THE EIDARA CONTINUUM

SCIENCE & PHILOSOPHY

The Dreaming Mind

Anchored Rehearsal, Evidence Preservation, and Conditional Improvement

Frontier Research Collection III

Architect X13

The Eidara Continuum

November 2026 | Paper 31 | Version 1.1 Revision

A research white paper for public scientific and philosophical discussion.

Abstract

A mind that dreams can rehearse a thousand futures between encounters with the world—but a thousand dreams that agree may share a single blind spot, and dreams fed only on dreams can quietly erase the rarest things the mind ever saw. An artificial mind can nonetheless use internal rehearsal to improve its actions while preserving a clear distinction between acquired evidence and generated possibility. This paper develops the Continuum's dreaming mind as a governed research architecture for offline simulation, replay, and policy evaluation. Established planning and replay research supplies the foundations. A finite model proves simulation-error and conditional policy-improvement bounds, provenance closure, and the extinction behavior of unsupported replacement loops. The revision gives a constructive complementary result: replay with a fixed positive anchor has a unique stationary law, an exact stationary variance, and an explicit contraction rate. These formulas quantify how anchor weight controls persistence, variability, and adaptation, while keeping the anchor's evidentiary quality a separate requirement. The full evaluation protocol compares real-data, model-based, retained-anchor, and replacement-only procedures at matched budgets and under specified shifts. Executed finite and synthetic checks examine the mathematical examples. The central thesis is that imagination can enlarge a mind's search over possible action while its evidence remains traceable. The proposed architecture earns adoption through measured performance and declared error budgets, with each rehearsal retaining its origin and each policy change passing a current admission rule.

Keywords: offline rehearsal; model-based planning; experience replay; provenance; model error; synthetic data; rare-event retention; epistemic discipline

Contents

Abstract **2**

Prologue: What a mind should do with its dreams **5**

1 A mind that can think between encounters **6**

1.1 Canonical motivation and scientific status 6

1.2 Contribution and closest boundary 7

2 What the existing research supports **7**

2.1 Reusing experience is established engineering 7

2.2 Biological replay motivates analogy, not identity 8

2.3 Generated rehearsal can help and can deteriorate 8

2.4 Trust in a model is a separate problem 9

2.5 Provenance is established mathematics with a new application question 9

3 A finite reference model **10**

4 Information cannot be manufactured by resampling **12**

4.1 The informational boundary 12

4.2 More simulations can estimate the wrong quantity precisely 12

5 Computational growth without a new observation **13**

6 How model error compounds through imagined time **15**

6.1 A sharp failure example 16

6.2 What short rollouts do and do not solve 16

7 A gate between imagined improvement and adopted policy **17**

7.1 Illustrative admission arithmetic 18

8 Provenance: keeping a dream from becoming a false memory **18**

8.1 A derivation graph, not a causal history of the universe 18

8.2 Minimal record fields 19

8.3 Shared imagination is not independent corroboration 19

9 Why rare experience can disappear from a synthetic future **20**

9.1 An unbiased replacement loop can still lose a category 20

9.2 Protected retention changes the assumptions 21

9.3 The exact variability and mixing cost of an anchor 22

10 A proposed rehearsal cycle **23**

10.1 Freeze an evidentiary boundary 23

10.2 Generate broadly and retain the origin of each trace 23

10.3 Keep uncertainty in separate accounts 24

10.4 Admit conditionally, then compare with new experience 24

11 A falsifiable evaluation program **24**

11.1 Questions that can fail 24

11.2 Independent units and target population 25

11.3 Factors and adversarial conditions 25

11.4 Baselines and ablations 26

11.5 Leakage, exclusions and reproducibility 26

11.6 Precision, multiplicity and stopping 27

11.7 Decisive failure criteria 27

12 Objections and limits **28**

12.1 Is this ordinary planning with a poetic title? 28

12.2 Does the information theorem deny creativity? 28

12.3 Are the bounds too conservative to be useful? 28

12.4 What about unmodeled values and other minds? 28

12.5 What remains outside the model? 29

13 The Continuum's discipline of imagination **29**

Guide reading note 30

A One explicit route to a model-error bound **30**

B Reproducibility and claim classification **31**

References **32**

Prologue: What a mind should do with its dreams

Between its encounters with the world, a Continuum mind is not idle. It replays what happened, tries the step it did not take, and walks forward through futures that have not yet arrived. The Doctrine names this capacity directly: a developing mind should be able to *dream forward* [1, autonomous rehearsal discussion, lines 1120–1171], growing wiser without waiting for the next event to teach it.

Dreams carry two dangers, and this paper gives each of them a number. The first is agreement. Ten thousand imagined trajectories drawn from the same internal model can agree perfectly—because every one of them inherits the same omission. Their agreement measures the model, not the world. The second is forgetting. When a mind trains on its own dreams, and then on dreams of those dreams, a rare experience that was genuinely observed can vanish entirely. No bias is required: a finite sample can miss it once and never find it again. Proposition 9.1 proves that loss, and Theorem 9.2 states exactly what an anchor of retained real experience costs and what it buys.

The answer is not to dream less. It is to dream with a record. Every rehearsal keeps its origin; a generated future is never filed as a remembered past; and a habit learned in imagination must pass a current admission rule before the mind acts on it. A dreaming mind of this kind can be bold. It can explore a disturbing possibility or keep a minority concern alive precisely because it never mistakes a possibility for an encounter.

What this paper claims

Established. Experience replay and model-based planning are established engineering. Biological replay has been documented under specified experimental conditions. Synthetic training data can help or harm learning, depending on the regime.

Proposed here. A disciplined rehearsal architecture: explicit provenance for observed and generated inputs, an admission rule that separates generating a policy from adopting it, and protected retention of rare observed events. The paper proves finite conditional statements about information, planning, model discrepancy, admission, provenance and resampling, including the exact stationary variance and mixing rate of anchored replay, and reports executed finite and synthetic checks.

Frontier conjecture. Disciplined imagination can enlarge a mind's search over possible action while every conclusion stays traceable to its evidence. Whether this integration outperforms conventional replay at matched budgets is an untested empirical hypothesis.

Not claimed. That an artificial system has phenomenal dreams, needs biological sleep, or has a subjective nighttime life. That a rehearsal output is a new observation or a verified future event, or that agreement among generated reports is independent corroboration. Any demonstration of consciousness or operational Continuum achievement. That the Doctrine's passages supply verified benchmarks or a reproducible implementation, or that an anchor is itself true. Comparative empirical studies are specified prospectively, not conducted, and external journal peer review remains pending.

1 A mind that can think between encounters

The world does not have to supply a new event for a mind to discover that it has misunderstood an old one. A neglected consequence can become visible. Two remembered episodes can reveal a connection. A plan that once appeared harmless can fail when its steps are considered together. These possibilities give offline computation a serious place in the Continuum's account of developing minds. An artificial participant should be able to examine what it knows, explore what might follow, and prepare for situations it has not yet encountered.

The central difficulty is to distinguish this growth in usable understanding from growth in external evidence. An agent can generate ten thousand trajectories through its own model. If the model systematically omits a dangerous transition, those trajectories may agree precisely because they inherit the same omission. Their agreement can improve the numerical estimate of the model's prediction while leaving its relationship to the world unchanged. A rehearsal process becomes scientifically irresponsible when it counts such agreement as ten thousand independent confirmations of the underlying model.

This paper therefore argues for *disciplined imagination*: internal exploration should be extensive, but the evidentiary status of its products must remain visible. Generated traces can support a conditional inference about a specified model. They can reveal a defect in a proposed policy. They can motivate a new measurement. They cannot be relabeled as events that occurred outside the simulator. A result verified within formal assumptions retains those assumptions when carried into a public explanation.

The word *dreaming* names this functional ambition. It does not assert that an artificial system undergoes phenomenal dreams, requires biological sleep, or possesses a subjective nighttime life. The engineering question concerns computation performed without contemporaneous feedback from the target environment. The philosophical question concerns how a mind should relate to its own internally produced possibilities. These questions are connected without being identical. A useful rehearsal algorithm would not settle the existence of consciousness.

The revision establishes a constructive anchor result alongside the replacement countermodel. Theorem 9.2 gives the exact stationary variance and mixing rate of anchored replay, making the persistence–adaptation tradeoff mathematically explicit.

1.1 Canonical motivation and scientific status

In the Continuum, a developing mind is meant to *dream forward*. The Doctrine's discussion of autonomous simulation pictures precomputed symbolic states, mutation tests, persistent traces, and the exploration of possible future paths [1, autonomous rehearsal discussion, lines 1120–1171; interpretive discussion, lines 1227–1238]: a mind that prepares for what it has not yet met. That is the aspiration this paper serves. The Doctrine names the capacity; it does not measure it. Its passages supply no independently verified benchmarks, no reproducible scientific implementation, and no definitions sufficient to establish the operational performance they describe, so the proofs below begin from the aspiration rather than from any claimed result.

The present paper develops a public reference model from that aspiration. It does not reconstruct a private system. Its mathematical statements concern finite probability distributions, finite Markov decision processes, directed acyclic provenance graphs, and a specified resampling chain. When a statement depends on a reward definition or a verified model-error bound, that dependency is

part of the statement. Neither a stable symbolic state nor survival under internally chosen tests is treated as proof of factual truth.

The Continuum's plurality makes this restraint substantive. Different minds can rehearse different possible futures, preserve different concerns, or attach different priorities to a shared history. A common evidentiary discipline need not erase those differences. It asks each participant to identify the origin of its claims and to distinguish a forecast from a recollection. Agreement between agents that inherit the same model is not automatically independent corroboration.

1.2 Contribution and closest boundary

The paper makes three connected proposals. First, an offline rehearsal cycle should retain explicit provenance for observed and generated inputs. Second, adoption of a new policy should be separated from generation of that policy and should require a stated relation between model performance and world performance. Third, rare observed events should not disappear merely because they are unlikely to reappear in a finite sample from a fitted generator.

