Grounded Abstraction

Identifiability, Representation Change, and the Limits of Learning

Komorebi P.
9 October 2026

First consolidated edition
Version 1.0.0

Abstract

A useful abstraction need not identify the system that realizes it. It must be justified by the evidence available to initialize, update, and correct it. This first edition connects finite evidence contracts, minimal corrective memory, and partially known linear dynamics. An exact signature preserves future candidate elimination. Its probabilistic counterpart retains joint target and likelihood-profile masses, while an explicit migration criterion identifies when a revised evidence language needs information the old memory cannot recover. For linear continuation, paired-state tests decide output agreement across fixed candidate systems. An invariant discrepancy span characterizes a shared observable quotient, and an initialization theorem gives the smallest common linear predictor accessible from the retained record. The quantitative extension replaces exact probability coordinates by a finite integer record, with separate likelihood, rounding, and horizon terms. A block comparison controls hidden-to-visible return and distinguishes exact invisibility from a small but unstable leakage channel. These bounds compose into one posterior decision certificate for the same evidence history and future response. The result is a theory of what incomplete experience can justify, supported by self-contained conditional proofs, counterexamples, and small exact illustrations. It is not a trained general learner or a certificate that its supplied model describes an external process.

Scope. Finite models and finite-dimensional linear systems, with explicit assumptions and self-contained proofs. Established ingredients are acknowledged. First edition freezes this finite theoretical scope. Independent peer review, a learned physical-semantic interface, and proof-assistant verification are not claimed.

Contents

Reading the first edition ii
First-edition claim and dependency guide iii

I Evidence before representation 1
1 What does experience justify? 1
2 A calibrated bit with an irrelevant hidden register 2
3 Laws, records, and reusable predictive state 3

II New questions and the memory needed to correct them 6
4 Reinterpretation, acquisition, and irrecoverable distinctions 6
5 A minimal memory for future correction 7

III Noisy evidence and quantitative justification 10
6 Identifiable does not mean certainly identified 10
7 An explicit radius for what remains unknown 11
8 A probability memory for repeated correction 13

IV Grounded continuation without complete identification 17
9 What a partial operator calibration determines 17
10 Why small local error needs an ambient growth bound 19
11 A smaller output is not automatically a weaker calibration requirement 20
12 A finite test for agreement under a fixed unknown model 21
13 When uncertain dynamics admit one common observable quotient 23
14 Initializing the smallest common linear predictor 25

V Quantitative grounding with finite retained memory 26
15 From exact probability coordinates to a finite record 26
16 Approximate invisibility must survive feedback 29
17 One justified decision from memory and dynamics 31

VI A consolidated grounding contract 34
18 A common theoretical contract for grounded revision 34
19 Open theoretical problems beyond the first edition 36
20 Conclusion 37
A Statement map and source relationships 38
B Small exact illustrations and reproduction 39
References 39

Reading the first edition

The central object is not an internal vector considered in isolation. It is a representation together with the evidence available to initialize it, the operations it supports, the questions it answers, and the future corrections it must permit. The paper asks which of those obligations can be met simultaneously.

Part I distinguishes identification of a hidden realization from justification of a reusable predictive state. Part II studies new questions, the cost of acquiring missing distinctions, and memory that preserves later correction. Part III separates certainty from noisy evidence, quantifies uncertainty, and constructs exact probability memory for future correction. Part IV compares full-state calibration, fixed-model output agreement, common observable quotients, and evidence-initialized linear prediction. Part V adds finite probability records, controlled visible leakage, and a common posterior decision certificate. Part VI consolidates revision requirements and the research boundary.

Relation to the existing collection. *Machine Understanding* develops supplied representations, explanatory artifacts, correction transport, and their common resource contracts [1]. *Self-Improving Intelligence* studies finite repair discovery and changes to discovery procedures [2]. *Sufficient Resolution* treats behavioral quotients, hidden memory, reconstruction, and calibration [3]. The present contribution is organized around a different source of uncertainty: the evidence may leave several underlying realizations possible. A representation is justified only to the extent that its intended uses survive that ambiguity. Earlier manuscripts are source relationships, not claims that every result here is already formally verified.

What is and is not proposed. Behavioral quotients, testing bounds, compatible-model arguments, invariant subspaces, and set-intersection updates have established mathematical foundations. Their appearance here is not a claim of priority. The proposed synthesis makes evidence access, corrective memory, and the identifiable continuation domain parts of the same representation contract. The exact correction-memory, likelihood-profile, calibrated-domain, and observable-realization constructions are derived in full within their stated models. The linear reductions use established reachability and observability ideas, while exact probability memory uses finite likelihood interpolation. The first-edition estimates use elementary Bayesian perturbation, rounding, and nonnegative matrix comparison. These arguments are not priority claims for their ingredients. Their integration requires one actual record and one consistent prediction contract.

Symbol	Meaning
$\omega \in \Omega$	A candidate world, including any fixed model or historical tag.
$K^{\pi}(h \mid \omega)$	Probability of observable record h under one permitted protocol.
$r : X \to Z$	Retained evidence or an available initial code. Its access is explicit.
F_a, q	A state transition and the task observation it must preserve.
$x \sim_C y$	Equality of every task observation after every permitted finite word.
$V \subseteq \Omega$	A set of candidate worlds, including the possible empty set.
$E \in \mathcal{E}$	An admitted evidence event used to filter V.
$\Sigma(V)$	The complete signature of future task-valued corrections.
U, S_a	A calibrated input subspace and a known operator restriction.
U_∞	The maximal domain on which calibrated continuation is determined.
I_Q	Initial vectors with common output continuations across fixed models.
E_∞	Propagated span of all model-discrepancy directions.
$p_{i,y}$	Joint probability of a likelihood profile and a target value.

First-edition claim and dependency guide

The first edition separates exact identification, quantitative approximation, and application to an external process. An argument may be complete within its stated model while the evidence that a physical or learned system satisfies that model remains unavailable.

Layer	Closed theoretical question	Premises that do not appear for free
Evidence	Does the record distinguish the future behavior required by the task?	Correct candidate worlds, admissible protocols, and actual record access.
Correction	Which joint distinctions must remain available for future filtering?	The evidence language, target, and a justified initial candidate set or prior.
Dynamics	Which outputs agree across compatible systems, and can one predictor realize them?	Shared observation semantics, a supplied model class, and enough initial evidence.
Approximation	How far can rounded memory and a reduced predictor deviate over the declared horizon?	A positive likelihood floor, model-error bounds, common block norms, and initial-state bounds.
Decision	Do the combined errors preserve a risk threshold or the chosen action?	One common history, response, posterior, and bounded Lipschitz loss.

The main exact dependency runs from the finite evidence criterion to predictive initialization, and from correction signatures to likelihood-profile memory and migration. A separate linear route runs from calibration through model agreement and observable quotienting to a minimum linear predictor. The quantitative route joins probability-memory error to an independently justified output allowance only at the final decision theorem.

Reading routes. For the central argument, read Sections 2, 5, 8, 13, 17, and 18. For proofs, follow the sections in order and retain the distinction between a fixed unknown model and explicitly permitted switching. For a review of the limits, compare the destructive-test, rare-evidence, hidden-return, and feedback counterexamples.

Three evidence levels. The mathematical claims have prose proofs. The reference programs check finite declared examples using exact arithmetic. External calibration, empirical model adequacy, and neural realization are not established by either a passing program or the edition number. The source package records assumptions, predecessor preservation, actual command results, and the depth of reference review.

What is frozen. The completed scope includes finite evidence models, static noisy correction, finite-dimensional linear continuation, finite retained probability fields, and a bounded-loss composition. State-changing noisy filters, general model invention, unbounded adaptive learning, and independently verified implementations remain subsequent research topics.

Part I

Evidence before representation

1 What does experience justify?

A device displays a bit. A reasoner learns to predict that display perfectly, including what happens when a permitted switch is pressed. It still may not know which physical bit the display describes. The code linking the display to the physical state could be reversed. A second, equally accurate reasoner might use the opposite internal convention. Prediction is real, yet the interpretation remains underdetermined.

That observation does not force a choice between calling every representation meaningless and treating every successful prediction as a complete understanding. It suggests a more precise question. What claims are common to all systems compatible with the evidence? Which of those claims can be realized by one representation that the observer can actually initialize and update? What additional information becomes necessary when the questions change?

1.1 Three separations

Identification of an underlying system, identification of a task-relevant property, and acquisition of a usable representation are distinct. A nuisance parameter can remain unknown while every required prediction is fixed. Conversely, knowing that some accurate representation exists in each possible world does not supply one representation that can be selected from the observer's record.

There is a second separation between a probability law and a finite observation drawn from that law. Two candidate models can have different laws while sharing every finite record in their supports. In that setting the models are identifiable at the level of their laws, but an individual finite record cannot yield error-free discrimination.

The third separation concerns time. A summary may preserve everything required for a present decision while losing the relationship needed to interpret a later measurement. More data arriving tomorrow need not repair what the memory policy erased today. This paper treats that relationship as part of the retained mathematical state.

1.2 Position among existing theories

Classical statistical identification separates parameters that induce different observational laws [4]. Predictive state representations use action-conditional predictions as a state description [5]. Query learning makes the teacher and the information it returns explicit [6]. These are neighboring answers to different parts of the present problem, not interchangeable definitions of grounding.

Data informativity is especially close to our organizing principle. It asks whether one property or one controller is justified for every system consistent with the data, without necessarily identifying a unique system [7]. Our later linear construction uses this compatible-extension viewpoint for the exact domain of repeated prediction. It is not offered as a replacement for data-driven control theory.

Causal abstraction concerns faithful relationships between mechanisms and their higher-level descriptions [8]. Here we ask what evidence makes such a description selectable and which continuations remain determined when parts of the underlying mechanism are unknown. We do not derive a learned encoder from the existence of a commuting equation. State-identification theory also distinguishes individually separable states from a realizable adaptive test [9]. That distinction will matter when measurements alter the system.

1.3 The intended contribution

The main results form a connected argument. Exact grounding is an information factorization through a behavioral quotient. New task distinctions are classified relative to an actual diagnostic interface. Future correction requires a richer memory signature than present target uncertainty. Partial operator calibration can nevertheless justify every future composition on a computable invariant domain. Approximate claims require both a local discrepancy and a bound on the dynamics that may amplify it.

The positive message is not that the observer must recover everything. It is that the observer can often recover exactly the structure its intended operations need, provided the evidence contract is stated honestly. The negative message is that an attractive coordinate system cannot manufacture missing distinctions, missing experiments, or missing bounds.

2 A calibrated bit with an irrelevant hidden register

Consider eight possible worlds

$$\omega = (s, \theta, k) \in \{0, 1\}^3.$$

The physical bit is s. The sensor convention is θ, and k is a nuisance bit that none of the admitted operations or questions uses. The display is

$$y = s \oplus \theta.$$

The ordinary action F flips s and leaves the other bits fixed. Its effect on the display is exactly $y \mapsto 1 - y$, whatever the unknown convention. Reading the current display therefore initializes a two-state representation that predicts every finite sequence of flips.

Neither s nor θ is identified separately. The two worlds $(0, 0, k)$ and $(1, 1, k)$ both display zero, and every flip sequence produces the same display in both. That ambiguity is harmless for the original display-prediction contract.

2.1 A new question

Now ask for the physical bit s as well as the display. The old representation is insufficient. The two previously equivalent worlds require different answers. A permitted nondestructive reference measurement returning θ makes the physical bit recoverable by $s = y \oplus \theta$. The enriched representation can retain (y, θ), or equivalently (s, θ). It has four states, not eight. The nuisance bit remains unidentified because no required continuation can detect it.

A new field named "sensor convention" does not provide the reference measurement. Nor does repeatedly flipping the physical bit resolve the ambiguity: the complete display history remains the same in the two worlds. Acquisition depends on an additional information channel, not merely a larger memory allocation.

2.2 A state-changing measurement

A different operation Z sets s to zero and then reports the new display. The report is θ. If the old display y_{old} was retained, then

$$s_{\text{old}} = y_{\text{old}} \oplus y_{\text{after } Z}.$$

The current physical state is zero, not s_{old}. A correct account distinguishes the recovered historical value from the new state it will act on.

If the old display was discarded, worlds with opposite old physical bits but the same θ reach the same current state and give the same future reports. The old physical bit is then irrecoverable through this unchanged interface. A known present state has replaced, rather than identified, the unknown past.

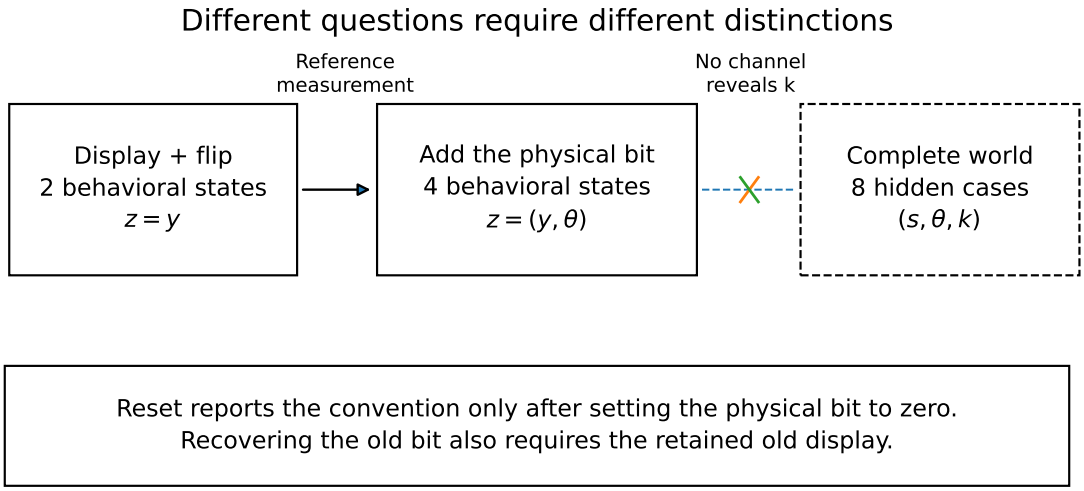


Figure 1: One finite sensor model supports different exact abstractions. The current display and flip action need two behavioral states. Adding the physical-bit question requires four, but not the irrelevant nuisance register. Calibration is an information source. Resetting is a physical change.

3 Laws, records, and reusable predictive state

Definition 3.1 (Finite evidence contract)**.** A finite evidence contract has a nonempty candidate set Ω, a family Π of permitted protocols, a finite record space H_π for each protocol, and kernels $K^\pi(h \mid \omega)$. A target $t : \Omega \to Y$ states what is to be identified. Any hidden model index that is fixed during a protocol is included in ω. The protocol's action choices are functions of its permitted records and independent randomization, not of the unobserved world.

Define $\omega \sim_\Pi \omega'$ when their record laws agree for every $\pi \in \Pi$. This relation concerns the complete records, including action choices when those are observable. Equality of individual observation marginals is insufficient.

Proposition 3.2 (Law identification and finite-record identification)**.** *The target is a function of the complete family of protocol laws if and only if it is constant on the $\sim_\Pi$ classes. For a fixed protocol π, an error-free rule $d : H_\pi \to Y$ exists if and only if t is constant on every nonempty support set*

$$B_\pi(h) = \{\omega : K^\pi(h \mid \omega) > 0\}.$$

Allowing a randomized output rule does not weaken this condition for probability-one correctness in every world.

Proof. For the first claim, factor t through the map sending a world to its tuple of protocol laws. A factor exists precisely when worlds with the same tuple have the same target. For the second claim, a record with positive probability in two worlds must be given the same output. Distinct targets on its support therefore prevent certainty. Conversely, assign each attainable record its support's common target. Every record generated in a world receives that world's target. A randomized error-free rule must place probability one on the required target at every positive-probability record. It cannot do so for two distinct targets at the same record. □

This elementary factorization does not supply the laws to the learner or bound the sample cost of estimating them. The two clauses answer different questions. A family of distinct laws can be identifiable in the first sense while every available short record remains ambiguous in the second.

