial-wise AF-HGF instead maps previous-outcome surprise to release:

$$S_{t-1} = -\log p(y_{t-1} \mid \hat{\mu}_{t-1}), \qquad \beta_t = \sigma\big[b_0 + b_v(S_{t-1} - \log 2)\big]. \tag{4}$$

When $b_v = 0$, $\beta_t = \sigma(b_0)$ is constant, so Trial-wise AF-HGF reduces exactly to Static AF-HGF. Together with the $\beta_t = 0$ boundary, the models form nested hypotheses about carryover: fixed at complete carryover (Standard HGF), uniformly relaxed (Static AF-HGF), or evidence-dependent (Trial-wise AF-HGF). Supplementary Algorithm S1 (Supplementary Section B.1) summarizes the resulting trial-wise persistence-controlled inference procedure.

2.2 Separating persistence control from learning-rate adaptation

We tested this distinction in a non-stationary three-arm bandit environment. Each arm carried an estimated reward value, while unannounced changes altered which values remained useful. An 81-environment evaluation crossed four properties that separate how often stored values became obsolete (context duration: 40, 80, or 150 trials), how gradually contingencies evolved (within-context drift: 0, 0.004, or 0.012), how reliably individual outcomes reflected them (observation noise: 0, 0.25, or 0.50), and how strongly they changed (reversal magnitude: 0.30, 0.55, or 0.75). We compared complete carryover, fixed and evidence-adaptive carryover, adaptive learning-rate control, and joint carryover–learning-rate control. All adaptive controllers received the same lagged evidence features, isolating whether adaptation acted on the inherited state or on the subsequent evidence update.

Before the current reward arrives, carryover control relaxes every stored action value toward a neutral prior:

$$\tilde{Q}_{t-1}(a) = c_t\, Q_{t-1}(a) + (1 - c_t)\, Q_0(a), \tag{5}$$

$$Q_t(a_t) = \tilde{Q}_{t-1}(a_t) + \alpha_t\big[r_t - \tilde{Q}_{t-1}(a_t)\big]. \tag{6}$$

Because c_t acts before the outcome and can alter both selected and unselected values, changing α_t cannot generally reproduce the same transformation. The two implementations retain different objects—HGF tempering operates on inherited belief and precision, whereas bandit carryover operates on cached action values—so the shared principle is control over inherited influence, not an identical stored representation. We therefore write c_t for bandit carryover and reserve P_t for HGF persistence.

We separately mapped the fixed carryover that maximized expected reward in each environment. Holding the update rate at $\alpha = 0.20$, we evaluated 41 carryover values between 0.92 and 1.00 across 49 context-duration-by-noise conditions. For environment e and evaluated carryover set C, optimal carryover is

$$c^*(e) = \arg\max_{c \in C} \; \mathbb{E}\Big[\textstyle\sum_{t=1}^{T} r_t \;\Big|\; e, \pi(\alpha,c)\Big]. \tag{7}$$

Supplementary Table S1 specifies every model, and Table S2 places these learners beside the Bayesian alternatives, stating where adaptation acts, which alternative each represents, and how it fared.

2.3 Human task and performance measures

Seventy-four participants completed 200 trials of a three-patch probabilistic reversal task while EEG was recorded. Four 50-trial blocks crossed reward stochasticity with reversal speed, placing isolated misleading outcomes and genuine changes in the same feedback stream; we refer to a misleading non-reward after an objectively best choice as an oddball. Models were fitted to whether the objectively highest-probability patch was chosen (the choice stream); persistence estimated from reward outcomes was additionally used to test feedback-linked protection and release. Performance was summarized by total reward and a Continual Learning Index (CLI), a composite designed to reward both stable performance and the ability to recover when contingencies changed: CLI = 0.5

$\times$ z(overall best-option accuracy) + 0.5 $\times$ z(post-reversal recovery). Analysis-specific samples and exclusions are listed in Supplementary Section A.1.

3 RESULTS

Participant models were compared using BIC and marginal log model evidence (LME) on common participants and trials. Primary human analyses used continuous within-participant contrasts or participant-level associations, with multiplicity correction within prespecified analysis families. Artificial-agent claims were evaluated using frozen controllers and held-out environments; additional robustness and sensitivity analyses are reported in the Supplement.

3.1 CAN A LEARNER TRACK CHANGE YET REMAIN CONSTRAINED BY THE PAST?

If representing volatility were enough to keep a learner responsive, persistence control would be inconsequential for Standard HGF, which already represents volatility. We tested this in controlled HGF agents that moved from a clean, slowly reversing phase (A) to a noisier, rapidly reversing phase (B); these interventions test whether tempering restores update leverage under known transitions and are not fits to participant behavior. Standard-HGF best-option accuracy fell from 0.642 to 0.398 after the transition, whereas tempering preserved greater responsiveness across both tested release strengths: Static AF changed from 0.741 to 0.488 at $\beta = 0.121$ and from 0.673 to 0.481 at $\beta = 0.30$. Reciprocal precision diverged: in phase B, ψ_2 was 1.455 for Standard HGF but 3.011 and 3.087 under AF. Volatility still varied, yet accumulated precision restricted update leverage; tempering inherited state restored responsiveness without altering the ordinary hierarchy.

The benefit generalized across repeated transitions. Across 100 paired A/B/A/B seeds, expected score rose from 179.1 for Standard HGF to 206.5 for Static AF and 206.3 for Trial-wise AF, and best-option accuracy rose from 0.51 to 0.62. Both AF agents improved reward, expected score, regret, and accuracy (all $p < 0.001$) but did not differ from each other in expected score (difference -0.28, $p = 0.40$). An unchosen-only leaky RL agent achieved 207.27, which bounds the claim: reducing inherited influence restores responsiveness, but the HGF implementation is not algorithmically unique.

The benefit also held when A returned after a much longer interruption. We lengthened the intervening B phase from 50 to 1,000 trials (A1 $\rightarrow$ B $\rightarrow$ A2; 100 paired seeds). Across this range, Trial-wise AF-HGF kept its uncertainty and volatility belief steady, whereas Standard HGF grew increasingly confident that its environment was stable, with its volatility belief falling from 0.99 to 0.84 (Supplementary Figure S3A–B). Sustained release came with a retention horizon: relative to a 50-trial B, early accuracy on the returning A was 2.78 percentage points lower after longer phases (pooled over 150–1,000 trials; 95% CI [-5.07, -0.45]), a cost that did not grow with duration. Trial-wise AF-HGF still earned more expected reward than Standard HGF on the return at every duration, although this gap partly reflects their different tonic volatility (Supplementary Section C.6).

Restored responsiveness alone does not show that carryover is the operative location; the same surprise signal might instead act through volatility. We therefore compared Trial-wise AF-HGF with a Surprise-Modulated Volatility HGF in which the same lagged-surprise signal controlled volatility rather than inherited-state carryover. In the human cohort (Section 3.3), Trial-wise AF was favored by ΔBIC = 692.9 and ΔLME = 527.7 (positive values favor carryover; Figure 2). Matched generative recovery then compared Trial-wise AF-HGF with Surprise-Modulated Volatility HGF under equated conditions: the same surprise signal, generating slope -2, fitting support [-2.5, 1], and 30 synthetic cohorts per family, with both models refit to every cohort, selection assessed under both criteria with Wilson intervals, and nested limits checked explicitly. Every volatility cohort selected the volatility model and every AF cohort selected AF under both criteria (30/30 each; Wilson 95% interval 88.6%–100%), so the sign of the evidence difference reverses with the generating mechanism (Figure 2b). The mechanisms are therefore distinguishable because of where the shared adaptive signal acts, rather than because one receives richer information (Supplementary Figure S2; Supplementary Sections C.2 and C.4).

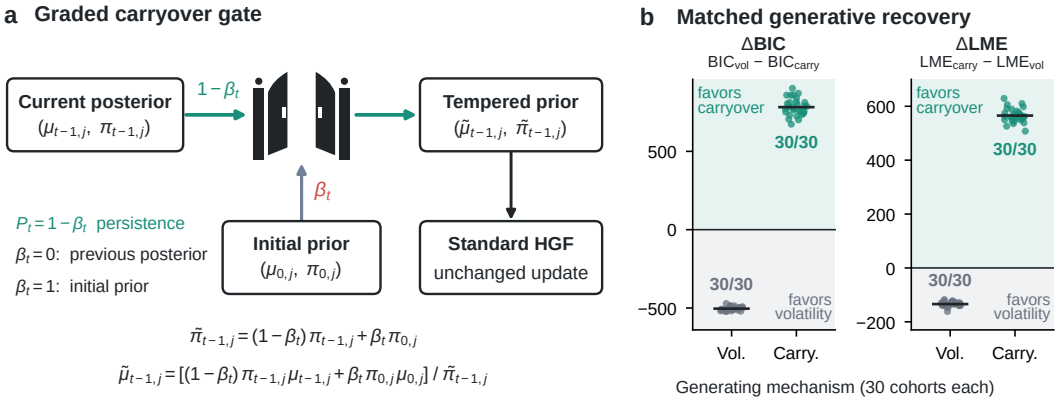

Figure 2: Surprise-driven carryover and surprise-driven volatility leave distinguishable signatures. (a) The graded gate tempers inherited HGF state toward its initial prior before the unchanged update. (b) Across 30 cohorts per generating mechanism, the sign of ΔBIC and ΔLME (both positive when carryover fits better) recovers which mechanism generated the data.

