Article

Not peer-reviewed version

Regret Is Weighted Forgetting

Michael Timothy Bennett *

Posted Date: 19 March 2026

doi: 10.20944/preprints202603.1546.v1

Keywords: regret decomposition; representation learning; state abstraction; selective forgetting; generalisation; partially observable decision processes; weakness maximisation; information bottleneck; bisimulation; causal models

 doi:10.20944/preprints202603.1546.v1

Article

Regret Is Weighted Forgetting †

Michael Timothy Bennett

Independent Researcher, Australia; michael.bennett@anu.edu.au

† This manuscript was revised with assistance from a large language model used as a writing and editing aid. The ideas, claims, formal statements, proofs, and citation judgments were checked by the author and remain the author's responsibility.

Abstract

How much of an agent's regret comes from a bad representation, and how much from a bad policy? This paper gives an exact answer. For a fixed representation M and a finite evaluation distribution over history-test pairs, the minimum average normalized regret over all M-based policies equals the minimum margin-weighted deletion cost needed to make the optimal bet single-valued on each representation-test cell $(M(h), T)$. A policy-wise decomposition then splits any actual policy's regret into irreducible aliasing cost plus avoidable within-cell misreporting. A Stack-Theoretic reformulation identifies the same quantity as a deficit in weighted weakness on a lifted task constructed from the evaluation support (where weakness is normally the degree to which a policy leaves open unseen diagnostic continuations). I use the identity to derive several direct corollaries, including a representation-convergence theorem in pure RL language, a regret-based partial order on abstractions, Lipschitz stability of K_ρ under margin estimation error, and connections to free energy and multi-agent coordination. A cross-framework corollary converts the regret floor into a generalisation probability. Under the canonical independent prior, the optimal M-based policy generalises with probability $e^{-K_\rho(M)}$. The multi-class generalisation to $K > 2$ diagnostic outcomes is proved. Controlled POMDP experiments confirm the decomposition is numerically exact and that K_ρ discriminates between representations where accuracy and raw impurity do not. The weakness-maximisation theorems predict optimal generalisation through least commitment, but their formal object (the extension of a policy in an embodied language) does not have a direct analogue in neural network function approximation. Bridging that gap is identified as an open problem.

Keywords: regret decomposition; representation learning; state abstraction; selective forgetting; generalisation; partially observable decision processes; weakness maximisation; information bottleneck; bisimulation; causal models

1. Introduction

When an agent compresses its observation history into a finite representation, it loses information. Some of that loss is harmless. Some of it makes the correct action ambiguous inside a representation cell, creating irreducible regret that no policy layered on top of the representation can fix. How much regret does a given representation force?

This paper gives an exact answer in the form of a decomposition. Fix a representation M and quotient history-test pairs by the representation-test cells $(M(h), T)$. The minimum average normalized regret over all M-based policies is exactly the minimum margin-weighted forgetting cost $K_\rho(M)$ needed to make the optimal bet single-valued on each such cell (Theorem 1). For any actual M-based policy, total regret decomposes exactly into $K_\rho(M)$ plus an avoidable within-cell misreporting term (Corollary 1). This is a representation-quality diagnostic analogous to the bias–variance decomposition in supervised learning: it tells you which part of the error is structural and which part is fixable by better optimisation.

The same quantity has a second life. In the Stack-Theoretic framework for generalisation-optimal learning [1–3], the natural defect variables are weakness, extension size, and selective forgetting on

 doi:10.20944/preprints202603.1546.v1

embodied task extensions. I prove that $K_\rho(M)$ is exactly a weighted weakness deficit on a lifted diagnostic task (Corollary 3). This connects RL regret to the broader Stack-Theoretic programme, which already links generalisation-optimal learning to selective memory, intervention-sensitive causal structure, free energy, and multiscale biological organisation [2,4]. It also connects Stack Theory to the standard RL and causal world-model literature [5–8]. The Stack-Theoretic side of the bridge is not a contribution of this paper as the EGRL programme, selective-memory clauses, and pair-proxy machinery were already theorem-level in Bennett [3,9,10,11]. The present result is a contribution to the mathematical core of that broader programme.

The identity is also a proof technique. I use it to derive several direct corollaries that illustrate the bridge's utility. A representation-convergence theorem in pure RL language says that any two minimal zero-regret representations must induce the same partition on the informative support (Theorem 3). A regret-based partial order on abstractions complements the existing bisimulation-metric order (Proposition 2). A Lipschitz stability result shows that K_ρ degrades gracefully under margin estimation error (Proposition 3). A cross-framework corollary converts the regret floor into a generalisation probability: under the canonical independent prior, the optimal M-based policy generalises with probability $e^{-K_\rho(M)}$ (Corollary 5). Further corollaries connect irreducible regret to a free-energy floor (Corollary 4) and quantify the coordination cost of shared representations in multi-agent settings (Proposition 4). The multi-class generalisation to $K > 2$ diagnostic outcomes is proved as Theorem 2.

The rest of the paper is organized as follows. Section 2 gives related work. Section 3 proves the exact reduction and policy-wise decomposition. Section 4 rewrites the same result in native Stack-Theoretic language. Section 5 derives new results via the bridge. Section 6 gives controlled POMDP experiments confirming the identity and testing K_ρ as a representation diagnostic. Section 7 discusses the prospects for using the bridge to guide representation learning in RL, identifies the central open problem of operationalising formal weakness in continuous function approximation, and summarises preliminary experimental findings.

2. Related Work

2.1. Selection Theorems and Genealogy

The broad claim that competence pressure forces agents to acquire internal structure has several independent lines. The Stack-Theoretic line proves weakness optimality under the maximally uninformative extension model (2023) and derives intervention-sensitive causal identities under explicit preconditions (2023–2025) [1–3,12]. Richens and Everitt derive robust-causal-model results under distributional shift (2024) [6]. Nayebi works in a standard POMDP setting and derives quantitative selection theorems for predictive state, memory, modularity, regime tracking, and recoding match (2026) [7]. These five structural selection theorems are distinctive contributions of Nayebi's programme that do not have direct Stack-Theoretic antecedents. The two frameworks also handle noise differently. Rabin and Scott showed that the languages recognisable by deterministic and non-deterministic finite automata coincide [13]. In the Stack-Theoretic framework, the role played by stochastic policies in Nayebi's setting is played by selective forgetting, which handles noise and inconsistency by discarding outlying data rather than by randomising over actions [3,12]. A detailed comparison is in Table 1; a claim-by-claim antecedent inventory is in Appendix C. The exact regret identity, the policy-wise decomposition, the multi-class generalisation, the measurable version, and the derived consequences in Section 5 are new to this paper.

The representational results sit inside a broader debate. Brooks pushed intelligence without explicit internal representation, while Marr and later work in representation learning and transfer emphasise structured internal states [14–17]. The causal representation learning programme argues that disentangled causal variables are the natural targets [18–21]. Nayebi's recoding theorem and Cao and Yamins' contravariance principle are best read against that background [7,22].

 doi:10.20944/preprints202603.1546.v1

Table 1. How the three lines relate.

	Bennett	**Richens & Everitt**	**Nayebi**
Main setting	Task extensions and policy weakness	Causal models under distributional shifts	POMDP betting tasks and predictive tests
Competence variable	Weakness, with selective forgetting for noise	Robust regret-bounded adaptation	Average-case normalised regret
Core structural conclusion	Optimal generalisation selects weak policies and, under preconditions, intervention-sensitive causal identities	Robust agents must learn an approximate causal model	Low regret selects predictive state, memory, modularity, regime tracking, and recoding match

2.2. State Abstraction and Bisimulation

The representation-test quotient in the bridge theorem is a form of state abstraction. The state abstraction literature in RL is mature and provides important context. Li, Walsh, and Littman give a unified treatment of five abstraction schemes for MDPs, ranging from model-irrelevance to π^*-irrelevance and bisimulation, and analyse which preserve optimal planning [23]. Givan, Dean, and Greig earlier studied equivalence notions and model minimization for MDPs, identifying bisimulation as the finest useful equivalence [24]. Ferns, Panangaden, and Precup introduced quantitative bisimulation metrics that measure state similarity continuously rather than through hard partitions, providing value-function bounds on the cost of aggregation [25,26]. Abel and collaborators extended these ideas to approximate and lifelong settings [27,28].

The quotient cells differ from bisimulation cells in an important respect. Bisimulation groups states by identical transition and reward structure. The representation-test cells group history-test points by what the representation M can see and which diagnostic test is applied. The bridge theorem then says exactly how much regret this grouping forces. In that sense it complements the bisimulation programme: bisimulation metrics bound value-function loss, while the bridge theorem identifies exact regret cost at a given quotient.

More recent deep RL work brings bisimulation ideas into the function-approximation regime. Zhang et al. learn representations that respect a bisimulation distance [29], Castro scales bisimulation computation to large deterministic MDPs [30], and Gelada et al. learn latent-space models that preserve MDP structure [31]. Nayebi's selection theorems and the bridge result operate at a more abstract level. They characterise when low regret forces a representation to preserve certain distinctions, regardless of how the representation is learned.

Remark 1 (Bisimulation distance and cell impurity)**.** *In an MDP with known transition and reward structure, if two states s, s' belong to the same representation-test cell C and the bisimulation metric $d_\sim(s, s')$ of Ferns et al. [25] is large, then s and s' are likely to carry different optimal actions, so their cell will be impure and will contribute positively to $K_\rho(M)$. Conversely, if every pair in a cell has $d_\sim(s, s') = 0$, then the cell is pure and contributes zero. So $K_\rho(M)$ can be read as an aggregate measure of how much bisimulation-relevant structure the representation has thrown away. The two formulations are complementary: bisimulation metrics bound value-function loss, while $K_\rho(M)$ gives exact regret cost at a given quotient.*

2.3. Predictive State and World Models

A parallel line of work asks what internal structure low-regret agents must have by reasoning about predictive sufficiency. Littman, Sutton, and Singh introduced predictive state representations, showing that observable predictions can serve as a complete state description without latent-variable modeling [32]. Singh, James, and Rudary extended the framework to controlled systems [33], and Boots, Siddiqi, and Gordon connected PSR learning and planning [34]. The approximate information state literature addresses partial observability more directly by identifying sufficient statistics for near-optimal planning [35].

 doi:10.20944/preprints202603.1546.v1

On the model-learning side, Ha and Schmidhuber's world models [36], Hafner et al.'s Dreamer agent [37], and Schrittwieser et al.'s MuZero [38] demonstrate that learned latent dynamics can support competitive planning. Richens, Everitt, and Abel give a theoretical counterpart, proving that general agents need world models under distributional shift [8]. The bridge theorem gives a different angle on the same family of questions. Rather than asking what structure a model must have, it asks how much regret a fixed representation forces, measured in exact weighted forgetting terms.