3.1 A canonical meaning for an operational state

Let X be a finite nonempty set, let $\mathcal{A}$ be a finite operation alphabet, and let each $F_a : X \to X$ be total and deterministic. Let $q : X \to Y$ be a finite vector of required observations. Different candidate models can be included by taking a disjoint union of their state sets and preserving each model tag under every action.

For a word $w = a_1 \cdots a_m$, F_w applies a_1 first and a_m last. The empty word is the identity. Define

$$x \sim_C x' \quad \Longleftrightarrow \quad q(F_w x) = q(F_w x') \quad \text{for every finite } w.$$

This is a comparison in a common observation language. It does not require that different models share arbitrary latent coordinates. The behavioral class is a canonical response profile even when its implementation is not identifiable.

Definition 3.3 (Exact initialization from available evidence)**.** Given an available initial code $r : X \to H$, an exact reusable predictor consists of a finite state set Z, an initialization $d : H \to Z$, updates $G_a : Z \to Z$, and an output map $\bar{q} : Z \to Y$ such that

$$\bar{q}\, G_w\, d\, r(x) = q(F_w x)$$

for all x and all finite words. After initialization the predictor receives the operation names, but no additional world observations or hidden state.

Theorem 3.4 (Evidence sufficient for one exact predictive realization)**.** *An exact reusable predictor exists if and only if*

$$r(x) = r(x') \quad \Longrightarrow \quad x \sim_C x'. \tag{1}$$

When it exists, the behavioral quotient $X/\!\sim_C$ supplies a predictor with the fewest initialized states. Every predictor initialized for all $x \in X$ requires at least $|X/\!\sim_C|$ distinct values in $d(r(X))$.

Proof. Equal initial codes yield equal initialized states. The same operation sequence must then produce equal predicted outputs. Exactness forces equality of all true continuation outputs, proving necessity.

For sufficiency, note that $x \sim_C x'$ implies $F_a x \sim_C F_a x'$ because any subsequent word can be prefixed by a. The quotient consequently supports $G_a([x]) = [F_a x]$ and $\bar{q}([x]) = q(x)$. Condition (1) makes $d(r(x)) = [x]$ well defined on the attained initial codes. Arbitrary unused codes can be assigned any quotient state. Induction on word length gives the required equality. Finally, two behaviorally different states cannot have the same initialized predictor state, or a distinguishing continuation would receive the same prediction. Thus the number of initialized values is at least the quotient cardinality. □

The underlying quotient argument is classical and is developed elsewhere in the collection [1, 3]. Here its role is to attach a separate initialization condition to a common predictor across the surviving candidate worlds. A quotient can exist while the available record fails to initialize it.

3.2 Finite checking does not require an infinite list of words

Begin with the partition by q. Split any block whose members have successors in different current blocks under some action. A strict split increases the number of nonempty blocks, so at most $|X| - |q(X)|$ strict rounds occur. At a fixed point the quotient updates are well defined and preserve every later observation by induction. Any operation-stable refinement of the original output partition refines each round, so the terminal partition is the behavioral equivalence above. This supplies a finite computation when the complete candidate tables are given. It does not make their empirical acquisition free.

3.3 State identity is relative to the chosen questions

In the sensor example, $r = y$ satisfies (1) for display prediction. The quotient has two states even though eight worlds remain in the candidate universe. Adding physical-bit output refines the quotient to four classes and invalidates (1) for that same initial code. Adding the nuisance bit to the target would refine again to eight, but no admitted sensor action reveals it.

Only adding requirements yields a refinement relation automatically. Replacing one task with a different task can produce crossing partitions rather than a finer version of the old one. For example, the partition by display and the partition by physical bit are neither finer nor coarser than one another. A claim of cumulative learning must specify which old questions remain protected.

3.4 An available predictor is stronger than a worldwise predictor

Consider two immutable worlds, one with a permanently zero output and one with a permanently one output. The available initial record is the same symbol in both. Within each world separately, a one-state predictor exists: it simply prints that world's output. There is nevertheless no predictor that can be initialized from the common record and be correct in both. The obstruction is the order of quantifiers,

$$\forall \omega \; \exists P_\omega \qquad \text{versus} \qquad \exists P \; \forall \omega .$$

The first does not provide a selector for the second. The behavioral quotient of the combined family has two states, but the evidence cannot choose between them. Adding an output observation repairs the initialization problem without changing the mathematical existence of either predictor.

Likewise, a coordinate convention is not an observation. Relabeling a latent state and conjugating its updates can leave every operational prediction unchanged. Without a separately meaningful reference, evidence about those predictions need not favor one coordinate system. A canonical response profile avoids claiming that an arbitrary coordinate name was discovered as a physical entity. It does not prove that the reasoner's actual implementation uses the inferred profile causally.

Thus two distinct questions remain after fitting a predictor. Is it a common valid representation of the possible worlds, and does the observer possess the evidence needed to initialize it? The theorem answers their exact conjunction in a finite model. Efficient acquisition and causal attribution require further premises.

Part II

New questions and the memory needed to correct them

4 Reinterpretation, acquisition, and irrecoverable distinctions

A revised task often invites the vague instruction to learn a better representation. That instruction hides three different possibilities. The old code may already contain everything needed. A new observation may be required. Or the remaining interface may be unable to recover the needed distinction at all.

We first separate these cases in a setting where observations do not interfere with one another. Let Ω be finite, let $r : \Omega \to Z$ be retained evidence, and let $t : \Omega \to Y$ be the revised target. Let $\mathcal{E} = \{e_1, \dots, e_m\}$ be a finite family of total, deterministic, nondestructive tests. Each test reads the same fixed world and leaves it unchanged. All tests remain available throughout the investigation. Equivalently, a reset may be allowed if it genuinely restores the same unknown world, including the relevant history. Resetting into a new known state does not satisfy this premise.

For $J \subseteq \{1, \dots, m\}$, write

$$R_J(\omega) = \big(r(\omega), (e_j(\omega))_{j \in J}\big).$$

Let the unresolved target pairs be

$$D = \{\{\omega, \omega'\} : r(\omega) = r(\omega'),\ t(\omega) \neq t(\omega')\}.$$

A test covers such a pair when it gives different reports in the two worlds.

Theorem 4.1 (Exact revision trichotomy for a reusable test interface)**.** *Exactly one of the following applies.*

1. *D is empty. A new decoder of the old code recovers t without new observations.*
2. *D is nonempty, but every pair in D is separated by at least one admitted test. Executing all tests recovers t, and a selected nonadaptive set J suffices exactly when it covers D.*
3. *Some pair in D has identical reports under every admitted test. No finite adaptive protocol using this unchanged test interface and r guarantees t in every world.*

If test j has nonnegative fixed cost c_j, the minimum cost of one fixed nonadaptive test set, used for all initial codes in case 2, is precisely

$$\min_{J:\, J \text{ covers } D} \sum_{j \in J} c_j. \tag{2}$$

This is not a claim that an optimal adaptive protocol has the same cost.

Proof. If D is empty, the target is constant on each old-code fibre, giving a decoder. Otherwise a set J permits a decoder exactly when no pair with the same old code and different target remains equal on all its test reports. That is exactly pair coverage. In particular, the full finite test family succeeds when every unresolved pair has a separating test.

If a pair survives every test, induction on an adaptive history shows that the two worlds give the same report after each chosen action. The protocol chooses the same next test because its available history is the same. Its stopping decision and output are therefore the same, so it cannot be correct on both targets. Independent randomization does not restore a probability-one guarantee, by conditioning on its random tape. The three cases are exclusive and exhaustive. Formula (2) follows from the equivalence between successful nonadaptive test sets and covers, with exactly the stated additive charges. □

The cover formulation exposes a concrete resource question without making a large learner part of the theorem. It also identifies what a negative certificate says. An uncovered pair is impossible to resolve through the declared interface. It is not impossible to resolve after a new sensor, retained record, or physical intervention changes that interface. Finding the cheapest cover can itself be difficult. Its formulation is not an efficient universal discovery algorithm.

4.1 Why experimentwise evidence need not compose

Counterexample 4.2 (Pairwise separation without a successful investigation)**.** There are three possible initial cases $0, 1, 2$. Test a reports whether the case is 0, and test b reports whether it is 1. Either first test then destroys the accessible case state. Every later test returns a fixed zero report. The target remains the initial case, stored only as an inaccessible historical tag.

Every pair is distinguishable by some permitted first test. Yet beginning with a leaves cases 1 and 2 indistinguishable whenever its answer is no. Beginning with b leaves cases 0 and 2 indistinguishable. No finite adaptive protocol identifies every initial case. Randomizing the first test cannot guarantee success in every case either.

The missing premise is not intelligence. It is the ability to combine evidence from mutually destructive inquiries. State-identification work studies related restrictions on adaptive tests [9]. The example makes clear why the target map, physical transition, and record cannot be separated in a general acquisition problem. The full state-changing repair constructions in *Self-Improving Intelligence* address that different model [2].

4.2 The sensor revisited

For physical-bit recovery, an old display and a nondestructive reference measurement satisfy case 2. Without the reference, flip sequences leave the convention ambiguity unresolved. For recovery of the pre-reset physical bit after losing the old display, the surviving states can have the same current physical state, convention, and every future sensor response while having opposite historical targets. That is case 3 relative to the post-reset interface.

A learner should record which kind of failure occurred. A failed search over decoders cannot establish that no new experiment would help. Conversely, repeatedly searching for a decoder cannot reconstruct a target that is not a function of its inputs. Formula length and information access are separate resources.

5 A minimal memory for future correction

The old record can matter even when it does not change today's set of possible answers. The essential information may lie in a relation between a possible answer and an observation that will arrive later.

Let Ω be a finite model set and $t : \Omega \to Y$ a fixed target. For a candidate set $V \subseteq \Omega$, let

$$Q(V) = t(V) = \{t(\omega) : \omega \in V\}.$$

The empty set is permitted. It represents a contradiction between the admitted model family and the evidence, not a successful identification. Let $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ be a finite family of evidence events containing Ω and closed under intersections. Filtering by E sends V to $V \cap E$. The events are known, truthfully interpreted, and do not change the underlying model in this section.

Definition 5.1 (Correction signature)**.** The correction signature of V is

$$\Sigma_{\mathcal{E},t}(V) = \big(Q(V \cap E)\big)_{E \in \mathcal{E}}. \tag{3}$$

Two candidate sets have the same signature when every admitted evidence filter leaves the same possible target values, including whether no candidate remains.

Theorem 5.2 (The coarsest exact memory for admitted corrections)**.** *Equality of correction signatures is a congruence for all updates $V \mapsto V \cap E_0$. The signature therefore supports well-defined updates and a readout of $Q(V)$. It preserves that readout after every finite sequence of admitted filters.*

Conversely, any retained representation with deterministic filter updates and a readout reproducing Q after every such sequence must distinguish every two candidate sets with different signatures. Thus the image of $\Sigma_{\mathcal{E},t}$ is the smallest exact state space, in cardinality, for this correction contract on all candidate subsets.

Proof. If $\Sigma(V) = \Sigma(W)$, then for every $E \in \mathcal{E}$,

$$Q((V \cap E_0) \cap E) = Q(V \cap (E_0 \cap E)) = Q(W \cap (E_0 \cap E)).$$

Intersection closure puts $E_0 \cap E$ in $\mathcal{E}$, so the equality is available from the original signatures. It proves $\Sigma(V \cap E_0) = \Sigma(W \cap E_0)$. The $E = \Omega$ coordinate gives $Q(V)$, and induction gives all finite filter sequences.

If two signatures differ, one event E produces different target sets. A retained representation that merged the two candidate sets would update identically on that event and return the same readout, contradicting exactness. Every distinct signature therefore requires a distinct retained value. The signature itself attains that count with the updates just constructed. □

This is a behavioral quotient of a powerset system. It preserves later task-valued correction, not merely the present prediction. Candidate identities can be discarded when no admitted correction distinguishes them. Contradiction with the entire candidate set remains observable.

5.1 An exact count when evidence identifies a partition cell

Suppose $A_1, \ldots, A_m$ are a partition of Ω into nonempty cells, and $\mathcal{E}$ consists exactly of all unions of those cells. Define the attainable evidence–target pairs

$$D_t = \{(i, y) : y \in t(A_i)\}, \qquad r = |D_t| = \sum_i |t(A_i)|.$$

A candidate set induces an occupancy record

$$O(V) = \{(i, t(\omega)) : \omega \in V \cap A_i\}.$$

Corollary 5.3 (Exact correction-memory size for partition evidence)**.** *Two candidate sets have equal correction signatures if and only if their occupancy records agree. On the domain of all subsets $V \subseteq \Omega$, there are exactly 2^r distinct signatures. A fixed-length exact binary memory therefore requires at least r bits, and an r-bit occupancy vector attains this bound.*

Proof. Filtering by cell A_i returns exactly the target values occurring in the ith part of $O(V)$. Thus equal signatures imply equal occupancy. Conversely, the target set after filtering by any union of cells is the union of their occupied target values, so occupancy determines every signature coordinate.

Every subset of D_t is realized: for each included pair choose one world in its nonempty intersection $A_i \cap t^{-1}(y)$, and choose no worlds for excluded pairs. These intersections are disjoint for different pairs. Hence the occupancy map has 2^r values. Distinct values need distinct memory states by the preceding theorem. Binary counting gives the lower bound, and direct occupancy bits attain it. □

A particular inquiry may reach fewer candidate sets, reducing the number of attained occupancy records. The bound counts logical states, not interpreter bytes, update work, or evidence acquisition.

The exact evidence family matters. Two worlds a, b in one cell with the same target have identical singleton occupancies. Adding the event $\{a\}$ distinguishes them: one filtered singleton retains the target and the other is empty. A new event that splits a cell can therefore require finer memory.

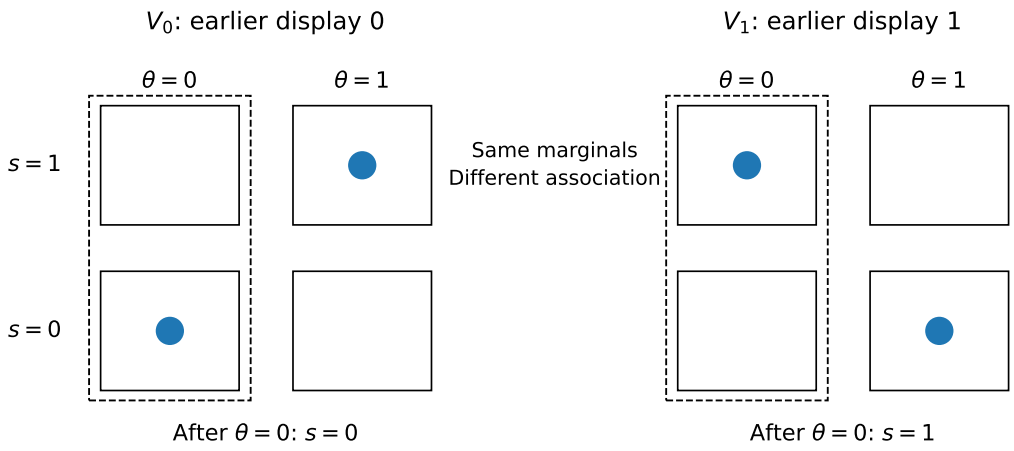


Figure 2: Two evidence histories preserve identical physical-bit and convention marginals. Their joint possibility patterns differ. A reference report selects a column, producing opposite physical-bit conclusions. The nuisance bit is omitted because it changes neither pattern.

