Acquisition, Forgetting, and Target Change: Information Accounting for Adaptive Systems

Brian Theory
AI-assisted research draft

November 13, 2026

D01 — Expository technical note. With finite-horizon certificates and a periodic tracking case study. Elementary results are proved for completeness; no mathematical priority is claimed. Reviews are internal AI reviews, not external peer review.

Abstract

An adaptive system may reduce uncertainty about a target, increase the variability of its memory, discard relevant evidence, or change the target through its actions. These events require separate accounting. For a finite target and a declared pre-observation record, we derive an exact decomposition into target change, information acquired, and information discarded by memory compression. We give the precise sufficiency condition for zero loss, extend the identity to randomized memory channels, and explain the missing term when memory has an undeclared source of target information. A sharp entropy bound shows how a finite memory forces some acquired information to be discarded. A two-bit example separates retained information from performance on later coordinate queries. Complete proofs, counterexamples to pathwise entropy monotonicity, and reproducible finite enumeration are supplied. We then specify costly sensing with a persistent finite memory, prove an exact finite-horizon synthesis recurrence as an established-control specialization, and compute a three-step binary tracking frontier in rational arithmetic. One-bit storage, three-label storage, and full history have different optimal losses. An exact parameter-box certificate shows selective one-bit sensing beats every fixed schedule even with unrestricted estimator memory. A six-step counterexample overturns a shorter-horizon pattern, and a periodic controller has an analytic long-run cost and an open region of improvement over always sensing and overwriting. The contribution is an explicit resource contract and reproducible finite benchmark, not a new information or control theory.

1 Problem, antecedents, and contribution

The phrase "the system gains information" is ambiguous unless one specifies a target, an available record, a retained record, and a task. An action can make a current state predictable by resetting it while erasing information about its past. A memory can become more variable while its uncertainty about a target decreases. Neither phenomenon is paradoxical.

Conditional entropy and its chain rule are standard information theory [11, Section 6]. Existing B-Theory observation topologies distinguish changing the test family from restricting a premise region [12, Sections 3 and 8]. The earlier MIND constructions already distinguish a combinatorial refinement from its information gain, and Boolean support from probabilistic dependence [1, Sections 3–4 and 9]. This note does not claim those distinctions as new. It organizes a dynamic record contract around them and makes memory loss explicit.

The accounting statements below are elementary consequences of conditional entropy, together with a finite coding example. "Proved here" means a proof is contained in this draft; it does not

mean a new result in the literature. The operational additions are the explicit comparison of targets at two times, the bookkeeping for randomized storage, and examples exposing unfair changes of the available record. A general average-cost tracking theorem is not asserted. Sections 7–9 add finite-horizon synthesis, exact tracking benchmarks, parameter certificates and a periodic-controller evaluation, without a priority claim.

2 Operational contract and quantities

All variables are finite and live on one specified probability space. Logarithms are base two, and $0 \log 0 = 0$. Conditional distributions are asserted only at events of positive probability. Let X_t be the target before a step, X_{t+1} the target after it, and U_t the information available immediately before the next observation, including the chosen action. The action is selected before Y_{t+1} arrives. A declared transition and observation law generates (X_{t+1}, Y_{t+1}); it may depend on X_t, U_t. The memory is updated after the observation:

$$R_t = (U_t, Y_{t+1}), \qquad M_{t+1} = f_t(R_t).$$

No Markov or stationarity assumption is needed for the identity. The law itself is an analyst's model; the note does not give the agent access to hidden X_t. U_t must mean the same record throughout a calculation. In a bounded controller it can be (M_t, A_t), whereas in a full-history controller it can include the whole transcript. Substituting one for the other changes the experiment.

The deterministic memory update can be implemented by a fixed transition table. Its code size and computation cost are outside the entropy identity; the memory capacity theorem below counts only the number of possible retained states. Actions may alter the target. We compare the current target at the declared step boundary, not the time at which a policy happens to stop.

Quantity	Operational meaning and limitation
$\mathrm{H}(X \mid M)$	Expected remaining uncertainty about a specified target.
$\mathrm{I}(X; M)$	Statistical dependence of target and memory under the declared law.
$\mathrm{H}(M)$	Variability of memory states, including possibly irrelevant noise.
$\log_2 \lvert\mathcal{M}\rvert$	Maximum discrete storage capacity; not typical entropy.
Bayes task risk	Best expected loss under a specified query/action interface.
Thermodynamic entropy	A physical quantity requiring a physical state model and units.

There is no universal equality among these quantities. In particular an abstract entropy identity does not price heat, work, or repair. Stochastic thermodynamics requires additional physical assumptions [4, Sections II–III]; D10 supplies one concrete implementation.

3 An exact accounting identity

Write $X = X_{t+1}$, $U = U_t$, $Y = Y_{t+1}$ and $M = M_{t+1}$ when the indices are clear. Define

$$D_t = \mathrm{H}(X_{t+1} \mid U_t) - \mathrm{H}(X_t \mid U_t), \tag{1}$$

$$A_t^{\mathrm{info}} = \mathrm{I}(X_{t+1}; Y_{t+1} \mid U_t), \tag{2}$$

$$F_t = \mathrm{I}(X_{t+1}; R_t \mid M_{t+1}). \tag{3}$$

The superscript distinguishes acquired information from the selected action. D_t can have either sign: it includes environmental changes and action-induced resetting. Calling it solely "innovation" would hide the latter.

Proposition 3.1 (Standard identity under deterministic compression)**.** *If M_{t+1} is a function of R_t, then*

$$\mathrm{H}(X_{t+1} \mid M_{t+1}) - \mathrm{H}(X_t \mid U_t) = D_t - A_t^{\mathrm{info}} + F_t. \tag{4}$$

Both A_t^{info} and F_t are nonnegative.

Proof. The chain rule gives $\mathrm{H}(X \mid U, Y) = \mathrm{H}(X \mid U) - \mathrm{I}(X; Y \mid U)$. Determinism gives $\mathrm{H}(X \mid R, M) = \mathrm{H}(X \mid R)$. Hence

$$\mathrm{H}(X \mid M) = \mathrm{H}(X \mid R) + \mathrm{I}(X; R \mid M).$$

Substitute the first equality and subtract $\mathrm{H}(X_t \mid U_t)$. Nonnegativity follows because conditional mutual information is the average divergence between a conditional joint law and the product of its conditional marginals. On zero-mass conditioning events no term is needed. □

Proposition 3.2 (Equality and sufficiency)**.** *Under the same assumptions, $F_t = 0$ if and only if X_{t+1} and R_t are conditionally independent given M_{t+1}. Equivalently, for each positive-mass record r,*

$$\Pr(X_{t+1} = x \mid R_t = r) = \Pr(X_{t+1} = x \mid M_{t+1} = f_t(r))$$

for all x. In that case a decision maker retaining M_{t+1} can attain the same Bayes risk as one retaining R_t for every finite action set and every loss depending only on X_{t+1} and the decision.