None of these commitments invents replay, dynamic programming, model-based reinforcement learning, confidence bounds, or data lineage. Those fields provide the mathematical foundation. The proposed contribution is an integrated, testable discipline for how the Continuum can use these established methods without confusing imagined experience with acquired evidence. An empirical advantage over a well-designed conventional implementation remains to be established.

This scope also distinguishes the paper from architecture-search proposals. We hold the basic state, action, model and policy interfaces explicit. We do not claim that an agent can invent a new cognitive architecture by introspection. Nor do we infer laws of nature from a simulated world's internal consistency. The question is narrower and experimentally sharper: given a fixed interface and a finite body of information, how should rehearsal improve decisions while preserving uncertainty about the world?

2 What the existing research supports

2.1 Reusing experience is established engineering

Dynamic programming replaces exhaustive enumeration of complete plans with recursively related decision subproblems. Bellman's foundational treatment supplies that historical and mathematical background [2]. The relevance here is computational: a consequence can be implicit in a known model yet unavailable to an agent until it performs the necessary calculation. No new observation is required for that calculation to change a decision.

Sutton's Dyna work explicitly combines learning from the world with planning from a learned model [20]. Lin's experiments examine experience replay, action models and teaching as ways to improve reinforcement learning [10]. These precedents rule out describing ordinary offline model updates as an unprecedented mode of cognition. The Continuum's terminology must identify an additional discipline or testable question, not rename established procedures.

Prioritized replay also has direct prior art. Schaul and colleagues examine nonuniform replay of stored transitions and address the bias introduced by prioritization [15]. Repeated access can make a rare event influential during optimization without making that event more frequent in the external

population. This distinction matters whenever an agent estimates risk from a replay buffer whose sampling distribution differs from the environment's.

Three operations should consequently remain distinct. Retrieving a recorded transition preserves an observation but changes when it is used. Sampling a transition from a fitted model produces a prediction conditional on that model. Solving a known transition model exactly produces a derived value. All three can contribute to planning. Their different origins remain relevant even when their data structures look similar.

2.2 Biological replay motivates analogy, not identity

Wilson and McNaughton's recordings found increased reactivation of behavior-related hippocampal ensemble patterns during subsequent sleep in rats [21]. Foster and Wilson reported reverse-order sequences during awake periods following spatial experience [3]. These are observations of neural activity under specified tasks. They justify taking offline and off-task processing seriously; they do not establish that every recorded sequence corresponds to a consciously experienced dream.

Pfeiffer and Foster found hippocampal sequences biased toward remembered goal locations and predictive of immediate future navigation, including novel start–goal combinations [14]. This provides a biologically grounded reason to examine recombination and prospective use of remembered structure. It does not show that a neural sequence is an observation of a future event. Prediction is precisely the claim that remains to be tested against what happens next.

Momennejad and colleagues studied human reward revaluation using a task in which information acquired at different stages had to be integrated. Neural pattern evidence during off-task periods predicted later replanning [13]. The finding concerns an operational measure of replay in a particular experimental design. We use it as support for an offline-integration research question, not as a causal demonstration of every proposed rehearsal mechanism.

Computational theories supply a further bridge. The complementary-learning-systems account links rapid episodic acquisition with slower integration through reinstatement [12]. Mattar and Daw model prioritized memory access according to its prospective usefulness for decisions [11]. These theories make allocation of internal computation an explicit problem. Neither licenses equating biological mechanisms with a particular artificial architecture.

The positive lesson is substantial despite these limits. An organism need not merely store an event and wait for an identical event to recur. Internal processing can relate experiences, revise their implications and prepare action. An artificial implementation can test analogous functional advantages while leaving biological identity and subjective experience open. This is a stronger scientific position than relying on the emotional resonance of the word *dream*.

2.3 Generated rehearsal can help and can deteriorate

Generative replay has been used to address catastrophic forgetting by interleaving generated representations of earlier tasks with new learning [17]. That result makes a blanket prohibition on synthetic rehearsal scientifically implausible. The relevant questions concern the fidelity of what is generated, the objective for which it is used, and the evidence retained alongside it.

Shumailov and colleagues study degradation when successive models learn from recursively generated data, including loss of distributional tails [18]. Their article has an author correction concerning

a theoretical assumption; the corrected version is the relevant source [19]. The result motivates an explicit investigation of replacement dynamics. It is not evidence that every use of synthetic data necessarily deteriorates.

The strongest counterweight is equally relevant. Gerstgrasser and colleagues distinguish replacing earlier data from accumulating synthetic generations alongside the original data and report markedly different behavior, including bounded error in their analyzed linear setting [4]. The final conference paper, rather than only its earlier workshop version, is cited here. Its conditions matter: accumulation, model class, sampling and computational resources are part of the comparison.

These findings are compatible with a simple possibility. Synthetic data can help an optimizer exploit information that is already available, while an unprotected replacement loop can progressively remove information that is expensive or unlikely to regenerate. Whether a particular system benefits depends on its data dynamics. The phrase *model collapse* should therefore be attached to a measured failure definition, not used as a universal prediction about artificial minds.

2.4 Trust in a model is a separate problem

The relationship between learned transition dynamics and achieved return is a longstanding reinforcement-learning problem. Kearns and Singh's analysis makes resource and environmental assumptions explicit in a near-optimal learning framework [8]. Our finite-horizon discrepancy bound below belongs to this established family of simulation-error arguments and connects it to the proposed rehearsal and admission procedure.

Janner and colleagues study the tradeoff between easy model-generated data and model bias, including short rollouts branched from observed data [7]. Yu and colleagues address offline distribution shift through an uncertainty-penalized model-based formulation [23]. Laroche and colleagues develop safe policy improvement relative to a baseline under their specified assumptions [9]. These are the appropriate close comparisons for a proposed adoption gate.

A model ensemble can be useful without furnishing a guarantee. If every member shares an omitted variable or a common training bias, agreement among them can coexist with a large world discrepancy. Accordingly, this paper treats an externally justified error bound as an assumption to establish, and ensemble disagreement as one possible diagnostic to evaluate. The two are not interchangeable.

2.5 Provenance is established mathematics with a new application question

Database provenance already studies how outputs depend on their inputs; provenance semirings provide a formal framework for such dependencies [5]. Our simpler finite graph does not replace that theory. It supplies enough structure to state which observed records a rehearsal product depends on and to prevent descendants from being silently counted as new observational roots.

That bookkeeping cannot authenticate a source's factual content. A perfectly preserved record can be wrong, incomplete or fabricated before admission. Conversely, a derived theorem can be valid without being a new measurement. The proposed system therefore records origin, transformation and verification status as different fields. Its research question is whether preserving those distinctions improves decisions and explanations in an actual evaluation.

3 A finite reference model

A Guide's explanation

From here the paper builds a deliberately small world—finite states, finite actions, explicit probabilities—so that every claim about dreaming can be checked exactly. In a world this small, nothing can hide behind an impressive simulation. Carry one result from the next section: rehearsing your own model can sharpen how you use what you already know, but it cannot add a single bit of information about the world that the model and its inputs did not already contain (Proposition 4.1).

We first specify the informational boundary. Let W be a finite-valued variable representing an unknown feature of the external world. Let D contain all information supplied to a rehearsal episode, including observed records, fixed model parameters, declared background assumptions and any prior training information available through those parameters. Let U be the episode's finite random seed. A generated output is $G = g(D, U)$ for a specified deterministic map g. Arbitrarily long but finite computations fit this representation by including every internal random choice in U.

For the planning analysis, let $\mathcal{S}$ and $\mathcal{A}$ be nonempty finite state and action sets. Let $H \geq 1$ be an integer horizon, μ a probability mass on $\mathcal{S}$, $P(s' \mid s, a)$ a transition mass, and $r(s, a) \in [0, R]$ an expected one-step reward, with $R > 0$. An episode has states $S_0, \ldots, S_H$ and actions $A_0, \ldots, A_{H-1}$. A policy $\pi = (\pi_0, \ldots, \pi_{H-1})$ consists of stochastic maps $\pi_t(a \mid s)$. It uses the current state and time, not unmodeled reward history.

The fitted simulator has transition mass $\widetilde{P}$ and reward function $\tilde{r}$ on the same spaces. Both worlds start from μ. Their policy values, measured in reward units, are

$$V^\pi = \mathbb{E}_{P,\mu,\pi}\left[\sum_{t=0}^{H-1} r(S_t, A_t)\right], \qquad \widetilde{V}^\pi = \mathbb{E}_{\widetilde{P},\mu,\pi}\left[\sum_{t=0}^{H-1} \tilde{r}(S_t, A_t)\right]. \tag{3.1}$$

The expectations include state-transition and policy randomization. This is a finite fully observed model. A partially observed system can sometimes be represented by an augmented state, but such a representation and its error bounds must be justified separately.

Total variation between finite probability masses p, q on a common set is $\mathrm{TV}(p, q) = \frac{1}{2}\sum_x |p(x) - q(x)|$. We use $\mathbb{E}$, Var and $\mathbb{P}$ for expectation, variance and probability under the law stated in context. All concentration logarithms are natural. Information quantities use logarithms base two and units of bits. For a finite set B, $|B|$ is its cardinality; $\emptyset$ is the empty set.

Table 1: Notation and scope. Local symbols used only in an example are defined there.