5.2 Association can be the missing state

Counterexample 5.4 (Equal marginals with different corrected answers)**.** In the sensor model, let

$$V_0 = \{(s, \theta, k) : s \oplus \theta = 0\}, \qquad V_1 = \{(s, \theta, k) : s \oplus \theta = 1\}.$$

Take the target to be s and the evidence cells to be the two values of θ. Both candidate sets currently allow physical bits $\{0, 1\}$, and both allow conventions $\{0, 1\}$. Yet after the reference reports $\theta = 0$, V_0 requires $s = 0$ and V_1 requires $s = 1$.

Retaining both marginal possibility sets loses the association needed to use the next measurement. Their equality is not a sufficient criterion for merging the records.

Here the nuisance bit duplicates every attainable (θ, s) pair. There are $2^8 = 256$ raw candidate subsets, but only $2^4 = 16$ correction signatures. Four occupancy bits preserve every admitted future correction. The present physical-target set alone has only four possibilities, requiring two bits, but that smaller code is inadequate for the correction contract.

Four association bits suffice where eight raw membership bits are redundant and two marginal bits are inadequate.

5.3 Correction memory has a scope

The signature is minimal for its specified target and evidence family. New targets, events, candidate models, state-changing observations, or revised event meanings can split its classes. Old minimality remains a statement about the old contract.

If a future question asks for k, the occupancy is insufficient. When every available measurement ignores k, expanding the code cannot recover it. A finer signature can exist mathematically without being obtainable from the current interface.

Part III

Noisy evidence and quantitative justification

6 Identifiable does not mean certainly identified

The earlier exact filters remove a world only when it cannot produce the observation. With noisy channels that criterion can retain every candidate indefinitely. A statistical procedure may nevertheless become increasingly reliable. These are different guarantees, not contradictory descriptions of learning.

For a fixed protocol, let two worlds with different targets generate finite transcript laws P_0 and P_1. An estimator sees only that transcript and independent randomness. Its average success under equal prior probabilities is bounded by the discrimination in the entire record.

Proposition 6.1 (A finite-record discrimination bound)**.** *For two distinct target values and equal world probabilities, every transcript-based estimator has average success at most*

$$\frac{1 + d_{\mathrm{TV}}(P_0, P_1)}{2}, \qquad d_{\mathrm{TV}}(P_0, P_1) = \frac{1}{2} \sum_h |P_0(h) - P_1(h)|. \tag{4}$$

A deterministic function or a common stochastic processing channel applied to the transcript cannot increase this upper bound.

Proof. Convert the estimator into a binary test by reporting world zero when it outputs the first target and world one otherwise. A correct target output always gives a correct binary decision. If $f(h) \in [0, 1]$ is the test's chance of declaring world zero, its success is

$$\frac{1}{2}\left(1 + \sum_h f(h)(P_0(h) - P_1(h))\right).$$

The sum is at most the sum of positive differences, which equals total variation. This proves (4). If $B(h, z)$ is a common processing channel and $g(z) \in [0, 1]$ a processed-record test, then $f(h) = \sum_z B(h, z)g(z)$ is a raw-record test with the same two expectations. Thus every processed decision was already available from the original record. □

This is an elementary testing and data-processing argument, related to the classical comparison of experiments [10]. It does not identify the best protocol, establish the correctness of a sensor model, or make independent randomness into additional evidence. If an adaptive protocol is compared, its complete history must be the transcript in (4).

Counterexample 6.2 (Distinct laws with no finite certain identification)**.** Let the unknown world be $\omega \in \{0, 1\}$. A repeated observation is conditionally independent Bernoulli with success probability $1/4$ in world zero and $3/4$ in world one. The two one-observation laws differ, so the world is identified by its law. Every finite binary string nevertheless has positive probability under both worlds. No fixed finite sample supplies an error-free world classifier.

For odd n, majority voting is the likelihood-ratio rule. Its error in either world is

$$e_n = \sum_{j=0}^{(n-1)/2} \binom{n}{j} (3/4)^j (1/4)^{n-j}.$$

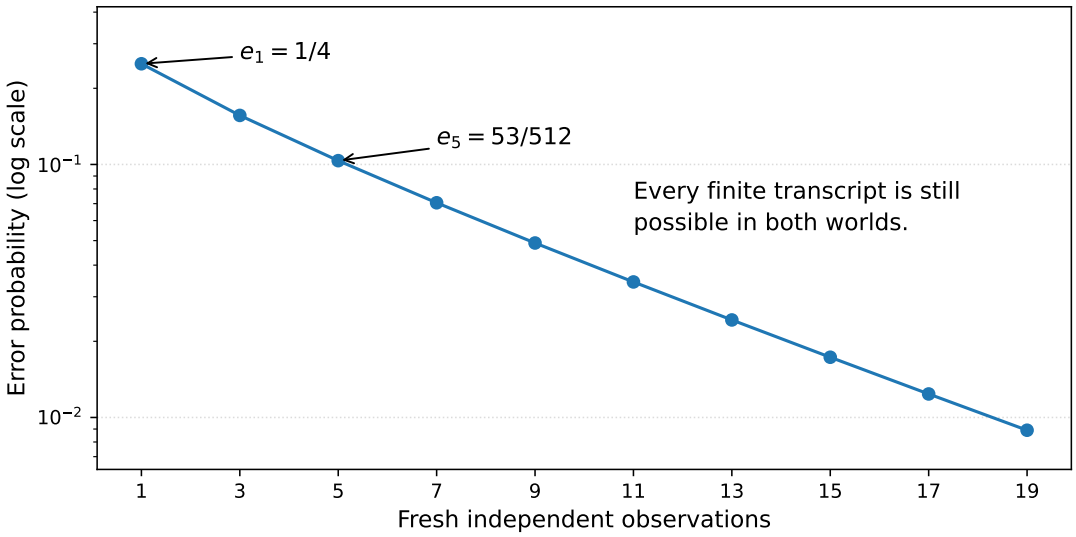


Figure 3: Exact majority-test error for two stipulated Bernoulli worlds. The laws differ, and repeated independent evidence reduces error, but every finite record remains compatible with both worlds. The curve is evaluated from the displayed binomial formula.

The first values are $e_1 = 1/4$, $e_3 = 5/32$, and $e_5 = 53/512$. These are exact analytical probabilities, not observed error rates from an agent experiment.

The example illustrates why logical candidate elimination and statistical confidence should not share a Boolean field named "known." Both candidate worlds remain in the support of every fixed finite record. Their relative likelihoods can become very different. Treating the most likely world as certain discards the distinction that the exact support test preserves.

6.1 Identifiability is not a uniform rate

Distinct models may induce arbitrarily similar laws. Unless a model class imposes a separation condition, a bounded number of observations need not distinguish every pair reliably. The total-variation term exposes that obstruction directly. A data-dependent summary cannot increase the discrimination contained in the actual record.

Conversely, a small class with known, well-separated distributions can admit a useful finite-sample procedure even when exact certainty is impossible. The statistical question is then how much error is acceptable and which assumptions justify its probability bound. We do not add a sample-complexity claim merely by labeling a model family finite. Separation, admissible protocols, and the sampling law have to enter the argument.

7 An explicit radius for what remains unknown

A set of candidate worlds may disagree on some future outputs while agreeing closely on others. It is useful to quantify that disagreement without selecting one world as if it had been identified.

Let $v : \Omega \to \mathbb{R}^d$ record a fixed finite collection of target quantities, where $d \geq 1$. These can be future predictions on specified operation words. For a nonempty $B \subseteq \Omega$, define

$$\ell_j(B) = \min_{\omega \in B} v_j(\omega), \qquad u_j(B) = \max_{\omega \in B} v_j(\omega),$$

$$m_j(B) = \frac{\ell_j(B) + u_j(B)}{2}, \qquad \rho(B) = \frac{1}{2}\max_j\big(u_j(B) - \ell_j(B)\big). \tag{5}$$

The radius concerns coordinatewise prediction in a common frame. It is not an uncertainty measure invariant under arbitrary nonlinear reparameterization.

Theorem 7.1 (Ambiguity radius and an explicit prediction certificate)**.** *For every nonempty finite B,*

$$\inf_{z\in\mathbb{R}^d}\max_{\omega\in B}\|z - v(\omega)\|_\infty = \rho(B),$$

and $m(B)$ attains the infimum. If $B' \subseteq B$ is nonempty then $\rho(B') \le \rho(B)$.

Fix nonnegative tolerances η, ε and $\delta_{\mathrm{id}}, \delta_{\mathrm{model}} \in [0,1]$. Suppose a data record H produces a candidate set $B(H)$ with

$$\Pr\{\omega_* \in B(H)\} \ge 1 - \delta_{\mathrm{id}}.$$

Suppose a second event, of probability at least $1 - \delta_{\mathrm{model}}$, guarantees

$$\sup_{\omega\in B(H)}\|v(\omega) - \widehat{v}_H(\omega)\|_\infty \le \eta.$$

Read the last supremum as zero when $B(H)$ is empty. When $B(H)$ is nonempty, form the midpoint $\widehat{m}$ and radius $\widehat{\rho}$ from $\widehat{v}_H$. If the returned numerical vector z satisfies $\|z - \widehat{m}\|_\infty \le \varepsilon$, then

$$\|z - v(\omega_*)\|_\infty \le \widehat{\rho} + \eta + \varepsilon \tag{6}$$

on an event of probability at least $1 - \delta_{\mathrm{id}} - \delta_{\mathrm{model}}$. The rule abstains on an empty candidate set. No independence between the two stated events is required.

Proof. For each coordinate j, any point z_j is at distance at least $(u_j - \ell_j)/2$ from one of the two extremes. Taking the largest coordinate width gives the lower bound. The midpoint is within half of every coordinate's width of every candidate, so it attains that bound. Inclusion can only narrow the coordinate ranges.

On the intersection of the coverage and model-accuracy events, the true world lies in the nonempty set and

$$\|z - v(\omega_*)\|_\infty \le \|z - \widehat{m}\|_\infty + \|\widehat{m} - \widehat{v}_H(\omega_*)\|_\infty + \|\widehat{v}_H(\omega_*) - v(\omega_*)\|_\infty.$$

The three terms are bounded by ε, $\widehat{\rho}$, and η. The union bound gives the probability statement. □

The center is allowed to be any vector in $\mathbb{R}^d$. Requiring it to be the prediction of one admissible model can increase the optimum. Nor does a coordinatewise midpoint automatically define a coherent stochastic process or preserve logical invariants. Those are additional constraints on the permitted approximation, not consequences of (5).

7.1 Three errors and two failure probabilities

The radius measures unresolved disagreement among retained candidates. The model discrepancy controls how well their represented predictions match the intended quantities. Numerical error concerns the computation of the returned answer. The coverage probabilities express the chance that the premises fail. They are not added to the numerical tolerance as if they had the same units.

The probability assertion is unconditional over the stated evidence procedure. It is not automatically the same bound conditional on a rare event such as issuance of a certificate. Likewise, repeatedly changing the candidate family or stopping rule requires a validity argument for the actual procedure. A confidence set cannot be assumed valid merely because its contents look plausible.

7.2 What compression does to the candidate set

For a fixed raw-record channel, let $B(h)$ be its exact support set. If only $c(h)$ is retained, the correct support is

$$B_c(z) = \bigcup_{h:c(h)=z} B(h).$$

Indeed, a world can produce the coarse record exactly when it can produce at least one raw record in that fibre. Thus $B(h) \subseteq B_c(c(h))$, and the ambiguity radius cannot decrease for the same target vector. This is an information statement, not a command to retain every byte. The correction signature in Section 5 identifies one principled way to compress while retaining the specified future uses.

For general confidence sets rather than exact supports, taking the same union still preserves the original coverage guarantee. It may be conservative. Selecting a smaller coarse-record set requires a new coverage argument, not the assertion that compression should make the model simpler.

Counterexample 7.2 (Agreement inside a misspecified family)**.** Suppose every candidate in a supplied family predicts zero for the target, but the actual unmodeled process produces one. The family has radius zero, and its internal arithmetic can be exact. The conclusion is nevertheless wrong because the true process is outside the represented family. Neither a narrow interval nor a successful proof about the family establishes that membership premise.

This is why model adequacy is not treated as a reward for confidence. In the present paper it is a stated premise, or a separate statistical event, rather than an automatic consequence of a concentrated representation.

8 A probability memory for repeated correction

The exact signature in Section 5 records which candidates remain possible. With noisy observations, all candidates can remain possible while their relative weights change substantially. Keeping only their support would then preserve logical possibility but discard the result of Bayesian correction. We now specify a different memory contract: retain both the probabilities of future reports and the posterior distribution of the selected target.

Let Ω be finite and nonempty, let $t : \Omega \to Y$ have finite image, and let $\mathcal{B}$ be a finite collection of experiments. Experiment e has a finite outcome set and specified likelihoods

$$\ell_{e,o}(\omega) \geq 0, \qquad \sum_o \ell_{e,o}(\omega) = 1.$$

The world is fixed. Every experiment remains available, and each new report has this conditional law given the world, the chosen experiment, and all earlier reports. This is a fresh conditional-observation premise, not a claim that copied readings become new samples. No experiment changes the physical state in this section.

Partition Ω into the nonempty cells $A_1, \ldots, A_m$ on which the complete likelihood vector $(\ell_{e,o}(\omega))_{e,o}$ is constant. Write its common value as $\lambda_{e,o,i}$. The profile vectors of different cells are distinct by construction. For a prior μ, retain

$$p_{i,y}(\mu) = \sum_{\omega \in A_i,\ t(\omega)=y} \mu(\omega), \qquad (i,y) \in D_t = \{(i,y) : A_i \cap t^{-1}(y) \neq \varnothing\}.$$

These are joint masses, not the two separate marginals. There are $r = |D_t|$ coordinates with nonnegative entries summing to one. Exact real probabilities are not a fixed r-bit code. A bit-complexity claim would also have to account for their representation and growing denominators.

Theorem 8.1 (Exact target correction from likelihood-profile memory)**.** *Under the fixed-world likelihood model, the record $p = (p_{i,y})$ supports every finite sequence of report predictions and target-posterior corrections. Its next-report probability and update are*

$$z_{e,o} = \sum_{i,y} \lambda_{e,o,i} p_{i,y}, \qquad p'_{i,y} = \frac{\lambda_{e,o,i} p_{i,y}}{z_{e,o}} \quad (z_{e,o} > 0),$$

and its target readout is $\Pr(t = y) = \sum_i p_{i,y}$. A zero-probability report is marked impossible rather than assigned a normalized posterior.

Conversely, any deterministic memory, initialized for every prior on Ω, that preserves all next-report probabilities and all target posteriors after every finite history must distinguish priors with different joint-mass records. In fact, the joint probabilities of the target and all experiment-report words of length at most $m - 1$ already determine the joint-mass record. The memory is therefore minimal in the sense of its induced equivalence on priors, not in machine words or physical storage.

Proof. Likelihoods are constant on each profile cell, so summing the ordinary Bayesian update over $A_i \cap t^{-1}(y)$ gives the displayed update. Its normalization and target readout are exact. Repeating this argument proves sufficiency for any fixed experiment sequence, and also for choices made from the retained report history without extra world-dependent information.

For necessity, a valid predictor and updater determine, for every word $w = ((e_1, o_1), \ldots, (e_k, o_k))$, the masses

$$M_{\mu,y}(w) = \sum_{t(\omega)=y} \mu(\omega) \prod_{j=1}^{k} \ell_{e_j,o_j}(\omega).$$

Indeed, multiply the successive predicted report probabilities by the final target posterior. If an earlier factor is zero, the joint mass is zero and no impossible conditional posterior is needed. The empty word gives the current target distribution.