Proof. Each divergence in the conditional mutual information is nonnegative and vanishes precisely at equality of the relevant distributions. Because $M = f(R)$, this conditional independence is exactly equality of the displayed posteriors. Expected loss for any action is a linear function of the posterior of X. Equality of the posteriors therefore preserves every conditional action risk and its minimum. This argument does not extend to losses involving an unretained historical variable unless that variable is included in the target. □

The condition is relative to the available record and target, not to every possible future question. A memory can be sufficient for the next bit and insufficient for an earlier fault or an intervention effect.

3.1 Randomized updates and undeclared side information

Determinism is sufficient but stronger than necessary. If M is generated from R with private randomness independent of X conditional on R, then $X - R - M$ is a Markov chain and Proposition 3.1 remains true. More generally the exact formula is

$$\mathrm{H}(X \mid M) - \mathrm{H}(X_t \mid U) = D_t - A_t^{\mathrm{info}} + \mathrm{I}(X; R \mid M) - \mathrm{I}(X; M \mid R). \tag{5}$$

To prove it, expand $\mathrm{I}(X; R, M)$ in the orders (R, M) and (M, R), obtaining $\mathrm{I}(X; R) - \mathrm{I}(X; M) = \mathrm{I}(X; R \mid M) - \mathrm{I}(X; M \mid R)$, and substitute entropy definitions. The last term is zero for a legitimate private coin; it is not zero for a hidden side channel.

For example, let X be a fair bit, R be constant and $M = X$. Then the memory has acquired one bit from an undeclared source. Treating this as ordinary compression predicts no improvement and is false: $\mathrm{I}(X; M \mid R) = 1$ is exactly the missing term. A comparison that gives one architecture such a side record is not an architecture comparison under matched evidence.

3.2 When a multi-step balance telescopes

If $U_t = (M_t, A_t)$ and A_t is conditionally independent of X_t given M_t, then $\mathrm{H}(X_t \mid U_t) = \mathrm{H}(X_t \mid M_t)$. Summing (4) for $t = 0, \ldots, T-1$ gives

$$\mathrm{H}(X_T \mid M_T) - \mathrm{H}(X_0 \mid M_0) = \sum_{t<T} (D_t - A_t^{\mathrm{info}} + F_t).$$

This covers actions chosen from bounded memory and an independent fresh coin. It need not cover actions chosen from an unretained full history, since the chosen action itself can reveal target information. The equality is between expected conditional entropies. It is not a statement about the sign of any realized trajectory increment.

4 Finite memory forces forgetting

Proposition 4.1 (Elementary sharp capacity obstruction)**.** *Let $X - R - M$ be a Markov chain and $|\mathcal{M}| \le 2^b$. Then*

$$\mathrm{I}(X; R \mid M) \ge \max\{0, \mathrm{I}(X; R) - b\}, \tag{6}$$

$$\mathrm{H}(X \mid M) \ge \max\{\mathrm{H}(X \mid R), \mathrm{H}(X) - b\}. \tag{7}$$

For integers $0 \le b \le n$, the bounds are attained when X is uniform on $\{0,1\}^n$, $R = X$, and M retains the first b coordinates.

Proof. The Markov condition and the chain rule give $\mathrm{I}(X; R \mid M) = \mathrm{I}(X; R) - \mathrm{I}(X; M)$. Since $\mathrm{I}(X; M) \le \mathrm{H}(M) \le \log_2 |\mathcal{M}| \le b$, the first bound follows, together with nonnegativity. The entropy inequality is the equivalent statement after subtracting mutual information from $\mathrm{H}(X)$. In the stated uniform example, $\mathrm{I}(X; R) = n$, $\mathrm{H}(M) = \mathrm{I}(X; M) = b$ and $\mathrm{H}(X \mid M) = n - b$, so equality holds. □

This is a sharp universal entropy bound, not a sharp performance theorem for every task and distribution. Equality in the numerical capacity inequality cannot specify *which* distinctions were retained.

Example 4.2 (Equal retained information, different task error)**.** Let $X = (X_1, X_2)$ be two independent fair bits. A one-bit memory is fixed before an independent uniform query $J \in \{1, 2\}$ arrives; the answer predicts X_J with zero-one loss. Storing $M = X_1$ yields error $1/4$. Storing $M = X_1 \oplus X_2$ yields error $1/2$, since either coordinate remains fair conditional on the parity. Both memories have entropy one and leave joint conditional entropy one.

The exact optimal error among all binary encoders is $1/4$. For a fixed encoder, each decoder label can be represented by a two-bit answer vector, and its optimal choices are coordinatewise majorities. There are at most two such vectors. Every source word not equal to one of these vectors has Hamming distance at least one from its assigned vector. Thus the total Hamming distance across four equiprobable words is at least two; averaging over the two coordinates gives $2/8 = 1/4$. Projection attains the bound. Randomized encoders cannot improve it: for each fixed decoder, expected loss is linear in the encoding distribution, so a minimizing deterministic encoder exists; a minimizing deterministic decoder also exists by linearity.

The proof uses a query drawn *after* storage and unavailable to the encoder. If the query arrives first, storing only the requested coordinate gives zero error. D06 and D07 study richer preservation and resource contracts; the present example is a convention-setting illustration, not a second claim to their central results.

5 Worked examples and failed implications

Noisy acquisition. For a static fair bit X, let $Y = X \oplus N$ with an independent Bernoulli(0.1) noise bit, and store $M = Y$. Starting from constant U, $D = 0$, $A^{\text{info}} = 1 - h_2(0.1) \simeq 0.531004$, and $F = 0$. Target uncertainty decreases to 0.468996 bits while memory entropy increases from zero to one bit. This is information acquisition without a claim about physical entropy.

Independent refreshing. If X_t is known in U_t and X_{t+1} is a fresh independent fair bit, with no observation, then $D = 1$, $A^{\text{info}} = F = 0$. Uncertainty increases despite a perfectly correct previous record. Identifying this increase with defective learning confuses a changing target with a failed estimator.

Resetting. A deterministic action sets $X_{t+1} = 0$, starting from a fair unknown X_t. With no further observation, $D = -1$ and both information terms vanish. Knowing the resulting state does not identify the historical state. Repair and historical diagnosis can have different objectives.

Acquiring and immediately discarding. For a static fair X, observe $Y = X$ and retain constant M. Then $A^{\text{info}} = F = 1$ and final uncertainty equals initial uncertainty. An external log that retains Y and the agent that discards it have different capabilities.

A realized observation can increase uncertainty. Let $\Pr(X = 1) = 0.1$, $\Pr(Y = 1 \mid X = 1) = 1$ and $\Pr(Y = 1 \mid X = 0) = 1/9$. The event $Y = 1$ has probability 0.2 and posterior $\Pr(X = 1 \mid Y = 1) = 1/2$. Its posterior entropy is one bit, larger than the prior $h_2(0.1)$. Nevertheless $Y = 0$ identifies $X = 0$, and the average posterior entropy is only 0.2. Nonnegative mutual information is an average statement.

Irrelevant complexity. Appending an independent fair noise string Z to a sufficient memory raises memory entropy by the length of Z and changes no posterior about the target. A score that rewards this increase alone can reward noise. Similarly, a finer internal alphabet need not improve a fixed task if its new distinctions are irrelevant.