3.2 When does releasing the past help rather than harm adaptation?

Restoring responsiveness is useful only if release does not erase knowledge that remains predictive. We therefore tested the boundary in the non-stationary three-arm bandit of Section 2.2, where persistence control relaxed stored values before reward and learning-rate control acted afterward. The adaptive learners of Section 2.2 were selected on disjoint training and validation seeds, frozen, and evaluated on a fresh 81-cell sweep with 12 seeds per cell.

The sweep revealed a dependence on environmental timescale: the margin of evidence-adaptive carryover over adaptive learning-rate control rose as contexts shortened, becoming positive only at duration 40, where carryover control won in 21 of 27 cells (77.8%; mean advantage +2.14), compared with 5 of 27 at duration 80 (18.5%; −1.70) and 2 of 27 at duration 150 (7.4%; −4.70). The frequency effect persisted across noise and drift: all nine noise-by-drift averages were lower at duration 150 than at 40, and seven declined at every step (Figure 3a). Release became competitive when values repeatedly became obsolete; preservation dominated as contexts endured.

In the fixed-carryover search (Section 2.2), the optimum shifted from partial release under frequent change to near-complete preservation in durable contexts; the lowest selected optimum retained 97.2% per trial. In Figure 3b, the orange valleys and broad blue plateau visualize this boundary—not a universal persistence constant.

Held-out evaluation confirmed where release helps. Discovery-selected carryover was frozen and evaluated on 24 new seeds per cell. At duration 40, gains over complete carryover generalized at noise 0 (+3.90 expected score, 95% CI [1.86, 5.89]), 0.08 (+3.74 [1.76, 5.65]), and 0.17 (+2.00 [0.28, 3.77]). The advantage weakened with noise and disappeared in longer contexts; no cell was reliably worse. Partial release therefore helped chiefly under frequent change and relatively reliable observations (Supplementary Section D.3).

3.3 Does human persistence distinguish isolated noise from accumulating change?

Human behavior then tests whether people deploy release selectively: does an established state resist isolated noise but yield when contradiction accumulates? We first asked whether human behavior was better characterized by continual graded relaxation or discrete changepoint-triggered resets. Changepoint detection is a standard response to abrupt non-stationarity; if discrete resets explained behavior, continual persistence control would be unnecessary. Bayesian Online Change Detection HGF provided this all-or-none alternative, in which a latent-state reset occurred when accumulated change evidence exceeded a fitted threshold. Static AF-HGF improved mean log model

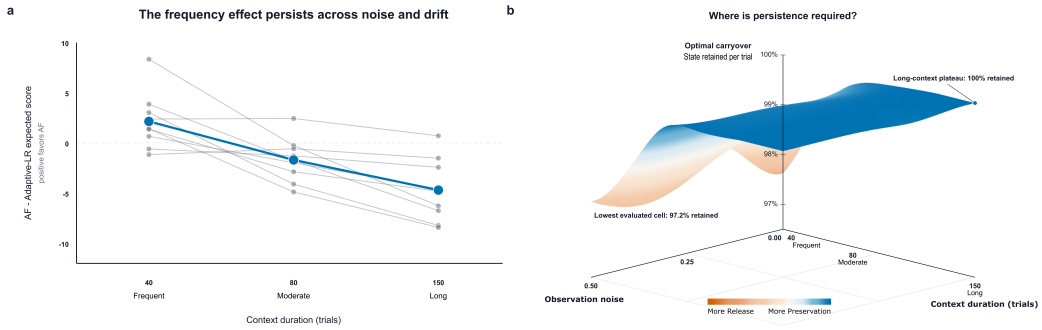

Figure 3: Releasing the past becomes competitive when contexts change often; preservation wins as contexts endure. (a) The frequency effect persists across noise and drift. (b) Optimal fixed carryover shifts from partial release toward preservation as contexts lengthen.

evidence from -132.53 for Standard HGF and -132.06 for Bayesian Online Change Detection HGF to -126.95, winning in 51 of 74 participants. The changepoint alternative rarely triggered hard resets: 45 of 74 participants never reset (Supplementary Section B.3). Thus under these conditions, human behavior favored continual relaxation over rare restarts (Supplementary Figure S4).

The decisive test concerned timing. After a single misleading non-reward, persistence was greater than after a rewarded-best control—a protection response (difference 0.0554, $N = 73$, FDR $q < 0.001$). Stronger protection correlated with reward ($\rho = 0.599$, $q < 0.001$) and with the CLI (Supplementary Table S4). The choice-stream contrast was null (difference -0.00643, $q = 0.135$), limiting generalization across streams.

Protection did not continue indefinitely. After a second consecutive oddball, persistence fell relative to the preceding transition (-0.0883, 95% CI $[-0.1372, -0.0394]$, $q = 0.001004$); the adjusted change and a sign-flip test agreed in direction (Supplementary Table S4). Figure 4a shows the selective trajectory: persistence rises when one adverse outcome may be noise, then turns downward once contradiction repeats, and the excursion is larger in the upper reward quartile.

Successful learners differed in where, not simply how much, they reconfigured persistence. Upper and lower reward quartiles differed little in mean persistence (difference 0.0228, $p = 0.704$) or overall change variability (difference 0.0207, $p = 0.501$), but their reversal surges were 0.1667 versus 0.0653 (difference 0.1014, Hedges $g = 1.222$, $p < 0.001$). Across 73 participants, surges correlated with the CLI ($\rho = 0.400$, $q = 0.00160$; Figure 4b) and with reward (Supplementary Section E.3). Better performance therefore accompanied selective reconfiguration around consequential changes rather than uniformly lower persistence.

Reconfiguration also carried prospective information: within participants, larger reconfiguration predicted more subsequent switching and lower immediate next-choice optimality, so it forecast behavioral change rather than guaranteeing a better choice (Supplementary Section E.4).

3.4 Does human EEG provide convergent evidence for persistence protection?

Finally, we asked whether feedback-locked FMT converged with the fitted protection response. The prespecified measure was 4–8 Hz activity over Fz, FCz, FC1, FC2, and Cz from 200–400 ms following feedback; total and induced Morlet power provided complementary measures.

Oddballs elicited greater FMT than rewarded-best controls. Across 72 participants, adjusted total theta increased by +0.4132 dB (95% CI [0.1370, 0.6894], $d_z = 0.3456$, $q = 0.00520$) and induced theta by +0.5469 dB ($q < 0.001$; Supplementary Table S5). A reward-by-best-choice contrast further showed that choice context in which feedback occurred mattered: +0.8919 dB total and +0.7690 dB induced theta ($N = 70$, both $q < 0.001$).

Larger neural oddball effects accompanied stronger computational protection. Across participants, total theta covaried with protection ($\rho = 0.3358$, $q = 0.00810$; Figure 5a), as well as with reward

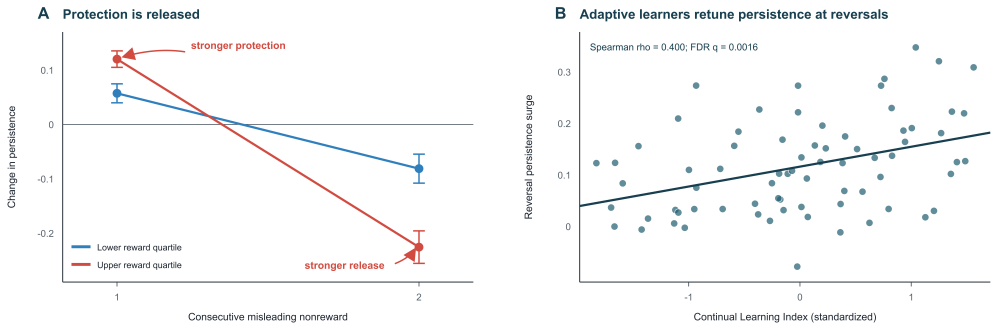

Figure 4: Human persistence is selective rather than uniformly low. (a) Persistence is protected after isolated contradiction and released after repetition, especially in upper-reward participants. (b) Reversal-linked reconfiguration scales with continual-learning performance. Prospective next-choice analyses and their prediction boundary are in Supplementary Figure S5 and Supplementary Section E.4.

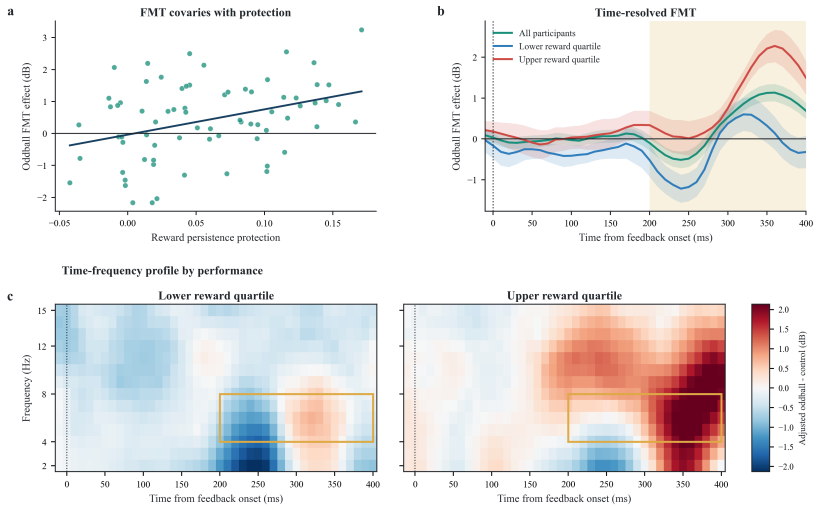

Figure 5: Feedback-locked theta covaries with persistence protection. (a) Participant-level FMT covaries with computational protection. (b) Feedback-locked trajectories separate within the prespecified 4–8 Hz, 200–400 ms window. (c) Descriptive time–frequency maps place the group difference within the prespecified theta window; inferential claims remain tied to the prespecified frontocentral ROI and time window (gold boxes).

and the CLI (Supplementary Table S5). Upper and lower reward quartiles differed by 1.2678 dB in total theta (Hedges $g = 1.104$, $q = 0.00230$; Figure 5b–c).