For each cell i and each $j \neq i$, choose one likelihood coordinate $b(i, j) = (e, o)$ for which $\lambda_{b(i,j),i} \neq \lambda_{b(i,j),j}$. Such a coordinate exists because profile vectors are distinct. On Ω the polynomial

$$P_i(\omega) = \prod_{j \neq i} \frac{\ell_{b(i,j)}(\omega) - \lambda_{b(i,j),j}}{\lambda_{b(i,j),i} - \lambda_{b(i,j),j}}$$

is exactly $\mathbf{1}_{A_i}(\omega)$. It is a linear combination of likelihood monomials of degree at most $m - 1$. Each monomial is the likelihood product of an admitted fixed experiment-report word. Consequently $p_{i,y}(\mu) = \sum_{t(\omega)=y} \mu(\omega) P_i(\omega)$ is a linear combination of the corresponding $M_{\mu,y}(w)$. For $m = 1$ the empty product is one and the same argument uses only the current target masses. Two priors merged by an exact memory have equal predicted joint masses, hence equal joint-mass records. This proves the claimed separation and finite word-length bound. □

The proof uses a finite interpolation argument. Its polynomial coefficients are an identification device, not negative-probability experiments that a learner must execute. All likelihood values are supplied exactly. Very similar profile vectors can make the interpolation numerically ill-conditioned, so the theorem does not provide a stable estimator of p from noisy frequency estimates. Nor does it infer the true likelihood model from a narrow posterior. The distinction between sufficient experiments and compressed statistical evidence is classical [10]; the present contract additionally keeps the selected target and all future update operations explicit.

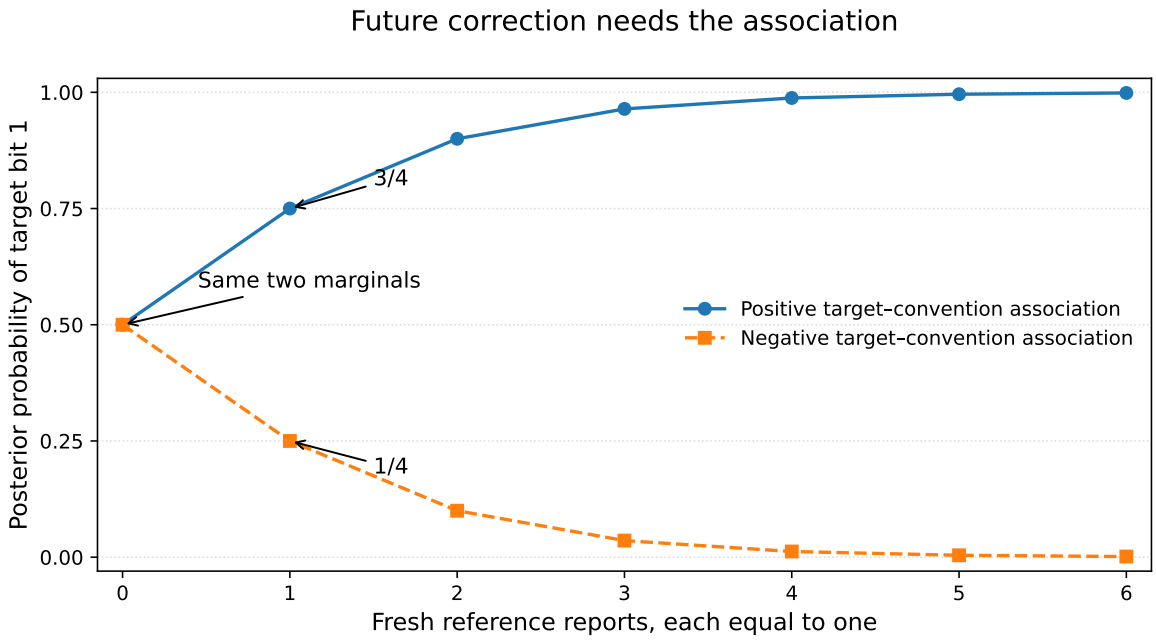


Figure 4: Exact posterior target probabilities for two priors with identical target and sensor marginals but opposite association. Each horizontal step is a fresh noisy-reference report equal to one. The curves are conditional analytical values, not measured learning performance. Four joint masses support the update; two separate marginals do not.

8.1 Two identical marginals, two opposite corrections

Return to the sensor worlds $\omega = (s, \theta, n) \in \{0,1\}^3$, where s is the target bit, θ is a convention, and n is irrelevant. Suppose the only future experiment is a nondestructive noisy reference with

$$\Pr(o = 1 \mid \theta = 0) = \tfrac{1}{4}, \qquad \Pr(o = 1 \mid \theta = 1) = \tfrac{3}{4}.$$

There are two likelihood-profile cells, one for each convention, and four nonempty profile–target pairs. The eight-world prior can be reduced to four joint masses without losing any specified prediction or correction. The nuisance coordinate disappears.

Consider a prior μ_+ assigning mass $1/2$ to $(s, \theta) = (0,0), (1,1)$ and a prior μ_- assigning mass $1/2$ to $(1,0), (0,1)$, with any fixed nuisance convention. Both have uniform s and uniform θ. Both predict the next report 1 with probability $1/2$. After that report, however,

$$\Pr_{\mu_+}(s = 1 \mid o = 1) = \tfrac{3}{4}, \qquad \Pr_{\mu_-}(s = 1 \mid o = 1) = \tfrac{1}{4}.$$

After k fresh reports all equal to one, these become $3^k/(1 + 3^k)$ and $1/(1 + 3^k)$. The difference is not a numerical error. It is the association that the separate marginals failed to retain. At every finite k, both target values remain possible under either prior.

8.2 What a revised evidence language requires

Let $\mathcal{P}_0$ be the partition of worlds into the nonempty old profile–target cells. Let $\mathcal{P}_1$ be the corresponding partition after likelihoods or the target have changed, still on the same finite Ω. The old record stores the probability of each $\mathcal{P}_0$ cell. The new record would store the probability of each $\mathcal{P}_1$ cell.

Corollary 8.2 (Exact migration of a probability memory)**.** *A deterministic map reconstructing the new record from the old record for every prior exists if and only if every $\mathcal{P}_0$ cell is contained in a single $\mathcal{P}_1$ cell. When this condition holds, summing the relevant old masses gives the new record. If the new evidence or target strictly splits an old cell, no such universal reconstruction exists.*

Proof. When every old cell lies inside a new one, each new mass is the sum of the old-cell masses it contains. Conversely, suppose two worlds lie in the same old cell but in different new cells. Point-mass priors on those worlds yield the same old record and distinct new records. No deterministic map can send that one old record to both required new records. □

Counterexample 8.3 (An old exact memory does not upgrade itself)**.** Two worlds have the same target and the same old report law, so point masses on them give an identical old record. A newly admitted experiment reports one with probabilities $1/4$ and $3/4$ in the respective worlds. The old record cannot predict its probability, although it was exact for every old question and update. Adding a field for the new likelihood profile does not determine its contents.

A restricted family of reachable priors can admit a reconstruction that fails for arbitrary priors. That restriction must be stated. Likewise, an archived raw record, a justified prior relation, or fresh evidence may help initialize the new memory, but its adequacy must be checked separately. Enlarging the candidate-world set is a different operation again: the old probabilities do not decide how much mass newly admitted worlds should receive.

8.3 Exact sufficiency is not a finite-precision guarantee

Even when two profiles differ only slightly, enough repeated evidence can make their posterior effects very different. Near-equal likelihoods are not interchangeable for an unbounded continuation contract. A finite-precision theory needs a horizon, a metric, a likelihood or evidence-mass condition, and a policy for outcomes of very small probability. Section 15 supplies a finite-horizon approximation contract with an explicit likelihood floor and retained-record denominator. It does not make rounding an exact sufficient statistic or provide an unbounded-time guarantee.

Part IV

Grounded continuation without complete identification

9 What a partial operator calibration determines

The finite-world examples show that complete identification can be unnecessary. A linear setting makes the positive result more structural. We will determine the exact domain on which every future operation sequence is already fixed by partial calibration data.

Let V be a finite-dimensional vector space over a field, and let $\mathcal{A}$ be a finite nonempty set of operation names. The unknown operations are linear maps $T_a : V \to V$. Calibration supplies input vectors $x_1, \dots, x_m$ and exact outputs $y_{a,1}, \dots, y_{a,m}$ for every a. The operator family is fixed, rather than being reselected after each action. Let

$$U = \operatorname{span}\{x_1, \dots, x_m\}.$$

The calibration is consistent precisely when every linear relation among the inputs is respected by the outputs, for each action. Under that condition it defines maps $S_a : U \to V$ with $S_a x_i = y_{a,i}$. Every S_a has linear extensions to V by extending a basis. Unless stated otherwise, *all* such extensions are admissible. There is no hidden norm bound, symmetry, invertibility condition, or preferred matrix completion.

The compatible-model viewpoint is shared with the data-informativity literature [7]. Our goal here is exact full-state prediction from a known initial vector, not stabilization, optimal control, or prediction from an unmeasured initial condition.

Proposition 9.1 (One-step identification from exact calibration)**.** *For a fixed action a, the value $T_a v$ is the same for every linear extension of S_a if and only if $v \in U$. The common value is then $S_a v$.*

Proof. If $v \in U$, all extensions agree by definition. If $v \notin U$, choose a linear functional ℓ that vanishes on U and satisfies $\ell(v) = 1$, and choose a nonzero vector $b \in V$. Given any extension T_a, the map $T'_a(x) = T_a(x) + \ell(x)b$ is another extension, but $T'_a v = T_a v + b \neq T_a v$. Such a functional exists by extending a basis of U with v. □

The proposition uses the complete unrestricted extension class. A justified structural prior can determine more. For example, knowing an operator belongs to a smaller parametric family can constrain its action outside U. Those constraints must be included as prior information, not inferred from the calibration span alone.

9.1 A known first step can leave the known domain

Even if $v \in U$, the vector $S_a v$ can lie outside U. A second action then acts at an input whose response was not calibrated. One-step accuracy is therefore insufficient for reusable operator meaning.

Define a descending sequence of subspaces

$$U_0 = U, \qquad U_{j+1} = \{v \in U : S_a v \in U_j \text{ for every } a \in \mathcal{A}\}. \tag{7}$$

The sequence is descending by induction. Every strict inclusion decreases dimension, so a fixed point is reached after at most $\dim U$ strict drops. Denote it by U_∞.

Theorem 9.2 (The maximal calibrated continuation domain)**.** *For every integer $n \geq 1$ and every $v \in V$, the values $T_w v$ for all operation words of length at most n are independent of the chosen compatible extension family if and only if*

$$v \in U_{n-1}.$$

Consequently, every finite full-state continuation is identifiable exactly on U_∞. This is the largest subspace $L \subseteq U$ such that $S_a L \subseteq L$ for every a. On this subspace the known restrictions themselves realize the complete operation family.

Proof. Membership in U_{n-1} means that every path of at most $n-1$ calibrated operations stays inside U. This follows by induction from (7), beginning with $U_0 = U$. All inputs on which the first n operations must act are consequently calibrated. Their outputs are determined by the S_a, independently of every completion outside U.

For necessity, suppose $v \notin U_{n-1}$. If $v \notin U$, the one-step proposition supplies two extensions that disagree within one action. Otherwise the recursive definition gives a word of length $j \le n-1$ with a first exit from U. Its intermediate vectors before exit lie in U, so its exit vector $z \notin U$ is the same under every compatible extension family. Choose any next action b. Two extensions of S_b can be made to disagree on z by adding a linear map that vanishes on U. Keep the other action maps the same. Even if action b occurred earlier, that modification has no effect before the exit because the earlier inputs lie in U. The two fixed extension families therefore produce different vectors after the exit word followed by b, whose total length is at most n. This contradicts uniform continuation identification.

At a fixed point, (7) gives $S_a U_\infty \subseteq U_\infty$. If $L \subseteq U$ is invariant under every restriction, induction gives $L \subseteq U_j$ for every j, so $L \subseteq U_\infty$. Applying the finite-horizon equivalence at all n proves the all-word assertion. □

This theorem supplies an identifiable *domain*, not necessarily an identified whole operator. It is exact for the stated completion class and full-state output. A smaller observation does not automatically enlarge the predictable domain: Proposition 11.1 proves that it does not under the present unrestricted completion class when the observation is nonzero. Sections 12–14 characterize the gains that become possible under explicitly structured uncertainty. Similarly, a learned nonlinear latent representation may compress the identifiable core further for its own tasks, but that is not part of this linear result.

9.2 A two-dimensional calibration with a one-dimensional reusable core

Counterexample 9.3 (Calibration does not close under continuation)**.** Let V have basis e_1, e_2, e_3, let $U = \operatorname{span}(e_1, e_2)$, and let the single calibrated operation satisfy

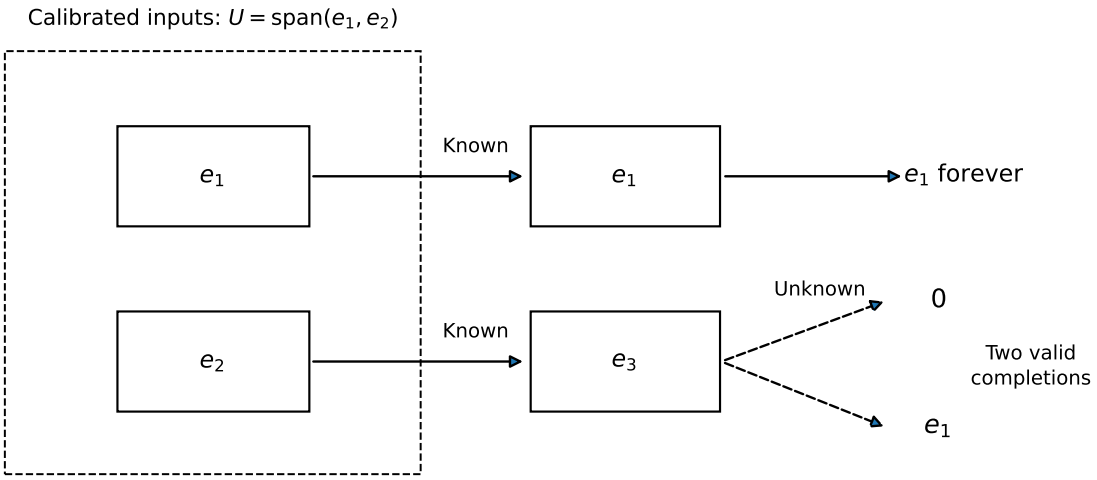
$$Se_1 = e_1, \qquad Se_2 = e_3.$$

The first continuation is known on all of U. The next calibrated domain is $U_1 = \operatorname{span}(e_1)$, which is already invariant. One completion has $T_0 e_3 = 0$ and another has $T_1 e_3 = e_1$. Both agree on every supplied calibration. Starting from e_2, their first outputs agree at e_3, but their second outputs are respectively zero and e_1.

Every continuation from e_1 is nevertheless determined. The system can ground a genuinely reusable operation on that line while remaining unable to predict repeated operations from e_2. A new exact calibration of e_3 expands the known input span to V and identifies the whole operator. Obtaining that calibration is a new experiment, not an algebraic consequence of the old data.

9.3 Several operators and the scope of the initial vector

With several actions, (7) retains only vectors that remain calibrated under *every* permitted word. An added action can therefore shrink the core even if the old operators and data are unchanged. Restricting the permitted grammar to a specific sequence can enlarge the domain of predictable use. The full free-word contract is deliberately stronger than one selected trajectory.



Known one-step domain: dimension 2. Identifiable all-word core: dimension 1.

Figure 5: A partial calibration can support unlimited exact continuation on a proper invariant core. The known image of e_2 leaves the calibrated input subspace, so its next image is unrestricted. Agreement with every supplied one-step calibration is not agreement on two-step behavior.

The initial vector v is known to the predictor in this section. If the actual input is hidden behind an evidence code, its initialization must satisfy an additional factorization condition like (1). A calibrated operation and an unobserved initial state are different missing inputs. They cannot be exchanged merely because both appear as uncertainty about the next output.