Changing alphabets. The entropy identity permits different finite alphabets for X_t and X_{t+1}, but the numerical difference then has no automatic semantic interpretation. A comparison between a bit, a two-bit pair, and a reset symbol must say which question each variable represents. Entropy is not an encoding-free measure of developmental progress across arbitrary changes of target.

6 Reproducible verification and its scope

The accompanying `code/verify.py` constructs 64 probability laws on four binary variables: all 16 point masses and 48 seeded sparse random laws represented by normalized floating-point arrays. It checks every one of the 16 deterministic maps from the four possible records (U, Y) to binary memory, giving 1,024 identity and capacity checks. The seed is 20261113. Degenerate distributions exercise zero-probability handling. The numerical tolerance is 10^{-12}. Mutual information is evaluated independently by its probability-ratio definition, while the left side uses entropy differences. These floating-point checks do not carry the general mathematical claims; the proofs and the explicitly specified finite examples do.

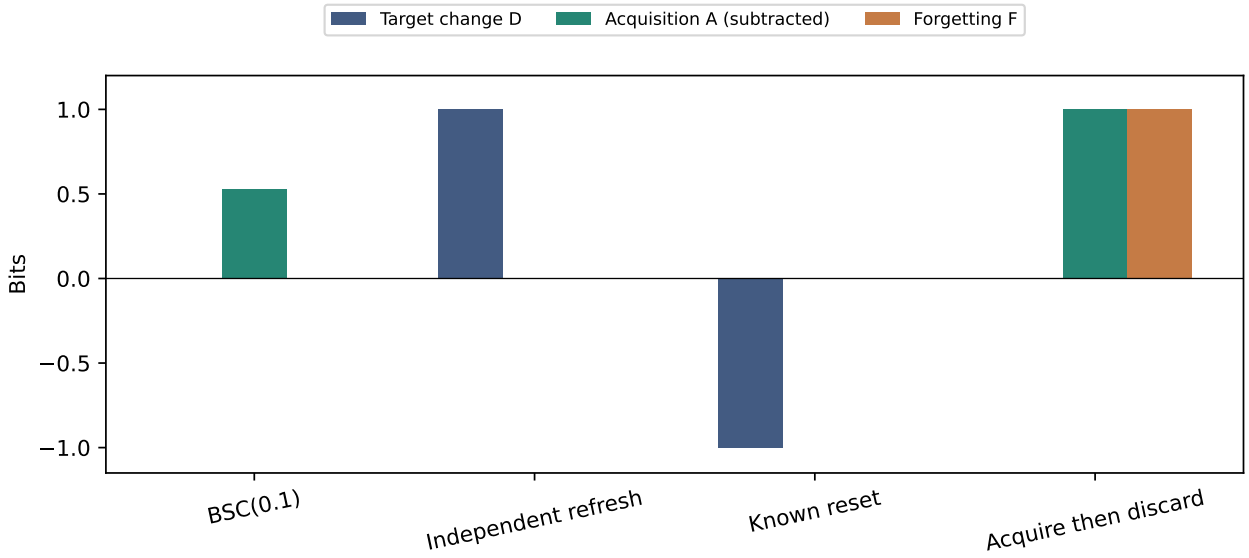


Figure 1: Exact finite examples. The target-uncertainty change equals the blue term minus the green term plus the orange term. Bars are analytic quantities evaluated numerically, not measurements or estimates. Independent refresh starts with the old target known; the other examples start with one bit of target uncertainty.

The same program enumerates all 16 two-bit-to-one-bit encoders, exports their conditional entropies and Bayes query errors, and reproduces the four named accounting examples. No sampled-data experiment or general empirical validation is claimed. The CSV, JSON output, plotting source, vector figure and editable SVG are included. The program has no network or hidden data dependencies.

7 A finite-horizon costly-sensing extension

The preceding identity accounts for a chosen controller; it does not select one. We now supply a small, completely specified tracking problem and an exact synthesis procedure. This is an *expository specialization* of finite-memory stochastic control, not a new general control principle. Limited-memory influence diagrams already optimize decisions with forgetting [7, Sections 1–3]; policy graphs already jointly represent action and memory updates [8, Section 2.2]. Distribution-space control of memory-limited systems is explicit in Tottori and Kobayashi [13, Section 4.1], whose continuous, finite-dimensional memory must not be confused with a finite alphabet. A contemporaneous preprint [6, Sections 2–3] addresses finite-bit POMDP control and tracking lower bounds; its current-observation interface differs from the post-compression interface below. These antecedents exclude a broad novelty claim.

7.1 Timing, accessible records, and priced resources

Fix a finite horizon $T \geq 1$, finite alphabets $\mathcal{X}, \mathcal{M}, \mathcal{A}, \mathcal{Y}, \mathcal{D}$, and $|\mathcal{M}| = K$. The initial joint law $q_0(x, m)$, transition kernels $P_t(x' \mid x, a)$, and observation kernels $O_t(y \mid x', a)$ are known to the

designer. At stage $t = 0, \dots, T-1$, the runtime controller performs

$$\begin{aligned} A_t = a_t(M_t), \quad X_{t+1} \sim P_t(\cdot \mid X_t, A_t), \quad Y_{t+1} \sim O_t(\cdot \mid X_{t+1}, A_t), \\ M_{t+1} = f_t(M_t, A_t, Y_{t+1}), \qquad \widehat{X}_{t+1} = d_t(M_{t+1}). \end{aligned} \tag{8}$$

The final decision alphabet need not equal $\mathcal{X}$; the notation $\widehat{X}$ simply denotes the decision. A skip action may return a fixed symbol $\perp$. Observation is paid for *before* its value is seen. Only the updated memory is available to the decoder: it cannot consult the old memory, current action, raw observation, or an external transcript. The decision has no subsequent state or observation effect and is not available later. Loss feedback is not supplied. More general control effects are already allowed through P_t; the numerical example uses action-independent target dynamics.

An external stage clock and preinstalled time-indexed tables are allowed and uncharged. The K bound applies to the persistent data record at every stage boundary, not transient input wires or scratch computation during one update. Code size, offline synthesis, online computation, and physical energy are excluded. This is not a bound on the total physical controller state. There is no persistent private seed, output log, or remembered action history outside M_t. Fresh independent randomization at the three decision kernels is permitted; it gives no improvement to the scalar optimum below. These choices differ from stationary controllers and from interfaces that let a zero-memory decoder use the current observation for free.

For bounded nonnegative task losses $\ell_t(x', d)$ and resource charges $c_t(a)$, minimize

$$J_{K,T}(\lambda) = \inf_{\pi} \mathbb{E}_{\pi} \sum_{t=0}^{T-1} \left[\ell_t(X_{t+1}, \widehat{X}_{t+1}) + \lambda c_t(A_t) \right], \qquad \lambda \geq 0. \tag{9}$$

This is a priced expected-cost problem, not a hard sample-count or expected-budget constraint. The former fixes λ; a constrained problem can require randomization and additional storage conventions.