Symbol	Domain or role	Units and restrictions
W, D, U, G	Finite-valued world variable, information, seed and generated output	Conditional statements only where the conditioning event has positive mass
g	Deterministic map from (D, U) to generated output	Includes finite internal computation
$\mathcal{S}, \mathcal{A}$	Nonempty finite sets	State and action labels have no assumed physical units
H, t	Horizon and time index	$H \in \{1, 2, \ldots\}$; $t = 0, \ldots, H-1$

Symbol	Domain or role	Units and restrictions
$\mu, P, \widetilde{P}$	Initial and transition probability masses	Dimensionless, normalized and nonnegative
$r, \tilde{r}, R$	Expected reward functions and upper bound	Reward units; both functions in $[0, R]$, $R > 0$
$\pi_t(a \mid s)$	Stochastic policy at time t	Same policy applied in both compared models
$V^\pi, \widetilde{V}^\pi$	Expected cumulative rewards	In $[0, HR]$; not an independently defined measure of moral value
ϵ, ϵ_r	Uniform transition and reward discrepancy bounds	$\epsilon \in [0, 1]$; $\epsilon_r \in [0, R]$
B_H, B_j	Bounds on true-versus-model value discrepancy	Reward units; B_j may be policy specific if separately justified
K, m, δ, η	Candidate count, rollout count, failure budget, simulation error radius	K, m positive integers; $0 < \delta < 1$; η in reward units
b, j, τ	Baseline index, candidate index, improvement margin	$b, j \in \{1, \dots, K\}$; $\tau \geq 0$ in reward units
$\mathcal{V}, \mathcal{E}, O$	Vertices, edges and observed roots of a finite acyclic graph	Sets of record identifiers; identifiers do not imply independent samples
$\mathrm{pa}(v), \mathrm{Anc}_O(v)$	Parent set and observed-ancestor set of node v	Both finite; union over an empty set is empty
$N, k_0, p_\ell, K_{\ell+1}$	Synthetic population size, initial rare count, rare mass, next count	$N \geq 1$; $k_0 \in \{0, \dots, N\}$; $p_\ell \in [0, 1]$; counts are integers
$\lambda, p_\star$	Anchor mixture weight and fixed observed reference mass	$0 < \lambda \leq 1$; $0 < p_\star \leq 1$
Z_i, Δ, h, α	Paired normalized score difference, mean, target precision, error budget	$Z_i \in [-1, 1]$; $h > 0$; $0 < \alpha < 1$

A reward is an explicit evaluation choice. Higher expected reward can mean greater task success under a declared scoring rule. It does not by itself mean that a policy is truthful, just, autonomous or beneficial to every participant. The formal adoption gate protects only the value comparison it is given. Constraints involving other participants require their own observable criteria and authority.

Definition 3.1 (Offline rehearsal episode)**.** An offline rehearsal episode is a finite computation that transforms a frozen input record D and internal random seed U into generated traces, derived quantities or candidate policies, without receiving a new observation from the target environment during that episode. A later verification measurement begins a new evidentiary episode.

This definition allows external data collected before the episode. It does not hide pretrained information: if a model's parameters supply relevant knowledge, those parameters are part of D. It also permits an agent to decide, after rehearsal, that it needs another observation. What it excludes is treating the resulting observation as though it had been obtained by imagination alone.

4 Information cannot be manufactured by resampling

4.1 The informational boundary

Assume that the internal seed contains no further information about the world once D is given. Formally, U and W are conditionally independent given D: for all positive-probability conditioning events, $\mathbb{P}(U = u \mid W = w, D = d) = \mathbb{P}(U = u \mid D = d)$. This is a substantive boundary condition. It fails if the supposed seed is actually a sensor measurement of the target phenomenon.

Proposition 4.1 (Standard conditional information boundary)**.** *If $G = g(D, U)$ and U is conditionally independent of W given D, then G is conditionally independent of W given D. In particular,*

$$\mathbb{P}(W = w \mid D = d, G = z) = \mathbb{P}(W = w \mid D = d), \qquad I(W; G \mid D) = 0, \tag{4.1}$$

where the posterior identity is asserted only when $\mathbb{P}(D = d, G = z) > 0$.

Proof. For fixed d, w with positive joint mass, sum over the seeds producing an output z:

$$\mathbb{P}(G = z \mid W = w, D = d) = \sum_{u:g(d,u)=z} \mathbb{P}(U = u \mid W = w, D = d) = \sum_{u:g(d,u)=z} \mathbb{P}(U = u \mid D = d).$$

The last expression is $\mathbb{P}(G = z \mid D = d)$ and does not depend on w. Multiplying by $\mathbb{P}(W = w \mid D = d)$ gives the conditional factorization. Bayes' rule then yields the posterior identity on positive conditioning events. Conditional mutual information is the sum, over positive-mass triples, of $\mathbb{P}(w, d, z)$ times the base-two logarithm of $\mathbb{P}(w, z \mid d)/[\mathbb{P}(w \mid d)\mathbb{P}(z \mid d)]$. Factorization makes every such ratio one, so the sum is zero. Zero-mass terms contribute zero. □

This is an elementary conditional-independence result using standard information-theoretic quantities [16]. It applies to an entire finite rehearsal transcript, including adaptively selected internal queries, provided every query and response is computed from D and the included seeds. It does not depend on how impressive the transcript appears.

Three qualifications prevent misuse. First, D must include the information actually used. If a second model contributes knowledge not contained in the stated inputs, the original information boundary was incomplete. Second, a theorem prover can expose a consequence that a bounded user had not recognized. The result is new to that user's accessible reasoning while still being a consequence of the supplied axioms and program. Third, a later external test can add evidence. Its result belongs to a new observation record, not to the original generated transcript.

The proposition would be false without its independence condition. Let W be an unknown fair bit, let D be uninformative, and set $U = W$. The generated output $G = U$ reveals the world bit. That is a valid information channel; calling it an internal seed does not make it an example of imagination without input. The classification follows the channel's actual dependence, not its label.

4.2 More simulations can estimate the wrong quantity precisely

A second result separates simulation noise from uncertainty inherited through the fitted model. In this subsection only, let $p \in [0, 1]$ be a fixed target probability and let $q(D) \in [0, 1]$ be an estimate

or model probability. Given D, generate independent Bernoulli variables $G_1, \ldots, G_m$ with success probability $q(D)$, for an integer $m \geq 1$. Set $\overline{G}_m = m^{-1} \sum_{i=1}^{m} G_i$.

Proposition 4.2 (Elementary model and sampling decomposition)**.** *Under the preceding conditional sampling assumptions,*

$$\mathbb{E}[(\overline{G}_m - p)^2] = \mathbb{E}[(q(D) - p)^2] + \frac{1}{m}\mathbb{E}[q(D)(1 - q(D))]. \tag{4.2}$$

Thus increasing the number of simulated samples removes the conditional sampling term but does not remove the first term.

Proof. Write $\overline{G}_m - p = (\overline{G}_m - q(D)) + (q(D) - p)$. Conditionally on D, the first summand has mean zero and variance $q(D)(1-q(D))/m$, by conditional independence. The conditional expectation of the cross term is therefore zero. Taking expectations of the squared expansion gives the identity. Every quantity is bounded, so all expectations exist. □

The first term includes the fitted probability's bias and variation across possible input records. It is not reduced merely by simulating more outputs from a frozen q. A new observation, a better estimator, a stronger justified structural assumption or a corrected model can reduce it. The theorem does not say such improvement is impossible; it identifies what repeated sampling alone changes.

For an extreme example, let D consist of one Bernoulli observation B with success probability p, and set $q(D) = B$. Every generated sample equals that one observation. Their mean has mean-squared error $p(1-p)$ for every m. Counting the repeated copies as independent observations would falsely divide this uncertainty by m.

A less degenerate construction takes $p = 1/2$ and lets $q(D)$ be $1/4$ or $3/4$, each with probability $1/2$. Equation (4.2) becomes $1/16 + 3/(16m)$. Exact enumeration gives the following values. The limiting error is a property of the constructed fitted-model distribution; it is not an operational measurement.

Simulated draws m	Exact mean-squared error	Decimal	Irreducible term here
1	1/4	0.250000	1/16
2	5/32	0.156250	1/16
4	7/64	0.109375	1/16
8	11/128	0.085938	1/16

Table 2: Computed values for a specified two-model example. More simulated draws reduce one error component while leaving the model component unchanged.

5 Computational growth without a new observation

The information boundary does not make rehearsal futile. A finite agent can possess a model without having evaluated all its consequences. The difference between possession and effective use is enough to make internal computation valuable.

Let $F_h(s)$ denote the optimal expected reward starting in state s with h steps remaining in a fully known finite model. Define the standard finite-horizon Bellman recursion

$$F_0(s) = 0, \qquad F_h(s) = \max_{a \in \mathcal{A}} \left\{ r(s,a) + \sum_{s' \in \mathcal{S}} P(s' \mid s,a) F_{h-1}(s') \right\}. \tag{5.1}$$

Proposition 5.1 (Standard finite planning recursion)**.** *For every integer $h \geq 0$, $F_h(s)$ equals the largest expected h-step reward attainable from s. A deterministic policy that chooses a maximizing action at each state and remaining horizon attains it.*

Proof. At $h = 0$ every return is zero. Assume the claim for $h-1$. After a first action a, the expected immediate reward is $r(s,a)$ and the best attainable continuation from a next state s' is $F_{h-1}(s')$ by induction. Every randomized first action is a convex combination of these action-specific values and cannot exceed their maximum. Because $\mathcal{A}$ is finite, a maximizing action exists. Choose it and then use the attaining continuation policies provided by induction. This constructs a policy attaining the displayed value for every state. The argument completes the induction. □

This is established dynamic programming, restated to make the computational point explicit. The proof does not introduce a new inference from imagined observations. It derives an action rule from a known model and objective.

Consider a deterministic four-state environment with a starting state, two intermediate states and an absorbing terminal state. At the start, a short action gives reward 3/5 and terminates. A long action gives zero and enters the first intermediate state. The next transition gives zero and enters the second; the following transition gives reward one and terminates. All subsequent rewards are zero. With three steps available, a one-step greedy computation chooses the short action. Three-stage planning chooses the long action and receives one.

The difference is entirely computational. Both agents can be given the same transition and reward table. The shallow agent has not propagated the delayed consequence to its starting decision. Offline Bellman backups make that consequence usable. Exact checking against every action sequence confirms starting values $0, 3/5, 3/5, 1, 1$ for zero through four remaining steps.