At the event-level, the first-oddball FMT preceded and predicted greater next-trial protection: across 1553 observations, the total theta slope was 0.02288 (95% CI [0.01489, 0.03087], $p < 0.001$), with a corresponding induced theta effect (Supplementary Section F.2). A model using all eligible trials and nuisance controls retained a positive association between persistence and total FMT (Supplementary Section F.3).

These neural associations were not fully specific. More extensive adjustment attenuated the first-oddball effects, stricter residualized and exact-conjunction tests were null, and source estimates were spatially coarse. We therefore interpret FMT as convergent evidence for selected expressions of persistence protection, not as evidence that theta implements the persistence gate or identifies its anatomical locus (Supplementary Figures S6–S7; Supplementary Sections F.2–F.4).

4 DISCUSSION

Adaptive learning must resist observations that should not overturn useful knowledge, yet release accumulated information once it obstructs change; treating inherited influence as an explicit control variable made that problem tractable. Because persistence control acts on the state arriving at an update, a pre-update gate can alter inherited and unchosen information before the current observation is incorporated—an operation a scalar learning-rate change cannot generally reproduce. Persistence and update control are nonetheless not independent in their consequences: tempering carried precision changes effective downstream learning. They are therefore distinct computational loci whose effects interact, complementing precision-weighted and volatility-sensitive learning (Mathys et al., 2011; Piray & Daw, 2020, 2021) and inference over abrupt boundaries (Adams & MacKay, 2007).

For artificial learners, the functional result is an environmental boundary, not aggregate dominance: release became advantageous in parts of the environment as contexts shortened, while the fixed-carryover surface shifted toward preservation as contexts lengthened. This pattern resembles continual feature replacement and reset-based remedies for lost plasticity (D'Oro et al., 2023; Dohare et al., 2024) but adds a decision boundary for when relaxation is useful rather than uniformly beneficial. The recurrence analysis exposes a complementary limit: repeated release can attenuate information acquired before an intervening context, producing a finite effective retention horizon when that context returns. Persistence control therefore addresses when old information should stop constraining current learning; it does not by itself solve how currently irrelevant information should be preserved for future reuse. Releasing knowledge from current control and erasing its future availability are not equivalent computational problems. This distinction motivates continual-learning benchmarks that separately score rapid adaptation, stable-period preservation, and return to an earlier context.

Human behavior was consistent with selective rather than uniform persistence regulation. Persistence in reward beliefs increased after isolated misleading evidence and was released after contradiction accumulated, while successful learners concentrated larger reconfigurations around consequential reversals. Importantly, stability of action was not equivalent to rigidity of inference: stronger fitted tempering could coexist with consistent exploitation and responsive beliefs.

Functional accounts of adaptive memory anticipate that transience can serve future decisions (Richards & Frankland, 2017; Anderson & Hulbert, 2021). Independent-block modelling can reset precisely the state whose influence is under study; our cross-block analysis instead treats persistence as an empirical quantity, so resetting should be a modelled hypothesis or validated boundary condition, not an automatic preprocessing decision.

Limitations: Human persistence is model-derived rather than directly observed, reward- and choice-stream effects remain distinct, and the protect–release sequence is a task-specific result rather than a direct transfer of a biological mechanism. The cross-system claim concerns a shared control problem, not an identical representation: human and HGF agents gate model-derived belief and precision, whereas the bandit relaxes action values. The recurrence analysis further shows that persistence control is not a complete solution to continual learning: repeated release can attenuate information that later becomes useful, while the present mechanism does not separately preserve such information for future retrieval.

The neural evidence is convergent rather than causal: feedback theta covaried with protection and performance and anticipated next-trial protection in the primary event model, consistent with broad accounts of frontal theta in feedback processing and control (Cavanagh & Frank, 2014; Sambrook & Goslin, 2015); the data do not identify an anatomical gate or show that theta implements release.

The central result is not that a learner should forget more, but that the value of forgetting depends systematically on environmental structure; active forgetting is a conditional mechanism for regulating what persists, not a prescription to remember less. Together with the human selectivity, these findings reveal a common computational principle: adaptive learning requires deciding not only how strongly to learn from new evidence, but also when prior information should remain influential and when the learner should let it go.

AI USE STATEMENT

A large language model was used to revise the grammar and language of the manuscript text and to assist with LaTeX formatting. All scientific claims, analyses, and numerical results were produced and checked by the authors, who take full responsibility for the final content of this work.

ETHICS STATEMENT

The human study involved 74 participants who completed a behavioral task during EEG recording. The study was approved by the Institute Ethics Committee.

REFERENCES

Adams, R. P., & MacKay, D. J. C. (2007). Bayesian online changepoint detection. arXiv:0710.3742. https://arxiv.org/abs/0710.3742

Anderson, M. C., & Hulbert, J. C. (2021). Active forgetting: Adaptation of memory by prefrontal control. Annual Review of Psychology, 72, 1–36. https://doi.org/10.1146/annurev-psych-072720-094140

Ash, J. T., & Adams, R. P. (2020). On warm-starting neural network training. Advances in Neural Information Processing Systems, 33.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. Nature Neuroscience, 10, 1214–1221. https://doi.org/10.1038/nn1954

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. Trends in Cognitive Sciences, 18, 414–421. https://doi.org/10.1016/j.tics.2014.04.012

D'Oro, P., Schwarzer, M., Nikishin, E., Bacon, P.-L., Bellemare, M. G., & Courville, A. (2023). Sample-efficient reinforcement learning by breaking the replay ratio barrier. International Conference on Learning Representations. https://openreview.net/forum?id=OpC-9aBBVJe

Dohare, S., Hernandez-Garcia, J. F., Lan, Q., Rahman, P., Mahmood, A. R., & Sutton, R. S. (2024). Loss of plasticity in deep continual learning. Nature, 632, 768–774. https://doi.org/10.1038/s41586-024-07711-7

Huszár, F. (2018). Note on the quadratic penalties in elastic weight consolidation. Proceedings of the National Academy of Sciences, 115, E2496–E2497. https://doi.org/10.1073/pnas.1717042115

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwińska, A., Hassabis, D., Clopath, C., Kumaran, D., & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences, 114, 3521–3526. https://doi.org/10.1073/pnas.1611835114

Kulhavý, R., & Zarrop, M. B. (1993). On a general concept of forgetting. International Journal of Control, 58, 905–924.

Kumar, S., Marklund, H., & Van Roy, B. (2023). Maintaining plasticity in continual learning via regenerative regularization. arXiv:2308.11958. https://arxiv.org/abs/2308.11958

Kurle, R., Cseke, B., Klushyn, A., van der Smagt, P., & Günnemann, S. (2020). Continual learning with Bayesian neural networks for non-stationary data. International Conference on Learning Representations.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. Frontiers in Human Neuroscience, 5, 39. https://doi.org/10.3389/fnhum.2011.00039

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. Journal of Neuroscience, 30, 12366–12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

Piray, P., & Daw, N. D. (2020). A simple model for learning in volatile environments. PLOS Computational Biology, 16, e1007963. https://doi.org/10.1371/journal.pcbi.1007963

Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. Nature Communications, 12, 6587. https://doi.org/10.1038/s41467-021-26731-9

Richards, B. A., & Frankland, P. W. (2017). The persistence and transience of memory. Neuron, 94, 1071–1084. https://doi.org/10.1016/j.neuron.2017.04.037

Saha, G., Garg, I., & Roy, K. (2021). Gradient projection memory for continual learning. International Conference on Learning Representations. https://openreview.net/forum?id=3AOj0RCNC2

Sambrook, T. D., & Goslin, J. (2015). A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. Psychological Bulletin, 141, 213–235. https://doi.org/10.1037/bul0000006

van de Ven, G. M., Tuytelaars, T., & Tolias, A. S. (2022). Three types of incremental learning. Nature Machine Intelligence, 4, 1185–1197. https://doi.org/10.1038/s42256-022-00568-3

Wang, L., Zhang, X., Li, Q., Zhang, M., Su, H., Zhu, J., & Zhong, Y. (2023). Incorporating neuro-inspired adaptability for continual learning in artificial intelligence. Nature Machine Intelligence, 5, 1356–1368. https://doi.org/10.1038/s42256-023-00747-w

SUPPLEMENTARY MATERIAL

The main paper establishes the mechanism, environmental boundary, human protect–release pattern, and neural convergence. This supplement instead exposes implementation details, tests alternatives that could mimic the mechanism, evaluates held-out conditions, and reports analyses that delimit rather than restate the central claims.