7.2 Exact synthesis on the designer's joint law

A deterministic prescription $p = (a, f, d)$ comprises all three tables for one stage. Define the linear propagation map

$$(F_{t,p}q)(x', m') = \sum_{x,m,y} q(x, m) P_t(x' \mid x, a(m)) O_t(y \mid x', a(m)) \mathbf{1}\{m' = f(m, a(m), y)\}. \tag{10}$$

Let

$$g^{\lambda}_{t,p}(q) = \sum_{x',m'} (F_{t,p}q)(x', m') \ell_t(x', d(m')) + \lambda \sum_{x,m} q(x, m) c_t(a(m)).$$

The variable q is a distribution over hidden state and memory, not the runtime posterior conditional on a realized history. The designer propagates it without seeing realized states or observations.

Proposition 7.1 (Finite-horizon synthesis; standard specialization)**.** *Under the declared contract, a deterministic time-indexed controller attains the infimum in* (9)*. Its value is given exactly by*

$$V_T(q) = 0, \qquad V_t(q) = \min_p \{ g^{\lambda}_{t,p}(q) + V_{t+1}(F_{t,p}q) \}, \qquad J_{K,T}(\lambda) = V_0(q_0).$$

Only finitely many prescriptions are minimized at each step. For fixed T and initial law, exact rational kernels, losses, and a rational price permit exact rational evaluation.

Proof. For a fixed prescription, summing the joint law of $X_t, M_t, X_{t+1}, Y_{t+1}$ gives (10). Its immediate expected cost is $g^{\lambda}_{t,p}$. No other past variable enters the future law or admissible decision kernels. Among deterministic controllers, enumerate the first prescription and apply the same argument to the remaining stages; induction proves the recursion.

For a controller with fresh private randomization, pre-sample the output for every entry of every time-indexed action, update, and decision table. Each entry is queried at most once in an episode because time never repeats. The resulting mixture over deterministic table sequences therefore has exactly the randomized controller's law. Its expected scalar cost is an average of their costs, so some deterministic sequence is at least as good. All such sequences have the same K-state runtime interface. Pre-sampling is a proof device, not an extra runtime seed. Finiteness ensures a minimum exists, and rational arithmetic is closed under the finite sums, products, and comparisons used by the recursion. □

An optimal recursion path supplies one preinstalled sequence of tables for the given q_0. The runtime does not calculate or store q_t. If an implementation instead feeds the full-history posterior to its action selector, it has changed the memory contract. Likewise, jointly selecting tables is not equivalent to optimizing each realized memory label independently: merging records changes future conditional laws. Reachable designer laws can grow exponentially; this is an exact finite procedure, not an efficiency guarantee.

Corollary 7.2 (Shape of the price value)**.** *Write E_π and N_π for the expected task loss and total resource use of each deterministic table sequence. Then*

$$J_{K,T}(\lambda) = \min_{\pi}(E_\pi + \lambda N_\pi).$$

It is continuous, nondecreasing, concave, and piecewise affine in $\lambda \geq 0$. Increasing K cannot increase it. For fixed λ, $V_t(q)$ is a finite minimum of linear functions of q, hence concave and piecewise linear.

Proof. A fixed sequence induces a cost linear in q and affine in λ, with slope $N_\pi \geq 0$. Take the finite minimum and use Proposition 7.1. Extra memory labels can be left unused. □

The lower envelope describes supported optima of the priced problem. It is not a claim that every point on its convexified risk–resource diagram is achievable with the same bounded-memory randomization interface.

7.3 What the determinism statement does not say

A stationary one-state controller using the same action rule at both stages gives a simple counterexample to extending purification to reused rules. Let the environment record the first of two binary actions and reward a mismatch at the second. Every deterministic stationary rule obtains zero reward, whereas fresh fair actions obtain 1/2. A time-indexed deterministic rule obtains one; thus unrolling the coin into different tables leaves the stationary class.

For a different boundary, a single fair bit, a perfect paid sensor, $K = 2$, and an expected sensing budget 1/2 give deterministic feasible risk 1/2 but randomized risk 1/4. Sense with probability 1/2, storing the observed bit; otherwise store zero. The decoder returns memory. This uses only the current action and fresh coin, and meets the expected budget. It does not contradict Proposition 7.1, which optimizes an unconstrained scalar price. An episode-level mixture of multi-stage controllers may need persistent mode information; the present contract does not grant it without counting storage.

8 An exact small tracking benchmark

Let $T = 3$, X_0 be fair, and $M_0 = 0$. The binary target flips independently with probability $1/10$ before every stage's observation. Sensing gives $Y = X_{t+1} \oplus Z$, with independent $Z \sim \text{Bernoulli}(1/5)$; skipping gives $\perp$. Actions do not change the target transition. Set $c(\text{sense}) = 1$, $c(\text{skip}) = 0$, and $\ell(x, d) = \mathbf{1}\{x \neq d\}$. All errors below are *total expected errors over three predictions*, not per-stage rates. The full-history comparator has the same sensing timing, target law, and loss; it retains its whole paid observation/action transcript.

Exact finite result. Exhaustive rational synthesis gives

$$J_{1,3}(\lambda) = 3/2, \tag{11}$$

$$J_{2,3}(\lambda) = \min\left\{\frac{3}{5} + 3\lambda,\ \frac{3093}{5000} + \frac{1089}{500}\lambda,\ \frac{96}{125} + \lambda,\ \frac{3}{2}\right\}, \tag{12}$$

$$J_{\text{hist},3}(\lambda) = \min\left\{\frac{1443}{2500} + \frac{589}{250}\lambda,\ \frac{96}{125} + \lambda,\ \frac{3}{2}\right\}. \tag{13}$$

Equation (11) follows analytically: the single retained state makes every post-compression decision constant, and action-independent dynamics keep each target fair. The other displayed numerical optima are **exact finite computational results**, supported by the exhaustive procedure below and independently checked computations; they are not parameter-uniform analytic theorems.

These are representative envelope segments, not every optimal table sequence. Policies or performance pairs that tie only at a breakpoint can be omitted without changing the value.

The active prices for (12) switch, in the displayed order, at

$$\lambda = \frac{31}{1370},\quad \frac{747}{5890},\quad \frac{183}{250}.$$

Adjacent policies tie at a switch. The full-history switches are $159/1130$ and $183/250$. At zero price, the one-bit penalty is $3/5 - 1443/2500 = 57/2500$. At $\lambda = 1/10$, the one-bit optimum is $2091/2500 = 0.8364$, while full history gives $508/625 = 0.8128$.

Executable witnesses. For all four one-bit vertices, label the memory by the predicted bit. Always sensing and overwriting with Y gives $(E, N) = (3/5, 3)$. Sensing only initially and retaining that bit gives errors $1/5, 13/50, 77/250$, hence $(96/125, 1)$. Never sensing gives $(3/2, 0)$. The remaining policy is genuinely selective:

Prediction stage	Sensing decision	New memory
1	Sense	Y
2	Sense iff $M = 1$	Y if sensed; otherwise 0
3	Sense iff $M = 0$	Y if sensed; otherwise 1

Its expected sample counts are $1, 1/2, 339/500$ and its errors are $1/5, 23/100, 943/5000$. Thus it realizes $(3093/5000, 1089/500)$. Its joint law after the second stage, with rows $X = 0, 1$ and columns $M = 0, 1$, is

$$\begin{pmatrix} 237/500 & 13/500 \\ 102/500 & 148/500 \end{pmatrix}.$$

Memory label 1 is more reliable than label 0, so the third-stage sensing choice reflects different confidence. A symmetric source does not justify restricting deterministic controllers to label-symmetric tables. The bit-complemented policy has the same value. Mixing these policies while retaining their mode is a different storage question.