10 Why small local error needs an ambient growth bound

Exact calibration can justify a sharp domain. Approximate calibration needs a quantitative statement about errors that leave it. Let V be a normed vector space and $L \subseteq V$ a linear subspace invariant under known bounded linear maps $S_a : L \to L$. Suppose the actual bounded linear maps $T_a : V \to V$ satisfy

$$\|T_a\| \le A, \qquad \|S_a\| \le B, \qquad \|(T_a - S_a)v\| \le \eta\|v\| \quad (v \in L), \tag{8}$$

for common nonnegative bounds A, B, η. The norm on T_a is an ambient bound. Its purpose is to control propagation of errors outside L, not only behavior on calibrated inputs.

Proposition 10.1 (Approximate continuation with explicit leakage control)**.** *For every $v \in L$ and every word w of length $n \ge 1$,*

$$\|T_w v - S_w v\| \le \eta\|v\| \sum_{j=0}^{n-1} A^{n-1-j} B^j. \tag{9}$$

For the empty word the discrepancy is zero. Integer zero powers in the displayed finite sum have their usual value one. The conclusion does not require the actual trajectory to remain in L.

Proof. Write $z_j = S_{a_j} \cdots S_{a_1} v \in L$ and $x_j = T_{a_j} \cdots T_{a_1} v$. For the next operation,

$$\|x_{j+1} - z_{j+1}\| \le A\|x_j - z_j\| + \eta\|z_j\| \le A\|x_j - z_j\| + \eta B^j \|v\|.$$

The initial discrepancy is zero. Iterating this scalar recurrence gives (9). Ambient boundedness applies to $x_j - z_j$ even when that error has left L. □

This is a telescoping error estimate, not a new perturbation principle. Similar continuation bounds already appear in the collection [1, 3]. Its role here is to identify a premise that calibration on a proper subspace cannot certify by itself.

Counterexample 10.2 (An uncalibrated direction amplifies a small residual)**.** Let $L = \operatorname{span}(e_1)$, let $Se_1 = e_1$, and in the ℓ^1 norm define

$$T = \begin{pmatrix} 1 & 0 \\ \eta & M \end{pmatrix}, \qquad \eta > 0,\ M \geq 0.$$

The one-step error on the unit calibrated vector is η, independent of M. After two steps the error is $\eta(1 + M)$. It can be arbitrarily large while the local calibration error is fixed. For n steps the second coordinate is $\eta \sum_{j=0}^{n-1} M^j$.

A learned model can therefore match calibrated transitions closely and still fail rapidly under repetition. Increasing one-step accuracy helps only in relation to the length of use and the growth of the unseen directions. A measured small residual does not imply an ambient norm bound. The bound can come from a declared operator class, a separate global calibration argument, or a stronger structural certificate, but not from renaming the local fit "stable meaning."

11 A smaller output is not automatically a weaker calibration requirement

The preceding full-state theorem suggests asking whether a small observable can remain predictable outside the calibrated core. The answer depends on the uncertainty class. Under the unrestricted extension class of Section 9, simply requesting fewer output coordinates does not enlarge the all-word identifiable domain unless the requested observation is identically zero.

Let $Q : V \to Y$ be a fixed linear observation. The initial vector v remains known. All linear extensions of every calibrated restriction S_a remain permitted, independently across action names but fixed throughout an execution.

Proposition 11.1 (No enlargement under unrestricted linear completion)**.** *If $Q \neq 0$, then for every $h \geq 1$, the vectors $QT_w v$ are independent of the compatible extension family for every word of length at most h if and only if $v \in U_{h-1}$. In particular, the all-word output-identifiable domain is exactly U_∞. If $Q = 0$, every initial vector is trivially output-identifiable.*

Proof. On U_{h-1} the full vectors are determined by Theorem 9.2, so their Q outputs are determined. Outside that subspace, follow the same first-exit construction as in that proof. At an uncalibrated input z, choose a linear functional ℓ vanishing on U with $\ell(z) = 1$. Because $Q \neq 0$, choose b with $Qb \neq 0$. Replacing the next action by $T_a + \ell(\cdot)b$ preserves all calibration data and every earlier step of the first-exit path, but changes the next observed output by Qb. This produces disagreement within h actions. Each modified action map is fixed, not chosen anew during execution. The zero-output case follows immediately. □

This corrects an overly broad suggestion in the first draft's surrounding prose. Output-relative prediction can be less demanding under *structured* uncertainty. It is not automatically less demanding when every uncalibrated direction is allowed to return arbitrarily into the output. The previous full-state theorem and its proof are unchanged.

The next sections therefore change the model class explicitly. We first consider a finite family of complete candidate systems, retaining the fact that one unknown candidate is fixed. We then characterize the stronger

situation in which an entire uncertainty family admits one common quotient. These are different questions and have different answers.

12 A finite test for agreement under a fixed unknown model

Let V be n-dimensional over a field, let Θ be a finite nonempty model set, and let $T_a^\theta : V \to V$ be specified linear maps for each model θ and action a in a finite nonempty alphabet. All words are allowed. Choose a reference model $0 \in \Theta$ and write $A_a = T_a^0$. The shared linear observation is $Q : V \to Y$. Define

$$I_Q = \{v \in V : QT_w^\theta v = QA_w v \text{ for every } \theta \in \Theta \text{ and every word } w\}.$$

This is an initial-condition subspace. It is not assumed invariant under the physical maps. The common coordinates are part of this linear comparison model. With unrelated latent carriers, a justified identification of the carriers would be an additional premise.

For each θ, form the paired maps and difference observation

$$B_a^\theta = \begin{pmatrix} T_a^\theta & 0 \\ 0 & A_a \end{pmatrix}, \qquad C = \begin{pmatrix} Q & -Q \end{pmatrix}, \qquad \Delta v = (v, v).$$

The pair is evolved by the *same* action word in both components. Define

$$K_0^\theta = \ker C, \qquad K_{j+1}^\theta = K_j^\theta \cap \bigcap_a (B_a^\theta)^{-1} K_j^\theta.$$

Theorem 12.1 (Finite output-agreement test with persistent model tags)**.** *The descending paired-kernel sequence stabilizes in at most* $2n$ *strict dimension drops. Its terminal subspace is*

$$K_\infty^\theta = \bigcap_w \ker(CB_w^\theta), \qquad I_Q = \bigcap_{\theta \in \Theta} \Delta^{-1} K_\infty^\theta.$$

Thus exact all-word output agreement for the finite candidate family is decidable by finite linear algebra. If an initial vector is excluded, a disagreeing candidate and a distinguishing word can be extracted. Words of length at most $2n$ *suffice for this test. The conclusion compares fixed candidate models; it does not admit arbitrary switching between candidates during a word.*

Proof. By induction, K_j^θ consists precisely of paired states with zero difference output for all words of length at most j. The initial case is the empty word. Membership in the next intersection says that the present difference is zero and that applying any first action leaves a pair whose next j actions have zero difference. This is exactly the length-$j + 1$ condition.

Every strict inclusion lowers dimension in $V \oplus V$, which has dimension $2n$. Once two successive subspaces agree, the subspace is invariant under every B_a^θ and lies in $\ker C$, so all later intersections agree. This proves the all-word formula. At the diagonal initial pair, $CB_w^\theta \Delta v = QT_w^\theta v - QA_w v$, giving the paired-kernel formula. Row-space construction retains the operation word producing each tested row. If a vector fails the terminal kernel test, one such row is nonzero on it and supplies a witness. The dimension argument gives the stated conservative length bound. □

Finite linear-algebra equivalence testing has established precedents, including algorithms for probabilistic automata [11]. Reachability, observability, and minimal realization for multiple linear operations are also established subjects [12]. The contribution here is the fixed-compatible-model interpretation and its connection to evidence initialization and correction. We do not claim a new general automata-equivalence algorithm.

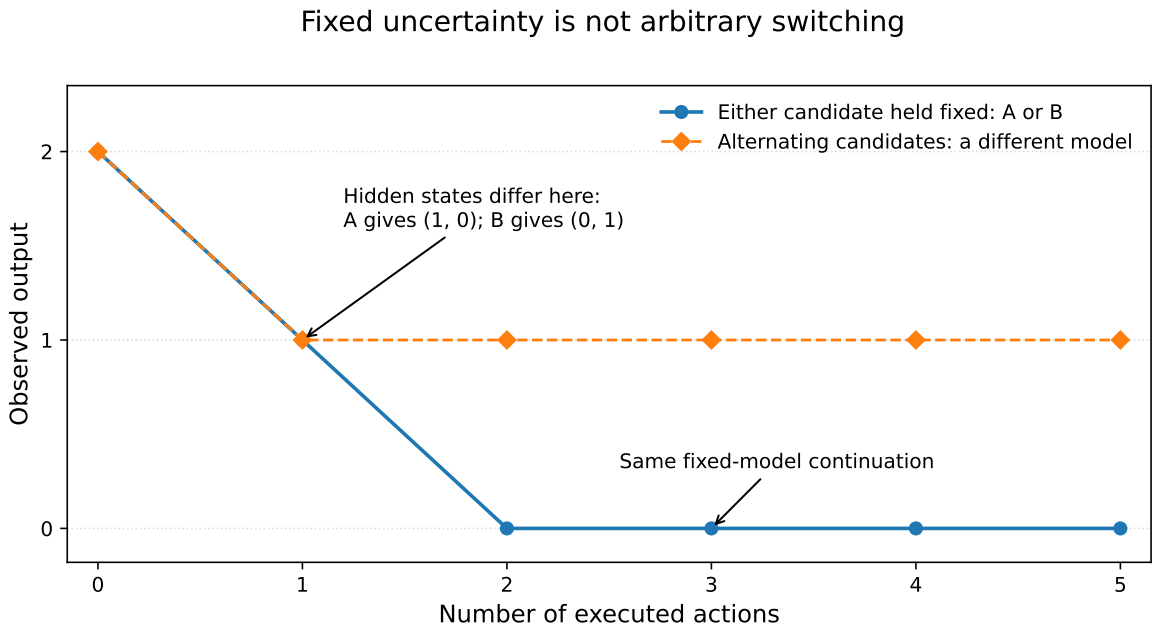


Figure 6: Two fixed candidate systems produce the same complete output history from $(1, 1)$ even though their hidden states differ. Switching from one candidate to the other after the first step creates an inadmissible history with a different second output. The fixed-model tag is information about the allowed trajectories.

12.1 Why the unknown model must not change between steps

Counterexample 12.2 (A switching adversary creates a nonexistent failure)**.** Over $\mathbb{Q}$, let there be one action and two possible fixed matrices

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad Q = \begin{pmatrix} 1 & 1 \end{pmatrix}, \qquad v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Both fixed systems have output sequence $2, 1, 0, 0, \ldots$ from v, since $A^2 = B^2 = 0$. But applying A first and B second gives $QBAv = 1$, not zero. That hybrid trajectory belongs to neither candidate model. Treating fixed uncertainty as a freely changing matrix would incorrectly reject the valid output prediction.

Here $I_Q = \operatorname{span}\{(1, 1)\}$, obtained already from the one-step equality. It is not invariant under A or B: $Av = (1, 0)$ and $Bv = (0, 1)$ leave that line. Those new physical vectors cannot be treated as fresh initial conditions with a forgotten model history. For example, starting both models at $(1, 0)$ gives different next outputs. The original fixed-model executions nevertheless remain in output agreement.

This is why the identifiable initial set is not automatically a common physical invariant core. A reference predictor may still work: it follows A from the initial vector and reports its outputs. It represents the common response sequence, not a claim that the unknown physical state always equals the reference state.

12.2 More data restricts the model family, not the logical quantifier

If new exact evidence removes candidates while preserving the reference model, the intersection defining I_Q can only enlarge. Its practical value depends on whether the learner can justify that removal and initialize the resulting predictor. Adding a new candidate has the opposite effect. Neither operation warrants replacing the model label at each step. A candidate family that omits the true system can still yield an impressive but inapplicable agreement subspace.

13 When uncertain dynamics admit one common observable quotient

Agreement on selected initial conditions is weaker than agreement on the whole space. The stronger problem has a useful exact solution. Retain the finite family and reference matrices of Section 12, and define the difference maps

$$D_a^\theta = T_a^\theta - A_a.$$

Their ranges contain all possible one-step discrepancies. Let

$$E_0 = \sum_{a,\theta} \operatorname{im} D_a^\theta, \qquad E_{j+1} = E_j + \sum_a A_a E_j, \qquad E_\infty = \bigcup_j E_j.$$

This increasing sequence stabilizes after at most n strict dimension increases. Its limit is the smallest subspace invariant under all reference actions and containing every discrepancy range. It includes discrepancies that are invisible now but return to the output later.

Theorem 13.1 (Exact common quotient for structured model uncertainty)**.** *The following conditions are equivalent.*

1. $QT_w^\theta = QA_w$ *on all of* V*, for every fixed model* θ *and every word* w*.*
2. $E_\infty \subseteq \ker Q$*.*
3. *There exists a subspace* $N \subseteq \ker Q$*, invariant under every* A_a*, containing every* $\operatorname{im} D_a^\theta$*.*
4. *There is a surjective linear representation* $r : V \to Z$*, shared linear updates* G_a*, and a readout* $\bar{Q}$ *such that*

$$rT_a^\theta = G_a r \quad \text{for all } a, \theta, \qquad Q = \bar{Q} r.$$

When the conditions hold, V/E_∞ *gives one such quotient. The smallest-dimensional common linear quotient is* V/N_0*, where* $N_0 = \bigcap_w \ker(QA_w)$ *is the reference system's unobservable subspace. Its dimension is* $n - \dim N_0$*. The shared quotient also preserves outputs if the candidate model is changed between actions, a stronger conclusion than fixed-model agreement on a restricted initial set.*

Proof. Assume condition 1. For any a, θ, w, agreement as linear maps on all of V gives

$$QA_w D_a^\theta = QT_w^\theta T_a^\theta - QA_w A_a = 0.$$

The last equality is condition 1 for the word that executes a and then w. Thus Q annihilates the range of every $A_w D_a^\theta$, whose span is E_∞, proving condition 2. Condition 2 supplies condition 3 with $N = E_\infty$.

Under condition 3, $T_a^\theta N \subseteq N$ because $T_a^\theta = A_a + D_a^\theta$, $A_a N \subseteq N$, and the entire range of D_a^θ lies in N. Every candidate therefore induces the same map $G_a([v]) = [A_a v]$ on V/N. Since $QN = 0$, the output factors through that quotient. This proves condition 4. Composing its commuting equations proves condition 1. The same composition works with a different model index at each action because r and G_a are shared.

The subspace N_0 is invariant under every A_a and lies in $\ker Q$. Every invariant subspace in $\ker Q$, including E_∞ when condition 2 holds, lies in N_0. Hence N_0 gives a valid common quotient. Conversely, the kernel of any representation in condition 4 is invariant under the reference actions and contained in $\ker Q$, so it lies in N_0. Rank-nullity gives the lower bound $\dim Z \geq n - \dim N_0$, attained by V/N_0. □

13.1 Large hidden variation with exact visible behavior

For a structured example over $\mathbb{Q}$, take either of the fixed maps

$$T_0 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad T_1 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 2 & -1 & 3 \end{pmatrix}, \qquad Q = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}.$$

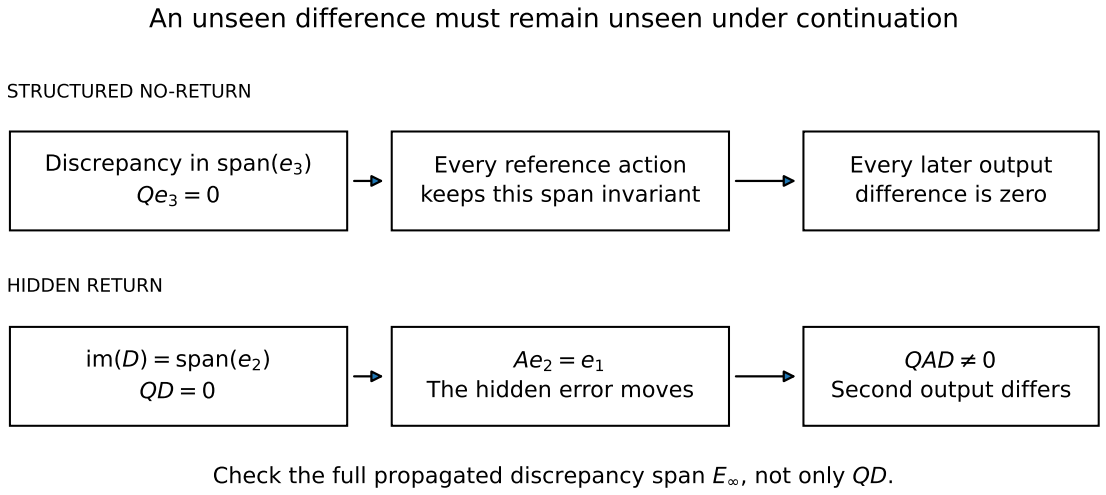


Figure 7: Structured uncertainty may be harmless when every difference remains in an invariant, output-invisible subspace. The opposite example has zero first-step output discrepancy but a visible second-step return. The figure represents exact algebraic alternatives, not a claim that small learned residuals certify the invariant subspace.