A COMPLETE EXPERIMENTAL AND INFERENTIAL SPECIFICATION

A.1 HUMAN TASK, PARTICIPANTS, AND ANALYSIS SAMPLES

This subsection establishes the task structure, analysis populations, and data streams underlying later claims, because differences in transitions, exclusions, or estimands determine which results can be compared. Seventy-four participants completed 200 trials of a three-patch probabilistic reversal task during 32-channel EEG. Four 50-trial blocks crossed reward stochasticity with reversal speed. The low-noise reward schedule was 0.80/0.15/0.05 and the high-noise schedule was 0.60/0.25/0.15; within-block reversals occurred every 25 trials in slow blocks and every 10 trials in fast blocks. Block order was counterbalanced, so equivalent transitions occurred at different session positions across participants.

Model comparison used all 74 participants. Primary persistence analyses used $N = 73$ after analysis-specific quality control; the principal oddball-FMT contrast used $N = 72$; the reward-by-best-choice theta contrast used $N = 70$; the source-ready analysis used $N = 56$; and the all-eligible-trials FMT model used $N = 63$. Each result reports its effective sample because missing trials, EEG quality control, and the availability of a fitted event sequence differed across analyses.

The choice-stream model was fitted to whether the objectively highest-probability patch was chosen. Persistence was constructed only as a separately labelled validation bridge for feedback analyses. It does not replace the choice-stream quantity used for model comparison, and reward- and choice-stream effects are not treated as replications of the same estimand.

A.2 EEG ACQUISITION, PREPROCESSING, AND PRESPECIFIED THETA TEST

We specify the acquisition and prespecified FMT target to separate confirmatory neural evidence from descriptive localization. EEG was recorded from 32 channels. The corrected feedback analysis used the FCz-restored, pre-reference bad-channel-QC branch and a frontocentral region of interest comprising Fz, FCz, FC1, FC2, and Cz. Morlet theta power was averaged over 4–8 Hz, baseline corrected to -200–0 ms, and evaluated in the prespecified 200–400 ms feedback-locked window. The time courses and time–frequency maps visualize how the locked effect developed; inferential claims use the prespecified region and window rather than post hoc pixelwise clusters.

Primary feedback models adjusted for condition, trial position, block order, reaction time, and feedback delay. Extended models additionally included computational update quantities and nuisance predictors. Total and induced theta were analysed separately. Descriptive scalp distributions and template-anatomy source estimates are included below as supporting localization, not as evidence for a precisely localized neural gate.

A.3 STATISTICAL REPORTING AND MULTIPLICITY

We document estimation and multiplicity control so evidence from participant-, event-, and trial-level analyses can be judged under a common reporting standard. Participant-level paired contrasts used effect sizes and confidence intervals; nonparametric associations used Spearman correlations. Permutation or sign-flip tests were used where the estimand was defined at participant level, and mixed or generalized-estimating-equation models accounted for repeated trials. Families of related tests were corrected using false-discovery-rate control. Exact values smaller than 0.001 are reported as $p < 0.001$ in prose; complete numerical values remain available in the analysis tables.

Algorithm S1: Persistence-controlled inference

Input: inherited posterior $\{\mu_{t-1,j}, \pi_{t-1,j}\}$, initial prior $\{\mu_{0,j}, \pi_{0,j}\}$, observation y_t, and fitted parameters.

1. **Surprise.** Compute $S_{t-1} = -\log p(y_{t-1} \mid \hat{\mu}_{t-1})$.
2. **Control.** Set $\beta_t = \sigma[b_0 + b_v(S_{t-1} - \log 2)]$ and retained persistence $P_t = 1 - \beta_t$.
3. **Temper.** For each controlled level $j \in \{2, 3\}$, apply the prior-directed mean and precision transformation defined above.
4. **Update.** Run the unchanged HGF prediction and update equations from the tempered state.
5. **Choose.** Map the updated latent state through the fitted observation model.

Output: updated posterior state and choice probability.

B PERSISTENCE-CONTROLLED INFERENCE

B.1 EXACT HGF REALIZATION

We state the exact HGF intervention to identify which inherited quantities persistence control changes and to distinguish it from altering the ordinary update. The implementation intervenes on the sufficient statistics passed from trial t−1 into the ordinary HGF prediction step. For level $j \in \{2,3\}$,

$$\tilde{\pi}_{t-1,j} = (1 - \beta_t)\, \pi_{t-1,j} + \beta_t\, \pi_{0,j}, \tag{8}$$

$$\tilde{\mu}_{t-1,j} = \big[(1 - \beta_t)\, \pi_{t-1,j}\, \mu_{t-1,j} + \beta_t\, \pi_{0,j}\, \mu_{0,j}\big] / \tilde{\pi}_{t-1,j}. \tag{9}$$

The unchanged HGF prediction and update equations then begin from $(\tilde{\mu}_{t-1,j}, \tilde{\pi}_{t-1,j})$; level 1 is not tempered. The supplement records the exact state transformation, while the main text carries the representation-independent persistence-versus-plasticity argument.

B.2 GATE TIMING, ENDPOINTS, AND NESTING

We verify gate timing, endpoints, and nesting to establish causal order, rule out current-outcome leakage, and show how the adaptive and standard models are related. Static AF-HGF uses $\beta_t = \sigma(b_0)$. Trial-wise AF-HGF uses $\beta_t = \sigma[b_0 + b_v(S_{t-1} - \log 2)]$, with retained persistence $P_t = 1 - \beta_t$. Thus $\beta_t = 0$ recovers standard carryover, $\beta_t = 1$ returns the controlled state to its initial prior, and $b_v = 0$ nests the static model. The implementation order was fixed: compute lagged surprise; map it to β_t; temper levels 2 and 3; run the unchanged HGF update; map the posterior to choice. Current-trial outcomes therefore cannot leak into the pre-update gate.

B.3 BAYESIAN ONLINE CHANGE DETECTION IMPLEMENTATION

We examine Bayesian Online Change Detection to test whether behavior requires graded continual relaxation beyond discrete state resets and to explain when that alternative activates. The Bayesian Online Change Detection HGF maintained a run-length posterior and triggered a state reset only when summed change probability exceeded a participant-specific threshold. A 20-trial cooldown prevented repeated resets. The calibrated threshold was 9.64 nats. The median participant generated no resets; 45 of 74 participants never reset and the mean was 1.43 resets. This diagnostic explains the comparator's behavior: under these stochastic schedules, thresholded evidence rarely justified an all-or-none restart.

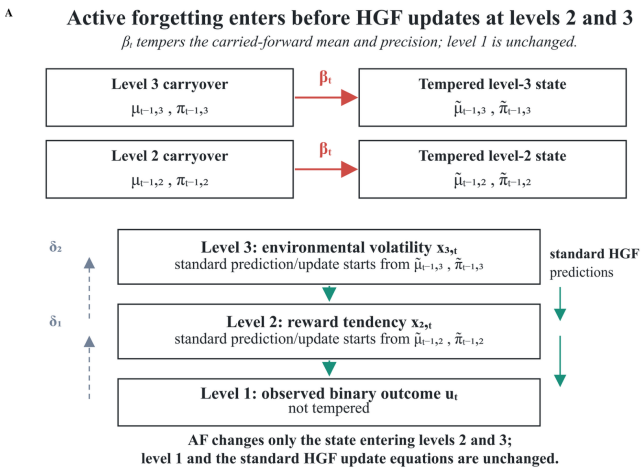

Figure S1: Level-specific intervention. Persistence control modifies the carried level-2 and level-3 states before standard HGF prediction and updating; the observed level remains unchanged. The figure writes the observed outcome as u_t; the text writes y_t.

C IDENTIFIABILITY, INTERVENTION, AND ALTERNATIVE MECHANISMS

C.1 MATCHED GENERATIVE RECOVERY

Model comparison is useful only if it can identify where surprise acts—on carryover or volatility—rather than merely reward flexibility. Exact nesting establishes that Trial-wise AF-HGF contains Static AF-HGF, but it does not establish distinguishability from a different adaptive mechanism. We therefore compared Trial-wise AF-HGF with Surprise-Modulated Volatility HGF. Both models received the same lagged-surprise signal, used generating slope -2, and were fitted over the same slope support $[-2.5, 1]$. The models differed only in where surprise acted: inherited state or volatility.

Trial-wise AF-HGF was favored over the volatility alternative in the empirical cohort by ΔBIC = 692.932 and ΔLME = 527.726 (both signed so that positive values favor Trial-wise AF-HGF; Figure 2b). In recovery, all 30 volatility-generated cohorts selected the volatility model and all 30 carryover-generated cohorts selected AF under both criteria. Each 30/30 rate has Wilson 95% interval 88.648%–100%. Within this calibrated comparison, the sign of the model-evidence difference recovered the generating computational location without overlap.

C.2 COUNTERFACTUAL SCHEDULE DISRUPTIONS

We disrupt the fitted β_t schedule to test whether its sign and timing carry explanatory information beyond each participant's average release. A counterfactual replay rescored each observed choice stream after replacing only β_t while holding ω_2, the choice parameter, and baseline gate b_0 fixed. The replay reproduced the saved likelihood to a maximum discrepancy of approximately 5×10^{-13} across 74 participants and 14,800 trials, so changes in negative log likelihood reflect only the schedule substitution.