2.4. Causal Inference and the Causal Hierarchy

The causal side of the genealogy involves more than Richens and Everitt alone. Pearl's structural causal model framework [39,40] and the Spirtes–Glymour–Scheines algorithmic tradition [41] provide the language in which "learning a causal model" is made precise. Bareinboim et al.'s causal hierarchy theorem formalises the separation between observational, interventional, and counterfactual reasoning [42]. Lattimore, Lattimore, and Reid study causal bandits, where the agent must learn which interventions are effective [43].

My emergent causality result [12] proved that under weakness maximisation, representations of causal interventions emerge as variables in the embodied language, without presupposing a do-operator. The key argument is that an agent which confuses intervention with passive observation will adopt a more committal (and therefore weaker in the weakness sense) policy than one which distinguishes them, so weakness pressure forces the distinction to emerge. That result, and the subsequent causal identity theorems in [2] and [3], are antecedent to the Richens-Everitt result that robust agents must learn approximate causal models under distributional shift [6].

The two lines are somewhat complementary. My argument was that since weakness maximisation is the optimal choice of learning heuristic [1], an agent that seeks to be optimal within the constraints of its body must use it. Such agents *must* learn causal models [12]. In contrast, Richens and Everitt claimed *robust* agents must learn causal models, basing their claim on reinforcement learning. These are approximately the same claim derived from two different formalisms, which suggests a bridge between Stack Theory and reinforcement learning may be beneficial. In this paper, the regret ordering (Proposition 2) gives a quantitative counterpart to the qualitative claim that robust agents must learn causal models. If the causal partition refines the diagnostic partition $\mathcal{P}^{\star}_{+}$, it sits at zero K_ρ and incurs no irreducible regret. If it does not, the bridge tells you exactly how much regret that costs. Richens and Everitt handle worst-case robustness across distributions. The bridge presented here handles the exact cost at a fixed evaluation distribution.

3. Regret as Exact Weighted Forgetting

I now give the bridge in finite form. The finite version is the natural one for Stack Theory because the original framework works over finite embodied vocabularies. A measurable version is given in Appendix A. The finite theorem is stated first because it keeps the reduction transparent and matches the original finite task-extension setting.

$$(\mathcal{X}, \mu, y, \rho) \xrightarrow{\text{quotient by } M} \mathcal{C}_M \xrightarrow{\text{drop the cheaper label in each mixed cell}} K_\rho(M)$$
$$K_\rho(M) \xrightarrow{\text{lift to a diagnostic Stack task}} Z - W^{\star}(M).$$

Figure 1. The reduction at a glance. Quotient history-test points by the representation-test cells induced by M. Drop the cheaper label class in each mixed cell. Read the same number as a weighted weakness deficit on the lifted task.

3.1. Setup

Fix a finite evaluation support

$$\mathcal{X} = \{x_1, \ldots, x_n\}, \qquad x_i = (h_i, T_i),$$

 doi:10.20944/preprints202603.1546.v1

with probability masses $\mu_i > 0$ and $\sum_i \mu_i = 1$. For each x_i, let

$$p_i := p_{T_i}(h_i), \qquad m_i := \left|p_i - \tfrac{1}{2}\right|.$$

Let $y_i \in \{0,1\}$ denote the optimal bet, where $y_i = 1$ when $p_i \geq \frac{1}{2}$ and $y_i = 0$ otherwise. When $p_i = \frac{1}{2}$, the choice of y_i is arbitrary because the weight below is then zero.

Fix a candidate memory representation M. Write $i \sim_M j$ iff $M(h_i) = M(h_j)$ and $T_i = T_j$. Let $\mathcal{C}_M$ be the resulting set of cells. I will call them the representation-test cells induced by M. An M-based policy chooses one report probability $q_C \in [0,1]$ of betting on outcome $y = 1$ for each representation-test cell $C \in \mathcal{C}_M$. This is a standard setup for analysing representation quality in partially observed decision problems [5,23,44]. For any policy π, write $\delta_i(\pi) := 1 - V_i^\pi / V_i^\star$ for the normalised regret on support point x_i, where $V_i^\star = \max\{p_i, 1-p_i\}$ is the optimal success probability and V_i^π is the success probability under π.

Definition 1 (Margin weight). *For each support point x_i, define*

$$\rho_i := \frac{2m_i}{\frac{1}{2} + m_i} \in [0,1].$$

Note that $\frac{1}{2} + m_i = \max\{p_i, 1-p_i\}$, so ρ_i is the normalised regret penalty per unit of wrong-action mass on test i. This is exactly the multiplier that converts wrong-action probability into the normalised regret $\delta_i = 1 - V_i^\pi / V_i^\star$ used in the RL selection-theorem literature [7], and the same multiplier arises as the natural prior weight w_P in the Stack-Theoretic programme [3]. When a test is nearly a coin flip, ρ_i is near zero. When a test is decisive, ρ_i is large. The role is analogous to the margin-based weighting that appears in information-theoretic accounts of lossy compression: uninformative distinctions are cheap to lose, informative ones are expensive [45].

Remark 2 (Why the bridge is natural). *The quotient cells come from fixing M exactly as in the memory setting and keeping the test variable visible. The deletion move is selective forgetting on the diagnostic support. The margin weights ρ_i, the quotient cells, and the deletion move all have native Stack-Theoretic antecedents in w_P, selective forgetting, and the diagnostic extension model [3]. The prior weight w_P appeared in the Stack-Theoretic generalisation-probability theorems [1,3]. The margin weight ρ_i appeared later as Nayebi's normalised regret multiplier [7]. The contribution of this paper is the precise identification of these as the same quantity. The cellwise structure that makes the bridge exact follows from that identification. Nayebi's betting framework provides the clean RL-side formalisation that gives the bridge its second endpoint.*

Remark 3 (Scope of the normalisation). *The exact identity depends on normalised regret $\delta_i = 1 - V_i^\pi / V_i^\star$ and the resulting margin weight ρ_i. Under unnormalised regret $V_i^\star - V_i^\pi$ or under alternative loss functions, the cellwise structure survives but the exact identification requires a different weight. Specifically, unnormalised regret replaces ρ_i with the constant $2m_i$, giving an unweighted deletion cost on the diagnostic support rather than a margin-weighted one. The bridge to weighted weakness (Corollary 3) requires the normalised form because it matches the prior-weighted quantity w_P already native to Stack Theory.*

Lemma 1 (Pointwise regret decomposition). *Let π be any M-based policy, and let q_C be its report probability on cell C. For $i \in C$,*

$$\delta_i(\pi) = \rho_i\Big((1-q_C)\mathbf{1}\{y_i = 1\} + q_C\mathbf{1}\{y_i = 0\}\Big).$$

Proof. If $y_i = 1$, then $p_i \geq \frac{1}{2}$ and the optimal report is the left bet. The success probability under report probability q_C is

$$V_i^\pi = q_C p_i + (1-q_C)(1-p_i),$$

 doi:10.20944/preprints202603.1546.v1

while the optimal success probability is

$$V_i^\star = p_i = \tfrac{1}{2} + m_i.$$

Hence

$$\delta_i(\pi) = 1 - \frac{V_i^\pi}{V_i^\star} = \frac{2m_i(1-q_C)}{\frac{1}{2}+m_i} = \rho_i(1-q_C).$$

If $y_i = 0$, the symmetric calculation gives $\delta_i(\pi) = \rho_i q_C$. □

Intuition.

Regret is just mistake probability scaled by how informative the test is. That scaling is the only reason the exact bridge needs a weighted, rather than raw, forgetting variable.

Definition 2 (Margin-weighted forgetting cost)**.** *The margin-weighted forgetting cost of M is*

$$K_\rho(M) := \min\Big\{ \sum_{i\in B} \mu_i\rho_i \;\Big|\; B \subseteq \{1,\dots,n\},\ \text{and for every } C \in \mathcal{C}_M,\ y \text{ is constant on } C \setminus B \Big\}.$$

The cost of forgetting point i is its evaluation mass times its margin weight. You forget just enough so that, inside every representation-test cell, the correct bet becomes single-valued. This is a weighted version of the selective deletion that appears in the state abstraction literature when one asks how much structure a given partition can preserve [23,24].

Theorem 1 (Exact reduction)**.** *For every fixed memory representation M,*

$$\inf_{\pi \text{ is } M\text{-based}} \sum_{i=1}^n \mu_i\,\delta_i(\pi) = K_\rho(M).$$

Proof. For each cell $C \in \mathcal{C}_M$, define

$$A_C := \sum_{i\in C,\ y_i=1} \mu_i\rho_i, \qquad B_C := \sum_{i\in C,\ y_i=0} \mu_i\rho_i.$$

By Lemma 1, any M-based policy with report probability q_C on cell C contributes

$$A_C(1-q_C) + B_C q_C$$

to expected regret on that cell. This is affine in q_C, so its minimum over $q_C \in [0,1]$ is

$$\min\{A_C, B_C\}.$$

Summing over cells gives

$$\inf_{\pi \text{ is } M\text{-based}} \sum_{i=1}^n \mu_i\,\delta_i(\pi) = \sum_{C\in\mathcal{C}_M} \min\{A_C, B_C\}.$$

Now compute the forgetting cost. To make y single-valued on a fixed cell C, one must delete either every $y_i = 1$ point in C or every $y_i = 0$ point in C. Deleting anything less leaves both labels in the cell. So the minimum forgetting cost on cell C is exactly

$$\min\{A_C, B_C\}.$$

 doi:10.20944/preprints202603.1546.v1

Summing over cells gives

$$K_\rho(M) = \sum_{C \in \mathcal{C}_M} \min\{A_C, B_C\}.$$

Comparing the two displays proves the theorem. □

Remark 4 (Computational cost)**.** *Despite the combinatorial phrasing of Definition 2, $K_\rho(M)$ is computable in $O(n)$ time once the representation-test cells are formed, because the minimum deletion on each cell decomposes independently into* $\min\{A_C, B_C\}$*. No search over subsets is required.*

Intuition.