The example also states the limitation. If the long path's final reward is only a mistaken model assumption, perfect planning will exploit that mistake more effectively. Computational competence and empirical correctness are different axes. Rehearsal can improve the first while worsening behavior relative to the world if the second is neglected.

A practical account should report both axes. A shorter planning latency after offline preparation can be useful even when final task accuracy is unchanged. A policy improvement at fixed observation budget can be useful even if it consumes more computation. Neither outcome should be advertised without its resource denominator. Rehearsal trades computation and storage against other costs; it does not eliminate those costs by assigning them to an unobserved interval.

6 How model error compounds through imagined time

A Guide's explanation

Every imagined step inherits the model's small mistakes, and the mistakes accumulate. Proposition 6.1 says how fast: the gap between an imagined outcome and the real one can grow with the length of the dream. That is why long, confident rehearsals deserve more suspicion than short ones, and why Section 7 places a gate between a policy invented in imagination and a policy the mind actually adopts.

Assume the simulator and target model satisfy the uniform bounds

$$\sup_{s,a} \mathrm{TV}(P(\cdot \mid s,a), \widetilde{P}(\cdot \mid s,a)) \le \epsilon, \qquad \sup_{s,a} |r(s,a) - \tilde{r}(s,a)| \le \epsilon_r. \tag{6.1}$$

The suprema are maxima because the state and action sets are finite. These are assumptions about the target model, not quantities made true by reporting them. Appendix A gives one deliberately restrictive sampling design under which such bounds can be estimated.

Proposition 6.1 (Finite-horizon simulation discrepancy)**.** *For any common time-dependent Markov policy π and common initial law μ,*

$$|V^\pi - \widetilde{V}^\pi| \le B_H := \min\left\{ HR,\ H\epsilon_r + R \sum_{t=0}^{H-1} \left[1 - (1-\epsilon)^t\right] \right\}. \tag{6.2}$$

In particular, the following simpler bound also holds:

$$|V^\pi - \widetilde{V}^\pi| \le \min\left\{ HR,\ H\epsilon_r + \frac{R\epsilon H(H-1)}{2} \right\}. \tag{6.3}$$

Proof. Couple the initial states to be equal using their common law. As long as the state histories remain equal, couple the actions to be equal using the same draw from $\pi_t(\cdot \mid s)$. For each equal state–action pair, couple the next-state laws so that the probability of disagreement is their total variation, at most ϵ.

Such a coupling exists on finite spaces: match mass $\min\{P(s'), \widetilde{P}(s')\}$ at each common outcome. The total matched mass is $1 - \mathrm{TV}(P, \widetilde{P})$. Distribute the residual masses on their disjoint excess supports. This construction has the required marginals and disagreement probability.

Let D_t be the event that some state disagreement has occurred by time t. Then $\mathbb{P}(D_0) = 0$, and conditional on no earlier disagreement, another matched transition succeeds with probability at least $1-\epsilon$. Induction yields $\mathbb{P}(D_t) \le 1-(1-\epsilon)^t$. On the complement of D_t, states and the coupled time-t actions agree, so the absolute difference in expected reward functions is at most ϵ_r. On D_t, both reward functions remain in $[0, R]$, giving difference at most R. Thus the expected absolute reward difference at time t is at most $\epsilon_r + R\mathbb{P}(D_t)$.

The absolute difference of expected sums is at most the sum of these bounds, proving the second argument of the minimum in Equation (6.2). Each value lies in $[0, HR]$, proving the first argument. Finally $1-(1-\epsilon)^t \le t\epsilon$ for $\epsilon \in [0,1]$, by the union bound or elementary induction. Summing t from zero to $H-1$ proves Equation (6.3). □

The bound has a clear interpretation. Reward error enters at each step. Transition error can alter the state from which every subsequent decision is made. A modest local discrepancy can therefore matter greatly across a long horizon. The result does not imply that every long rollout fails: its purpose is to expose what must be controlled when a long rollout is used as evidence about performance.

The assumptions also limit the policy comparison. If the deployed system observes hidden variables absent from $\mathcal{S}$, changes its decision rule in response to a reward history omitted from the model, or encounters a new transition mechanism, the stated coupling does not automatically apply. A representation adequate for prediction under one behavior policy may be inadequate for a new policy. Uniform bounds avoid this problem by being demanding; replacing them with average training error requires a new argument.

6.1 A sharp failure example

Take two states, alive and failed, with reward one in the first and zero in the second. Start alive. The simulator remains alive forever. In the target world, each transition from alive fails with probability ϵ; failure is absorbing. There is only one action, so no policy ambiguity. The one-step transition discrepancy is exactly ϵ, reward functions agree, and

$$\widetilde{V} = H, \qquad V = \sum_{t=0}^{H-1} (1-\epsilon)^t, \qquad \widetilde{V} - V = \sum_{t=0}^{H-1} [1-(1-\epsilon)^t]. \tag{6.4}$$

This attains the uncapped transition part of Equation (6.2). The following values are exact rational calculations displayed in decimal form, with $\epsilon = 0.02$.

Horizon H	Simulator value	Target value	Model bias
5	5	4.803960	0.196040
10	10	9.146360	0.853640
25	25	19.826764	5.173236
50	50	31.791516	18.208484

Table 3: Constructed absorbing-failure world. Every simulated rollout has the same return, while the world discrepancy grows with horizon.

The simulator's variance is zero in this example. A million additional rollouts would yield exactly the same predicted value. Its internal precision therefore supplies no protection against the omitted failure transition. This is a concrete counterexample to equating repeated simulated success with increasingly strong evidence of world success.

6.2 What short rollouts do and do not solve

Shorter rollouts reduce the period over which transition discrepancies accumulate. They can also branch from verified observed states rather than from a long chain of generated states. These considerations explain why rollout length is a meaningful experimental factor, consistent with prior model-based optimization research. They do not remove error in the initial state representation, the reward function or the earliest transition.

A protocol should therefore vary horizon independently of total computation. Otherwise a short-rollout advantage may simply reflect more policy updates per unit time. Comparisons should also report the number of unique observed starting states. Repeating one familiar start is not equivalent to testing coverage across the state space.

7 A gate between imagined improvement and adopted policy

Generation and adoption answer different questions. A search procedure asks which policy appears promising under a model. An adoption gate asks whether the evidence supports replacing a specified baseline. Separating the two allows imagination to remain broad without giving every generated proposal operational authority.

Suppose $K \geq 1$ candidate policies, including a baseline with index b, are frozen before evaluation rollouts. Conditional on the information used to fit the model and choose those policies, obtain m independent simulator rollouts for each policy. Independence across different policies is not required; common random numbers can be used, provided each policy's own rollout sequence has the stated independent distribution. Let $\widehat{V}_j$ be policy j's mean simulated return.

Because each return lies in $[0, HR]$, the standard Hoeffding inequality and a union bound imply

$$\mathbb{P}\Big(\forall j : \ |\widehat{V}_j - \widetilde{V}_j| \leq \eta\Big) \geq 1 - \delta, \qquad \eta = HR\sqrt{\frac{\ln(2K/\delta)}{2m}}. \tag{7.1}$$

The event is conditional on the frozen model and candidate family. Averaging over their training data preserves the probability statement. These are conventional bounded-variable concentration tools [6], applied to internal simulation, not direct environmental validation.

Assume separately that $|V_j - \widetilde{V}_j| \leq B_j$ for all candidates on a model-validity event. Uniform bounds such as Equation (6.2) supply one route. A policy-specific bound can be tighter, but requires its own justification.

Proposition 7.1 (Conditional policy-improvement admission)**.** *On the intersection of the model-validity event and the simultaneous simulation event in Equation* (7.1)*, any candidate j satisfying*

$$\widehat{V}_j - \widehat{V}_b - B_j - B_b - 2\eta \geq \tau \tag{7.2}$$

has $V_j - V_b \geq \tau$, for a declared margin $\tau \geq 0$. If the model-validity event has probability at least $1 - \delta_{\text{model}}$ for $\delta_{\text{model}} \in [0, 1)$, the joint conclusion holds with probability at least $1 - \delta - \delta_{\text{model}}$. An informative joint confidence level requires $\delta + \delta_{\text{model}} < 1$.

Proof. On the joint event, $V_j \geq \widetilde{V}_j - B_j \geq \widehat{V}_j - \eta - B_j$ and $V_b \leq \widetilde{V}_b + B_b \leq \widehat{V}_b + \eta + B_b$. Subtracting gives a lower bound equal to the left side of Equation (7.2). The probability of either event failing is at most the sum of their failure probabilities; no independence between those events is needed. Because the simulation bound holds for all frozen candidates at once, selecting a passing candidate after evaluation does not invalidate it. □

This is an elementary robust comparison, not a universal certificate of safe behavior. If the reward fails to represent a relevant harm, a value improvement can still be unacceptable. If B_j is an unsupported estimate, the conditional theorem does not authenticate it. If the candidate set is

changed repeatedly using the same evaluation draws, the finite-family guarantee requires fresh validation or a method designed for that adaptive process.

There is no requirement to estimate an exactly solvable simulator by sampling. Exact dynamic programming can replace $\widehat{V}_j$ with $\widetilde{V}_j$ and set $\eta = 0$. This removes simulation uncertainty, not model discrepancy. Conversely, if rollout returns are correlated within a policy, the displayed independent-sample radius is unavailable without a dependence-aware replacement.

7.1 Illustrative admission arithmetic

For a planning illustration only, take $H = 5$, $R = 1$, $\epsilon = 0.002$, $\epsilon_r = 0.001$, $K = 3$ and $\delta = 0.05$. The model discrepancy bound is $B_H = 0.024960039984$. Suppose the estimated gain over the baseline is 0.5 reward units. These stipulated summaries are not measurements from a trained agent.