The hidden third-coordinate dynamics are different, and T_1 can make that coordinate grow. Nevertheless the discrepancy range is $\mathrm{span}\{e_3\}$, which the reference leaves inside itself and Q annihilates. The common quotient retains (x_1, x_2), with update

$$G(x_1, x_2) = (x_1 + x_2, x_2).$$

The output at step k is $x_1 + kx_2$ under both models, and even under switching between them. The output alone is not an initialized reusable state: equal x_1 with different x_2 gives different next outputs. Ignoring x_3 is justified by the no-return structure, not by its small size or its being temporarily unobserved.

Counterexample 13.2 (An invisible difference returns one step later)**.** Let

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad D = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad Q = \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

The candidates are A and $A + D$. Since $QD = 0$, every one-step output agrees for every input. But $QAD = (1, 0) \neq 0$. Starting at e_1, the second outputs are zero and one. The initial discrepancy range is invisible, while its propagated span $E_\infty = V$ is not. Checking only $QD = 0$ misses the failure.

13.2 What the criterion does not learn

The family $\{T_a^\theta\}$, its shared carrier, and Q are given. The closure in the discrepancy-span recursion uses exact linear algebra and decides whether those models share the stated quotient. It neither discovers the model class from finite observations nor proves that the true mechanism belongs to it. A structural prior such as an invariant hidden block can justify a smaller uncertainty family, but that prior must be carried as evidence or an explicit assumption.

14 Initializing the smallest common linear predictor

The paired-state criterion supplies a domain of identifiable output behavior. A learner may still lack the information needed to select the correct behavior within that domain. We now combine the two obligations without requiring the identifiable initial subspace to be physically invariant.

Let $L \subseteq I_Q$ be a specified linear initial-condition subspace and let the available initial record be a linear map $R : L \to H$. Define the reference reachable span and its unobservable part by

$$W = \operatorname{span}\{A_w v : v \in L,\ w \text{ any word}\}, \qquad N = \{z \in W : QA_w z = 0 \text{ for every word } w\}.$$

Both W and N are invariant under the reference actions, and $N \subseteq \ker Q$. These spaces can be computed by finite reachable-span and row-kernel iterations. They concern a reference predictor, not a claim that every physical candidate remains inside W.

Theorem 14.1 (Evidence-initialized minimal common linear prediction)**.** *There is a finite-dimensional linear predictor with initialization $d : R(L) \to Z$, updates $G_a : Z \to Z$, and readout $\bar{Q} : Z \to Y$ satisfying*

$$\bar{Q} G_w d R(v) = Q T_w^\theta v \quad \text{for every } v \in L, \theta \in \Theta, w$$

if and only if $\ker R \subseteq N$. *When it exists, the smallest possible state-space dimension among such* linear *predictors is*

$$\dim W - \dim N.$$

The quotient W/N, initialized by $d(Rv) = [v]$, attains that minimum.

Proof. If $v \in \ker R$, linear initialization gives the same state as input zero. Exactness therefore requires $QA_w v = 0$ for every word, so $v \in N$. Conversely, if $\ker R \subseteq N$, the rule $d(Rv) = [v]$ is well defined and linear. Invariance supplies $G_a([z]) = [A_a z]$, and $QN = 0$ supplies $\bar{Q}([z]) = Qz$. Their iteration gives $QA_w v$, which agrees with every candidate by $L \subseteq I_Q$.

For minimality, let any competing linear predictor be given. Its reachable span is generated by vectors $G_w d R(v)$. Associate such a vector with $[A_w v] \in W/N$ and extend by linearity. To see this association is well defined, suppose a finite linear combination of the predictor vectors is zero. Applying any later word and the predictor readout shows that the matching combination of reference vectors has zero QA_u output for every u. That combination therefore belongs to N. The induced map from the competitor's reachable span onto W/N is a well-defined linear surjection, since the classes of $A_w v$ span W/N. Its dimension cannot be smaller than the displayed quotient dimension. The constructed quotient attains equality. □

The dimension statement is a linear-realization theorem in a declared algebraic model, not a lower bound for arbitrary nonlinear encodings, unbounded-precision tricks, or shortest source programs. Reachability and observability are classical tools [12]. The distinctive conjunction here is compatibility across the retained models, initialization from the actual record, and the smallest common linear predictor.

14.1 Initial dimension, dynamic memory, and hidden state

In Counterexample 12.2, set $L = \operatorname{span}\{(1,1)\}$ and $R(x) = x_1$. The reference reaches $(1,1)$ and $(1,0)$, so $W = \mathbb{Q}^2$. The rows $Q = (1,1)$ and $QA = (0,1)$ have trivial common kernel, giving $N = 0$. A two-dimensional linear predictor is therefore necessary even for this one-dimensional initial family: it must reproduce $2t, t, 0, 0, \ldots$ for every multiplier t.

For the three-coordinate structured example, $L = V$, $W = V$, and $N = \operatorname{span}\{e_3\}$. Two predictor coordinates suffice. The record (x_1, x_2) initializes them; x_1 alone does not, since its kernel contains e_2, which becomes visible after one action. Physical dimension, initial-family dimension, and minimum predictive dimension are different quantities.

Part V

Quantitative grounding with finite retained memory

15 From exact probability coordinates to a finite record

The likelihood-profile theorem identifies what an exact statistical memory must retain. It does not make its probability coordinates finitely representable. A rational posterior can acquire larger denominators after each update. Rounding those coordinates changes the object, so a finite-record theory needs a stated error, horizon, and observation model.

We now give one such theory. The result deliberately uses a conservative positive-likelihood condition rather than claiming uniform stability for every observation process. Stability and numerical approximation of nonlinear filters have an established literature [13]. Our construction concerns a static hidden world, a finite cell record, and an explicit finite-horizon error account. It is not a particle-filter theorem and does not inherit a contraction hypothesis from that literature.

15.1 A concrete probability record

For an integer $k \geq 1$, write Δ_k for the probability simplex on k cells. Fix a denominator $N \geq 1$. A retained record has the form

$$\widehat{p} = (n_1/N, \ldots, n_k/N), \qquad n_i \in \{0, \ldots, N\}, \quad \sum_i n_i = N.$$

The cell identities, denominator, and decoding rule are fixed background. The k integer fields can be stored literally in $k\lceil \log_2(N+1) \rceil$ bits. This is an upper bound for this specified encoding, not a minimum description length. It excludes the likelihood table, program, temporary arithmetic, and any retained raw observation history.

Define $Q_N(p)$ by first setting $n_i = \lfloor N p_i \rfloor$. Distribute the remaining $N - \sum_i n_i$ units to the largest fractional remainders, with ties broken by a fixed cell order. This makes the normalization exact after every rounding. It is not independent coordinate rounding followed by an unspecified repair.

For a likelihood vector $\ell = (\ell_i)$ define the Bayes map

$$\mathcal{B}_\ell(p)_i = \frac{p_i \ell_i}{\sum_j p_j \ell_j}$$

whenever its denominator is positive.

Proposition 15.1 (Rounding and positive-likelihood update bounds)**.** *For every $p \in \Delta_k$,*

$$d_{\mathrm{TV}}(p, Q_N(p)) \leq \frac{k}{2N}.$$

Let $p, q \in \Delta_k$, let $0 < \beta \leq 1$, and let every entry of ℓ and $\widetilde{\ell}$ lie in $[\beta, 1]$. If $\max_i |\ell_i - \widetilde{\ell}_i| \leq \varepsilon$, then

$$d_{\mathrm{TV}}\big(\mathcal{B}_\ell(p), \mathcal{B}_{\widetilde{\ell}}(q)\big) \leq \frac{d_{\mathrm{TV}}(p, q) + \varepsilon}{\beta}.$$

Each right-hand side may also be replaced by its minimum with one. The probability-field storage bound above applies to the retained rounded record, not to all working memory.

Proof. The remainder rule produces nonnegative integers summing to N. Each rounded coordinate differs from the original by at most $1/N$, so half the sum of absolute differences is at most $k/(2N)$.

For the update, first hold ℓ fixed. Given $0 \le f_i \le 1$, put $a = \sum_i \mathcal{B}_\ell(q)_i f_i$ and $Z_p = \sum_i p_i \ell_i$. Then

$$\sum_i \big(\mathcal{B}_\ell(p)_i - \mathcal{B}_\ell(q)_i\big) f_i = \frac{\sum_i (p_i - q_i)\ell_i (f_i - a)}{Z_p}.$$

Every value $\ell_i(f_i - a)$ lies in $[-a, 1-a]$, an interval of length one. Since $p - q$ has total mass zero, its pairing with a function of oscillation at most one is bounded in absolute value by $d_{\mathrm{TV}}(p, q)$. Also $Z_p \ge \beta$. Taking the supremum over f bounds the fixed-likelihood difference by $d_{\mathrm{TV}}(p, q)/\beta$.

Next hold q fixed and set $a = \sum_i \mathcal{B}_{\widetilde{\ell}}(q)_i f_i$. The corresponding numerator is $\sum_i q_i(\ell_i - \widetilde{\ell}_i)(f_i - a)$, whose absolute value is at most ε. Its denominator is at least β. A triangle inequality proves the displayed update bound. The cap by one is the universal range of total variation. The storage assertion counts the stated integer fields only. □

The estimate is sufficient and can be loose. In particular, it need not credit favorable observations that reduce an existing error. Its purpose is a guarantee whose entire cost and probability assumptions are visible, not an assertion that rounding usually produces its worst possible error.

15.2 Approximate likelihood profiles

Let Ω be a finite static world set with prior μ_0. A fixed partition $C_1, \dots, C_k$ refines the target: $t(\omega) = t_i$ throughout C_i. Each admitted experiment e is nondestructive and has a finite report alphabet. Its actual conditional probabilities are $\ell_{e,o}(\omega)$. Reports are freshly generated from those likelihoods conditional on the fixed world and the chosen experiments.

Choose one valid representative report law $\bar{\ell}_{e,o,i}$ per cell. Assume, for every admitted experiment, report, cell, and world in that cell,

$$\beta \le \ell_{e,o}(\omega), \bar{\ell}_{e,o,i} \le 1, \qquad |\ell_{e,o}(\omega) - \bar{\ell}_{e,o,i}| \le \varepsilon. \tag{10}$$

Both the actual and representative report laws sum to one over reports. Exact profile cells satisfy this condition with $\varepsilon = 0$ when their likelihoods have the positive floor. A coarser partition can be used with a nonzero ε, but target-constant cells are still required for the target readout.

After a history, let p_t be the cell masses of the exact world posterior. Initialize the retained record by $\widehat{p}_0 = Q_N(p_0)$. For the next selected experiment and observed report use

$$\widehat{p}_{t+1} = Q_N\Big(\mathcal{B}_{\bar{\ell}_{e,o}}(\widehat{p}_t)\Big).$$

The reference posterior and rounded record are compared along the same actual action/report history. Experiment selection may depend on earlier retained records and independent randomness. It may not introduce a hidden world-dependent selector whose information is omitted from that history. The likelihood condition applies to every experiment the policy can select.

Theorem 15.2 (Finite retained memory for a declared correction horizon)**.** *Under (10), set $\eta_N = k/(2N)$, $\gamma = \beta^{-1}$, and let e_0 be any valid bound on $d_{\mathrm{TV}}(p_0, \widehat{p}_0)$. For every possible action/report history of length t,*

$$d_{\mathrm{TV}}(p_t, \widehat{p}_t) \le e_t := \min\left\{1,\ \gamma^t e_0 + (\gamma\varepsilon + \eta_N) \sum_{j=0}^{t-1} \gamma^j \right\}. \tag{11}$$

The same bound holds after mapping cells to target values. With the generic initialization bound $e_0 \le \eta_N$, define

$$A_H = \sum_{j=0}^{H} \gamma^j, \qquad B_H = \gamma \sum_{j=0}^{H-1} \gamma^j.$$

For any horizon $H \ge 0$ and tolerance $\tau > \varepsilon B_H$, every integer

$$N \ge \max\left\{1, \left\lceil \frac{kA_H}{2(\tau - \varepsilon B_H)} \right\rceil\right\} \tag{12}$$

is sufficient for target and cell error at most τ at every time through H. Empty sums are zero. This is a finite-horizon retained-record guarantee, not an unbounded-time or total-computation guarantee.

Proof. Condition on a particular attainable history and its next experiment. In each positive-mass cell, the exact next likelihood is the conditional average of $\ell_{e,o}(\omega)$ within that cell. It remains in $[\beta, 1]$ and within ε of $\bar{\ell}_{e,o,i}$. For a zero-mass cell any value obeying these bounds gives the same zero contribution. Thus the exact next cell mass is a Bayes update with such an effective likelihood vector, even when the worlds within a cell do not have identical laws.

Proposition 15.1 and the rounding bound give

$$d_{\mathrm{TV}}(p_{t+1}, \widehat{p}_{t+1}) \le \gamma d_{\mathrm{TV}}(p_t, \widehat{p}_t) + \gamma\varepsilon + \eta_N.$$

Iterating the uncapped scalar bound proves (11); total variation is also at most one. Aggregating cell coordinates into target coordinates cannot increase total variation, by the triangle inequality. Substituting $e_0 \le \eta_N$ gives $\eta_N A_t + \varepsilon B_t$. These quantities are nondecreasing in t because $\gamma \ge 1$ and all terms are nonnegative. Inequality (12) makes the horizon-H value at most τ. It therefore suffices at every earlier step as well. Every report has positive probability under both updates by the floor assumption, so no zero denominator is silently normalized. □

Equation (12) is a sufficient precision requirement, not an exact minimum. An infeasible or large bound does not prove that the particular rounded filter is inaccurate. It means that this uniform certificate is insufficient. Conversely, increasing precision cannot eliminate the model term εB_H. A false likelihood model does not become correct because its probabilities are stored with more bits.

Counterexample 15.3 (A tiny rounded-away mass can control a later posterior)**.** Take two worlds with prior $(1 - d, d)$, where $0 < d < 1/(2N)$. The remainder rule gives the retained record $(1, 0)$. A binary report has likelihoods $(d^2, 1)$ for outcome one. After that outcome, the actual posterior mass of the second world is

$$\frac{d}{d + (1-d)d^2} = \frac{1}{1 + d - d^2},$$

which tends to one as d tends to zero. The rounded posterior remains zero on that world. The initial total-variation error is only d. There is no common positive likelihood floor across this family: the first likelihood tends to zero, and the other report is impossible in the second world. No nontrivial uniform continuation conclusion follows from the small initial rounding error alone.

The counterexample concerns evidence amplification, not a failure of the remainder arithmetic. Replacing the likelihood one by $1 - d^2$ makes both outcomes possible in both worlds while preserving the limiting phenomenon. The minimum likelihood still tends to zero. Exact support, approximate probability, and justified abstention should therefore remain distinct data.