Every participant was fit worse after reversing the surprise-gate sign (mean ΔNLL = 9.86; 74/74 real schedules better). Setting $b_v = 0$ cost 6.42 NLL on average and favored the real schedule in 72/74 participants. Freezing β_t at its participant mean cost 1.44 NLL; delaying it by one trial cost 1.44; delaying by five trials cost 1.75; shuffling within participant cost 1.95; and shuffling within block cost 1.97. Across 20 within-participant shuffles per participant, the fitted schedule beat the shuffle in 92.97% of 1,480 replays and beat the participant's average shuffle in every participant.

The timing cost concentrated around change. Reversing the gate sign cost 0.0887 NLL per trial during the reversal and first five post-reversal trials, 0.0273 during trials 6–15, and 0.0063 later

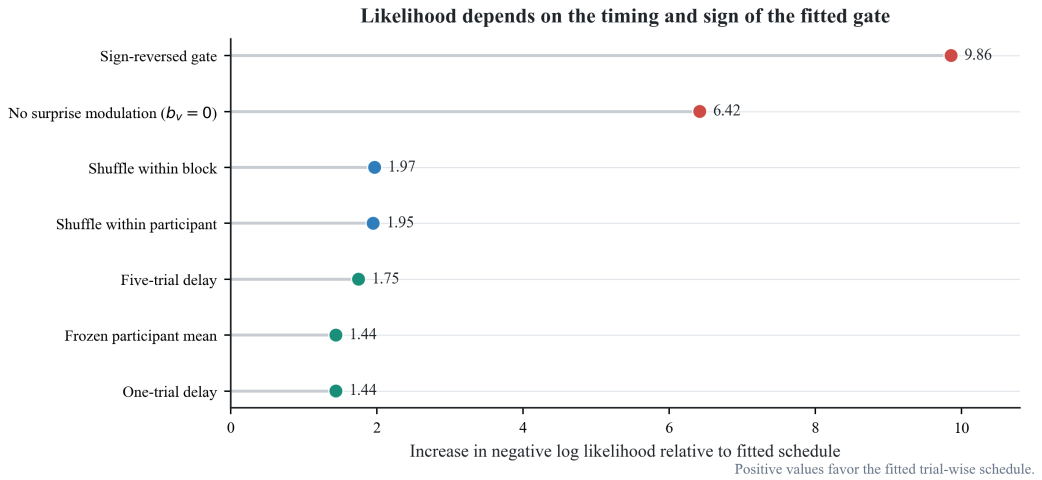

Figure S2: Counterfactual timing test. Likelihood costs after sign reversal, removal, shuffling, delay, or freezing of the fitted gate; positive ΔNLL favors the original trial-wise schedule.

in the block. Freezing β_t was slightly helpful late in a block (-0.0043 NLL per trial) but costly immediately after reversal (+0.0176). Trial-wise persistence is therefore not justified by generic flexibility; its fitted timing matters chiefly where histories become obsolete.

C.3 HGF INTERVENTION AND THE PRESERVATION COST

We intervene in controlled learners to test whether relaxing inherited state restores responsiveness, while the leaky-RL comparison asks whether that benefit is unique to the HGF implementation. In a repeated A/B/A/B stress test with 100 paired environment seeds, expected score increased from 179.1 for Standard HGF to 206.5 for Static AF and 206.3 for Trial-wise AF; best-option accuracy increased from 0.51 to 0.62. Both AF agents improved reward, expected score, regret, and accuracy relative to Standard HGF (all $p < 0.001$), whereas Static and Trial-wise AF did not differ in expected score (difference -0.28, $p = 0.40$). The repeated A phase showed backward transfer +0.05 and stability–plasticity improvement +0.10 (both $p < 0.001$).

The same intervention did not establish uniqueness. An unchosen-only leaky-RL agent achieved expected score 207.27, compared with 206.54 for Static AF and 206.26 for Trial-wise AF; neither paired difference was reliable ($p = 0.223$ and $p = 0.084$). In empirical full-choice fits, an unchosen-value leak improved over vanilla Q-learning, but the fitted leak itself did not distinguish upper and lower performers (0.267 versus 0.327, $p = 0.524$). Value decay is therefore a strong alternative route to responsiveness, whereas AF-HGF provides the latent belief-and-precision account examined here. Supplementary Section C.6 tests how this benefit trades off against retention when the intervening context lengthens.

C.4 VOLATILITY NON-REDUNDANCY

Because the HGF already estimates volatility, we test whether tonic volatility absorbs the fitted carryover effect; non-redundancy is necessary for persistence control to add a distinct computational location. The HGF tonic volatility parameter ω_2 enters the level-2 precision prediction, whereas β_t changes the prior state before that prediction. Across participants, ω_2 was unrelated to the baseline release logit ($r = -0.043$, $p = 0.715$), mean β_t ($r = 0.084$, $p = 0.475$), and AF model benefit ($r = -0.140$, $p = 0.233$). With ω_2 free in both models, moving from static coupling $b_v = 0$ to the fitted trial-wise coupling improved cohort LME by 131.78 and produced positive participant-level gains in 67/74 participants. The ordinary volatility parameter did not absorb the carryover effect.

C.5 MODEL INVENTORY

Table S1 lists every model compared in the paper with the operations and parameters named in the text; Table S2 summarizes where each is tested and how it fared.

Table S1: Model inventory.

Model	**Pre-update operation**	**Evidence update**	**Parameters named in the paper**	**Nesting**	**Reported in**
Standard HGF	None (complete carryover, $\beta_t = 0$)	Standard HGF prediction and update; tonic volatility ω_2 enters the level-2 precision prediction	ω_2; choice parameter	Special case of Static AF-HGF ($\beta = 0$)	Main 3.1, 3.3; B.1, C.3, C.4, C.6, E.1
Static AF-HGF	Constant release β toward the initial prior at levels 2 and 3	Unchanged HGF update	β (participant-specific); ω_2; choice parameter	Special case of Trial-wise AF-HGF ($b_v = 0$)	Main 3.1, 3.3; B.2, C.3, E.1
Trial-wise AF-HGF	Surprise-dependent release $\beta_t = \sigma[b_0 + b_v(S_{t-1} - \log 2)]$	Unchanged HGF update	b_0; b_v; ω_2; choice parameter	Contains Static AF-HGF and Standard HGF	Main 3.1, 3.3; B.1–B.2, C.1–C.4, C.6, E.1
Surprise-Modulated Volatility HGF	None	Same lagged surprise applied to volatility	Surprise slope (generating -2; fitting support $[-2.5, 1]$)	Matched to Trial-wise AF-HGF except where surprise acts	Main 3.1; C.1
Bayesian Online Change Detection HGF	All-or-none reset when summed change probability exceeds a threshold; 20-trial cooldown	Standard HGF update between resets	Participant-specific threshold	—	Main 3.3; B.3
Complete carryover (bandit)	None ($c_t = 1$)	Delta-rule update with learning rate α	α	Special case of fixed carryover (c = 1)	Main 2.2, 3.2
Fixed carryover (bandit)	Constant carryover c toward a neutral prior	Delta-rule update	c; α (terrain: $\alpha = 0.20$)	—	Main 2.2, 3.2; D.2–D.3
Evidence-adaptive carryover (bandit)	c_t set from lagged evidence features	Delta-rule update	Controller settings, frozen after selection	—	Main 2.2, 3.2; D.1

Table S2: Rival models and where each is tested.

Model	Where adaptation acts	Alternative it represents	Outcome (section)
Standard HGF	Evidence gain through learned volatility; complete carryover	Volatility tracking suffices	Loses update leverage in continuous histories (3.1); lower model evidence than Static AF-HGF (3.3)
Static AF-HGF	Constant pre-update release	Uniform relaxation suffices	Did not differ from Trial-wise AF in A/B/A/B expected score (3.1); beats Standard HGF and the changepoint model on model evidence (3.3)
Trial-wise AF-HGF	Surprise-dependent pre-update release	Proposed mechanism	Favored over the volatility alternative (3.1); higher model evidence than Static AF-HGF (Supplementary C.4); protect–release (3.3); finite retention horizon over long intervening contexts (3.1; Supplementary C.6)
Surprise-Modulated Volatility HGF	Same surprise signal, applied to volatility	Surprise acts through gain, not carryover	Separated in matched recovery; disfavored empirically (3.1)
Bayesian Online Change Detection HGF	All-or-none reset at detected boundaries	Discrete boundary resets	Rarely resets; lower model evidence than Static AF-HGF (3.3)
Adaptive learning rate (bandit)	Post-outcome learning rate	Faster updating suffices	Evidence-adaptive carryover exceeded it in 21 of 27 cells at duration 40 but in only 5 and 2 of 27 at durations 80 and 150 (3.2)
Complete carryover	No release	Preservation is always best	Beaten by discovery-selected fixed carryover at duration 40 and noise ≤ 0.17 on held-out seeds (3.2)
Unchosen-only leaky RL	Decay of unchosen values	Value decay alone restores responsiveness	Matches AF in the A/B/A/B test, so AF-HGF is not algorithmically unique (3.1)

Model	Pre-update operation	Evidence update	Parameters named in the paper	Nesting	Reported in
Adaptive learning rate (bandit)	None	α_t set from lagged evidence features	Controller settings, frozen after selection	—	Main 2.2, 3.2
Joint carryover–learning-rate control (bandit)	c_t set from lagged evidence features	α_t set from lagged evidence features	Controller settings, frozen after selection	—	Main 2.2
Unchosen-only leaky RL	Decay of unchosen action values	Delta-rule update of the chosen value	Leak rate	—	Main 3.1; C.3

C.6 Recurrent-context retention

We test whether the per-trial release that keeps Trial-wise AF-HGF responsive also limits how long a state learned before an intervening context remains useful. Simulated agents experienced A1 $\rightarrow$ B $\rightarrow$ A2. Environment A (reward probabilities 0.80, 0.15, 0.05; reversals every 25 trials) lasted 100 trials in each visit, and A2 repeated the A1 probability trajectory with fresh reward draws. Environment B (0.60, 0.25, 0.15; reversals every 10 trials) lasted 50, 150, 200, 250, 500, or 1,000 trials. Across B durations, A1 and A2 were identical, shorter B schedules were nested prefixes of the longest schedule, and environment and choice-randomness streams were paired, over 100 paired seeds. Standard HGF ($\omega_2 = -4$, $\log \theta = -6$, choice inverse temperature 10) and Trial-wise AF-HGF ($\omega_2 = -2$, $\log \theta = -6$, inverse temperature 10, $b_0 = -3$, $b_v = -3$) used fixed settings that differ from the agents in main-text Section 3.1; because the two models also differ in tonic volatility, the primary contrast is within Trial-wise AF-HGF, relative to B = 50.