A full-history low-price witness senses at stages 1 and 2, then senses at stage 3 exactly when the two readings disagree. The disagreement probability is $89/250$, giving $N = 589/250$; exact Bayes decisions yield total error $1443/2500$. At zero sensing price, always sensing also attains that error, but the displayed policy uses fewer samples. Four memory labels suffice to implement this witness: retain the first reading, then the ordered pair of readings, and finally the decision bit. Together with the history lower bound this proves $J_{4,3} = J_{\mathrm{hist},3}$ for every nonnegative price, using the analogous one-sample and no-sample witnesses on the other intervals. This is an all-time storage construction with a free external clock, not D07's offline terminal encoder.

For completeness the same rational algorithm gives the following three-label lower envelope, recording the supported (E, N) pairs in increasing-price order:

$$\left(\frac{2943}{5000}, \frac{1339}{500}\right), \quad \left(\frac{759}{1250}, \frac{589}{250}\right), \quad \left(\frac{3093}{5000}, \frac{1089}{500}\right),$$

$$\left(\frac{3363}{5000}, \frac{839}{500}\right), \quad \left(\frac{96}{125}, 1\right), \quad \left(\frac{3}{2}, 0\right).$$

In particular, three labels still have zero-price penalty $57/5000$ relative to full history. Label count is the primary resource here: three labels require two fixed-length binary storage bits, just as four labels do. The three-label result must not be described as a fractional physical-bit implementation.

Immediate error reduction is not the control objective. At $\lambda = 1/10$, a controller minimizing each current expected error plus current sample charge, ignoring future value, senses at stages 1 and 3 and skips stage 2. Its errors are $1/5, 13/50, 1/5$, with two samples, giving $43/50 = 0.86$. The selective one-bit optimum is smaller by $59/2500$. The comparison uses the same clock, memory, and sensing access. It establishes a counterexample for this immediate-risk rule, not a universal failure bound for all information-gain heuristics.

8.1 Why the computation is exhaustive and how to reproduce it

For each memory label there are K skip/update choices and K^2 sense/update choices (one next label for each binary observation). Thus each stage has $(K + K^2)^K$ action/update tables: 2, 36, and 1728 for $K = 1, 2, 3$. Off-action table entries never used by the chosen action are irrelevant. For a propagated law q', the best post-compression binary decision has error

$$e(q') = \sum_m \min\{q'(0, m), q'(1, m)\}.$$

It can be optimized separately at each stage because decisions do not affect future state, memory, or observations. This evaluates all decoder choices without explicitly multiplying the table count by 2^K.

The included `code/tracking.py` enumerates every action/update prescription, propagates integer probability masses with denominator $2 \cdot 50^t$, and memoizes identical (t, q) states. At each state it retains the exact lower envelope of future (E, N) lines on $\lambda \geq 0$. A line dominated for every nonnegative price can be removed without changing any ancestor's priced optimum. All intersections and comparisons use integer arithmetic or Python `Fraction`; plotted decimal curves

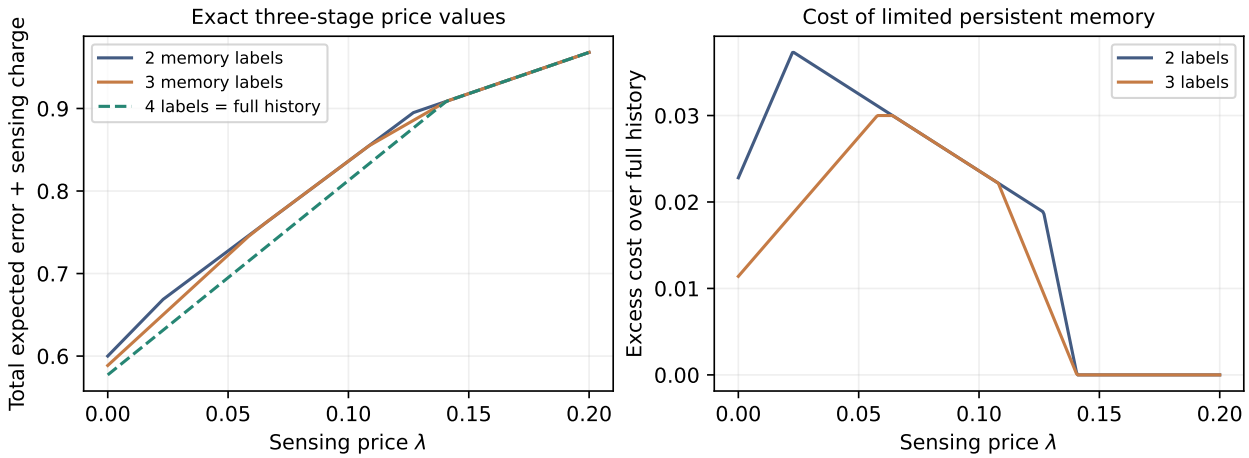


Figure 2: Exact supported price values and the memory penalty in the declared three-stage binary model. Curves visualize rationally computed envelopes; decimal rendering supplies no additional numerical evidence. Four labels implement the full-history optimum. One-label cost is constantly 3/2 and is omitted for readability.

are not certificates. The full-history comparator separately recurses over exact Bayesian beliefs, considering skip and sense and all combinations of continuation policies on the two observation branches. The output saves a representative policy witness for each retained envelope segment, its stage laws, errors, sample counts, breakpoints, and cache counts. This procedure is bounded and reproducible; no random search or learned optimizer is used.

An independent implementation enumerates all $36^3 = 46{,}656$ two-label table sequences (109 distinct loss/sample pairs) and all 42 full-history sensing trees (19 distinct pairs), testing each affine pair against every competitor without the author's Bellman recursion or hull pruning. It agrees on all nonzero-width envelope segments and records additional breakpoint-only ties. A second independent program propagates three-label occupancies forward over all 1728 prescriptions with per-law Pareto pruning, rather than the author's backward envelope pruning; it obtains 9, 81, and 3471 reachable laws after successive stages and confirms all six three-label segments.

The independent specialist and hostile AI reviews, their separate code, and author responses are retained with this revision. These are internal model-assisted checks, not human peer review or formal verification. Reproduce from this directory with `python3 code/tracking.py`; Python 3.10 or later and its standard library suffice. The earlier entropy checks still use NumPy and Matplotlib. Computational global-optimality evidence is confined to this declared finite instance.