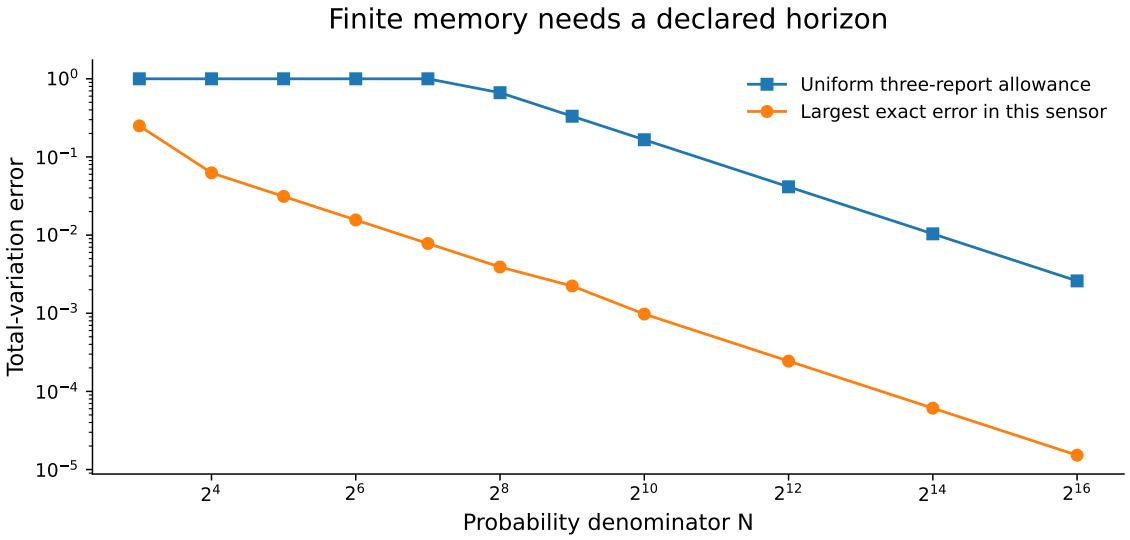


Figure 8: The noisy sensor with four retained cell masses, three future reports, and several denominator choices. The actual maximum error is enumerated over all eight three-report histories and their prefixes. The upper curve is the uniform positive-likelihood bound using generic rounded initialization. The gap is conservatism, not evidence of an uncounted learning gain. The installed likelihoods are exact; retained-record size does not count the program or temporary arithmetic.

16 Approximate invisibility must survive feedback

The exact common-quotient theorem allows arbitrarily large hidden dynamics when those directions never return to the selected output. Approximation changes that fact. A small hidden-to-visible coupling can make hidden growth consequential. This section states the relevant dependence explicitly rather than replacing exact invariance by a small residual with no continuation bound.

Approximate simulation and bisimulation offer established ways to compare observable behavior with a declared precision [14]. The following is a narrower block comparison argument for discrete linear operations. It does not claim that its matrix bound is a minimal simulation metric or that stability of every candidate separately supplies a common switching bound.

Write the physical state as (v_t, h_t) in a direct sum of two finite-dimensional normed spaces. The required output is the visible component v_t. Along an admitted execution suppose

$$\begin{pmatrix} v_{t+1} \\ h_{t+1} \end{pmatrix} = \begin{pmatrix} G_t & B_t \\ C_t & H_t \end{pmatrix} \begin{pmatrix} v_t \\ h_t \end{pmatrix}, \qquad \bar{v}_{t+1} = G_t \bar{v}_t. \tag{13}$$

The reduced predictor knows the chosen G_t and receives the same action sequence. It does not observe h_t. The blocks may come from a fixed unknown candidate or from an explicitly admitted varying family. Assume common operator-norm bounds

$$\|G_t\| \le g, \quad \|B_t\| \le b, \quad \|C_t\| \le c, \quad \|H_t\| \le d, \qquad g, b, c, d \ge 0.$$

Let $\|v_0\| \le V_0$, $\|h_0\| \le H_0$, and $\|v_0 - \bar{v}_0\| \le e_0$. These initial bounds are supplied evidence or declared model assumptions. Put

$$M = \begin{pmatrix} g & b \\ c & d \end{pmatrix}, \qquad u_0 = \begin{pmatrix} V_0 \\ H_0 \end{pmatrix}, \qquad u_j = M^j u_0.$$

Powers of this nonnegative comparison matrix are bounds on norms. They are not a replacement physical dynamics inferred from measurements.

Theorem 16.1 (Visible-error bounds from a common block comparison)**.** *Every admitted execution satisfying (13) obeys*

$$\|v_t - \bar{v}_t\| \le d_t := g^t e_0 + b \sum_{j=0}^{t-1} g^{t-1-j} (u_j)_2. \tag{14}$$

In particular, when $b = 0$ and $e_0 = 0$, the visible prediction is exact for every finite time, without a stability requirement on d or c.

If $g < 1$, $d < 1$, and

$$D = (1-g)(1-d) - bc > 0,$$

then the total visible discrepancy is bounded by

$$\sum_{t=0}^{\infty} \|v_t - \bar{v}_t\| \le \frac{e_0}{1-g} + \frac{b\big(cV_0 + (1-g)H_0\big)}{(1-g)D}. \tag{15}$$

These are sufficient common bounds. Failure of the inequality for D is not a proof that a particular signed system is unstable. Conversely, the two diagonal conditions alone are insufficient for a uniform conclusion over the stated block class.

Proof. The norm triangle inequality and the common block bounds give

$$\begin{pmatrix} \|v_{t+1}\| \\ \|h_{t+1}\| \end{pmatrix} \le M \begin{pmatrix} \|v_t\| \\ \|h_t\| \end{pmatrix} \le M^{t+1} u_0$$

componentwise. The error satisfies

$$v_{t+1} - \bar{v}_{t+1} = G_t (v_t - \bar{v}_t) + B_t h_t.$$

Induction gives (14), including its empty sum at $t = 0$. If $b = 0$ and the initial error is zero, the recurrence has no forcing term and remains zero, regardless of hidden growth.

Under the three strict inequalities, $(I - M)^{-1} = D^{-1} \begin{pmatrix} 1-d & b \\ c & 1-g \end{pmatrix}$ has nonnegative entries. Let $a = (I - M)^{-1}(1,1)^T$. Both coordinates are positive and $Ma = a - (1,1)^T < a$. Hence M is a strict contraction in the weighted maximum norm with weights a, and its geometric series converges to $(I - M)^{-1}$. Summing the nonnegative right-hand side of (14) gives

$$\frac{e_0}{1-g} + \frac{b}{1-g} \sum_{j \ge 0} (M^j u_0)_2.$$

The second coordinate of $(I - M)^{-1} u_0$ is $(cV_0 + (1-g)H_0)/D$, proving (15). The common comparison applies at each step and therefore also covers switching when that switching is explicitly allowed by the contract. It need not cover a family supplied only with unrelated per-model stability certificates. □

The result connects exact and approximate abstraction without collapsing them. Exact no-return can protect an output even when the hidden coordinate grows rapidly. With nonzero return, hidden growth and two-way feedback enter the error account. A small b is useful only in relation to the horizon, the initial hidden bound, and the other blocks.

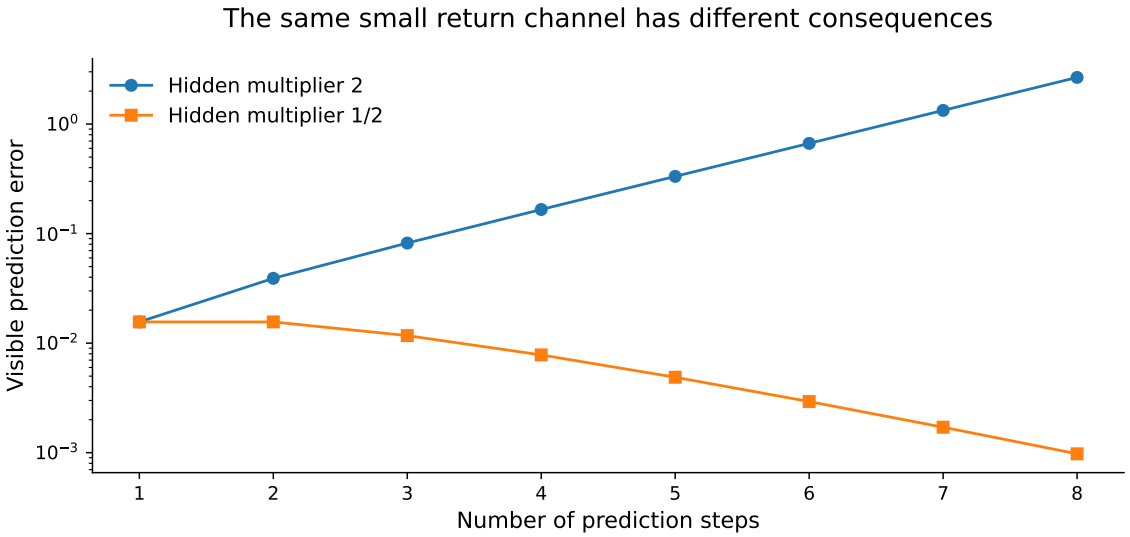


Figure 9: Exact scalar output discrepancies with the same small return coupling $1/64$ and visible multiplier $1/2$. The stable hidden multiplier $1/2$ gives $t2^{-(t-1)}/64$; the hidden multiplier 2 gives increasing error. The vertical axis is logarithmic. At exactly zero return coupling, the discrepancy is zero at every time in both cases and is not plotted on the log scale. These are explicit linear models, not neural activation traces.

Counterexample 16.2 (An arbitrarily small return channel exposes an unstable hidden mode)**.** Let $v_0 = \bar{v}_0 = 0$, $h_0 = 1$, and use scalar blocks $G = 1/2$, $B = \delta > 0$, $C = 0$, $H = 2$. The reduced output remains zero, while

$$v_t = \delta \sum_{j=0}^{t-1} (1/2)^{t-1-j} 2^j = \frac{\delta(2^t - 2^{-t})}{3/2}.$$

The first error is δ, yet the later error is unbounded for every fixed positive δ. At $\delta = 0$, every output is exactly zero. Thus a tiny positive coupling is not uniformly equivalent to exact no-return on an unbounded horizon.

Counterexample 16.3 (Stable diagonal blocks can have unstable feedback)**.** Take scalar blocks $G = H = 1/2$ and $B = C = 3/4$. Both diagonal subsystems are contractions. The full matrix sends $(1, 1)$ to $(5/4)(1, 1)$ and consequently grows on that direction. Here $D = 1/4 - 9/16 < 0$. A certificate that checks only the two diagonal bounds misses the feedback loop.

17 One justified decision from memory and dynamics

A probability-memory error and an output error concern different objects. The former is a distance between distributions on retained cells. The latter is a distance between a candidate response and its surrogate. They cannot be added without a loss function connecting both to the same decision.

Fix an actual evidence history. Let μ be the exact posterior on the stipulated finite world set and let $p_i = \mu(C_i)$. Suppose the retained distribution is $\widehat{p}$ with $d_{\mathrm{TV}}(p, \widehat{p}) \le e$. For a declared later action word or response horizon, let $y(\omega)$ be the actual model response and let $\bar{y}_i$ be the response used for cell i. Assume

$$\sup_{\omega \in C_i} \|y(\omega) - \bar{y}_i\| \le d \quad \text{for every cell } i. \tag{16}$$

The cell need not identify the physical state or candidate operator. It must support this common response allowance. A block bound such as (14) can supply it when the initial records, common visible updates, and hidden-state bounds match that cell.

Let the permitted decision set $\mathcal{D}$ be finite and nonempty. Its loss $L(y, a)$ takes values in $[0, 1]$ and is K-Lipschitz in y, uniformly over a. Define

$$R(a) = \sum_{\omega} \mu(\omega) L(y(\omega), a), \qquad \widehat{R}(a) = \sum_{i} \widehat{p}_i L(\bar{y}_i, a).$$

These are risks under the stated model posterior. They are not frequentist coverage probabilities or guarantees on the loss of a single realized world.

Theorem 17.1 (A common quantitative grounding certificate)**.** *Under (16), put* $\xi = \min\{1, e + Kd\}$. *Then*

$$|R(a) - \widehat{R}(a)| \le \xi \quad \textit{for every } a \in \mathcal{D}.$$

If $\widehat{a}$ *minimizes* $\widehat{R}$*, its true posterior regret is at most* 2ξ*. If it is the unique approximate minimizer and*

$$\widehat{R}(a) - \widehat{R}(\widehat{a}) > 2\xi \quad \textit{for every } a \ne \widehat{a},$$

then it is the unique true minimizer. A decision issued only when $\widehat{R}(a) + \xi \le \tau$ *has true posterior risk at most the declared threshold* τ.

The conclusion can combine (11) and (14) only when they refer to the same history, cell meanings, future action word, initialization, and admitted candidate family. No guarantee that the true external process belongs to that family is supplied by this composition.

Proof. Insert the intermediate risk $\sum_i p_i L(\bar{y}_i, a)$. Its difference from $R(a)$ is at most Kd by (16), the Lipschitz condition, and summing posterior masses. Its difference from $\widehat{R}(a)$ is at most e because the function $i \mapsto L(\bar{y}_i, a)$ lies in $[0, 1]$. A triangle inequality gives $e + Kd$; both risks also lie in $[0, 1]$.

Let a_* minimize R. Then

$$R(\widehat{a}) \le \widehat{R}(\widehat{a}) + \xi \le \widehat{R}(a_*) + \xi \le R(a_*) + 2\xi.$$

For any other decision, the approximate gap loses at most 2ξ when the two risks are replaced by their true values, proving the strict-gap assertion. The threshold assertion is the upper risk bound itself. Each step uses the same posterior and surrogate responses, which is why separately valid certificates about different histories or targets cannot be substituted. □

This is a bounded-loss composition of two error accounts, not a newly derived general decision principle. Related recipient-relative bounds appear in *Machine Understanding* [1]. The additional content here is the instantiated chain from a finite retained probability record and a separately controlled hidden-return error to one declared decision.

17.1 A complete sensor-to-predictor example

Use the eight worlds $(s, \theta, n) \in \{0, 1\}^3$. The four joint (s, θ) masses in the order $(0, 0), (0, 1), (1, 0), (1, 1)$ are initially

$$p_0 = (3, 1, 1, 3)/8,$$

and n is conditionally uniform. A noisy nondestructive reference has probability $1/4$ of report one under $\theta = 0$ and $3/4$ under $\theta = 1$. It does not depend on n. After three fresh reports all equal to one, the exact target probability is $\Pr(s = 1 \mid 111) = 41/56$.

The learner retains the four masses using $N = 65\,536$ and the stated remainder rule after each update. There are 68 literal probability-field bits because each of the four counts includes values through 65 536. The initial masses are exactly representable, so $e_0 = 0$. With $\beta = 1/4$, $\varepsilon = 0$, and three observations, (11) gives the sharper bound

$$e_3 \le \frac{1 + 4 + 16}{32\,768} = \frac{21}{32\,768}.$$

This uses the actual initialization, not the larger generic bound in (12).

The later response is produced by three applications of one fixed candidate matrix,

$$T_n = \begin{pmatrix} 1/2 & (-1)^n/64 \\ 1/8 & 1/2 \end{pmatrix}, \qquad (v_0, h_0) = (s, 1). \tag{17}$$

The predictor for cell (s, θ) uses $\bar{v}_t = 2^{-t}s$ and therefore reports $\bar{y}_i = s/8$. The nuisance bit n affects the physical response but remains unidentified by the reference observations. With $g = d = 1/2$, $b = 1/64$, $c = 1/8$, and $V_0 = H_0 = 1$, the comparison determinant is $127/512 > 0$. Directly evaluating (14) at three prediction steps yields

$$d_3 = \frac{481}{32\,768}.$$

Choose between numerical predictions $a = 0$ and $a = 1/8$, with loss $L(y, a) = \min\{1, |y - a|\}$. It is bounded and 1-Lipschitz. The common radius is consequently

$$\xi = \frac{21 + 481}{32\,768} = \frac{251}{16\,384}.$$

The exact companion computes the actual retained masses, approximate risks, both true risks, and their comparison. It verifies that the retained predictor selects $a = 1/8$, that its approximate advantage exceeds 2ξ, and that $\widehat{R}(1/8) + \xi < 3/50$. Thus the same finite construction supports a unique posterior-optimal decision and a posterior-risk threshold of 0.06 under the supplied model. The complete prior, sensor laws, candidate matrices, bit record, and loss function all belong to this one example.