Once the task is quotiented by representation-test cell, regret is exactly the cheapest weighted deletion needed to make the correct action a function of that cell. The two quantities are not merely analogous. They are equal.

Remark 5 (The optimal bet is selective forgetting)**.** *The identification is not merely numerical. The optimal M-based policy's per-cell binary choice (bet on $y = 1$ or $y = 0$) is itself a deletion operator. In each cell, the policy commits to one label class and discards the other. The discarded class is the cheaper one, and its margin-weighted mass is the cell's contribution to $K_\rho(M)$. So the RL-side optimal policy is not merely computing a cost equal to selective forgetting. It is* performing *selective forgetting on the diagnostic support.*

Remark 6 (Relation to weighted Bayes error)**.** *The cellwise quantity* $\min\{A_C, B_C\}$ *is recognisable as a weighted Bayes error on the representation-test partition. From a statistical decision theory perspective, this is the conditional Bayes risk of the partition under a margin-weighted 0-1 loss. The contribution is the three-way identification. It equals Nayebi's normalised regret exactly, not merely up to bounds. It equals the minimum margin-weighted deletion cost in the selective-forgetting sense. And it equals a weighted weakness deficit on a lifted Stack-Theoretic task (Corollary 3). The first gives it RL semantics, the second gives it a forgetting and compression interpretation, and the third connects it to the generalisation-optimal learning programme.*

Corollary 1 (Policy-wise decomposition)**.** *For each cell $C \in \mathcal{C}_M$, choose any*

$$q_C^\star \in \arg\min_{q \in [0,1]} \big(A_C(1-q) + B_C q\big).$$

Then every M-based policy π satisfies

$$\sum_{i=1}^{n} \mu_i\, \delta_i(\pi) = K_\rho(M) + \sum_{C \in \mathcal{C}_M} |A_C - B_C|\, |q_C - q_C^\star|.$$

In particular,

$$K_\rho(M) \le \sum_{i=1}^{n} \mu_i\, \delta_i(\pi),$$

with equality exactly when the policy is cellwise optimal on every non-tied cell.

Proof. If $A_C > B_C$, then the minimum is attained at $q_C^\star = 1$ and

$$\begin{aligned} A_C(1-q_C) + B_C q_C &= B_C + (A_C - B_C)(1-q_C) \\ &= \min\{A_C, B_C\} + |A_C - B_C|\, |q_C - q_C^\star|. \end{aligned}$$

If $A_C < B_C$, the minimum is attained at $q_C^\star = 0$ and the same identity becomes

$$\begin{aligned} A_C(1-q_C) + B_C q_C &= A_C + (B_C - A_C) q_C \\ &= \min\{A_C, B_C\} + |A_C - B_C|\, |q_C - q_C^\star|. \end{aligned}$$

 doi:10.20944/preprints202603.1546.v1

If $A_C = B_C$, both sides reduce to A_C for every q_C. Summing over cells and using Theorem 1 proves the result. □

Intuition.

This separates two kinds of regret. The first part is representation cost. It is the aliasing penalty you cannot remove without changing M. The second part is execution cost. It is the avoidable penalty from betting suboptimally even after M is fixed. That decomposition is useful for the same reason the bias–variance decomposition is useful in supervised learning: it tells you which part of the error is structural and which part is fixable by better optimisation.

Example 1 (A tiny quotient task). *Suppose one mixed cell C_1 has $A_{C_1} = 0.30$ and $B_{C_1} = 0.20$, while a second cell C_2 is pure with $A_{C_2} = 0$ and $B_{C_2} = 0.25$. Then*

$$K_\rho(M) = \min\{0.30, 0.20\} + \min\{0, 0.25\} = 0.20.$$

If an actual policy uses $q_{C_1} = 0.8$ and $q_{C_2} = 0$, then its regret is

$$0.30(1 - 0.8) + 0.20(0.8) + 0.25(0) = 0.22.$$

The extra 0.02 is exactly

$$|0.30 - 0.20|\,|0.8 - 1| = 0.02.$$

The example makes the decomposition concrete. The 0.20 is the irreducible cost of aliasing opposite bets inside C_1. The extra 0.02 is not representational at all. It is just the cost of choosing the wrong within-cell report.

Corollary 2 (Unweighted forgetting on the informative region). *In practice one may wish to ignore near-tie diagnostic tests whose margin m_i is close to zero, for instance when empirical estimates of p_i are noisy and tests near $\frac{1}{2}$ carry little signal. The following shows that restricting to sufficiently informative tests recovers an unweighted forgetting picture up to fixed constants.*

Fix $\gamma \in (0, \frac{1}{2}]$ and let

$$\mathcal{X}_\gamma := \{i : m_i \geq \gamma\}.$$

Define the unweighted informative-region forgetting cost

$$K_\gamma(M) := \min\Big\{ \sum_{i \in B \cap \mathcal{X}_\gamma} \mu_i \;\Big|\; B \subseteq \{1, \dots, n\},\ \text{and for every } C \in \mathcal{C}_M,\ y \text{ is constant on } (C \cap \mathcal{X}_\gamma) \setminus B \Big\}.$$

Then, with $c(\gamma) = \frac{4\gamma}{1+2\gamma}$,

$$c(\gamma) K_\gamma(M) \leq \inf_{\pi \text{ is } M\text{-based}} \sum_{i \in \mathcal{X}_\gamma} \mu_i\, \delta_i(\pi) \leq K_\gamma(M).$$

Proof. On $\mathcal{X}_\gamma$ we have $\rho_i \in [c(\gamma), 1]$. Apply Theorem 1 on the restricted support $\mathcal{X}_\gamma$. The resulting weighted deletion cost lies between $c(\gamma)$ times the unweighted deletion mass and the unweighted deletion mass itself. □

Intuition.

This recovers the intuitive raw-forgetting picture. If you only look at tests that matter by at least γ, then exact weighted forgetting and ordinary forgetting differ only by fixed constants.

 doi:10.20944/preprints202603.1546.v1

3.2. Multi-Class Generalisation

The binary case connects directly to Nayebi's betting setup, but the algebraic structure survives for any finite number of outcome classes.

Theorem 2 (Multi-class exact reduction). *Fix a finite evaluation support as before, but let $y_i \in \{1, \dots, K\}$ for $K \geq 2$ outcome classes. For each cell $C \in \mathcal{C}_M$ and class k, define*

$$A_C^{(k)} := \sum_{i \in C,\, y_i = k} \mu_i \rho_i.$$

Then

$$\inf_{\pi \text{ is } M\text{-based}} \sum_{i=1}^{n} \mu_i \, \delta_i(\pi) = \sum_{C \in \mathcal{C}_M} \Big(\sum_{k=1}^{K} A_C^{(k)} - \max_k A_C^{(k)} \Big).$$

The right-hand side is the minimum margin-weighted deletion cost needed to make y single-valued on each representation-test cell, and it equals the cellwise weighted Bayes error.

Proof. An M-based policy assigns a distribution $q_C \in \Delta^{K-1}$ over the K classes for each cell C. Its expected regret on cell C is

$$\sum_{k=1}^{K} A_C^{(k)} (1 - q_C^{(k)}) = \Big(\sum_k A_C^{(k)} \Big) - \sum_k A_C^{(k)} q_C^{(k)}.$$

To minimise this over $q_C \in \Delta^{K-1}$, one must maximise $\sum_k A_C^{(k)} q_C^{(k)}$, which is achieved by putting all mass on the class with the largest $A_C^{(k)}$. So the minimum per-cell regret is $\sum_k A_C^{(k)} - \max_k A_C^{(k)}$. To make y single-valued on C, one must delete every point not in the dominant class, costing $\sum_k A_C^{(k)} - \max_k A_C^{(k)}$. Summing over cells gives both sides. □

4. A Stack-Theoretic Reformulation

The previous section gives the exact bridge in Nayebi's language. I now translate it into the Stack-Theoretic language. The finite diagnostic support becomes the unseen region of a lifted task, and margin weights become a weighted weakness score on the same representation-test quotient.

4.1. Notation and Prerequisites

This subsection gives self-contained definitions sufficient to verify the Stack-Theoretic form of the bridge (Corollary 3). The reader familiar with Stack Theory may skip ahead.

A *lifted diagnostic task* is defined from the finite evaluation support as follows. Construct a set $U = \{u_1, \dots, u_n\}$ of *unseen outputs*, one for each support point $x_i = (h_i, T_i)$. Each unseen output u_i inherits the label $y_i \in \{0, 1\}$ and weight $w_i := \mu_i \rho_i$ from its corresponding support point. A *policy on the lifted task* selects, for each representation-test cell $C \in \mathcal{C}_M$, which label class to retain. Its *extension*

$$\mathrm{Ext}(\vartheta) \cap U := \{u_i \in U : y_i \text{ equals the label class retained in cell } C \ni i\}$$

is the set of unseen outputs whose label matches the policy's selection. The policy is *admissible* if it retains outputs from at most one label class per cell.

This vocabulary is standard in the Stack-Theoretic framework [1,3] but is used here only in the restricted form above. Table 2 gives the correspondence.

 doi:10.20944/preprints202603.1546.v1

Table 2. Notation mapping between the Stack-Theoretic and RL sides.

Stack-Theoretic side	RL side
Truth set of a diagnostic statement	Representation-test cell C
Child policy ϑ	Policy π
Label class selection per cell	Report probability q_C
Weight not retained: $w_i \mathbf{1}\{i \notin \mathrm{Ext}(\vartheta)\}$	Normalised regret δ_i
Weakness deficit $Z - W^\star(M)$	Min regret $K_\rho(M)$
Weakness-maximising admissible policy	Optimal policy per cell

4.2. Weighted Weakness and the Bridge

Definition 3 (Prior-weighted weakness)**.** *Let $U = \{u_1, \ldots, u_n\}$ be a finite unseen region with positive weights $w_1, \ldots, w_n$. For any policy ϑ, let $\mathrm{Ext}(\vartheta)$ denote its extension, i.e. the set of all outputs it is compatible with in the host language [1,3]. Define its prior-weighted weakness on U by*

$$W(\vartheta) := \sum_{u_i \in \mathrm{Ext}(\vartheta) \cap U} w_i.$$

Ordinary weakness counts how many unseen continuations a policy leaves open. Weighted weakness counts how much weighted unseen future it leaves open. The idea is analogous to the information bottleneck: discard what is cheap, keep what is expensive [45].