8.2 Connection to the accounting identity and portfolio boundaries

For any synthesized policy, take $U_t = (M_t, A_t)$, $R_t = (U_t, Y_{t+1})$, and the update in (8). The original identity and its telescoping condition apply. The quantities $D_t, A_t^{\text{info}}, F_t$ describe that policy's information flow, while $J_{K,T}$ prices its task errors and sensing. None alone determines the optimum. In particular, a low-entropy or high-mutual-information summary need not encode the confidence distinctions useful to future sensing.

The added material owns one finite streaming benchmark and its explicit synthesis contract. It neither replaces D02's restricted-experiment minimax theorem nor duplicates D07's noisy offline terminal coding frontier. D03 treats an aging exact reference under an ancillary-observation assumption; here noisy readings create informative confidence differences. D05 owns physical

resource viability, absent here, and D10 owns implementation-dependent energy costs. The baseline D01 accounting identity remains a shared expository introduction.

9 Parameter ranges, longer horizons, and a periodic controller

The three-step experiment is insufficient evidence for a parameter-uniform or long-run optimum. This section strengthens it in three different senses: a robustness bound for optimized values, a certified comparison on a parameter box, and an analytic long-run cost for one executable periodic controller. These statements have different quantifiers. None claims that the three-step optimizer remains globally optimal throughout a box or at every horizon.

9.1 An elementary finite-horizon robustness bound

Write p for target flip probability and ϵ for sensor crossover, with a fair initial target, fixed initial memory, action-independent target dynamics, and the timing of Section 7. Fix a price $\lambda \geq 0$ and a policy class with the same interface under both parameter settings. For $\delta_p = |p - p_0|$, $\delta_\epsilon = |\epsilon - \epsilon_0|$, put

$$B_T = \left[\frac{T(T-1)}{2} + \lambda\frac{(T-1)(T-2)}{2}\right]\delta_p + \left[\frac{T(T+1)}{2} + \lambda\frac{T(T-1)}{2}\right]\delta_\epsilon. \tag{14}$$

Proposition 9.1 (Finite-horizon coupling bound; standard method)**.** *For any fixed admissible policy, and also for the infimum over any fixed such policy class, the absolute change in total expected error plus sensing charge is at most B_T.*

Proof. The first scored target X_1 is fair for both models, irrespective of p, so couple it identically. Couple later Bernoulli flips and all potential sensor noises by common uniform random variables, with disagreement probabilities δ_p and δ_ϵ. Generate potential sensor noise even on skipped stages, where it is ignored, and couple any private policy coins identically. Before a primitive disagreement the two executions have identical observations, memories, actions and predictions. By a union bound, prediction-stage s errors can differ with probability at most $(s-1)\delta_p + s\delta_\epsilon$. Sensing at stage s is chosen before current noise, so its indicator can differ only after at most $\max\{s-2, 0\}$ flip disagreements and $s-1$ sensor disagreements. Sum these bounds for $s = 1, \ldots, T$. Each error or sample indicator has range one. Taking an infimum over the same policies preserves a uniform two-sided bound. No continuity of the optimizer itself is required. □

For example, at $T = 3, \lambda = 0$, the two-label versus full-history gap obeys

$$J_{2,3}(p, \epsilon; 0) - J_{\text{hist},3}(p, \epsilon; 0) \geq \frac{57}{2500} - 6|p - 1/10| - 12|\epsilon - 1/5|.$$

It is therefore positive on an explicit open diamond, not only at one rational parameter point. The corresponding three-label bound replaces 57/2500 by 57/5000. These bounds may be loose and grow quadratically with horizon. They are not infinite-horizon robustness bounds, nor do they preserve an optimizer's identity. Explicit POMDP model-mismatch and robust-controller theory substantially predates this specialization [2, 3, 5].

9.2 A uniform adaptive advantage against every fixed schedule

Let π_A be the selective two-label policy from Section 8: sense initially; at stage 2 sense iff $M = 1$; at stage 3 sense iff $M = 0$; store Y when sensing, hold M otherwise, and predict M. Let $\mathcal{O}$ contain

all sensing schedules chosen independently of the target and sensor outcomes. The estimator for a schedule has *unrestricted* retained history and knows the schedule and true model parameters. Thus the comparator is stronger than a one-bit nonadaptive estimator. Schedule randomization is allowed, and its realized schedule may be disclosed to the estimator.

Proposition 9.2 (Exact parameter-box comparison)**.** *Throughout the closed box*

$$p \in [9/100, 11/100], \qquad \epsilon \in [19/100, 21/100], \qquad \lambda \in [9/100, 11/100],$$

the three-step cost satisfies

$$\inf_{\pi \in \mathcal{O}} J_\pi(p, \epsilon, \lambda) - J_{\pi_A}(p, \epsilon, \lambda) \geq \frac{86123}{12500000} > 0.$$

This is a finite polynomial certificate for a policy-class comparison, not a global optimality claim for π_A throughout the box.

Proof by explicit polynomial certificate. Define $r = p + (1-2p)\epsilon$ and $r_2 = p + (1-2p)r$. Direct path expansion gives the adaptive policy's expected error and sample count as

$$\begin{aligned} E_A &= (2p-1)\epsilon^3 + (3/2 - 3p)\epsilon^2 + (p^2 - p + 5/2)\epsilon - p^2/2 + p, \\ N_A &= (2p-1)\epsilon^2 + (1-2p)\epsilon + p/2 + 2. \end{aligned}$$

For the all-sense schedule write

$$h = (4p-2)\epsilon^3 + (3-6p)\epsilon^2 + (2p^2+2)\epsilon - p^2 + p.$$

The eight schedules, their optimal errors on this box, and certified lower bounds for $E_s + \lambda N_s - E_A - \lambda N_A$ are:

Schedule	Error E_s	Samples N_s	Minimum Bernstein coefficient
000	$3/2$	0	1896723/3125000
001	$1 + \epsilon$	1	1334223/3125000
010	$1/2 + \epsilon + r$	1	156837/781250
011	$1/2 + 2\epsilon$	2	771723/3125000
100	$\epsilon + r + r_2$	1	86123/12500000
101	$2\epsilon + r$	2	8106/390625
110	$2\epsilon + r$	2	8106/390625
111	h	3	24927/781250

Here 1 means sensing and 0 skipping. The Bayes errors in the table are certified on the stated box; in particular, the polynomial h is not asserted to be the optimal error for arbitrary noise parameters.

For reproducibility, affinely transform the box to $[0,1]^3$ and write each difference in the tensor-product Bernstein basis. If its ordinary power coefficients are c_{abc} with degree bounds (d_1, d_2, d_3), the Bernstein coefficient at (i, j, k) is

$$b_{ijk} = \sum_{a \leq i, b \leq j, c \leq k} c_{abc} \frac{\binom{i}{a}\binom{j}{b}\binom{k}{c}}{\binom{d_1}{a}\binom{d_2}{b}\binom{d_3}{c}}.$$

Bernstein basis functions are nonnegative and sum to one, so the polynomial lies between the smallest and largest coefficient. The eight differences have degree bounds $(2, 3, 1)$; all 192 coefficients are

positive, with minima shown above. The exact file `data/parameter_certificate.json` gives every coefficient, not just these minima. It also records all 50 nonzero posterior-mass sign polynomials needed to certify the displayed history-based Bayes decoders throughout the box. Their Bernstein coefficients have the correct sign; identically tied fair posteriors permit either decision. Thus no better estimator for any fixed schedule was excluded. A randomized open-loop schedule is a mixture of these eight conditional problems, and cannot beat their minimum. Taking the least of the eight displayed bounds proves the claim. □

The certificate was independently reconstructed by symbolic stage propagation and exact interpolation in the Bernstein basis, rather than the author's primitive-path expansion and power-basis conversion. This is exact computational algebra, not a grid or Monte Carlo argument. At the central parameters, the best open-loop cost is 43/50 (schedules 101 and 110), versus 2091/2500 for π_A; the gap is 59/2500.