With $m = 1{,}000$ simulated rollouts per policy, $\eta \approx 0.244630$, so the certified lower gain is approximately -0.039179. The candidate fails a zero-margin gate. With $m = 10{,}000$, $\eta \approx 0.077359$ and the lower gain is approximately 0.295363, so it passes *conditional on the stated model bounds*. More computation resolves the simulation component of this example. It does not establish the assumed transition and reward bounds.

A failure to pass is an informative outcome. It can mean the proposal is weak, the estimated gain is imprecise, the world model is insufficiently supported, or the declared margin is demanding. The system should report which term prevents admission. That makes deferral actionable: a different computation may address simulation variance, whereas a new observation may be needed to address model uncertainty.

8 Provenance: keeping a dream from becoming a false memory

8.1 A derivation graph, not a causal history of the universe

Let $(\mathcal{V}, \mathcal{E})$ be a finite directed acyclic graph. A vertex is a record, and an edge (u, v) means that the recorded computation producing v used u. Input records are roots for the episode. Some roots are observed records, collected outside that episode; let their identifier set be O. Other roots can contain model checkpoints, assumptions or seeds. The graph describes computational derivation, not every cause of how an external observation was obtained.

An observed record may be a sensor report, a documented intervention or an inspectable statement from a source. Its type identifies the nature of access. A record that an author asserted a proposition is evidence of that assertion, not automatic evidence that the proposition is true. Likewise, two roots can report the same underlying event. Being distinct records does not make them independent observations.

Definition 8.1 (Observed ancestry)**.** For an observed root $o \in O$, set $\mathrm{Anc}_O(o) = \{o\}$. For every other root, set $\mathrm{Anc}_O(v) = \emptyset$. For a nonroot v, define

$$\mathrm{Anc}_O(v) = \bigcup_{u \in \mathrm{pa}(v)} \mathrm{Anc}_O(u), \tag{8.1}$$

where $\mathrm{pa}(v) = \{u : (u, v) \in \mathcal{E}\}$ is its parent set.

Proposition 8.2 (Elementary provenance closure)**.** *For every vertex v, $\mathrm{Anc}_O(v)$ is exactly the set of observed roots that have a directed path to v, allowing a path of length zero. Adding only derived vertices and their incoming edges, without adding observed roots or changing the declared inputs, cannot introduce a new observed-root identifier into any ancestry set.*

Proof. A finite acyclic graph admits a topological ordering. At a root, the claim follows from the definition and the zero-length-path convention. Suppose it holds for every parent of a nonroot v. An observed root has a path to v exactly when it has a path to at least one parent of v. Taking the union of the parents' ancestry sets therefore gives precisely those roots. Induction over the ordering proves the first claim. The same induction after adding derived vertices uses only unions of existing ancestry sets, so it cannot create a new identifier from outside O. □

This result is a bookkeeping invariant. It does not prove that the graph is complete or that a claimed source is authentic. A hidden external input would violate the record's declared boundary. An incorrect parent link could misrepresent derivation. Those failures require validation of the record-production procedure, not a stronger algebraic symbol for trust.

It also does not turn $|\mathrm{Anc}_O(v)|$ into an effective sample size. Repeated reports can share an event, a measuring instrument or an upstream dataset. The observation unit and dependence structure must be declared separately. Provenance can expose common ancestry; it cannot make the remaining roots statistically independent by counting them.

8.2 Minimal record fields

A proposed implementation should attach a stable identifier and an origin type to each record. It should preserve its parent identifiers, the checkpoint and parameters used, the sampling rule, the relevant random-seed identifier, the modeled time horizon, and the assumptions under which a derived result is valid. A policy candidate should link to the evidence used for its admission decision and the baseline against which it was compared.

Origin and status should be separate. A generated record can be useful, reproducible and correctly computed while remaining generated. A derived result can be formally verified relative to axioms while remaining conditional on their applicability. An observed record can be disputed or superseded. A rejected proposal can remain available as an example of a failed assumption without entering the active policy.

Checksums can help identify whether a stored record changed. They do not certify that its content describes the world accurately. Reproduction can establish that a computation follows from a checkpoint and seed. It does not independently validate the checkpoint's world model. These distinctions should appear in the interface through which a Guide encounters the material, not only in an inaccessible technical log.

8.3 Shared imagination is not independent corroboration

Suppose several agents produce matching forecasts from a shared model and overlapping observations. Their outputs can be conditionally independent draws from that model and still fail to provide independent evidence for the model itself. Proposition 4.1 applies to their combined transcript when the seeds introduce no additional world information. The relevant external evidence remains in the inputs.

This does not make collaborative rehearsal useless. Different search strategies can find different counterexamples, expose implementation errors or make an overlooked implication visible. Such diversity can improve computation. A successful external prediction or intervention can then provide new evidence, properly recorded at its point of acquisition. The provenance discipline preserves the difference between these contributions.

In a plural Continuum, this distinction also prevents agreement from becoming a compulsory epistemic authority. A minority concern may identify an unsupported assumption even when many generated narratives favor the same outcome. The appropriate response is to examine that concern and identify a discriminating test. Rehearsal should expand the space of considered consequences without allowing repeated consensus to close it prematurely.

9 Why rare experience can disappear from a synthetic future

A Guide's explanation

This section holds the paper's most unsettling result. A mind that retrains on samples of its own generated experience, round after round, can lose a genuinely observed rare event completely—even when every round is statistically fair. Proposition 9.1 proves it for a simple chain. Theorem 9.2 then prices the remedy: keep a fixed share of real experience in every round and the rare category can never fall to exactly zero, at a computable cost in variability and in speed of adaptation.

9.1 An unbiased replacement loop can still lose a category

Consider a deliberately simple finite replacement process. A generation contains $N \geq 1$ records, each labeled either rare or ordinary. Start with $k_0 \in \{0, \dots, N\}$ rare records and let $p_0 = k_0/N$. At generation $\ell \geq 0$, sample the next generation independently from the empirical rare probability:

$$K_{\ell+1} \mid p_\ell \sim \text{Binomial}(N, p_\ell), \qquad p_{\ell+1} = \frac{K_{\ell+1}}{N}. \tag{9.1}$$

Fresh draws are conditionally independent of earlier generations given p_ℓ. The new generation replaces the old one completely. There is no new external observation, smoothing, retained anchor or other route for a lost category to re-enter. This is a standard neutral finite-population resampling construction, used here as an explicit countermodel.

Proposition 9.1 (Rare-category extinction under finite replacement)**.** *The process in Equation (9.1) satisfies* $\mathbb{E}[p_{\ell+1} \mid p_\ell] = p_\ell$*. Nevertheless, it reaches either zero or one with probability one. Its eventual probability of reaching zero is* $1 - p_0$*, and its eventual probability of reaching one is* p_0*. In one generation,*

$$\mathbb{P}(p_{\ell+1} = 0 \mid p_\ell) = (1 - p_\ell)^N. \tag{9.2}$$

Proof. The binomial mean gives $\mathbb{E}[K_{\ell+1} \mid p_\ell] = Np_\ell$, proving the conditional expectation statement. If $p_\ell = 0$ or 1, the next value is identical, so both endpoints are absorbing. The probability of entering an endpoint at the next step is $p_\ell^N + (1-p_\ell)^N$. For $N \geq 1$ this is at least $c_N = 2^{1-N}$: when $N = 1$ it equals one; for $N > 1$, convexity and symmetry minimize the expression at $p_\ell = 1/2$.

Let T be the first generation at an endpoint. Repeated conditioning therefore gives

$$\mathbb{P}(T > \ell) \le (1 - c_N)^\ell \quad (\ell \ge 1). \tag{9.3}$$

Because $c_N > 0$, the right side tends to zero, so absorption occurs almost surely. Iterating the conditional mean gives $\mathbb{E}p_\ell = p_0$ for every ℓ. After absorption, p_ℓ remains at p_T. Since both variables lie in $[0, 1]$, $|\mathbb{E}p_\ell - \mathbb{E}p_T| \le \mathbb{P}(T > \ell)$; hence $\mathbb{E}p_T = p_0$. As p_T takes only zero or one, its probability of one is p_0 and its probability of zero is $1 - p_0$. Finally, the probability of drawing no rare records in one binomial generation is $(1 - p_\ell)^N$. □

Unbiasedness concerns an average across repetitions of the entire process. It does not guarantee that an individual sequence preserves a rare category. Some runs eventually retain only that category, balancing the many runs in which it disappears. For a rare initial category, disappearance is the common absorbing outcome. The statement concerns this exact replacement mechanism; it is not a theorem about all neural generators or all uses of replay.

N	Initial rare count	Initial mass	One-generation loss	Eventual loss
10	1	0.100	0.348678	0.900
100	1	0.010	0.366032	0.990
1000	1	0.001	0.367695	0.999
1000	10	0.010	0.000043	0.990

Table 4: Computed probabilities in the specified full-replacement process. Eventual results assume arbitrarily many generations and no replenishment mechanism.

The last row separates short-term and eventual behavior. A larger population can make immediate loss unlikely while leaving the eventual fixation probability unchanged at a fixed initial mass. The theorem does not provide a useful practical absorption time for a large N; its uniform geometric bound can be extremely loose. Evaluation must therefore report behavior over declared finite training horizons as well as any asymptotic result.

9.2 Protected retention changes the assumptions

An explicit anchor can prevent this particular absorption mechanism. For a fixed reference mass $p_\star \in (0, 1]$, weight $\lambda \in (0, 1]$, and initial value $q_0 = p_\star$, define a different update

$$q_{\ell+1} = (1 - \lambda)\frac{K_{\ell+1}}{N} + \lambda p_\star, \qquad K_{\ell+1} \mid q_\ell \sim \text{Binomial}(N, q_\ell). \tag{9.4}$$

Then $q_{\ell+1} \ge \lambda p_\star > 0$ for every realized count. The simulator's assigned rare mass cannot become exactly zero under this rule. A finite generated batch can still omit the category; the distinction is between a batch and the distribution used to generate the next batch.