Remark 7 (This is not a new Stack-Theoretic primitive)**.** *The appendix already defines the more general quantity*

$$w_P(\pi) = P\big(S \subseteq \mathrm{Ext}(\pi) \cap U\big)$$

for an arbitrary prior P over subsets $S \subseteq U$. The score $W(\vartheta)$ above is just the finite product-prior specialisation that becomes convenient after taking logs [3]. So this section is a translation of existing Stack-Theoretic machinery, not a replacement for it.

Proposition 1 (Independent nonuniform priors give weighted weakness)**.** *Assume each unseen output $u_i \in U$ becomes relevant independently with probability $r_i \in (0,1)$. Then, for any correct child policy ϑ,*

$$\log \Pr(\vartheta \textit{ generalises}) = \sum_{u_i \notin \mathrm{Ext}(\vartheta) \cap U} \log(1 - r_i).$$

Equivalently, maximising generalisation probability is the same as maximising

$$\sum_{u_i \in \mathrm{Ext}(\vartheta) \cap U} \big(-\log(1 - r_i)\big).$$

So the Bayes-optimal rule is weighted weakness maximisation with weights

$$w_i := -\log(1 - r_i).$$

Proof. Generalisation occurs exactly when every unseen output omitted by ϑ fails to become relevant. Independence gives the product

$$\Pr(\vartheta \text{ generalises}) = \prod_{u_i \notin \mathrm{Ext}(\vartheta) \cap U} (1 - r_i).$$

Taking logs yields the display. The remaining term depends on ϑ only through the retained weighted support. □

 doi:10.20944/preprints202603.1546.v1

This is the weighted version of the earlier counting theorem [1]. Uniform priors recover ordinary weakness. Biased priors recover weighted weakness. The uniform case is the no-free-lunch baseline: if you know nothing about which unseen outputs matter, you keep as many open as possible [46].

Corollary 3 (Exact Stack-Theoretic form of the bridge)**.** *Let $w_i := \mu_i \rho_i$. Build a lifted diagnostic task whose unseen outputs are $u_1, \ldots, u_n$, one for each support point x_i. Restrict admissible policies so that, on each representation-test cell $C \in \mathcal{C}_M$, the policy retains outputs from at most one label class. Let*

$$Z := \sum_{i=1}^{n} w_i, \qquad W^\star(M) := \max_{\vartheta \text{ admissible}} W(\vartheta).$$

Then

$$\inf_{\pi \text{ is } M\text{-based}} \sum_{i=1}^{n} \mu_i \, \delta_i(\pi) = Z - W^\star(M).$$

Moreover, choosing independent prior probabilities

$$r_i := 1 - e^{-w_i}$$

turns $W^\star(M)$ into the Bayes-optimal weighted weakness score of the lifted task.

Proof. On the lifted task, retaining a support point u_i means not forgetting the corresponding diagnostic point x_i. Admissibility means each cell keeps at most one label class. So maximising retained weighted mass is the same optimisation problem as minimising deleted weighted mass. Hence

$$W^\star(M) = Z - K_\rho(M).$$

Now apply Theorem 1. The final statement follows from Proposition 1 and the choice $r_i = 1 - e^{-w_i}$. □

Intuition.

Regret is the weighted unseen future that the representation forced you to discard. The earlier counting theorem says generalisation is about how much unseen future you leave open; the bridge says regret is the part of that weighted future lost to aliasing.

4.3. Continuity with the Earlier EGRL and Pair-Proxy Programme

This bridge is not a detached 2026 repair. EGRL here means the appendix's embodied GRL translation from reward-labelled interaction histories into the task-extension language. A pair proxy is the rule used to rank the admissible positive-negative policy pairs extracted from that history. The weakness pair proxy ranks such pairs by the size of their joint extension. The appendix already constructs instantiated history tasks, reward predicates, admissible pair policies, selective memory for inconsistent histories, and EGRL-style translations from interaction histories into the task-extension language [3,9–11]. The present theorem is the sharp mathematical core of that broader programme. It fixes a diagnostic goal family, quotients the history support by representation-test cell, and proves that the resulting defect variable is weighted selective forgetting.

Intuition.

The earlier programme translated the whole interaction loop. This paper isolates the reusable kernel: the exact defect variable at the representation-test quotient.

5. Consequences of the Bridge

The exact identity is a proof technique, not just a dictionary. This section derives results via the bridge to illustrate its utility. I separate results that follow from the exact reduction alone (Section 5.1) from those that additionally import prior Stack-Theoretic machinery (Section 5.2).

 doi:10.20944/preprints202603.1546.v1

5.1. Direct Corollaries of the Exact Reduction

The following results use only Theorem 1 and basic partition combinatorics. No Stack-Theoretic imports are required.

5.1.1. A Regret-Based Partial Order on Abstractions

The bridge induces a natural ordering on representations.

Proposition 2 (Regret ordering). *Define $M_1 \preceq_\rho M_2$ iff $K_\rho(M_1) \geq K_\rho(M_2)$. This partial order is consistent with the partition refinement lattice: if $\mathcal{C}_{M_2}|_{I_+}$ refines $\mathcal{C}_{M_1}|_{I_+}$, then $K_\rho(M_2) \leq K_\rho(M_1)$. The inequality is strict whenever the refinement splits at least one mixed cell on I_+ into subcells in which the minority class flips (i.e. the label achieving $\min\{A, B\}$ differs between subcells).*

Proof. If $\mathcal{C}_{M_2}$ refines $\mathcal{C}_{M_1}$, then every cell $C_1 \in \mathcal{C}_{M_1}$ is partitioned into subcells $C_2^{(j)} \in \mathcal{C}_{M_2}$. By the cellwise formula,

$$\min\{A_{C_1}, B_{C_1}\} \geq \sum_j \min\{A_{C_2^{(j)}}, B_{C_2^{(j)}}\},$$

since $\min\{a + a', b + b'\} \geq \min\{a, b\} + \min\{a', b'\}$ for non-negative reals. Summing gives $K_\rho(M_1) \geq K_\rho(M_2)$. For strictness, suppose a mixed cell C_1 with $A_{C_1} > B_{C_1}$ splits into subcells C_2, C_2' where $A_{C_2} > B_{C_2}$ but $A_{C_2'} < B_{C_2'}$. Then

$$\min\{A_{C_1}, B_{C_1}\} = B_{C_2} + B_{C_2'} > B_{C_2} + A_{C_2'} = \min\{A_{C_2}, B_{C_2}\} + \min\{A_{C_2'}, B_{C_2'}\},$$

so the contribution from this cell is strictly smaller after refinement. □

Intuition.

This gives the state-abstraction and bisimulation communities a new tool: a regret-based partial order on abstractions that complements the existing bisimulation-metric order of Ferns et al. [25]. Finer representations have lower or equal K_ρ, and the ordering tells you exactly when refinement helps. The ordering also complements the Richens-Everitt qualitative characterisation [6]. Their result says robust agents must learn causal models. The regret ordering says how much it costs if the learned representation falls short of the causal partition.

5.1.2. Representation Convergence in RL Language

The regret ordering has a clean endpoint. At zero K_ρ, every cell is pure, and the coarsest pure partition is unique. This gives a representation convergence theorem stated entirely in RL language.

Theorem 3 (Representation convergence). *Fix a finite evaluation support $(\mathcal{X}, \mu, y, \rho)$ and let $I_+ = \{i : \mu_i \rho_i > 0\}$ be the informative support. Define $\mathcal{P}_+^\star := \big\{\{i \in I_+ : T_i = t,\, y_i = b\} : t \in \mathcal{T}_+,\, b \in \{0, 1\}\big\} \setminus \{\varnothing\}$, the coarsest partition of I_+ that separates test values and optimal labels.*

1. *$K_\rho(M) = 0$ if and only if $\mathcal{C}_M|_{I_+}$ refines $\mathcal{P}_+^\star$.*
2. *If M is minimal under partition coarsening on I_+ among the zero-regret representations, then $\mathcal{C}_M|_{I_+} = \mathcal{P}_+^\star$.*
3. *Any two minimal zero-regret representations induce the same partition on I_+ and differ there only by a relabelling of representation states.*

Proof. By Theorem 1, $K_\rho(M) = 0$ iff $\min\{A_C, B_C\} = 0$ for every cell C. On I_+ this means each cell contains only one optimal label at its fixed test value, which is exactly $\mathcal{C}_M|_{I_+} \preceq \mathcal{P}_+^\star$. Conversely, points outside I_+ carry zero deletion weight, so refinement there does not affect K_ρ. For part 2, $\mathcal{P}_+^\star$ is itself zero-cost because each of its cells is pure. If a minimal zero-cost M were strictly finer than $\mathcal{P}_+^\star$ on I_+, coarsening to $\mathcal{P}_+^\star$ would preserve zero cost, contradicting minimality. Part 3 follows because equality

 doi:10.20944/preprints202603.1546.v1

of quotients on I_+ means there is a bijection between nonempty representation states, which is a relabelling. □

Intuition.

Any two minimal zero-regret representations must carve the informative support the same way. The proof uses only the cellwise decomposition from Theorem 1 and basic partition logic.

5.1.3. Rate-Distortion Interpretation

$K_\rho(M)$ has a natural information-theoretic reading. The representation M is a lossy compression of histories. The margin-weighted forgetting cost $K_\rho(M)$ is the residual relevant information that the bottleneck has destroyed, weighted by diagnostic importance. In the language of the information bottleneck [45], $K_\rho(M)$ is the Bayes error of the sufficient statistic induced by M, expressed as a margin-weighted deletion cost. The bridge theorem says this is the exact regret cost of that compression. Representations with lower K_ρ are better compressions in the sense that matters for decision-making: they preserve the distinctions that actually affect regret and discard the ones that do not.

5.1.4. Stability under Margin Estimation Error

In practice the conditional probabilities p_i are estimated from data, so the margin weights ρ_i are known only approximately. The following shows that K_ρ is Lipschitz in the margin estimates, so small estimation errors produce small diagnostic errors.