This comparison shows that a little persistent information can be useful for *choosing when to acquire evidence*, even when the nonadaptive comparator has much more storage. It does not show that limited memory intrinsically outperforms full memory: unrestricted adaptive history attains a lower cost. Adaptive-versus-open-loop advantages and paid Markov-state tracking are established research topics [9, 14]; the addition here is the explicit certified box under the present interface.

9.3 Longer horizons overturn a tempting extrapolation

The same exact synthesis and a separately written forward Pareto computation extend the nominal two-label results to $T = 6$:

Horizon	Minimum two-label error, $\lambda = 0$	Always sense and overwrite
1	1/5	1/5
2	2/5	2/5
3	3/5	3/5
4	4/5	4/5
5	1	1
6	29944419/25000000	6/5

Thus the pattern $J_{2,T}(0) = T/5$ fails at six, by 55581/25000000. The winning witness senses first and then alternates the two conditional-sensing rules of π_A, with mean sample count 10037403/2500000 over six stages. Full-history and two-label price envelopes through $T = 6$, including witnesses, are saved in `data/longer_horizon.json`. Independent forward propagation confirms every two-label segment without the author's backward-envelope pruning. This is exact finite evidence, not a theorem about every horizon.

Free observations can be ignored. The comparator is *always sensing and overwriting*, not all controllers that always sense. When $\lambda = 0$, any selective policy in this action-independent target model can be simulated by always sensing and ignoring readings on its virtual skip stages. The update computes the virtual sensing choice from the same memory and clock. It thereby preserves exactly the memory and error law with no extra persistent state. Consequently removing free observations cannot improve the optimum over all allowed updates. The six-stage result exposes a failure of overwriting every reading, not a benefit from destroying available evidence.

9.4 A parameter-uniform long-run policy evaluation

Now repeat the alternating conditional-sensing rule indefinitely after one initial sample. This policy uses one persistent bit *and a free alternating phase*. Define

$$a = \epsilon(1-\epsilon)(1-2p), \qquad D = 1 + p - a.$$

Proposition 9.3 (Achievable long-run cost; no optimality claim)**.** *For $0 < p < 1/2$ and $0 < \epsilon < 1/2$, the limiting expected average prediction error and sensing rate of this periodic policy are*

$$\bar{E} = \frac{2\epsilon + p - a}{2D}, \qquad \bar{N} = \frac{1+2p}{2D}.$$

It uses fewer samples than always sensing. If

$$p < \frac{\epsilon(1-\epsilon)}{1+2\epsilon(1-\epsilon)}, \tag{15}$$

then $\bar{E} < \epsilon$, so its average error-plus-price cost is strictly below that of always sensing and overwriting for every $\lambda \geq 0$.

Proof. Let the phase ϕ be the memory label that triggers sensing on the next step. It alternates after each step. In phase-transformed coordinates $Z = X \oplus \phi$ and $B = M \oplus \phi$, sensing occurs iff $B = 0$. The next transformed target changes with probability $1 - p$; if sensing, $B' = Z' \oplus N$, while if skipping, $B' = 0$. The homogeneous transition matrix in the order $(Z, B) = 00, 01, 10, 11$ is

$$P = \begin{pmatrix} p(1-\epsilon) & p\epsilon & (1-p)\epsilon & (1-p)(1-\epsilon) \\ p & 0 & 1-p & 0 \\ (1-p)(1-\epsilon) & (1-p)\epsilon & p\epsilon & p(1-\epsilon) \\ 1-p & 0 & p & 0 \end{pmatrix}.$$

For interior parameters, every state reaches a $B = 0$ state, from which every state has positive one-step probability; the 00 self-loop is positive. Hence the chain is irreducible and aperiodic. Its unique stationary row vector, verified by multiplication, is

$$\frac{1}{2D}\big(1 + p - \epsilon,\ \epsilon[\epsilon + 2p(1-\epsilon)],\ \epsilon + p,\ (1-\epsilon)[1 - \epsilon + 2p\epsilon]\big).$$

Error is the event $Z \neq B$, and sensing is $B = 0$, so summing the appropriate coordinates gives the formulas. Convergence makes the initial sample asymptotically negligible. Finally

$$\epsilon - \bar{E} = \frac{(1-2\epsilon)(a-p)}{2D}, \qquad 1 - \bar{N} = \frac{1-2a}{2D} > 0.$$

Condition (15) is equivalent to $a > p$; this proves both strict comparisons. The result evaluates one policy and supplies an upper bound on optimal average cost, not a matching lower bound. □

At the nominal point, $\bar{E} = 31/162 \simeq 0.191358$ and $\bar{N} = 50/81 \simeq 0.617284$. For arbitrary time-indexed policies use the expected upper-average criterion

$$J_2^{\mathrm{av}}(\lambda) = \inf_{\pi} \limsup_{T\to\infty} \frac{1}{T} \mathbb{E}_\pi \sum_{t=0}^{T-1} \left[\mathbf{1}\{X_{t+1} \neq \widehat{X}_{t+1}\} + \lambda A_t \right],$$

where $A_t = 1$ for sensing and zero for skipping. For example, a clocked policy can alternate never-sense/reset-to-zero blocks (error $1/2$) and always-sense/overwrite blocks (error ϵ), each much

longer than all predecessors; its expected average need not converge. The two policies exhibited here do have limits, and therefore

$$J_2^{\mathrm{av}}(\lambda) \le \min\left\{\frac{1}{2},\ \frac{31}{162}+\frac{50}{81}\lambda\right\}.$$

The constant-memory policy supplies $1/2$. No equality is claimed. For example at $p=\epsilon=1/5$, the periodic policy has $\bar{E}=21/92>1/5$, so it loses to overwrite at zero price; the parameter condition cannot be omitted.

The phase is a real resource. Without an external clock, storing (M,ϕ) uses four states, or two bits. If an implementation charges an extra average overhead κ relative to overwrite, the nominal improvement persists only when

$$\kappa < \frac{7}{810}+\frac{31}{81}\lambda.$$

This is a sensitivity calculation in task-cost units, not an energy measurement. It prevents translating the free-clock result into an unqualified one-bit hardware advantage. Periodic finite-state controllers themselves are established [10]. The specific rational evaluation and parameter condition are the contribution investigated here, without historical-priority certification.