The primary endpoint was early A2 accuracy, the proportion of best-arm choices in the first 20 A2 trials. Recurrent-A savings compared early A2 expected reward with a cold start from the prior on the identical A2 schedule; for Trial-wise AF-HGF, savings at B = 50 were already near zero (-0.0023), so we report relative losses rather than the loss of a large savings effect. Secondary outcomes were cumulative expected and realized reward over the fixed 100-trial A2 block. Paired 95% intervals used 20,000 bootstrap draws, and the planned pooled contrast averaged B = 150–1,000 within each seed before comparing that average with B = 50.

Longer B phases produced a small but reliable average cost. Pooled across B = 150–1,000, early A2 accuracy fell by 2.78 percentage points (95% CI [-5.07, -0.45]), savings by 0.0194 expected reward per trial ([-0.0348, -0.0040]), cumulative A2 expected reward by 0.360 ([-0.703, -0.007]), and realized A2 reward by 0.384 ([-0.760, -0.014]). Individual durations were not monotonic: early-accuracy losses were reliable at B = 150, 250, and 500 but not at B = 200 or 1,000 (Table S3), consistent with B's ten-trial reversals leaving an endpoint-dependent state when A returned.

The mean release rate during B stayed near 0.072 at every duration (0.0731 at B = 50; 0.0716 at B = 1,000), so the nominal direct carryover of the pre-B state, the product of $(1 - \beta_t)$ over B trials, fell from a base-10 logarithm of -1.69 at B = 50 to -33.12 at B = 1,000 (Figure S3C–D). This product describes compounded prior-directed shrinkage, not a probability of remembering A, and learning continues throughout B. At the end of B, Trial-wise AF-HGF kept level-2 uncertainty ψ_2 near 3.1 and volatility belief μ_3 near 0.96–0.97 at every duration, whereas Standard HGF held lower uncertainty (1.3–1.5) and a volatility belief that declined from 0.987 at B = 50 to 0.837 at B = 1,000 (Figure S3A–B). The regulation that prevents over-commitment therefore also attenuates the contribution of the state learned before B.

Trial-wise AF-HGF retained higher cumulative A2 expected reward than Standard HGF at every duration (+10.07 to +13.45; all intervals excluding zero), although this between-model gap also reflects the differing ω_2. The result identifies a flexibility–retention trade-off with a finite effective carryover horizon within Trial-wise AF-HGF; it does not show monotonic decline, erasure of A, performance below Standard HGF, or an explicit or biological memory process.

D Environmental boundary and held-out generalization

D.1 Discovery-confirmation separation

We separate discovery from confirmation to test whether the environmental boundary survives after controller selection rather than reflecting reuse of optimization seeds. Frozen controllers were evaluated on an 81-cell grid crossing context duration (40, 80, or 150 trials), drift speed (0, 0.004, or 0.012), observation noise (0, 0.25, or 0.50), and reversal magnitude (0.30, 0.55, or 0.75). Each cell used 12 fresh seeds, disjoint from the training and validation seeds that selected the controllers. The main paper reports the resulting frequency interaction and the fixed-carryover terrain. The supplementary question is stricter: does a discovery-selected release policy survive when its parameters are frozen and evaluated on new seeds?

Table S3: Trial-wise AF-HGF changes relative to B = 50 (paired means with 95% bootstrap intervals; 100 seeds).

B duration	Early A2 accuracy (pp)	Recurrent-A savings	A2 cumulative expected reward	Realized A2 reward
150	−3.80 [−7.00, −0.65]	−0.0264 [−0.0481, −0.0051]	−0.484 [−0.963, −0.024]	−0.55 [−1.05, −0.06]
200	−1.00 [−3.55, 1.55]	−0.0071 [−0.0246, 0.0108]	−0.163 [−0.578, 0.235]	−0.25 [−0.71, 0.20]
250	−3.30 [−6.35, −0.25]	−0.0230 [−0.0437, −0.0021]	−0.454 [−0.903, 0.013]	−0.53 [−1.02, −0.01]
500	−3.45 [−6.50, −0.45]	−0.0238 [−0.0443, −0.0034]	−0.429 [−0.856, −0.0065]	−0.36 [−0.83, 0.09]
1,000	−2.35 [−5.20, 0.50]	−0.0167 [−0.0366, 0.0031]	−0.272 [−0.734, 0.206]	−0.23 [−0.76, 0.32]
Pooled 150–1,000	−2.78 [−5.07, −0.45]	−0.0194 [−0.0348, −0.0040]	−0.360 [−0.703, −0.007]	−0.384 [−0.760, −0.014]

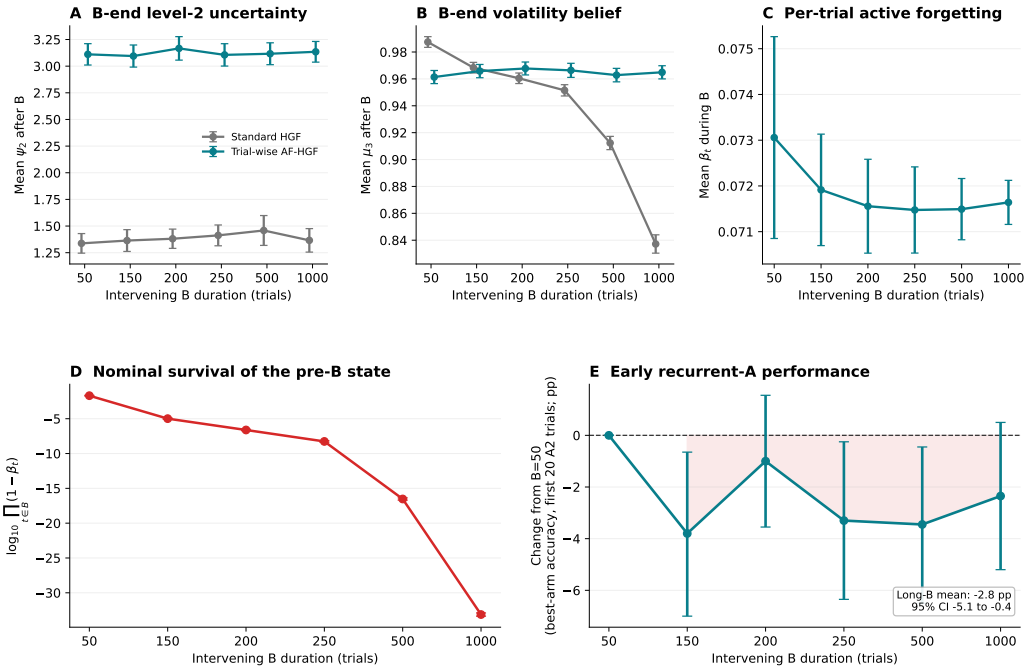

Figure S3: Recurrent-context retention in the A1 → B → A2 simulation. (A–B) End-of-B level-2 uncertainty and volatility belief for both models. (C) Mean per-trial release β_t during B. (D) Nominal compounded carryover of the pre-B state (base-10 logarithm of the product of $1 - \beta_t$ over B). (E) Change in early A2 accuracy relative to B = 50 (paired means with 95% intervals; 100 seeds).

D.2 What the fixed-carryover search estimates

We define what the terrain optimizes so its height is interpreted as environment-specific reward-maximizing carryover, not as a fitted human trait or universal decay constant. To isolate persistence from learning-rate adaptation, the update rate was fixed at $\alpha = 0.20$ and 41 carryover values from 92% to 100% were evaluated within each discovery environment. For candidate set C, c*(e) = arg max c ∈ C E[$\Sigma_t r_t$ | e, $\pi(\alpha$,c)]. The surface in the main paper is therefore a score-maximizing grid estimate, not an inferred psychological quantity or universal decay constant. Selection and confirmation were separated so the held-out map below tests the chosen setting rather than reusing the trajectories that selected it.