Proposition 3 (Lipschitz stability of K_ρ)**.** *Let $\hat{\rho}_1, \ldots, \hat{\rho}_n$ be perturbed margin weights. Define $\hat{K}_\rho(M)$ by replacing ρ_i with $\hat{\rho}_i$ in Definition 2. Then*

$$\left|K_{\hat{\rho}}(M) - K_\rho(M)\right| \;\le\; \sum_{i=1}^{n} \mu_i \, |\hat{\rho}_i - \rho_i|.$$

Moreover, since ρ_i is a function of p_i with $|d\rho_i/dp_i| \le 4$ everywhere on $(0,1)$, estimation error in the conditional probabilities propagates as

$$\left|K_{\hat{\rho}}(M) - K_\rho(M)\right| \;\le\; 4 \sum_{i=1}^{n} \mu_i \, |\hat{p}_i - p_i|.$$

Proof. For each cell C, define $\hat{A}_C := \sum_{i \in C,\, y_i = 1} \mu_i \hat{\rho}_i$ and $\hat{B}_C$ analogously. Since $|\min\{a', b'\} - \min\{a, b\}| \le |a' - a| + |b' - b|$ for non-negative reals,

$$\left|\min\{\hat{A}_C, \hat{B}_C\} - \min\{A_C, B_C\}\right| \le |\hat{A}_C - A_C| + |\hat{B}_C - B_C| \le \sum_{i \in C} \mu_i |\hat{\rho}_i - \rho_i|.$$

Summing over cells gives the first bound. For the second, note that for $p > \frac{1}{2}$ we have $\rho = (2p-1)/p$, so $d\rho/dp = 1/p^2 \le 4$; for $p < \frac{1}{2}$ we have $\rho = (1-2p)/(1-p)$, so $|d\rho/dp| = 1/(1-p)^2 \le 4$. Hence $|\hat{\rho}_i - \rho_i| \le 4|\hat{p}_i - p_i|$. □

Intuition.

The diagnostic $K_\rho(M)$ degrades gracefully under estimation noise. With N samples per support point, standard concentration gives $|\hat{p}_i - p_i| = O(N^{-1/2})$, so $|K_{\hat{\rho}} - K_\rho| = O(N^{-1/2})$ as well.

5.2. Cross-Framework Consequences

The following results combine Theorem 1 with prior Stack-Theoretic machinery. Each one is flagged with the additional assumption it requires.

 doi:10.20944/preprints202603.1546.v1

5.2.1. Free-energy Floor from Irreducible Regret

Combining Corollary 3 with the Law of the Stack [4], which proves that under a uniform viability prior at layer $i+1$, the base-2 free energy proxy satisfies $F_2 \geq \log_2 |\mathrm{Ext}(\mu)| - |\mathrm{Ext}(\pi^i)|$, I obtain:

Corollary 4 (Free-energy floor from regret)**.** *In the setting of Corollary 3, suppose the lifted diagnostic task is embedded as a layer in a Stack-Theoretic hierarchy, with $W^\star(M)$ playing the role of the extension size of the admissible policy at that layer. Then the free-energy proxy at the layer above satisfies*

$$F_2 \geq \log_2 |\mathrm{Ext}(\mu)| - W^\star(M) = \log_2 |\mathrm{Ext}(\mu)| - (Z - K_\rho(M)).$$

In particular, any positive irreducible regret $K_\rho(M) > 0$ raises the free-energy floor relative to the zero-regret case by exactly $K_\rho(M)$.

Proof. By the Law of the Stack [4], $F_2 \geq \log_2 |\mathrm{Ext}(\mu)| - |\mathrm{Ext}(\pi)|$ where $|\mathrm{Ext}(\pi)|$ is the extension size of the realised policy at the diagnostic layer. The maximum extension size of any admissible policy on the lifted task is $W^\star(M)$. By Corollary 3, $W^\star(M) = Z - K_\rho(M)$. Substituting gives the display. □

Intuition.

This links RL regret to variational free energy through a chain of exact identities, not analogy. A bad representation at a lower layer puts a floor under free energy at the layer above.

5.2.2. Generalisation Probability from Regret

The free-energy corollary imports Stack-Theoretic structure but stays in the language of bounds. The following corollary converts the regret floor into a generalisation probability, giving K_ρ a direct predictive semantics that the RL identity alone does not provide.

Corollary 5 (Generalisation probability from regret)**.** *In the setting of Corollary 3, choose independent relevance probabilities $r_i = 1 - e^{-\mu_i \rho_i}$ as in Corollary 3. Under this prior, the probability that the optimal M-based policy generalises to a new set of diagnostic demands is*

$$\Pr(\textit{optimal policy generalises}) = e^{-K_\rho(M)}.$$

Proof. By Proposition 1, generalisation occurs exactly when no deleted unseen output becomes relevant. The optimal admissible policy deletes the cheaper label class in each cell, removing total weight $K_\rho(M)$ from the unseen region. Under the independent prior with $r_i = 1 - e^{-w_i}$, the generalisation probability of any admissible policy ϑ is

$$\prod_{u_i \notin \mathrm{Ext}(\vartheta) \cap U} (1 - r_i) = \prod_{u_i \notin \mathrm{Ext}(\vartheta) \cap U} e^{-w_i} = e^{-\sum_{u_i \notin \mathrm{Ext}(\vartheta) \cap U} w_i}.$$

For the optimal policy, the exponent is exactly $K_\rho(M)$. □

Intuition.

This is the payoff of the bridge as a bidirectional connection, not merely a dictionary. The RL side provides $K_\rho(M)$ as a regret floor. The Stack-Theoretic side, via the independent-prior generalisation model, converts that regret floor into a generalisation probability. Neither framework alone gives both quantities. A representation with $K_\rho = 0.02$ generalises with probability $e^{-0.02} \approx 0.98$ under the canonical prior; one with $K_\rho = 0.20$ generalises with probability $e^{-0.20} \approx 0.82$. The exponential sensitivity to K_ρ makes the diagnostic practically meaningful: small differences in irreducible regret translate into measurable differences in generalisation reliability.

 doi:10.20944/preprints202603.1546.v1

5.2.3. Multi-Agent Coordination Cost

Combining Theorem 1 with the contravariance result (Appendix B), I quantify the cost of sharing a representation across subsystems with incompatible diagnostic needs.

Proposition 4 (Coordination cost)**.** *Let N subsystems each face a diagnostic family on the same support with the same single test T. Let $\mathcal{P}_j^\star$ be the coarsest pure partition for subsystem j, with $|\mathcal{P}_j^\star| = k_j$ cells. A shared representation that achieves zero regret for all N subsystems simultaneously requires at least $|\mathcal{P}_1^\star \vee \cdots \vee \mathcal{P}_N^\star|$ cells, where $\vee$ denotes the common refinement. In the worst case this is $\prod_{j=1}^N k_j$.*

Proof. By Theorem 1, zero regret for subsystem j requires the shared partition to refine $\mathcal{P}_j^\star$. The coarsest partition refining all N is the common refinement $\mathcal{P}_1^\star \vee \cdots \vee \mathcal{P}_N^\star$. Its size is bounded above by $\prod_j k_j$ because each cell is determined by its class in each subsystem's partition. The bound is achieved when the partitions are independent (no two agree on the grouping of any pair of support points). □

Intuition.

Subsystems with incompatible diagnostic needs impose a combinatorial tax on shared representations. This connects to federated learning, where client drift [47] has the same structure, and to the biological literature on cancer-like coordination failure [4,48–50].

Example 2 (Coordination cost in a two-subsystem scenario)**.** *Consider a domain with 6 support points under one common test T. Subsystem A requires the partition $\mathcal{P}_A^\star = \{\{1,2\},\{3,4,5,6\}\}$ ($k_A = 2$), while subsystem B requires $\mathcal{P}_B^\star = \{\{1,2,3\},\{4,5,6\}\}$ ($k_B = 2$). Each subsystem needs only 2 representation states, but a shared representation that achieves zero regret for both must have at least $|\mathcal{P}_A^\star \vee \mathcal{P}_B^\star| = |\{\{1,2\},\{3\},\{4,5,6\}\}| = 3$ states. In the worst case with N such subsystems, the shared representation may require up to $\prod_j k_j$ states, while any individual subsystem needs only k_j. This gap is the coordination tax.*

6. Diagnostic Experiments

The theorems above are exact algebraic identities, so they do not require empirical validation. The purpose of this section is to show that $K_\rho(M)$ is informative in practice: that the bridge identity holds under native Stack-Theoretic computation, that K_ρ discriminates between representations where simpler diagnostics do not, and that the aliasing–execution decomposition tracks representation quality during training.

6.1. Two-Sided Verification on Boolean Domain

I encode the diagnostic task natively using the Stack Theory Suite [3]. The environment has $n_{\text{bits}} = 5$ Boolean variables, giving $2^5 = 32$ states. The vocabulary contains 10 programs (bit-tests: b_i=0 and b_i=1 for each bit i), giving an induced language $|L_{\mathfrak{v}}| = 243$. For each trial, a random structured Boolean label $y : \{0,1\}^5 \to \{0,1\}$ is generated that depends on 2–3 randomly chosen bits. Representations are defined by which bits the agent observes (0 through 5).

The bridge is verified two-sided through independent code paths. *Side A (RL):* cells are computed by direct bit inspection; K_ρ is computed via the cellwise formula $\sum_C \min\{A_C, B_C\}$. *Side B (Stack Theory):* for each possible observation pattern, the corresponding STS statement is built by conjoining the appropriate bit-test programs; its truth set is computed via `Statement.truth_set()`; the states in the truth set define the cell membership. K_ρ is then recomputed from the STS-derived cells. Both the cell structures and the resulting K_ρ values are compared. Over 100 trials × 6 representation levels = 600 checks, cell membership agrees exactly in every case, and K_ρ matches to numerical precision. The two-sided agreement confirms that the Stack-Theoretic reformulation is not a post-hoc relabelling: the STS truth-set computation and the RL cellwise computation produce identical cell structures through entirely independent formal machinery.

Table 3 shows that K_ρ and raw impurity behave differently. K_ρ reaches zero at 4 observed bits: the remaining unobserved bit creates cells that are impure but whose impurity carries zero margin weight

 doi:10.20944/preprints202603.1546.v1

(the minority class within each cell has $\rho_i \approx 0$ because the posterior is close to a coin flip on that bit). Raw impurity remains positive until all 5 bits are observed. In a separate comparison of 200 pairs of 2-bit versus 3-bit representations, K_ρ discriminated between representations that had similar accuracy ($< 5\%$ difference) in 21% of cases.