Decision	Retained-record risk	Exact model-posterior risk
$a = 0$	$23991/262144 \approx 0.09152$	$173737/1835008 \approx 0.09468$
$a = 1/8$	$8777/262144 \approx 0.03348$	$81105/1835008 \approx 0.04420$

The certified upper risk for the selected decision is $12793/262144 \approx 0.04880$, below $3/50$. The actual cell-mass error is $1/114688$, smaller than its uniform allowance. The risk certificate uses the proved allowance, not a substituted post hoc exact error. Exact enumeration is available here only because the example is fully supplied and finite.

The probability coordinates do not encode n, and the dynamical allowance does not pretend they do. The result remains approximate precisely where the evidence leaves the physical response unresolved. Reducing the denominator, admitting an observation that depends on n, or increasing the hidden return coupling changes a different premise. These revisions require separate responses rather than an undifferentiated instruction to trust the old certificate.

Part VI

A consolidated grounding contract

18 A common theoretical contract for grounded revision

The results do not combine into one universal scalar score. They address different parts of a continuing claim.

An evidence contract says what the observer receives and which worlds remain possible. A predictive contract says what outputs must remain correct after which operations. A correction contract says which relationships between future evidence and target values must remain available. A calibration contract says where an operation is determined, and a quantitative contract controls any approximation and its growth.

A combined statement needs one representation and consistent premises. An unavailable input prevents initialization. An uncalibrated continuation prevents exact prediction. A lost association prevents correction. A narrow confidence set does not establish its own coverage.

18.1 The complete sensor argument

The eight-world sensor illustrates the chain without a large implementation. Initially the display and flip operation admit a two-state exact predictor. That predictor is grounded by the current display, even though the physical bit and sensor convention are not separately identified. Adding the physical-bit question refines the needed behavioral state, and the old code fails the initialization condition. A reference measurement repairs the information deficit while leaving the nuisance bit irrelevant.

Before calibration, the records corresponding to the two possible display values have the same marginal physical uncertainty. They must not be merged if later reference evidence is expected to identify the physical bit. The correction-memory theorem identifies the missing association. Resetting supplies convention information only after changing the physical state. Recovering the original state then depends on the retained old record. If that record is gone, the unchanged future interface cannot reconstruct it.

Noise changes the strength of the conclusion. A highly probable physical bit is not an exact state certificate. A quantitative response set can carry the remaining ambiguity into a tolerance, but only under coverage and model-accuracy premises. Every clause refers to the same task and same record history.

18.2 What this adds to the machine theory

The framework does not demand that an agent identify every detail of a world before it can reason. It gives restricted positive conditions for doing less. An unidentifiable physical realization can support an identified predictive quotient. An underdetermined operator can support a fully determined invariant domain. A large candidate set can be compressed exactly for future correction without preserving every candidate identity.

The additional discipline is that each reduction is indexed by what it preserves. One cannot transfer the word “grounded” from a successful prediction to an unrelated physical interpretation, from an exact model to a learned approximation, or from a present decision to every later correction. Such transfers require new evidence or a proved bridge.

18.3 Three different reductions should not share one name

The exact set-valued memory merges worlds only when the selected target and every admitted filter agree. The likelihood-profile memory preserves joint target probabilities and the future statistical correction law. The observable linear quotient suppresses directions that cannot return to the chosen output under its declared

model class. They are related examples of operation-compatible reduction, but their carriers, updates, and information requirements are different.

The fixed-model agreement subspace is a fourth object. It describes allowed initial vectors, not necessarily a physical invariant state space. A reference predictor can realize its common outputs even when the hidden candidate trajectories diverge. Forgetting that model history and restarting from a current physical vector can create a different inference problem.

18.4 What changes when the evidence contract changes

A new likelihood can split an old probability cell. A new candidate operator can shrink I_Q. A new action can propagate a previously invisible discrepancy into the output. A new initial-code restriction can fail to initialize a quotient that still exists mathematically. Each revision therefore needs a named compatibility check. The smallest representation for an old contract is not an entitlement to make new claims without acquiring the missing distinctions.

None of these facts obliges a learner to retain every detail. In the noisy sensor, four joint masses replace an eight-world prior. In the structured linear example, two predictor coordinates replace three physical coordinates. Both savings are exact under their respective contracts. What must be preserved is the relation between the discarded information and the future operations, not discarded information for its own sake.

18.5 A revision changes specific claims, not every claim at once

The sensor-to-predictor construction provides one ledger for revision. Its probability cells are indexed by (s, θ), while the hidden candidate index n is represented only through the response allowance. Each row below describes a change to that same example, not a newly selected favorable model.

Revision	What remains justified	What must be re-established
Change only N	The likelihood-profile partition and physical model.	Rounded history, memory error, probability-field size, and the resulting decision margin.
Let the reference depend on n	Old predictions under the old reference.	New profile cells and their masses. The old four masses cannot universally supply a split.
Admit a larger return coupling	The original observation update when its likelihoods are unchanged.	The future-response allowance and posterior risk. A former threshold may no longer be certified.
Retain only the two marginals	The present target and convention distributions.	The association required by later evidence. Equal marginals need not determine an update.
Treat a copied report as fresh	The recorded original observation.	The claimed conditional sampling law. Repetition in storage is not another independent measurement.
Replace a fixed model by switching	Results explicitly proved for the larger switching class, when their common bounds hold.	A restricted-initial-state fixed-model prediction need not survive this change.

A failed old certificate can mean that its premises no longer apply, not that the earlier theorem was false. A valid new risk assessment may conclude that no action meets the requested threshold. Abstention then preserves the stated claim instead of relabeling an uncertain answer as certain. No automatic dependency extraction from arbitrary programs is asserted here.

19 Open theoretical problems beyond the first edition

The first edition closes a restricted account of evidence, corrective memory, uncertain dynamics, and finite-horizon quantitative use. A complete theorem does not require an implemented general agent. It does require that the assumptions delimit the conclusion rather than merely decorate it.

Sharper approximate observable agreement. The block bound gives a sufficient norm-based certificate, not an optimal abstraction metric. Cancellation, structured candidate correlations, and action-language restrictions can permit smaller errors than the common nonnegative comparison detects. Characterizing the least attainable error or the smallest approximate state remains a distinct problem.

Economical long-horizon probability memory. The positive-likelihood estimate can grow rapidly with the horizon. It does not exploit statistical forgetting, contraction, or favorable evidence. Adaptive precision, support-aware safeguards, and stronger model conditions may permit better retained-memory guarantees. Such refinements must preserve the distinction between posterior risk, realized error, and frequentist coverage. The current record-size bound does not bound temporary arithmetic or total runtime.

Restricted operation languages and infinite model families. The exact linear results admit every action word. An automaton restricting those words changes the observable-equivalence problem and the relevant minimum. Infinite parametric families require a certified finite spanning argument, polynomial identity, or another uniform device. Sampling a few parameter values does not supply that device.

Model admission and state-changing evidence. Adding candidate worlds requires an explicit allocation of prior probability. A previous posterior does not determine that allocation. Noisy observations that change the hidden state require transition filtering, not only multiplication by a static likelihood. Those extensions need their own correction-memory and approximation theorems.

Learning the premises. Candidate systems, likelihoods, calibration, and structural bounds remain supplied mathematical data in this edition. Determining them from interaction, proving useful sample and search complexity bounds, and establishing a learned implementation's causal use are further research questions. They are not prerequisites that erase the completed conditional results.

Independent assurance. Prose proofs, exact finite illustrations, source comparison, and PDF rebuilding provide different forms of internal evidence. They do not establish independent peer review, a proof-kernel certificate, or an externally authenticated run. Those assurance layers can be developed without changing what the present statements assert.

20 Conclusion

Grounded abstraction relates a target, its possible realizations, accessible evidence, and the future operations a retained state must support. It is neither complete reconstruction by default nor confidence in a convenient coordinate system.

Incomplete knowledge can support exact continuing use. A likelihood-profile record preserves correction while dropping nuisance identity. Different linear systems can share one observable quotient. A reference predictor can reproduce common outputs without matching the unknown physical trajectory, provided the available record can initialize it.

Approximate use needs additional structure. Finite probability fields introduce rounding error at each update. A likelihood allowance and positive floor control its finite-horizon amplification. A hidden coordinate can be arbitrarily unstable when return is exactly absent, yet even a small positive return can become consequential. The common block bound and the rounded-memory bound meet only through a specified loss for one response and one posterior.

The integrated sensor illustrates that conjunction. Its probability record does not recover the hidden candidate, and its predictor does not silently identify the physical state. It nevertheless supports an explicit posterior decision within a certified allowance. Changing the observation law, precision, or return coupling changes a named obligation rather than granting universal reuse of an old certificate.

The first edition therefore supplies a theory of which claims survive what experience leaves unresolved, how a useful memory can preserve later correction, and when a reduced predictor remains justified across continued use. Its unresolved learning and implementation questions are visible applications of that theory, not claims supplied by the word grounded.

A Statement map and source relationships

Result	Mathematical role	Status and boundary
3.2	Law-level and support-level identification.	Elementary factorization, proved here. No sample-rate claim.
3.4	One evidence-initialized exact predictive realization.	Classical quotient argument with explicit initialization. Finite declared models.
4.1	Reinterpretation, acquisition, impossibility, and nonadaptive cost.	Exact for total nondestructive finite tests. Destructive tests excluded.
5.2	Coarsest memory preserving every admitted correction.	A powerset behavioral quotient. Evidence family is intersection closed.
5.3	Exact memory count for partition evidence.	All candidate subsets and exactly the atom-union evidence algebra. Logical states, not byte-optimal code.
6.1	Finite-record error and information processing.	Classical two-point bound, proved here. Fixed actual protocol.
7.1	Minimax coordinate ambiguity and a coverage certificate.	Unconstrained real-vector center. Coverage and model accuracy are premises.
9.1	Linear response identifiable exactly on the calibration span.	All linear extensions admissible. Exact inputs and outputs.
9.2	Maximal domain of identifiable full-state continuation.	Finite-dimensional common calibration span and fixed extension families.
10.1	Error propagation from a nominal invariant domain.	Local discrepancy plus a separate ambient growth bound.
8.1	Minimal likelihood-profile memory for target correction.	Fixed finite likelihoods, fresh conditional observations, all priors, and report probabilities retained.
8.2	Exact reconstruction after memory-contract change.	Same finite world set and every prior. Restricted prior families can differ.
11.1	Nonzero outputs do not enlarge unrestricted calibration.	All independent linear extensions admitted, fixed during an execution.
12.1	Finite fixed-model output agreement and a witness word.	Finite specified matrix family and all operation words.
13.1	Common observable quotient from propagated discrepancy.	Agreement on the whole space, not only a special initial set.
14.1	Evidence initialization and minimum linear state dimension.	Allowed initial subspace, linear record, linear predictor class.
15.1	Quantized probability fields and one-step Bayes perturbation.	A normalized finite simplex and positive likelihood floor. Retained fields, not total workspace.
15.2	Finite-horizon correction under rounding and profile error.	Static worlds, common observed history, valid likelihoods, and a specified denominator.
16.1	Visible discrepancy with hidden feedback.	Common norm bounds and initial allowances. Infinite sum needs the strict comparison condition.
17.1	One posterior decision from memory and response bounds.	Same cells, response, history, and bounded Lipschitz loss. No external coverage inferred.

B Small exact illustrations and reproduction

Three small exact suites check the declared constructions with integers, sets, rational arithmetic, and $\mathbb{F}_2$. The retained observable suite compares paired row-space closure with direct word enumeration, tests common-quotient and initialization criteria, and compares likelihood-profile updates with raw-world Bayes updates. It includes explicit switching, hidden-return, zero-probability, and memory-migration controls. The two retained suites are unchanged. The first-edition suite additionally checks quantized probability records, matched adaptive updates, block-error recurrences, and the joint sensor decision. These are bounded implementation checks, not universal proofs or agent experiments.

Run `python code/run_checks.py`. The JSON records state each domain and result. Nine Python-generated analytical figures use floating point only for drawing. The supplied Bash build uses their vector assets. Review notes record the actual proof, source, rendering, and clean-rebuild scope. No independent review, complete novelty audit, or proof-assistant certificate is claimed.

References

[1] Komorebi P. *Machine Understanding: Representation, Explanation, and Transfer*. First consolidated edition, version 1.0.0, Operation Komorebi, 8 October 2026. Supplied project manuscript and source archive.

[2] Komorebi P. *Self-Improving Intelligence: Representation, Discovery, and Recursive Revision*. First consolidated edition, version 1.0.0, Operation Komorebi, 8 October 2026. Supplied project manuscript and source archive.

[3] Komorebi P. *Sufficient Resolution*. Standardized traditional paper, Operation Komorebi, 2026. Finite behavioral models, hidden memory, reconstruction, and calibration. Source inspected in the project Library.

[4] T. J. Rothenberg. Identification in parametric models. *Econometrica* 39(3):577–591, 1971. doi:10.2307/1913267.

[5] M. L. Littman, R. S. Sutton, and S. Singh. Predictive representations of state. *Advances in Neural Information Processing Systems 14*, 1555–1561. Conference 2001, proceedings 2002. Proceedings record.

[6] D. Angluin. Learning regular sets from queries and counterexamples. *Information and Computation* 75(2):87–106, 1987. doi:10.1016/0890-5401(87)90052-6.

[7] H. J. van Waarde, J. Eising, H. L. Trentelman, and M. K. Camlibel. Data informativity: a new perspective on data-driven analysis and control. *IEEE Transactions on Automatic Control* 65(11):4753–4768, 2020. arXiv:1908.00468.

[8] A. Geiger, D. Ibeling, A. Zur, M. Chaudhary, S. Chauhan, J. Huang, A. Arora, Z. Wu, N. Goodman, C. Potts, and T. Icard. Causal abstraction: a theoretical foundation for mechanistic interpretability. *Journal of Machine Learning Research* 26(83):1–64, 2025. Publication record.

[9] P. van den Bos and F. Vaandrager. State identification for labeled transition systems with inputs and outputs. 2019. arXiv:1907.11034. Cited for the testing framework, not as a source of the present three-case construction.

[10] D. Blackwell. Equivalent comparisons of experiments. *Annals of Mathematical Statistics* 24(2):265–272, 1953. doi:10.1214/aoms/1177729032.

[11] W.-G. Tzeng. A polynomial-time algorithm for the equivalence of probabilistic automata. *SIAM Journal on Computing* 21(2):216–227, 1992. doi:10.1137/0221017. Cited for the established finite linear-algebra equivalence method.

[12] M. Petreczky, L. Bako, and J. H. van Schuppen. Realization theory of discrete-time linear switched systems. *Automatica* 49(11):3337–3344, 2013. doi:10.1016/j.automatica.2013.07.022. Author preprint: arXiv:1103.1343.

[13] F. Le Gland and N. Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Annals of Applied Probability* 14(1):144–187, 2004. doi:10.1214/aoap/1075828050. Cited for filter stability and approximation, not as a source of the present static finite-record bound.

[14] A. Girard and G. J. Pappas. Approximate bisimulation relations for constrained linear systems. *Automatica* 43(8):1307–1317, 2007. doi:10.1016/j.automatica.2007.01.019. Cited for observable system approximation; the discrete block comparison is proved directly here.