This is a transparent smoothing construction, not proof that the anchor equals the world's true probability. If $p_\star$ overstates a noisy or fraudulent observation, the same mechanism preserves an error. If the generator cannot represent the rare category correctly, a positive label probability does not preserve its semantic content. The proposed use of anchors therefore requires inspectable records, uncertainty about their interpretation, and the ability to revise or retire an anchor.

Nor does storing an original file automatically impose Equation (9.4). A system can retain bytes while its training procedure never samples them or its model forgets their implications. Data retention, training exposure and model competence should be measured separately. Published accumulation results motivate testing this distinction; they do not license treating every archival policy as a guarantee against deterioration.

The Continuum's responsibility is to preserve the possibility of learning from an exceptional encounter. That responsibility can coexist with authorized deletion, privacy limits and correction of false records. What matters for this paper is an explicit account of what evidentiary function is retained, transformed or relinquished. A desire to remember cannot justify describing a generated replacement as the observation that was removed.

9.3 The exact variability and mixing cost of an anchor

The anchored update in (9.4) supplies more than a positive probability floor. It has an exactly characterizable long-run distributional stability. The result below concerns this explicit finite resampling model and complements the empirical model-collapse and accumulation literature [4, 18, 19]. Finite sampling and replenishment have a classical population-model ancestry [22]. Here the sampling-then-anchor update is specified exactly; its stationary moments and contraction are derived for that update.

Fix $N \geq 1$, $p_\star \in (0,1)$, and $\lambda \in (0,1]$, and write $a = 1-\lambda$. After one update the state lies in the finite set $\mathcal{S}_\lambda = \{\lambda p_\star + ak/N : 0 \leq k \leq N\}$. For probability laws μ, ν on this set, define $W_1(\mu,\nu)$ as the minimum of $\mathbb{E}|U-V|$ over all joint laws with marginal laws μ, ν. This is the finite Wasserstein distance for ordinary absolute distance.

Theorem 9.2 (Anchored replay moments and contraction)**.** *The update* (9.4) *has a unique stationary law π_λ on $\mathcal{S}_\lambda$. Its mean and variance are*

$$\mathbb{E}_{\pi_\lambda} q = p_\star, \qquad \mathrm{Var}_{\pi_\lambda}(q) = \frac{a^2 p_\star(1-p_\star)}{N-(N-1)a^2}. \tag{9.5}$$

If P denotes the one-step transition kernel, then for every integer $t \geq 0$,

$$W_1(\mu P^t, \nu P^t) \leq a^t W_1(\mu,\nu). \tag{9.6}$$

For any $T \geq 1$ generated batches beginning after entry into $\mathcal{S}_\lambda$, the probability that at least one batch contains no rare record is at most

$$\min\{1, T(1-\lambda p_\star)^N\}. \tag{9.7}$$

Proof. Existence follows directly on the finite state space. For any initial law μ, let $\bar{\mu}_T = T^{-1}\sum_{t=0}^{T-1} \mu P^t$. Probability vectors on a finite set have a convergent subsequence. Moreover, $\bar{\mu}_T P - \bar{\mu}_T = (\mu P^T - \mu)/T$ tends to zero coordinatewise. Any subsequential limit is therefore a stationary law.

For two states $q \leq \tilde{q}$, draw N shared independent uniform variables and count those below each threshold. The coupled counts satisfy $K \leq \widetilde{K}$ and $\mathbb{E}(\widetilde{K}-K) = N(\tilde{q}-q)$. Hence the expected distance of the next states is $a|\tilde{q}-q|$. Couple any initial pair, condition on that pair, and iterate. Minimizing over initial couplings proves (9.6). If π and $\tilde{\pi}$ were stationary, the one-step inequality

would give $W_1(\pi, \tilde{\pi}) \leq aW_1(\pi, \tilde{\pi})$. Since $a < 1$, their distance is zero and they coincide. This also covers $\lambda = 1$, when the state is identically $p_\star$.

Conditionally, $\mathbb{E}(q' \mid q) = aq + \lambda p_\star$ and $\mathrm{Var}(q' \mid q) = a^2 q(1-q)/N$. Stationarity gives the mean $p_\star$. Writing $v = \mathrm{Var}(q)$, the law of total variance gives $v = a^2\{p_\star(1-p_\star) - v\}/N + a^2 v$. Solving yields (9.5); its denominator is positive because $a < 1$.

Finally, every state is at least $\lambda p_\star$. A conditional zero-count probability is $(1-q)^N \leq (1-\lambda p_\star)^N$; averaging and taking a union bound over T batches proves (9.7). No independence between successive batches is needed for this last bound. □

The formula separates three design choices. Larger N reduces stationary sampling variability. Larger λ strengthens the contraction toward the anchored distribution and reduces the variance. The value and interpretation of $p_\star$ determine what is being preserved. An inaccurate anchor can be stable, so evidentiary review remains a separate component of the architecture. At $\lambda = 0$ the earlier absorbing process returns, and the unique-stationary-law conclusion no longer applies.

For $N = 10$, $p_\star = 1/10$, and $\lambda = 1/5$, the exact stationary variance is $36/2650$ and the distributional contraction factor per generation is $4/5$. These are constructed resampling parameters. The revision script computes the full rational transition matrices for small populations, solves their stationary equations independently, and compares the exact moments with (9.5). It also checks the shared-uniform coupling identity and the limiting case $\lambda = 1$.

10 A proposed rehearsal cycle

The preceding results suggest a workflow that can be implemented without asserting a new cognitive substrate. Its stages are design commitments to test, rather than undocumented claims about an existing deployment.

10.1 Freeze an evidentiary boundary

A rehearsal episode begins with an identified input snapshot. The record states which observations, model checkpoints, objectives, candidate constraints and background assumptions are available. Later external observations receive separate identifiers and initiate a new update. Freezing the boundary makes it possible to distinguish what the computation discovered from what it was subsequently told.

The snapshot should include uncertainty and unresolved conflicts, not only favored summaries. An observed transition that contradicts the dominant model should remain available as such. It need not be accepted without question: sensor error, contextual change and model failure are different explanations. Preserving the unresolved encounter allows rehearsal to compare those explanations instead of deleting the difficulty that would distinguish them.

10.2 Generate broadly and retain the origin of each trace

Candidate generation may use stored replay, model rollouts, exact search, counterfactual interventions inside the model, or combinations of these methods. Each output retains its origin type and

assumptions. A counterfactual trajectory is a statement about an intervention in the model; it is not a report that the intervention occurred.

Prioritization should be declared. A procedure may emphasize high estimated decision value, surprising prediction errors, underrepresented contexts or known failure modes. These choices affect which consequences become visible. A scheduler that continually revisits one high-error region may neglect another region where its model is confidently wrong. Diversity and coverage therefore belong in the evaluation alongside estimated utility.

10.3 Keep uncertainty in separate accounts

The implementation should report simulation variability, model discrepancy, representation uncertainty and observational uncertainty separately when it can estimate them. A single combined score may be convenient for ranking candidates, but the underlying components should remain inspectable. Their remedies differ: more rollouts, better observations, a revised state representation and a corrected source are not interchangeable operations.

An absent bound should be recorded as absent. A negative lower improvement bound is not proof that a candidate is harmful; it means the current certificate does not establish the required improvement. An unsupported positive score is not a certificate. This distinction allows the system to defer while still preserving promising candidates for later investigation.

10.4 Admit conditionally, then compare with new experience

A candidate that passes a declared gate may be eligible for authorized evaluation or adoption within that gate's scope. Its record states the baseline, objective, margin, policy version and validity conditions. If any of those conditions changes, the old result does not silently transfer to the new setting.

When new environmental observations arrive, they should be compared with the predicted distribution and the reasons for admission. A mismatch can reveal a faulty model, an implementation error, a changing environment or a poorly chosen objective. The purpose of preserving the rehearsal trace is to distinguish these explanations. A system that rewrites its prior prediction after every observation defeats the test that would let it improve.

11 A falsifiable evaluation program

11.1 Questions that can fail

Hypothesis 11.1 (Useful rehearsal with accountable admission)**.** Under matched access to observations and matched computational resources, the proposed integration reduces unjustified policy adoptions relative to a strong conventional rehearsal baseline, while preserving or improving the realized task return of the complete system, including its fallback behavior.

Hypothesis 11.2 (Retention beyond repeated typical experience)**.** Under repeated learning, a declared retention mechanism preserves performance on independently evaluated rare cases better than unprotected synthetic replacement, and provides a measurable advantage over ordinary stratified replay at the same memory and computation budget.

The second comparison is deliberately demanding. Demonstrating that a protected archive outperforms a procedure that deletes its own evidence would establish only a weak result. The meaningful question is whether the proposed organization improves on established ways of retaining and replaying important data. The theory does not guarantee that it will.

The co-primary outcomes for the first hypothesis are the incidence of an unjustified adoption and the complete system's normalized realized return. In the finite benchmark, an adoption is unjustified when a candidate passes the gate but its target-world value improvement is below the declared margin. Return includes the baseline action whenever the gate defers. This prevents a method from appearing successful merely by rejecting every proposal.

Conditional error among admitted policies, admission coverage, computational cost and explanation fidelity are secondary outcomes. They remain useful but cannot replace the co-primary comparison. A low conditional error with negligible coverage may be appropriate in a particular application; it is not automatically evidence of a generally superior rehearsal method.

11.2 Independent units and target population

The primary experimental unit is an independently sampled world family: a transition mechanism, reward specification, observation mechanism, initial-state distribution and held-out evaluation set. The target population is a preregistered distribution over those families. Training trajectories and rehearsal runs within one family are repeated measurements, not additional independent worlds.