Table 3. Boolean classification experiment (5 bits, 32 states, 100 trials). K_ρ is computed via RL cellwise arithmetic (Side A) and independently via STS truth-set enumeration (Side B); both agree to numerical precision in all 600 checks. Cell counts are deterministic given the number of observed bits.

Bits observed	$K_\rho(M)$	Raw impurity	Accuracy	Cells
0	0.102 ± 0.008	0.318 ± 0.015	0.68 ± 0.01	1
1	0.079 ± 0.007	0.285 ± 0.014	0.71 ± 0.01	2
2	0.047 ± 0.007	0.247 ± 0.015	0.75 ± 0.02	4
3	0.019 ± 0.005	0.191 ± 0.015	0.81 ± 0.01	8
4	0.000 ± 0.000	0.124 ± 0.016	0.88 ± 0.02	16
5	0.000 ± 0.000	0.000 ± 0.000	1.00 ± 0.00	32

6.2. Discretised Encoder at Varying Granularity

To demonstrate the diagnostic value of K_ρ at moderate scale, I construct a structured POMDP with $|\mathcal{S}| = 16$ hidden states, $|\mathcal{O}| = 6$ observations, and observation sequences of length $L = 8$. The first 8 states have label $y = 1$; the remaining 8 have label $y = 0$. Observation distributions are block-structured so that states in the same class share similar emission profiles, with cross-block transitions occurring infrequently.

I train a two-layer MLP (hidden dimension 4, tanh activation) to convergence on each of 20 random POMDPs, then evaluate the learned representation at 8 levels of discretisation granularity (from 2 to 15 bins per hidden dimension). Finer discretisation yields more cells and a more refined partition of the observation-history space.

Table 4 shows that K_ρ is consistently smaller than raw impurity across all granularities, because it discounts aliasing on near-tie diagnostic tests. At 6 bins per dimension, $K_\rho = 0.002$ while raw impurity is still 0.012, a $5.5\times$ ratio. Both quantities eventually approach zero at fine granularity, but the gap at intermediate levels is where the diagnostic is most useful: a representation that looks impure by raw vote may already have near-zero aliasing cost on the tests that actually matter for decision-making.

Table 4. Discretised MLP encoder on a 16-state POMDP (20 trials). Finer discretisation reduces both K_ρ and raw impurity, but K_ρ drops faster: at 6 bins, K_ρ is a factor of $5.5\times$ smaller than raw impurity.

Bins	$K_\rho(M)$	Raw impurity	Accuracy	Cells
2	0.0097 ± 0.0004	0.041 ± 0.001	0.959 ± 0.001	166
3	0.0101 ± 0.0008	0.038 ± 0.002	0.962 ± 0.002	172
4	0.0057 ± 0.0004	0.026 ± 0.001	0.974 ± 0.001	343
5	0.0036 ± 0.0003	0.017 ± 0.001	0.983 ± 0.001	378
6	0.0022 ± 0.0002	0.012 ± 0.001	0.989 ± 0.001	517
8	0.0009 ± 0.0002	0.005 ± 0.001	0.995 ± 0.001	630
10	0.0003 ± 0.0000	0.002 ± 0.000	0.998 ± 0.000	700
15	0.0001 ± 0.0000	0.000 ± 0.000	1.000 ± 0.000	763

6.3. Architecture Comparison

To test whether K_ρ provides actionable information beyond accuracy, I compare four representation architectures on the same 16-state structured POMDP: a GRU (hidden dimension 4, trained for 300 epochs on sequences), an MLP (hidden dimension 4, trained for 400 epochs on flattened one-hot inputs), a last-observation-only baseline, and a random projection (tanh of a fixed random linear map). All representations are discretised with 5 bins per hidden dimension.

Table 5 reveals that the GRU achieves the lowest irreducible regret ($K_\rho = 0.003$), while the MLP achieves the highest accuracy (0.982). The MLP compensates for its slightly worse representation

 doi:10.20944/preprints202603.1546.v1

with a much finer-grained partition (371 cells vs. 50), which gives its output layer more degrees of freedom. But K_ρ isolates the representational quality from this policy effect: the GRU's sequential processing better separates the two state classes, even though the MLP's richer partition yields higher classification accuracy overall. In 35% of trials, the architecture with the best K_ρ differs from the one with the best accuracy. This confirms that K_ρ captures a distinct aspect of representation quality—the structural aliasing cost—that accuracy conflates with policy optimisation.

Table 5. Architecture comparison on a 16-state POMDP (20 trials). The GRU has the lowest K_ρ but the MLP has the highest accuracy. In 35% of trials, the architecture with the best K_ρ differs from the one with the best accuracy.

Architecture	$K_\rho(M)$	Accuracy	Raw impurity	Cells
GRU	0.0031 ± 0.0004	0.978 ± 0.002	0.022 ± 0.002	50
MLP	0.0039 ± 0.0004	0.982 ± 0.001	0.018 ± 0.001	371
Last-obs	0.0683 ± 0.0026	0.838 ± 0.004	0.162 ± 0.004	6
Random	0.1679 ± 0.0057	0.755 ± 0.006	0.245 ± 0.006	212

6.4. Learned Representation During Training

To demonstrate the decomposition as a training diagnostic, I train a two-layer MLP classifier (hidden dimension 4, tanh activation, discretised to 5 bins per hidden unit) on the 16-state structured POMDP with $|\mathcal{O}| = 6$ observations and observation sequences of length $L = 8$. The discretised hidden-layer activations define the representation-test cells.

Table 6 reports $K_\rho(M_\theta)$, execution cost, and their ratio at selected training epochs, averaged over 20 trials.

Table 6. Decomposition for a learned MLP during training on the 16-state POMDP ($L = 8$, hidden dim 4, 20 trials). The ratio $K_\rho/(K_\rho + \text{exec. cost})$ drops from 0.30 to 0.03 in the first 100 epochs, making the transition from representational bottleneck to policy bottleneck visible.

Epoch	$K_\rho(M_\theta)$	Execution cost	K_ρ / total	Accuracy
0	0.112 ± 0.005	0.268 ± 0.005	0.30	0.49 ± 0.01
50	0.012 ± 0.001	0.158 ± 0.005	0.07	0.81 ± 0.00
100	0.003 ± 0.000	0.097 ± 0.002	0.03	0.83 ± 0.00
150	0.003 ± 0.000	0.090 ± 0.001	0.03	0.83 ± 0.00
250	0.003 ± 0.000	0.086 ± 0.001	0.03	0.83 ± 0.00
500	0.003 ± 0.000	0.085 ± 0.001	0.04	0.83 ± 0.00

K_ρ drops by a factor of 37× in the first 100 epochs (from 0.112 to 0.003) as the hidden layer learns to separate the two state classes. After that, K_ρ plateaus near zero and further improvements come from execution cost reduction. The ratio $K_\rho/(K_\rho + \text{exec. cost})$ falls from 0.30 to 0.03 during this period, making the bottleneck transition quantitatively visible. Early in training, the representation is the bottleneck. Later, the policy is the bottleneck. The decomposition makes that transition readable from a single diagnostic, which is the practical contribution of the bridge for representation evaluation.

7. Towards Representation Learning Via the Bridge

The bridge theorem gives exact tools for computing and decomposing the representational component of regret. A natural question is whether these tools can guide representation learning during RL training. This section discusses the prospects and obstacles.

7.1. Auxiliary Losses Derived from the Bridge

The identity suggests a family of auxiliary losses for training recurrent encoders on partially observable tasks. Given a hidden-state predictor layered on top of the encoder, one can weight the prediction loss by the margin weight ρ_i, focusing the encoder on states where the diagnostic distinction is decision-relevant. Alternatively, one can apply an MDL compression penalty (KL divergence toward a maximum-entropy prior) to suppress noise in spare representational capacity. These strategies

 doi:10.20944/preprints202603.1546.v1

operate on the *representation*, not on the policy's extension, and are therefore representation-level surrogates for the formal weakness-maximisation principle rather than implementations of it.

Preliminary experiments on four small partially observable environments (a binary-context corridor, T-Maze, NoisyCartPole with hidden velocities, and a symbol-recall task) suggest that the optimal strategy is task-dependent: margin weighting helps when the diagnostic and control objectives are aligned (as in T-Maze, where the hidden label *is* the decision-relevant distinction), while MDL compression helps when the encoder has excess capacity relative to the task's information content. No single strategy dominates. This is consistent with the expectation that representation-level surrogates cannot capture the policy-level quantity that the weakness theorems optimise. Full experimental code and results are available in the Technical Appendices [3].

7.2. The Open Problem: Operationalising Weakness

The weakness-maximisation theorems [1,3] prove that the weakest correct policy—the one that maximises $|\mathrm{Ext}(\pi)|$ in an embodied language $L_{\mathfrak{v}}$—is optimal for generalisation under the maximally uninformative prior. A neural network policy is a total function from observations to action distributions. It specifies an output for every input. There are no "unset bits" whose completions could be counted, so the extension is trivial and weakness is undefined.

The central difficulty is that weakness is defined relative to a language $L_{\mathfrak{v}}$ that determines which completions count. A neural network does not have a native analogue of this structure. One possible approach is a hybrid architecture in which a neural encoder discretises observations into a finite vocabulary, and exact weakness computation is then performed in the discrete domain using existing Stack-Theoretic machinery. Developing and evaluating such an architecture, and scaling the auxiliary-loss strategies to standard partially observable benchmarks [51], are the natural next steps.

8. Conclusions

At fixed representation-test quotient, minimum average normalised regret is exactly margin-weighted forgetting. For an actual policy, total regret decomposes into irreducible aliasing cost plus avoidable within-cell misreporting. In Stack-Theoretic form, the same quantity is a weighted weakness deficit on a lifted diagnostic task.

The identity yields a representation-convergence theorem in pure RL language, a regret-based partial order on abstractions, Lipschitz stability under margin estimation error, and connections to free energy and multi-agent coordination, along with the multi-class generalisation to $K > 2$ outcomes. A cross-framework corollary converts the regret floor into a generalisation probability ($e^{-K_\rho(M)}$ under the canonical independent prior), giving the bridge bidirectional utility: the RL side supplies the regret floor, the Stack-Theoretic side converts it into a generalisation guarantee. The decomposition also serves as a practical training diagnostic: on controlled POMDPs, K_ρ isolates representational quality where accuracy does not, and the aliasing–execution ratio makes the transition from representational bottleneck to policy bottleneck quantitatively visible.