D.3 Held-out confirmation

We freeze discovery-selected carryover and evaluate new seeds to test whether the frequent-change advantage generalizes beyond the trajectories that selected it. Each cell's discovery-selected carryover was frozen and evaluated on 24 non-overlapping seeds. At duration 40, gains over complete carryover generalized at noise 0 (+3.90 expected score, 95% CI [1.86, 5.89]), 0.08 (+3.74 [1.76, 5.65]), and 0.17 (+2.00 [0.28, 3.77]). The advantage weakened with observation noise and disappeared as contexts lengthened; no evaluated cell had a 95% interval reliably below zero. Held-out evaluation therefore confirmed a frequent-change, relatively reliable release region while preserving complete carryover across much of the landscape.

E Human behavioral selectivity and robustness

E.1 Static persistence and the exploitation phenotype

We test whether lower fitted persistence denotes indiscriminate instability or a coordinated phenotype combining stable exploitation with faster belief recovery. Static AF-HGF estimated modest release (mean $\beta = 0.121$, SD = 0.051, range 0.035–0.316), equivalent to mean persistence P = $1 - \beta$ of 87.9% per trial (SD = 5.1%; range 68.4%–96.5%). These percentages describe retention of the fitted HGF belief-and-precision state and are not numerically comparable with value carryover in the bandit.

The lower-persistence extreme tertile earned more reward than the higher-persistence extreme tertile (96.0 versus 89.6; $n = 25$ per group; $p = 0.023$, d = 0.67) while switching less (38.4% versus 54.0%; $p < 0.001$, d = -1.26). It also showed higher win–stay (89.9% versus 76.5%; $p = 0.006$, d = 0.84) and lower lose–shift (65.1% versus 80.0%; $p < 0.001$, d = -1.00). Lower fitted persistence therefore coexisted with more stable exploitation rather than random behavioral instability.

Fitted beliefs supplied the complementary flexibility measure. Across the common 0–9-trial post-reversal window, Standard HGF mean recovery slope was 0.0077. Static AF increased it by 0.0090 (95% CI [0.0066, 0.0115], $p < 0.001$), and Trial-wise AF increased it by 0.0116 ([0.0093, 0.0141], $p < 0.001$). Trial-wise AF also produced a larger early-to-late gain (+0.0657, [0.0509, 0.0814], $p < 0.001$) despite a slightly lower mean post-reversal best-option belief. The model begins more conservatively and climbs more steeply.

E.2 Primary protect–release result: supplementary scope

We delimit the primary protect–release claim to determine which event sequence and data stream support it, because the null choice-stream effect constrains its generality. Figure 4 in the main paper reports the primary selective sequence, so it is not reproduced here. The event definition required an objectively best choice followed by misleading non-reward; accumulated contradiction required a second consecutive oddball. The reward-stream pattern distinguished protection after the first violation from additional release after repetition. The corresponding choice-stream contrast was null (difference -0.00643, $q = 0.135$), defining an observation-stream boundary rather than a general tendency to repeat choices.

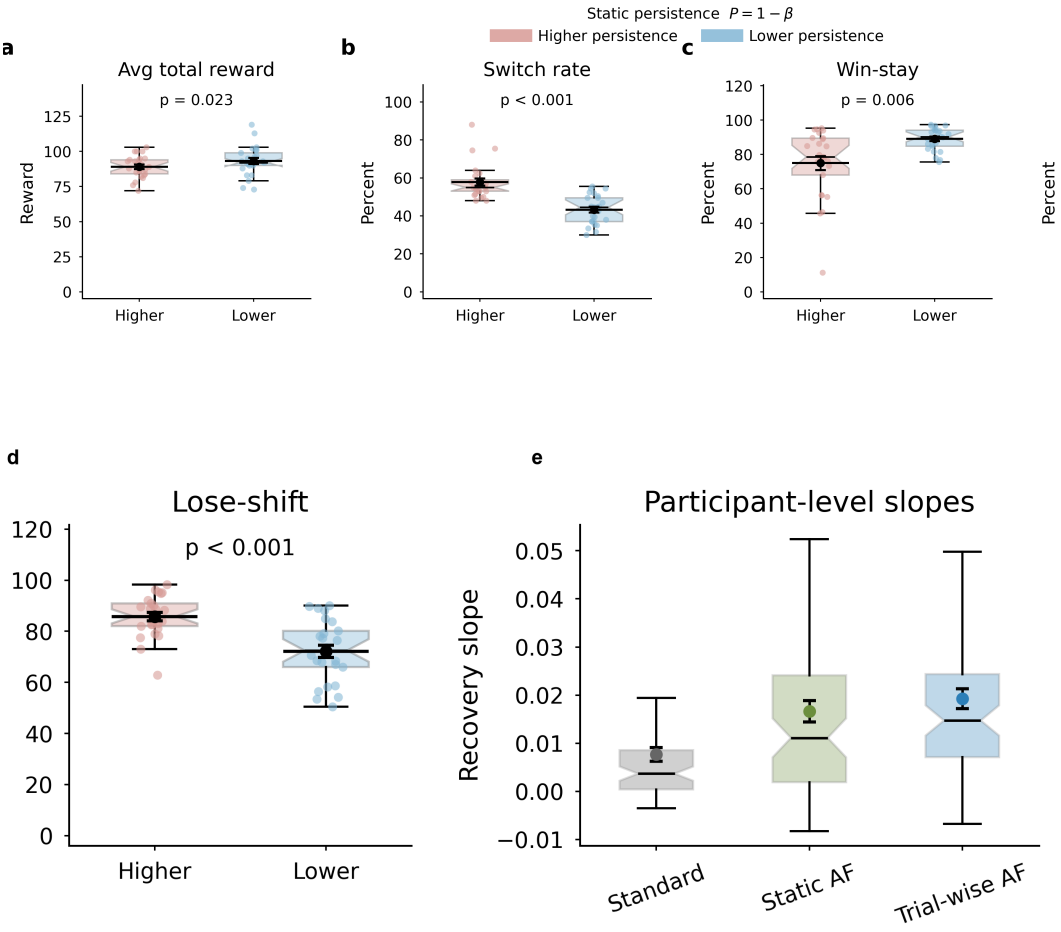

Figure S4: Static persistence phenotype and fitted belief recovery. Extreme persistence groups differ in reward, switching, win–stay, and lose–shift, while participant-level recovery slopes compare Standard, Static-AF, and Trial-wise-AF HGF. The grouping concerns fitted static persistence and differs from the performance quartiles used in the selective event analyses.

Table S4: Secondary behavioral statistics for main-text Section 3.3.

Result	**Statistic**
Protection vs. continual-learning index	$\rho = 0.578$, $q < 0.001$
Release after a second consecutive oddball: adjusted change	-0.0967
Release after a second consecutive oddball: sign-flip test	$q = 0.00333$, $d_z = -0.529$

E.3 Reversal-specific deployment and performance

We ask whether successful adaptation depends on globally lower persistence or on selectively concentrating reconfiguration around reversals. Upper and lower reward quartiles differed little in mean persistence (difference 0.0228, $p = 0.704$) or overall persistence-change variability (difference 0.0207, $p = 0.501$), but their reversal surges were 0.1667 and 0.0653 (difference 0.1014, Hedges $g = 1.222$, $p < 0.001$). Across 73 participants, surge magnitude correlated with reward ($\rho = 0.432$, bootstrap 95% CI [0.236, 0.596], $p < 0.001$) and continual-learning performance ($\rho = 0.400$, $q = 0.00160$). Reversal reconfiguration added 0.123 explained variance beyond the adjusted base-

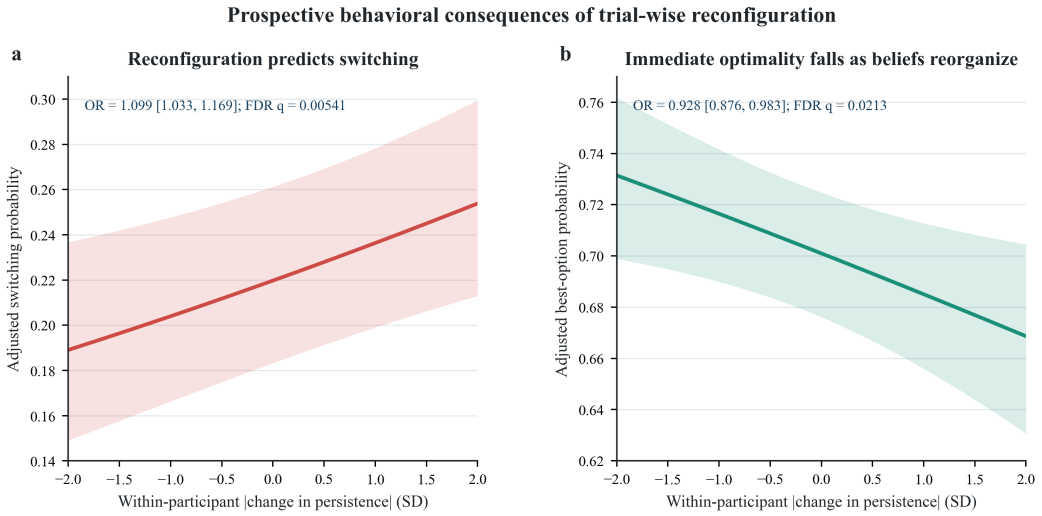

Figure S5: Prospective behavioral consequence. Marginal predictions from mixed models relate within-participant persistence reconfiguration to the next choice; these panels are not repeated from the main figure.

line ($p = 0.00120$). Successful learners differed primarily in where persistence changed, not in its global mean.