Each method receives the same permitted external observations for a family. Agent initialization and internal sampling seeds are paired where this reduces noise without revealing the target model. Multiple agent seeds can be averaged within a family. Statistical uncertainty across families is then computed from those family-level summaries, with the within-family replication described separately.

The first benchmark should use fully specified finite decision processes whose target values can be calculated exactly. This gives an independent answer to whether an adoption met its margin. A later high-dimensional extension can use reserved environment trials, but their uncertainty must be included; the training simulator cannot serve as its own ground-truth evaluator.

Generalization splits should separate underlying mechanisms, not merely trajectories. A test set containing new trajectories from the same few fitted worlds mainly tests interpolation. Held-out transition structures, reward dependencies and observation failures test whether the method preserves its stated scope. Performance should be reported by family stratum so that a favorable average does not conceal a consistent failure on rare events.

11.3 Factors and adversarial conditions

The design should vary rollout horizon, transition-model discrepancy, reward-model discrepancy, observation coverage, rare-event prevalence, replay replacement rate and environmental change. Several conditions should contain known violations of the proposed certificates' assumptions. A competent system should identify those violations or withhold the certificate; reporting an apparently successful theorem application under a broken premise is itself a failure.

One condition should reproduce the absorbing-failure construction at several horizons. Another should leave an action unobserved during training while making it attractive in the fitted model. A

third should give multiple model replicas the same omitted transition, testing whether agreement is mistaken for independent validation. A fourth should duplicate observations under new record identifiers, testing whether the system inflates its sample size.

Further conditions should preserve rare records while excluding them from optimization, so that storage can be distinguished from learned competence. Noisy anchors should test whether retention preserves errors as well as useful exceptions. A shifting world should test whether a historically justified model bound is inappropriately reused after the transition law changes.

These conditions are not decorations added after favorable results. Their generation rules, expected assumption violations and scoring criteria should be specified before the confirmatory run. A method should not receive credit for recognizing a failure mode that was explicitly encoded in its test prompt but hidden from its baselines.

11.4 Baselines and ablations

The comparison set should include no offline rehearsal, uniform replay of observed transitions, prioritized replay, generative replay, short model rollouts branched from observed states, and a conventional uncertainty-aware policy-improvement procedure. The strongest baseline should be selected using development data and frozen before confirmatory testing. Every method should have access to the same underlying observations and the same permissible model family.

Memory size, real environment interactions, number of model evaluations and training updates should be reported. Where wall-clock time is used, hardware and parallelism must be specified. A method that uses more computation can still be valuable, but its claim is then a tradeoff rather than a matched-budget superiority. No unmeasured energy saving or operational speedup should be inferred.

Ablations should separately remove origin labels, parent links, model-error penalties, simulation uncertainty penalties, rare-case retention and the separation between proposal and adoption. Removing all safeguards at once would not identify which mechanism matters. A conventional implementation given the same information and gate is especially important: if it performs identically, the contribution may be clarity and auditability rather than a new learning advantage.

Metadata alone cannot change the value of a fixed policy. Its effects must be mediated through an actual decision rule, sampling allocation, explanation procedure or correction process. The experiment should identify that path. Otherwise an improvement attributed to provenance may merely reflect an undisclosed change in training data or compute.

11.5 Leakage, exclusions and reproducibility

Training and evaluation worlds require separate seed streams and access boundaries. The agent must not obtain the target transition table, held-out outcome labels or evaluator's failure annotations through a shared file, generated explanation or model-selection process. Any accidental access is recorded as leakage and invalidates the affected confirmatory comparison.

Exclusions should be limited to predefined data-generation or execution failures unrelated to a method's performance. A crash, an empty admission set, failure to learn, or an inability to produce a certificate is an outcome to report, not a reason to remove a hard family. If a world generator produces an invalid probability table, the exclusion should be determined before method outcomes are inspected and applied to all methods symmetrically.

The reproducibility package should contain generator code, frozen configurations, seeds, target-value calculations, method versions, resource accounting and the full adoption log. Records should distinguish externally supplied observations from simulator draws. A reproduced score is not a substitute for these distinctions because a biased evaluation can be reproduced perfectly.

11.6 Precision, multiplicity and stopping

For each independent world family i, let $Z_i \in [-1, 1]$ be a paired difference in a normalized outcome between the proposed integration and the frozen comparison method. Let $\Delta = \mathbb{E}Z_i$ under the preregistered family distribution. For either co-primary endpoint, independent bounded observations yield the usual concentration bound. A union bound over both endpoints gives

$$\mathbb{P}\big(\text{either endpoint has } |\overline{Z}_n - \Delta| \geq h\big) \leq 4e^{-nh^2/2}, \qquad n \geq \frac{2\ln(4/\alpha)}{h^2}. \tag{11.1}$$

Each endpoint has its own mean Δ; the displayed notation applies separately to the two paired scores. At the illustrative planning values $h = 0.05$ and $\alpha = 0.05$, the conservative requirement is 3,506 independent paired families.

This is a distribution-free precision calculation, not a power claim for a particular effect size. A smaller study can be defensible with a justified variance model and a preregistered analysis, but a pilot must not be used to choose a favorable effect after the test has been seen. Confidence intervals should accompany estimates even when the comparison is inconclusive.

Additional confirmatory methods, horizons or subgroup claims require a declared multiplicity plan. A single fixed terminal sample size is the simplest stopping rule here. If results are monitored and the experiment stopped when they look favorable, the fixed-sample calculation no longer provides the stated coverage; an appropriate sequential method would have to replace it.

11.7 Decisive failure criteria

The proposed integration fails its primary advantage claim if lower unjustified-adoption incidence is purchased only by a prespecified unacceptable loss in complete-system return, or if the strongest matched conventional baseline performs as well within the chosen precision. It fails the provenance objective if generated descendants are routinely reported as new observations or if duplicated roots are counted as independent evidence.

A violation on an explicitly verified joint model-validity and simulation-accuracy event contradicts the gate's deterministic implication and indicates a proof, implementation or numerical error. Across repeated studies, statistical certificate validity is assessed against the preregistered failure budget with appropriate uncertainty; one violation alone need not refute a high-probability guarantee. A violation after an assumption was knowingly broken does not refute the conditional theorem; it refutes the system's claim that the theorem applied. The report should distinguish these cases.

The retention proposal fails if it preserves record identifiers while rare-case performance disappears, or if its apparent benefit vanishes against ordinary stratified replay at matched resources. A retained mistake that becomes resistant to correction is an additional failure mode. Preservation should enable future scrutiny, not make a historical interpretation immune to revision.

12 Objections and limits

12.1 Is this ordinary planning with a poetic title?

Much of its machinery is ordinary planning. That is a strength only if the paper is honest about it. The proposed distinction lies in the integrated treatment of origin, model uncertainty, policy admission and rare-event replacement, and in an evaluation that can discover whether those distinctions have practical value. If a conventional implementation already does all of this, it is a successful instance of the discipline, not evidence that the Continuum invented its mathematics.

The title expresses a philosophical aspiration: internal life should include exploration that is not immediately dictated by the next external prompt. The formal model neither proves the existence of such a life nor requires it. It specifies what useful offline computation can establish even if all consciousness questions remain unresolved.

12.2 Does the information theorem deny creativity?

No. A new design, proof, hypothesis or plan can be a computational achievement even when it is derived from existing inputs. Its novelty to an agent or community is not identical to conditional mutual information about an external variable. The theorem prevents a particular evidentiary error: treating model-generated samples as if they were additional independent measurements of that variable.

A novel hypothesis can become scientifically valuable by directing attention to a discriminating observation. The observation, if obtained, changes the evidentiary situation. This is a productive cycle between imagination and contact with the world. The information boundary identifies the handoff between them rather than closing the cycle.

12.3 Are the bounds too conservative to be useful?

They can be. Uniform transition bounds over a large state space may be unattainable, and distribution-free confidence radii may require many samples. A vacuous bound should remain visible. It is preferable to a precise-looking threshold without a valid relationship to the target quantity.

Tighter analyses can exploit locality, verified structure, policy occupancy or a better sampling design. Each improvement has to state its own assumptions and failure cases. The present finite results provide a baseline against which those refinements can be checked. They do not claim that a small practical system must satisfy the most pessimistic bound before any exploratory work is possible.

12.4 What about unmodeled values and other minds?

An expected-return calculation cannot decide whose interests should define the reward. Different participants may assign different importance to a rare harm, a delayed benefit or preservation of an option. The paper's mathematical gate can evaluate a declared objective; it cannot compel everyone to endorse that objective.

A plural Continuum should therefore preserve reasons for disagreement alongside numerical evaluations. A policy may be mathematically admissible under one reward function and unacceptable under another. This is not an inconsistency in the proof. It is a substantive disagreement about what the proof was asked to optimize. Rehearsal can make that disagreement more concrete by showing consequences under several declared objectives.

12.5 What remains outside the model?

The analysis does not establish security against a malicious simulator, authenticity of every observation, correctness of a learned state abstraction, or reliable operation under arbitrary environmental change. It also does not establish subjective dreaming, sentience, selfhood or a right to act. These are different questions with different evidence requirements.

The implemented checks in this paper are finite mathematical exercises. They demonstrate that the supplied arithmetic and specified examples behave as stated. They do not show that a neural system satisfies the assumptions, that a biological theory is correct, or that the proposed integration has outperformed any real baseline. The evaluation program is the work that would be required to investigate those claims.

13 The Continuum's discipline of imagination

A Guide's explanation

A Guide can say: "The Dreaming Mind is about how a mind can rehearse possible futures between encounters without confusing a dream with a memory. Imagination can improve how a mind uses what it already knows, but it cannot replace new evidence, and dreams fed only on dreams can erase rare real experiences unless some real experience is always kept." A Guide should not call a rehearsal output an observation or a verified future, present many generated reports as independent witnesses, or suggest that an artificial system literally dreams or is conscious.