The weakness-maximisation theorems [1,3] predict that the least-committal correct policy maximises generalisation probability. However, the formal object they optimise—the extension of a policy in an embodied language—does not have a direct analogue in neural network function approximation. Preliminary experiments with bridge-derived auxiliary losses (margin-weighted state prediction, MDL compression) suggest that the optimal representation-level strategy is task-dependent and that no single surrogate dominates. Operationalising formal weakness in continuous function approximation, and developing the hybrid discrete–neural architecture needed to do so, remain the central open problems.

Acknowledgments: This paper is primarily theoretical, with controlled POMDP experiments on small structured environments. It does not introduce a deployed capability. Its primary positive impact is conceptual and methodological: the decomposition serves as a diagnostic tool for evaluating learned representations, and the derived results connect RL, active inference, and biological coordination through one shared defect variable.

 doi:10.20944/preprints202603.1546.v1

Bridge papers can overstate equivalence, novelty, or priority. I mitigate that risk by keeping the formal claims at the level of the representation-test quotient, stating explicitly what earlier work did and did not prove, noting where corollaries follow by direct substitution, and distinguishing representation-level strategies from the formal weakness-maximisation principle.

Appendix A. Measurable Exact Reduction

The finite theorem has a direct measurable analogue. Let $(\mathcal{X}, \Sigma, \mu)$ be a probability space of history-test points $x = (h, T)$. Let

$$\zeta := (M(h), T)$$

be the observable representation-test variable. Assume $y : \mathcal{X} \to \{0,1\}$ and $\rho : \mathcal{X} \to [0,1]$ are measurable. Define

$$a(z) := \mathbb{E}\big[\rho\,\mathbf{1}\{y = 1\} \mid \zeta = z\big], \qquad b(z) := \mathbb{E}\big[\rho\,\mathbf{1}\{y = 0\} \mid \zeta = z\big].$$

Any M-based policy is determined by a measurable map $q : \mathcal{Z} \to [0,1]$. Its expected normalised regret is

$$R(q) := \mathbb{E}\Big[\rho\big((1 - q(\zeta))\mathbf{1}\{y = 1\} + q(\zeta)\mathbf{1}\{y = 0\}\big)\Big].$$

Proposition A1 (Measurable exact reduction)**.** *With the notation above,*

$$\inf_{q:\mathcal{Z}\to[0,1]\ \text{measurable}} R(q) = \mathbb{E}\big[\min\{a(\zeta), b(\zeta)\}\big].$$

An optimal measurable selector is given by

$$q^\star(z) = \begin{cases} 1 & a(z) > b(z), \\ 0 & a(z) < b(z), \\ \text{any value in } [0,1] & a(z) = b(z). \end{cases}$$

Proof. By the tower property,

$$R(q) = \mathbb{E}\big[a(\zeta)(1 - q(\zeta)) + b(\zeta)q(\zeta)\big].$$

For each fixed z, the integrand is affine in $q(z)$. Its pointwise minimum over $q(z) \in [0,1]$ is therefore $\min\{a(z), b(z)\}$. Choosing $q^\star$ pointwise attains that minimum almost surely. Integrability is immediate because $0 \le a, b \le \mathbb{E}[\rho \mid \zeta] \le 1$. □

This is the conditional-expectation form of the finite theorem. In the finite support case, ζ takes one value per representation-test cell, a and b become the cell sums A_C and B_C, and Proposition A1 reduces exactly to Theorem 1. In measurable language, $\mathbb{E}[\min\{a(\zeta), b(\zeta)\}]$ is the weighted impurity that remains after quotienting by representation-test cell. It is the natural measurable version of $K_\rho(M)$.

Appendix B. Contravariance and Splintering on the Diagnostic Quotient

The bridge theorem already contains a small contravariance result. Let

$$I_+ := \{i : \mu_i \rho_i > 0\}, \qquad \mathcal{T}_+ := \{T_i : i \in I_+\}.$$

Define the *coarsest pure diagnostic partition on the informative support*

$$\mathcal{P}_+^\star := \big\{\, \{i \in I_+ : T_i = t,\ y_i = b\} \neq \varnothing : t \in \mathcal{T}_+,\ b \in \{0,1\} \,\big\}.$$

So $\mathcal{P}_+^\star$ groups positive-weight support points only by test and optimal label. Write $\mathcal{C} \preceq \mathcal{D}$ when every cell of partition $\mathcal{C}$ is contained in some cell of $\mathcal{D}$.

 doi:10.20944/preprints202603.1546.v1

Proposition A2 (Minimal zero-cost quotients are unique up to recoding). *For a fixed support and label map y, the following hold.*

1. *$K_\rho(M) = 0$ if and only if the restriction of $\mathcal{C}_M$ to I_+ satisfies $\mathcal{C}_M|_{I_+} \preceq \mathcal{P}^\star_+$.*
2. *If M is minimal among the zero-cost representations under partition coarsening on I_+, then $\mathcal{C}_M|_{I_+} = \mathcal{P}^\star_+$.*
3. *Consequently, any two minimal zero-cost representations induce the same quotient on the informative support and differ there only by a recoding of representation states.*

Proof. If $K_\rho(M) = 0$, then every cell $C \in \mathcal{C}_M$ must satisfy $\min\{A_C, B_C\} = 0$ by Theorem 1. So on the positive-weight support I_+ each cell contains only one optimal label at its fixed test value, which is exactly the statement that $\mathcal{C}_M|_{I_+} \preceq \mathcal{P}^\star_+$. The converse is immediate from the definition of $K_\rho(M)$ because points outside I_+ carry zero deletion weight.

For part 2, note that $\mathcal{P}^\star_+$ itself is zero-cost because every one of its cells is pure by construction. If a zero-cost representation M were minimal but strictly finer than $\mathcal{P}^\star_+$ on I_+, then coarsening $\mathcal{C}_M|_{I_+}$ to $\mathcal{P}^\star_+$ would preserve zero cost, contradicting minimality. So $\mathcal{C}_M|_{I_+} = \mathcal{P}^\star_+$. Part 3 follows because equality of quotients on the informative support means there is a bijection between the nonempty representation states induced by the two minimal representations there. That bijection is exactly a recoding. □

Intuition.

Shared task pressure drives away weighted impurity. At zero impurity there is one coarsest way to preserve the required distinctions. Minimal competent agents therefore converge on the same informative quotient, even if they use different internal names for its cells.

Example A1 (Shared pressure versus splintering). *Take three support points x_1, x_2, x_3 under one common test T. Subsystem A has optimal labels*

$$y^A(x_1) = 1, \qquad y^A(x_2) = 0, \qquad y^A(x_3) = 0,$$

so its coarsest pure quotient is

$$\mathcal{P}^\star_{A,+} = \{\{x_1\}, \{x_2, x_3\}\}.$$

Subsystem B has labels

$$y^B(x_1) = 0, \qquad y^B(x_2) = 1, \qquad y^B(x_3) = 0,$$

so its coarsest pure quotient is

$$\mathcal{P}^\star_{B,+} = \{\{x_2\}, \{x_1, x_3\}\}.$$

These minimal zero-cost quotients do not coincide. If each subsystem separately minimises its own weighted forgetting cost, they converge to different recoding classes. If they are forced to share one common zero-cost quotient, they must instead move to the finer common refinement

$$\{\{x_1\}, \{x_2\}, \{x_3\}\}.$$

This is the abstract pattern behind the contrast used in the main text. Shared task pressure yields contravariant convergence to one common quotient. Split task pressure yields divergent locally sufficient quotients, which is the toy formal template for splintering and cancer-like coordination failure.

Appendix C. Detailed Stack Theory Antecedents

This appendix records named results from the earlier Stack Theory corpus that are antecedent to the broad selection-style claims shared across the three lines. The purpose is to make the genealogy precise, so that readers can trace which components of each line were already theorem-level and which are new, so that the novelty and positioning of this particular paper is clearer. This is not an exhaustive comparison across all three programmes. Nayebi's structural selection theorems for predictive state, memory, modularity, regime tracking, and recoding match [7] are distinctive contributions of his

 doi:10.20944/preprints202603.1546.v1

programme that do not have direct Stack-Theoretic antecedents. The Richens-Everitt robust-causal-model result under distributional shift [6] is likewise distinctive.

Claim 1 (Earlier Stack Theoretic antecedents)**.** *The earlier Stack Theory corpus already contains the following.*

1. *"Sufficiency of weakness under a maximally uninformative prior" (2023), "Necessity of weakness under the same prior" (2023), and "Weakness maximisation beyond the uniform prior" (2023) prove that weakness is the generalisation-optimal proxy in the task-extension setting [1,3].*
2. *The EGRL history-task construction (2024), its selective-memory clauses (2024–2025), and "Weakness pair proxy optimality" (2025) supply a formal positive-negative pair machinery for noisy or inconsistent histories [3,9–11].*
3. *"Causal identity candidates" (2023), "w-maximised causal identities" (2023), "Existence of w-maximised causal identities" (2023), "Do-operator as conditioning on an intervention indicator" (2025), and "n^{th}-order self convergence" (2025) give the intervention-sensitive causal line, with the later self result stated under explicit representation and incentive preconditions [2,3,12].*

Justification. Each item is a direct summary of named definitions or propositions in the cited sources. The point of listing them is to make the genealogy of the broad selection-style claim traceable at the level of individual theorem statements.