9.5 Distinction from existing control literature

A difference in timing does not put this model outside finite-memory control theory. The sensing, update and decision phases can be encoded as successive stages of a finite POMDP or limited-memory influence diagram with only the declared information parents. The present recurrence is a specialization of that framework. Likewise, the coupling proof and Bernstein basis are standard techniques, and periodic controllers are established constructions.

Antecedent	Established overlap	Narrow addition retained here
LIMIDs and FSCs	Finite-memory policy design	Auditable after-compression sensing benchmark
Active sensing	Feedback can improve acquisition	One-bit versus full-history open-loop box certificate
Robust POMDPs	Perturbation and uncertainty guarantees	Elementary finite-horizon constants and exact box data
Periodic FSCs	Infinite repetition of layered policies	Closed-form binary controller costs and validity region
Finite-bit tracking	Memory–performance separation	Matched timing and explicit memory/clock accounting

The source ledger records inspected passages and preprint versions. Robust synthesis over uncertainty sets [2], kernel-mismatch guarantees [3], and fixed-policy observation robustness [5] have stronger general aims than the local certificate above, under their own assumptions. The new results establish neither a first adaptive-sensing advantage nor a new universal finite-memory theorem. They make this particular example less dependent on one point and one short horizon. No physical measurements, learned-model validation, human peer review, or exhaustive historical search is claimed.

10 Being, Becoming, and The MIND

Operational interpretation. Being can refer here to the distinctions and predictive capabilities retained at a given step. Becoming can include acquisition, forgetting, a change of target, or a change of available memory. The MIND programme can use this decomposition to ask whether an internal update preserves the evidence needed for a later decision. The useful claim is that those events can be measured separately under an explicit contract.

These interpretations neither define subjective experience nor show that increased information is increased awareness. Information loss can be a sensible design choice; resetting can restore function without historical identification; a task-beneficial simplification can improve a developing controller. The identity does not decide which of these goals is desirable.

11 Open questions, distinctness, and readiness

The original proposal is now partially resolved at a deliberately finite scope: the transition, loss, sensing interface, persistent memory, and horizon are fixed; Proposition 7.1 gives an exact finite-horizon design procedure; and the three-stage example has executable witnesses and exact enumerative optima. The periodic construction now supplies an analytic achievable long-run cost, but does not solve the original optimal-average-tracking programme. The general recursion is established-control synthesis and the numerical instance has no certified novelty. A targeted primary-source search through November 13, 2026 included the recent finite-memory-control preprint [6]; neither its theorems nor the wider literature have been exhaustively audited.

The parameter-uniform comparison and periodic-policy evaluation in Section 9 partially resolve the previous short-horizon limitation. They do not solve global long-run optimization. Remaining research problems include matching lower bounds for average cost, scalable certified synthesis, larger memories and longer horizons, unknown dynamics, and joint accounting for code, clocks, scratch space and physical resources. The three-step adaptive policy has a certified advantage over open-loop schedules on a box; its global optimality on that box is not asserted. The periodic result is an achievable cost formula, not an optimal frontier. These distinctions must survive any future integration or submission.

The manuscript remains a common expository chapter with a methods appendix. It should be integrated with the acquisition and memory work using the stated interfaces, and receive human specialist assessment before submission. None of the reviews supplied here establishes historical priority, external peer review, or claims about consciousness.

References

[1] B-Theory research programme. The MIND: Five finite constructions, 2026. URL `https://www.b-theory.com/manuscripts/mind-five-finite-constructions-r1-2026-10-05.pdf`. First research edition with P1 and P2 addenda, October 5, 2026. Printed byline: Research prepared for Brian Theory; AI-assisted mathematical synthesis. Sections 1–4 and 9 inspected.

[2] Murat Cubuktepe, Nils Jansen, Sebastian Junges, Ahmadreza Marandi, Marnix Suilen, and Ufuk Topcu. Robust finite-state controllers for uncertain POMDPs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021. URL `https://arxiv.org/abs/2009.11459`. Author preprint consulted.

[3] Yunus Emre Demirci, Ali Devran Kara, and Serdar Yüksel. Computable robustness bounds to transition and measurement kernel perturbations and approximations in partially observable stochastic control, 2025. URL https://arxiv.org/html/2508.10658v1. arXiv:2508.10658v1; relevant model and robustness sections inspected.

[4] Jordan M. Horowitz and Massimiliano Esposito. Thermodynamics with continuous information flow. *Physical Review X*, 4:031015, 2014. doi: 10.1103/PhysRevX.4.031015. URL https://arxiv.org/abs/1402.3276.

[5] Benjamin Kraske, Qi Heng Ho, Federico Rossi, Morteza Lahijanian, and Zachary Sunberg. Robustness analysis of POMDP policies to observation perturbations, 2026. URL https://arxiv.org/html/2604.21256v1. arXiv:2604.21256v1, April23; relevant definitions and results inspected.

[6] Emrecan Kutay, Atilla Eryilmaz, and Ness Shroff. Finite-memory control of POMDPs: Fundamental limits and efficient design, 2026. URL https://arxiv.org/html/2611.03414v1. arXiv:2611.03414v1, November 2. Preprint; relevant model and examples inspected, theorem proofs not independently certified.

[7] Denis D. Mauá and Cassio P. de Campos. Solving decision problems with limited information. In *Advances in Neural Information Processing Systems 24*, 2011. URL https://proceedings.neurips.cc/paper/2011/file/f85454e8279be180185cac7d243c5eb3-Paper.pdf.

[8] Nicolas Meuleau, Leonid Peshkin, Kee-Eung Kim, and Leslie Pack Kaelbling. Learning finite-state controllers for partially observable environments. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 427–436, 1999. URL https://arxiv.org/abs/1301.6721. Author manuscript archived as arXiv:1301.6721.

[9] Sirin Nitinawarat, George K. Atia, and Venugopal V. Veeravalli. Controlled sensing for multihypothesis testing. *IEEE Transactions on Automatic Control*, 2013. URL https://arxiv.org/html/1205.0858v6. Primary author version, September4,2013.

[10] Joni K. Pajarinen and Jaakko Peltonen. Periodic finite state controllers for efficient POMDP and DEC-POMDP planning. In *Advances in Neural Information Processing Systems 24*, 2011. URL https://proceedings.neurips.cc/paper_files/paper/2011/file/1f3202d820180a39f736f20fce790de8-Paper.pdf.

[11] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x. URL https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf. Part I; corrected full reprint inspected, Section 6.

[12] Brian Theory. Observation topologies for finite propositional semantics: Distinguishability, distance, and inference, 2026. URL https://www.b-theory.com/manuscripts-html/logical-topology/. Working manuscript v0.9, reading copy 2; October 5 copyedits. Sections 3 and 8 inspected. Accessed November 13, 2026.

[13] Takehiro Tottori and Tetsuya J. Kobayashi. Memory-limited partially observable stochastic control and its mean-field control approach. *Entropy*, 24(11):1599, 2022. doi: 10.3390/e24111599. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9689590/.

[14] Daphney-Stavroula Zois and Urbashi Mitra. Nonlinear POMDPs for active state tracking with sensing costs, 2014. URL https://arxiv.org/html/1405.5892v1. arXiv:1405.5892v1; this preprint version cited rather than conflating later journal versions.