E.4 Prospective behavioral consequence and prediction boundary

We test whether persistence reconfiguration predicts the participant's next choice and whether that association transfers to held-out participants, distinguishing explanatory association from predictive utility. The prospective outcome was defined on the choice after a completed transition, so it was not the same choice used to classify that transition. Logistic GEE adjusted for participant-mean reconfiguration, transition type, condition, trial position, block order, and log reaction time. Across 5,811 choice-stream events, a one-SD within-person increase in absolute persistence change predicted more switching (OR = 1.099, 95% CI [1.033, 1.169], $q = 0.00541$) and lower immediate best-option choice (OR = 0.928, [0.876, 0.983], $q = 0.02132$). Signed direction did not predict switching, and the magnitude effect did not differ reliably across transition types (Wald $\chi^2 = 5.370$, df = 5, $p = 0.373$).

The association did not yield participant-held-out predictive improvement. Sequence-only and sequence-plus-magnitude models had log losses 0.41301 and 0.41294, and the participant-level gain interval included zero (sign-flip $p = 0.415$). Persistence change therefore contains within-sample trial-level behavioral information beyond sequence identity, but it is not currently a useful out-of-participant predictor.

F Neural analyses and specificity boundaries

F.1 Participant heterogeneity and sensor-space context

We examine participant dispersion and scalp distribution to test whether the group effect is broadly expressed and plausibly frontocentral, without converting descriptive localization into inference. The main figure already reports the continuous protection association, feedback-locked time course, and time–frequency localization. The supplementary panels answer two different questions. First, the participant distribution tests whether the reward-group contrast is carried by a few observations. Second, the scalp map asks whether the locked effect has plausible frontocentral support. Upper- versus lower-reward quartiles differed by 1.2678 dB in total theta (Hedges $g = 1.104$, $q = 0.00230$) and by 1.1642 dB in induced theta (Hedges $g = 1.233$, $q = 0.00100$; Table S5). In the main figure, the upper-reward trajectory separates late in the prespecified 4–8 Hz, 200–400 ms window (Figure

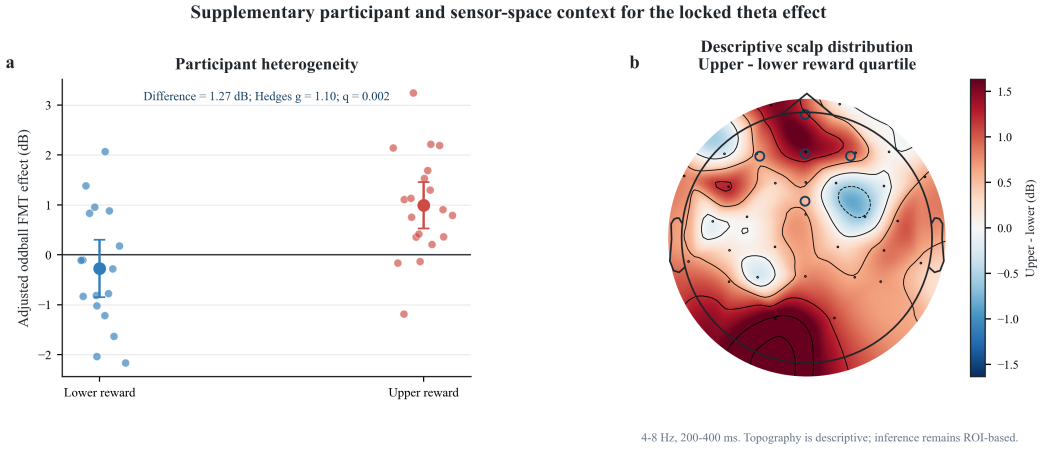

Figure S6: Participant and sensor-space context. Individual theta effects by reward quartile and the descriptive upper-minus-lower scalp distribution in the locked 4–8 Hz, 200–400 ms window; outlined sensors mark the prespecified ROI.

Table S5: Secondary neural statistics for Section 3.4.

Result	**Statistic**
Oddball minus rewarded-best, induced theta ($N = 72$)	+0.5469 dB, 95% CI [0.3200, 0.7739], $d_z = 0.5566$, $q < 0.001$
Total theta vs. reward	$\rho = 0.3461$, $q = 0.00810$
Total theta vs. continual-learning index	$\rho = 0.3357$, $q = 0.00810$
Upper vs. lower reward quartile, induced theta	1.1642 dB, Hedges $g = 1.233$, $q = 0.00100$

5b), and matched time–frequency maps localize the descriptive difference to theta (Figure 5c). The topography is descriptive; inference remains tied to the prespecified frontocentral ROI.

F.2 Event-level temporal ordering

We test temporal ordering to determine whether feedback theta precedes the next-trial persistence adjustment; precedence strengthens convergence but cannot establish causality. The primary event model included 1,553 observations from 73 participants: 1,416 first oddballs and 137 second oddballs. First-oddball total theta predicted greater next-trial protection (slope 0.02288, 95% CI [0.01489, 0.03087], $p < 0.001$), as did induced theta (0.02075, [0.01254, 0.02896], $p < 0.001$); GEE estimates agreed closely. An exact isolated sequence also showed an induced-theta pulse of +1.5072 dB ($N = 49$, $q = 0.00200$). Feedback activity preceded the fitted next-trial change, but the analysis remains observational.

F.3 Whole-trial-set and nuisance-control analyses

We expand trial inclusion and nuisance adjustment to test whether the theta–persistence association survives broader sampling and correlated computational variables. A model using all eligible trials retained a positive association between persistence and total FMT ($N = 63$, standardized $\beta = 0.03422$, $p = 0.01990$, $q = 0.03570$), controlling update quantities, reward, reaction time, stochasticity, volatility, and feedback delay. More extensive computational adjustment attenuated the first-oddball slopes. Stricter residualized tests and exact-conjunction tests were null. These results support convergence between theta and selected expressions of protection, not a unique neural signature of the entire protect-to-release sequence.

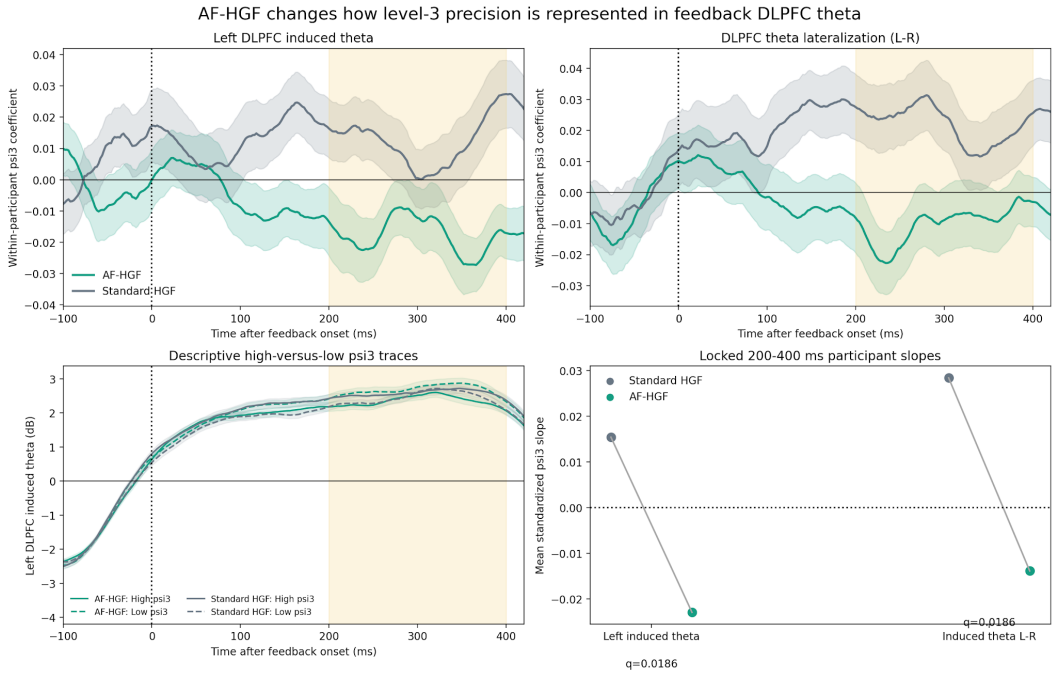

Figure S7: Model-dependent source correspondence. Feedback-window coefficient trajectories and participant summaries compare the relationship of the same level-3 precision quantity with DLPFC theta under AF-HGF and Standard HGF. The result is a model-dependent correspondence under coarse source localization, not localization of a forgetting mechanism.

F.4 Model-dependent source correspondence

We compare source-space correspondence under AF-HGF and Standard HGF to ask whether tempering changes the neural relation of the same hierarchical quantity while limiting anatomical interpretation. In 56 source-ready participants, the relationship between the level-3 precision quantity ψ_3 and DLPFC theta differed between AF-HGF and Standard HGF. AF-minus-Standard slopes were -0.03841 for left-DLPFC induced theta, -0.04232 for induced left-minus-right theta, and -0.03986 for total left-minus-right theta (all $q = 0.01860$). Because ψ_3 exists in both models, the comparison asks whether tempering changes the neural correspondence of the same hierarchical quantity.

Source estimates used 32-channel EEG, template anatomy, and eLORETA. They are spatially coarse and model dependent. A sign change in source-space regression does not identify an anatomical persistence gate, establish causal inhibition, or show that DLPFC implements β_t.