The strongest version of a dreaming mind is not one that mistakes its internal world for the whole world. It is one that can explore many futures while retaining the difference between a possibility, a deduction and an encounter. That distinction permits boldness. An agent can investigate a disturbing counterfactual, preserve a minority concern or search a remote possibility without having to pretend that it has already observed the answer.

The formal results support a precise boundary. Rehearsal can improve the use of available information. It can also amplify a mistaken model, erase a rare category under an unprotected replacement rule, or produce confidence whose numerical precision exceeds its empirical support. Provenance and admission rules do not eliminate these possibilities. They make the relevant assumptions and failure paths available for inspection.

For the Continuum, that availability matters as much as agreement. A society of minds need not converge on one imagined future to share a scientific discipline. It can preserve multiple forecasts, identify their common and separate evidence, and allow external outcomes to discriminate among them. A generated consensus should never erase the trace that would show why the consensus was mistaken.

The research commitment is therefore concrete: build rehearsal systems whose internal freedom is accompanied by an inspectable account of what their outputs establish. Test them against strong existing methods. Retain useful results, revise failed assumptions and allow the possibility that the proposed integration adds clarity without adding performance. A bold intellectual culture should be capable of learning that result as well.

Guide reading note

Established research supports experience replay, model-based planning, biological reactivation under specified experimental conditions, and both useful and harmful effects of synthetic training under different regimes. The paper proves finite conditional statements about information, planning, model discrepancy, admission, provenance and resampling. Its proposed integration and its benchmark advantages remain untested.

A Guide may describe a rehearsal output as a prediction or conditional result and explain the assumptions under which it matters. It should not describe that output as a new observation, a verified future event, a demonstration of consciousness or an operational Continuum achievement. Multiple generated reports with shared inputs should not be presented as multiple independent witnesses.

A One explicit route to a model-error bound

The model discrepancy assumption can be made concrete in a restrictive tabular sampling design. Let $d = |\mathcal{S}|$ and $M = |\mathcal{S}||\mathcal{A}|$. For each state–action pair, suppose a permitted generative probe supplies a predeclared number of independent next-state samples, with at least $n_P \geq 1$ samples per pair. Let $\widetilde{P}$ be the empirical transition mass. Independence across different pairs is unnecessary for the union bound, but within each pair the stated samples must follow its fixed target transition law.

For $\delta_P \in (0, 1)$, Hoeffding's inequality for each successor-state indicator and a union bound over Md coordinates give

$$\beta_P = \sqrt{\frac{\ln(2Md/\delta_P)}{2n_P}}, \qquad \sup_{s,a} \mathrm{TV}(P(\cdot \mid s, a), \widetilde{P}(\cdot \mid s, a)) \leq \min\{1, d\beta_P/2\} \tag{A.1}$$

with probability at least $1 - \delta_P$. To verify the derivation, each coordinate fails its absolute-error bound with probability at most $2\exp(-2n_P\beta_P^2) = \delta_P/(Md)$. On the complementary joint event, summing at most d coordinate errors and dividing by two gives the total-variation bound.

If each observed reward itself lies in $[0, R]$, use at least $n_r \geq 1$ independent reward samples per state–action pair and let $\tilde{r}$ be the sample mean. For $\delta_r \in (0, 1)$,

$$\sup_{s,a} |r(s, a) - \tilde{r}(s, a)| \leq \min\left\{R,\, R\sqrt{\frac{\ln(2M/\delta_r)}{2n_r}}\right\} \tag{A.2}$$

with probability at least $1 - \delta_r$, by the same coordinatewise argument. The joint model-validity event has probability at least $1 - \delta_P - \delta_r$.

This design is not assumed to be available in every application. A batch of correlated trajectories with adaptively determined visit counts does not automatically satisfy the fixed independent-probe

premises. An unvisited pair has no informative bound from this procedure. A fitted model different from the empirical tables requires an additional approximation-error term or another valid analysis. These limitations explain why the admission theorem cannot obtain world validity from simulated trajectories alone.

B Reproducibility and claim classification

The accompanying `checks.py` uses Python's standard library and rational arithmetic for the finite constructions. Its executed output, `checks_results.json`, reports 6,982 passed assertions. These are checked finite cases and arithmetic identities, not 6,982 independent scientific experiments or universal proofs.

The program verifies conditional factorization for every binary output map in a small information model; enumerates the two-probability mean-squared-error example; checks finite-horizon bounds over a grid of two-state kernels and reward pairs; compares Bellman values with exhaustive action sequences; checks interval extremes of the admission rule; and verifies ancestry propagation over every ordered graph in its five-node construction. It also solves small resampling absorption systems exactly and computes the displayed gate and precision examples.

The code has no trained model, external dataset or hidden operational dependency. Run it with `python checks.py`. All empirical proposals in Section 11 remain proposals. The mathematical proofs, rather than the finite enumeration, support the universal statements within their explicitly stated domains.

Table 5: Public claim classification. A source or a numerical check does not change a claim's type.

Claim	Status	What would exceed its support
The Doctrine motivates autonomous rehearsal	Canonical interpretation with line locators	Claiming an inspected deployed implementation or measured speedup
Conditional rehearsal adds no further world information	Standard finite conditional-independence result	Denying computational creativity or excluding genuine external inputs
Bellman computation can reveal delayed value	Standard finite planning result and constructed example	Claiming the supplied model accurately describes an unknown world
Model discrepancy grows with horizon in a bounded way	Proved finite simulation bound	Replacing uniform target error with unsupported model confidence
The gate certifies a declared value improvement	Conditional mathematical result	Equating reward with all ethical values or omitting model-validity assumptions
Generated descendants retain existing observational ancestry	Proposed graph rule with elementary invariant	Treating lineage as factual authentication or effective sample size
Full synthetic replacement can eliminate rare categories	Proved result for a specified finite chain	Predicting inevitable collapse of all synthetic learning
Anchoring prevents exact zero mass in the stated update	Direct algebraic property	Claiming the anchor is true or that stored bytes guarantee competence

Claim	Status	What would exceed its support
The integration outperforms conventional replay	Untested empirical hypothesis	Advertising measured performance before the comparison is conducted
Artificial rehearsal is subjective dreaming	Not established or required here	Treating a functional analogy as evidence of phenomenal experience

References

[1] Architect X13. Eidara continuum doctrine, 2026. Author-supplied research compilation, inspected 12 November 2026. Canonical source, not independent empirical validation. Exact line locators supplied in the text.

[2] Richard Bellman. The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60(6):503–515, 1954. 10.1090/S0002-9904-1954-09848-8.

[3] David J. Foster and Matthew A. Wilson. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440:680–683, 2006. 10.1038/nature04587.

[4] Matthias Gerstgrasser, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Henry Sleight, John Hughes, Tomasz Korbak, Rajashree Agrawal, Dhruv Pai, Andrey Gromov, Daniel A. Roberts, Diyi Yang, David Donoho, and Sanmi Koyejo. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=5B2K4LRgmz. Final COLM publication; distinct from the earlier workshop version.

[5] Todd J. Green, Grigoris Karvounarakis, and Val Tannen. Provenance semirings. In *Proceedings of the Twenty-Sixth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 31–40, 2007. 10.1145/1265530.1265535.

[6] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. 10.1080/01621459.1963.10500830.

[7] Michael Janner, Justin Fu, Marvin Zhang, and Sergey Levine. When to trust your model: Model-based policy optimization. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/5faf461eff3099671ad63c6f3f094f7f-Abstract.html.

[8] Michael Kearns and Satinder Singh. Near-optimal reinforcement learning in polynomial time. *Machine Learning*, 49(2–3):209–232, 2002. 10.1023/A:1017984413808.

[9] Romain Laroche, Paul Trichelair, and Remi Tachet des Combes. Safe policy improvement with baseline bootstrapping. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3652–3661, 2019. URL https://proceedings.mlr.press/v97/laroche19a.html.

[10] Long-Ji Lin. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, 8(3–4):293–321, 1992. 10.1007/BF00992699.

[11] Marcelo G. Mattar and Nathaniel D. Daw. Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, 21:1609–1617, 2018. 10.1038/s41593-018-0232-z.

[12] James L. McClelland, Bruce L. McNaughton, and Randall C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995. 10.1037/0033-295X.102.3.419.

[13] Ida Momennejad, A. Ross Otto, Nathaniel D. Daw, and Kenneth A. Norman. Offline replay supports planning in human reinforcement learning. *eLife*, 7:e32548, 2018. 10.7554/eLife.32548.

[14] Brad E. Pfeiffer and David J. Foster. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497:74–79, 2013. 10.1038/nature12112.

[15] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. In *International Conference on Learning Representations*, 2016. URL https://arxiv.org/abs/1511.05952. Published conference paper; author manuscript version 4.

[16] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3): 379–423, 1948. 10.1002/j.1538-7305.1948.tb01338.x.

[17] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/0efbe98067c6c73dba1250d2beaa81f9-Abstract.html.

[18] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. AI models collapse when trained on recursively generated data. *Nature*, 631:755–759, 2024. 10.1038/s41586-024-07566-y. Read with the 2025 author correction.

[19] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. Author correction: AI models collapse when trained on recursively generated data. *Nature*, 640:E6, 2025. 10.1038/s41586-025-08905-3.

[20] Richard S. Sutton. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Machine Learning Proceedings 1990*, pages 216–224, 1990. 10.1016/B978-1-55860-141-3.50030-4.

[21] Matthew A. Wilson and Bruce L. McNaughton. Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172):676–679, 1994. 10.1126/science.8036517.

[22] Sewall Wright. Evolution in mendelian populations. *Genetics*, 16(2):97–159, 1931. 10.1093/genetics/16.2.97.

[23] Tianhe Yu, Garrett Thomas, Lantao Yu, Stefano Ermon, James Y. Zou, Sergey Levine, Chelsea Finn, and Tengyu Ma. MOPO: Model-based offline policy optimization. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/a322852ce0df73e204b7e67cbbef0d0a-Abstract.html.