References

1. Bennett, M.T. The Optimal Choice of Hypothesis Is the Weakest, Not the Shortest. In Proceedings of the 16th International Conference on Artificial General Intelligence. Springer, 2023, Lecture Notes in Computer Science, pp. 42–51. https://doi.org/10.1007/978-3-031-33469-6_5.
2. Bennett, M.T. How To Build Conscious Machines. PhD thesis, The Australian National University, 2025. https://doi.org/10.25911/ta58-p428.
3. Bennett, M.T. Technical Appendices, 2025. Archived release on Zenodo. Source repository: https://github.com/ViscousLemming/Technical-Appendices, https://doi.org/10.5281/zenodo.7641741.
4. Bennett, M.T. Are Biological Systems More Intelligent Than Artificial Intelligence?, 2024, [arXiv:cs.AI/2405.02325]. In press, 2026, Philosophical Transactions of the Royal Society B: Biological Sciences. Special issue on Hybrid agencies: crossing borders between biological and artificial worlds, https://doi.org/10.48550/arxiv.2405.02325.
5. Sutton, R.S.; Barto, A.G. *Reinforcement learning: An introduction*; MIT press: MA, 2018.
6. Richens, J.; Everitt, T. Robust agents learn causal world models. In Proceedings of the The Twelfth International Conference on Learning Representations, 2024, [2402.10877]. https://doi.org/10.48550/arxiv.2402.10877.
7. Nayebi, A. What Capable Agents Must Know: Selection Theorems for Robust Decision-Making under Uncertainty, 2026, [arXiv:cs.AI/2603.02491]. https://doi.org/10.48550/arXiv.2603.02491.
8. Richens, J.; Everitt, T.; Abel, D. General Agents Need World Models. In Proceedings of the Proceedings of the 42nd International Conference on Machine Learning. PMLR, 2025, Vol. 267, *Proceedings of Machine Learning Research*, pp. 51659–51687.
9. Bennett, M.T. Computational Dualism and Objective Superintelligence. In Proceedings of the 17th International Conference on Artificial General Intelligence. Springer, 2024, Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-031-65572-2_3.
10. Bennett, M.T. Optimal Policy Is Weakest Policy. In Proceedings of the Artificial General Intelligence. Springer, 2025, Vol. 16057, *Lecture Notes in Computer Science*, pp. 43–56. https://doi.org/10.1007/978-3-032-00686-8_5.
11. Bennett, M.T. A Formal Theory of Optimal Learning with Experimental Results. *IJCAI* **2025**, pp. 10967–10968. https://doi.org/10.24963/ijcai.2025/1238.
12. Bennett, M.T. Emergent Causality and the Foundation of Consciousness. In Proceedings of the 16th International Conference on Artificial General Intelligence. Springer, 2023, Lecture Notes in Computer Science, pp. 52–61. OUCI metadata page: https://ouci.dntb.gov.ua/en/works/7BoXJMW4/, https://doi.org/10.1007/978-3-031-33469-6_6.
13. Rabin, M.O.; Scott, D. Finite Automata and Their Decision Problems. *IBM Journal of Research and Development* **1959**, *3*, 114–125.

 doi:10.20944/preprints202603.1546.v1

14. Brooks, R.A. Intelligence without representation. *Artificial Intelligence* **1991**, *47*, 139–159. https://doi.org/10.1016/0004-3702(91)90053-m.
15. Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*; W.H. Freeman, 1982.
16. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2013**, *35*, 1798–1828.
17. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? In Proceedings of the Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2, Cambridge, MA, USA, 2014; NIPS'14, p. 3320–3328.
18. Schölkopf, B.; Locatello, F.; Bauer, S.; Ke, N.R.; Kalchbrenner, N.; Goyal, A.; Bengio, Y. Toward Causal Representation Learning. *Proceedings of the IEEE* **2021**, *109*, 612–634.
19. Goyal, A.; Bengio, Y. Inductive Biases for Deep Learning of Higher-Level Cognition. *Proceedings of the Royal Society A* **2022**, *478*, 20210068.
20. Ke, N.R.; Bilaniuk, O.; Goyal, A.; Bauer, S.; Larochelle, H.; Schölkopf, B.; Mozer, M.C.; Pal, C.; Bengio, Y. Systematic Evaluation of Causal Discovery in Visual Model Based Reinforcement Learning. In Proceedings of the Advances in Neural Information Processing Systems, 2021, Vol. 34.
21. Bengio, Y.; Deleu, T.; Rahaman, N.; Ke, N.R.; Lachapelle, S.; Bilaniuk, O.; Goyal, A.; Pal, C. A Meta-Transfer Objective for Learning to Disentangle Causal Mechanisms. In Proceedings of the International Conference on Learning Representations, 2020.
22. Cao, R.; Yamins, D. Explanatory models in neuroscience, Part 2: Functional intelligibility and the contravariance principle. *Cognitive Systems Research* **2024**, *85*, 101200.
23. Li, L.; Walsh, T.J.; Littman, M.L. Towards a Unified Theory of State Abstraction for MDPs. In Proceedings of the International Symposium on Artificial Intelligence and Mathematics, 2006.
24. Givan, R.; Dean, T.; Greig, M. Equivalence Notions and Model Minimization in Markov Decision Processes. *Artificial Intelligence* **2003**, *147*, 163–223.
25. Ferns, N.; Panangaden, P.; Precup, D. Metrics for Finite Markov Decision Processes. In Proceedings of the Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence. AUAI Press, 2004, pp. 162–169.
26. Ferns, N.; Panangaden, P.; Precup, D. Bisimulation Metrics for Continuous Markov Decision Processes. *SIAM Journal on Computing* **2011**, *40*, 1662–1714. https://doi.org/10.1137/10080484X.
27. Abel, D.; Hershkowitz, D.E.; Littman, M.L. Near Optimal Behavior via Approximate State Abstraction. In Proceedings of the Proceedings of the 33rd International Conference on Machine Learning. PMLR, 2016, Vol. 48, *Proceedings of Machine Learning Research*, pp. 2915–2923.
28. Abel, D.; Arumugam, D.; Lehnert, L.; Littman, M.L. State Abstractions for Lifelong Reinforcement Learning. In Proceedings of the Proceedings of the 35th International Conference on Machine Learning. PMLR, 2018, Vol. 80, *Proceedings of Machine Learning Research*, pp. 10–19.
29. Zhang, A.; McAllister, R.; Calandra, R.; Gal, Y.; Levine, S. Learning Invariant Representations for Reinforcement Learning without Reconstruction. In Proceedings of the International Conference on Learning Representations, 2021.
30. Castro, P.S. Scalable Methods for Computing State Similarity in Deterministic Markov Decision Processes. *Proceedings of the AAAI Conference on Artificial Intelligence* **2020**, *34*, 10069–10076.
31. Gelada, C.; Kumar, S.; Buckman, J.; Nachum, O.; Bellemare, M.G. DeepMDP: Learning Continuous Latent Space Models for Representation Learning. In Proceedings of the Proceedings of the 36th International Conference on Machine Learning. PMLR, 2019, Vol. 97, *Proceedings of Machine Learning Research*, pp. 2170–2179.
32. Littman, M.L.; Sutton, R.S.; Singh, S. Predictive Representations of State. In Proceedings of the Advances in Neural Information Processing Systems, 2001, Vol. 14, pp. 1555–1561.
33. Singh, S.; James, M.R.; Rudary, M.R. Predictive State Representations: A New Theory for Modeling Dynamical Systems. In Proceedings of the Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence. AUAI Press, 2004, pp. 512–519.
34. Boots, B.; Siddiqi, S.M.; Gordon, G.J. Closing the Learning-Planning Loop with Predictive State Representations. *International Journal of Robotics Research* **2011**, *30*, 954–966.
35. Subramanian, J.; Sinha, A.; Seraj, R.; Mahajan, A. Approximate Information State for Approximate Planning and Reinforcement Learning in Partially Observed Systems. *Journal of Machine Learning Research* **2022**, *23*, 1–83.

 doi:10.20944/preprints202603.1546.v1

36. Ha, D.; Schmidhuber, J. World Models. In Proceedings of the Advances in Neural Information Processing Systems, 2018, Vol. 31.
37. Hafner, D.; Lillicrap, T.; Ba, J.; Norouzi, M. Dream to Control: Learning Behaviors by Latent Imagination. In Proceedings of the International Conference on Learning Representations, 2020.
38. Schrittwieser, J.; Antonoglou, I.; Hubert, T.; Simonyan, K.; Sifre, L.; Schmitt, S.; Guez, A.; Lockhart, E.; Hassabis, D.; Graepel, T.; et al. Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. *Nature* **2020**, *588*, 604–609.
39. Pearl, J. Causal Diagrams for Empirical Research. *Biometrika* **1995**, *82*, 669–688.
40. Pearl, J. *Causality: Models, Reasoning, and Inference*, 2 ed.; Cambridge Uni. Press: United Kingdom, 2009.
41. Spirtes, P.; Glymour, C.; Scheines, R. *Causation, Prediction, and Search*, 2 ed.; MIT Press: Cambridge, MA, 2000.
42. Bareinboim, E.; Correa, J.D.; Ibeling, D.; Icard, T. On Pearl's Hierarchy and the Foundations of Causal Inference. In *Probabilistic and Causal Inference: The Works of Judea Pearl*; ACM, 2022; pp. 507–556. https://doi.org/10.1145/3501714.3501743.
43. Lattimore, F.; Lattimore, T.; Reid, M.D. Causal Bandits: Learning Good Interventions via Causal Inference. In Proceedings of the Advances in Neural Information Processing Systems, 2016, Vol. 29, pp. 1181–1189.
44. Kaelbling, L.P.; Littman, M.L.; Cassandra, A.R. Planning and Acting in Partially Observable Stochastic Domains. *Artificial Intelligence* **1998**, *101*, 99–134.
45. Tishby, N.; Pereira, F.C.; Bialek, W. The Information Bottleneck Method. *arXiv preprint physics/0004057* **2000**.
46. Wolpert, D.; Macready, W. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* **1997**, *1*, 67–82. https://doi.org/10.1109/4235.585893.
47. Karimireddy, S.P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; Suresh, A.T. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. In Proceedings of the Proceedings of the 37th International Conference on Machine Learning, 2020, pp. 5132–5143.
48. Davies, P.C.W.; Lineweaver, C.H. Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors. *Physical Biology* **2011**, *8*. https://doi.org/10.1088/1478-3975/8/1/015001.
49. Levin, M. The Computational Boundary of a "Self": Developmental Bioelectricity Drives Multicellularity and Scale-Free Cognition. *Frontiers in Psychology* **2019**, *10*, 2688.
50. Fields, C.; Levin, M. Scale-Free Biology: Integrating Evolutionary and Developmental Thinking. *BioEssays* **2020**, *42*. https://doi.org/10.1002/bies.201900228.
51. Morad, S.; Kortvelesy, R.; Bettini, M.; Liwicki, S.; Prorok, A. POPGym: Benchmarking Partially Observable Reinforcement Learning. In Proceedings of the International Conference on Learning Representations, 2023.