WHEN THE BODY COMPUTES: THE INFORMATION GEOMETRY OF CHEAP CONTROL

Anonymous authors
Paper under double-blind review

ABSTRACT

A body can sustain useful motion that a controller would otherwise have to produce. But when does exploiting that motion make control cheaper, and can an information-based reward encourage it? We develop an account of cheap control that separates what a policy encodes from what it physically imposes. A common divergence geometry connects two complementary views of embodied dynamics: action influence through empowerment, and state propagation through morphological computation. Rewarding the latter induces policy compression gated by the body, targeting precisely the controller information that the body's own dynamics render redundant. A causal decomposition extends the result to feedback. We then separate informational from physical cheapness. Physical savings depend on an alignment between what the body sustains, what the task permits, and what control costs. Experiments recover the selective compression mechanism across linear and nonlinear control. The resulting theory explains how embodiment simplifies control, and when information savings become physical ones.

1 INTRODUCTION

A passive dynamic walker descends a shallow slope without motors, sensors, or a controller. Gravity, inertia, and the geometry of its feet carry each leg through its swing, and the gait is a stable limit cycle of the unactuated mechanics (McGeer, 1990; Garcia et al., 1998). Changing the foot geometry can disrupt this gait and create a need for active correction, as illustrated schematically in Figure 1. Conversely, adding compliance or tuning inertia can let the mechanics sustain more of the motion and simplify the controller that remains (Collins et al., 2005; Pfeifer & Bongard, 2006).

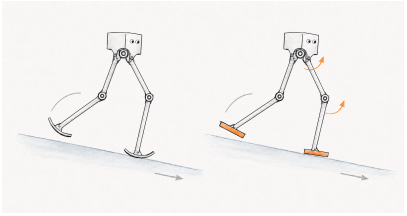

Figure 1: **The body does part of the work.** Schematic contrast between a gait sustained by passive mechanics and one requiring active correction. Foot geometry can change which motions the body supports.

This observation motivates a question that is easy to state and surprisingly hard to make precise: *when, and in what precise sense, can embodied dynamics reduce what a controller must do?* We call control *cheap* when an agent achieves its task with less active intervention in the system's own evolution. The body may sustain momentum, coordinate motion, or suppress disturbances that would otherwise require feedback. In this operational sense, it "computes" by carrying structure relevant to control; the claim does not require treating mechanics as symbolic computation (Müller & Hoffmann, 2017). Embodied intelligence has long emphasized this possibility (Pfeifer & Bongard, 2006; Paul, 2006), and earlier cheap control theory relates embodiment to the complexity of controllers needed to realize behavior (Montúfar et al., 2015; Ay, 2015).

For reinforcement learning, the natural next question is how to encourage such behavior. One approach rewards *morphological computation* (MC), an established information measure of how much the current state predicts the successor state once the executed action is known (Zahedi & Ay, 2013; Ghazi-Zahedi, 2019). A second applies *policy compression*, penalizing the information that actions carry about the state, encouraging a simpler policy (Tishby & Polani, 2011; Rubin et al., 2012; Fox & Tishby, 2016a;b; Shafieepoorfard et al., 2016; Tanaka et al., 2022). A third penalizes *control in-*

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

tervention, the departure of the controlled dynamics from their passive counterpart (Todorov, 2006). Each can be added to the task reward, yet they need not encourage the same behavior. Understanding that difference is necessary before reading an information bonus as an incentive for cheap control.

Conditioning on the executed action is what makes MC plausible as a measure of the body, and also what makes it ambiguous. In closed loop, an action is computed from the state. Knowing that action therefore reveals something about the state, even before considering what the actuator does. A controller that reads more finely leaves less for the current state to predict about its successor, while a coarser one leaves more. In neither case have the mechanics changed. MC therefore conflates physical propagation with controller encoding and cannot be read as a property of the body alone.

We separate these effects by introducing the *counterfactual passive future*: the successor that would have occurred from the same state and disturbance had the action been passive. The controller's state dependence then splits into a component shared with this future and a complementary component. Under additive actuation with independent process noise, an MC bonus penalizes the shared component alone when the state distribution is fixed. The objective is therefore not generic policy compression of the kind studied in information-limited control (Rubin et al., 2012; Tiomkin & Tishby, 2017; Tanaka et al., 2022): the body selects which controller information the penalty targets. A gait can carry a change in phase forward while damping a small deviation from its cycle. A body-aware penalty can distinguish these even when the controller encodes both equally well.

Selectivity says which controller information a regularizer targets. It does not say that reducing it saves anything physical. Passive motion is not automatically useful motion: gravity carries the walker's swing leg forward, and it carries a toppling walker to the ground just as reliably. Predictability alone does not distinguish their value to the task. The gap can open even when the passive motion is useful, because, in an additive system, a constant action offset can increase control intervention without changing the controller information. Physical savings therefore require more than informational redundancy: the directions a regularizer relieves must also be directions the task can spare, and directions where control is costly. We make the following contributions:

- **Common divergence geometry.** We place state propagation, action influence, and control intervention in one KL geometry, clarifying their complementary roles in cheap control.
- **Body-gated policy compression.** We derive the shared/complementary information split, identify its Gaussian directional operator, and extend it to causal controller information under feedback.
- **Conditions for physical cheapness.** An information no-go theorem shows that controller information alone cannot guarantee low control intervention. We identify a local alignment condition linking what the body sustains, what the task permits, and where control is costly.
- **Empirical tests of the distinction.** Controlled and learned policy experiments recover selective compression, demonstrate physical savings in an aligned system, and show why the information mechanism and its physical consequence require separate evaluation.

2 Embodied control through three information views

A single controlled transition can be read in three ways. We can ask what the current state still tells us about the successor, what the action contributes, or how far the executed dynamics have moved from what the body would have done unaided. We make these three views precise below.

Control problem and reference dynamics. We consider a discounted MDP with state S, action A, successor S', transition kernel $P(s'|s,a)$, and external reward $r(s,a)$. A stationary policy $\pi(a|s)$ has task return $J_{\mathrm{ext}}(\pi) \doteq \mathbb{E}_\pi[\sum_{t\geq 0}\gamma^t r(S_t,A_t)]$, with discount $\gamma \in [0,1)$ and initial state distribution ρ_0. The transition kernel P represents the body's mechanics and actuator coupling, its interaction with the environment, and process noise. The *passive kernel* $P_0(s'|s) \doteq P(s'|s,a_0)$ describes zero actuation a_0; a nonzero baseline would instead define a nominal reference. By default, information quantities use the states visited by the policy, represented by its normalized discounted distribution $d_\pi(s) \doteq (1-\gamma)\sum_{t\geq 0}\gamma^t \Pr_\pi(S_t = s)$. When comparing policies on the same states, we instead use a declared reference distribution d_{ref}. For either choice d, the one-step law is $\mathbb{P}^d_\pi(s,a,s') \doteq d(s)\pi(a|s)P(s'|s,a)$. These are *on-policy* and *fixed state ensemble* comparisons, respectively. Unless stated otherwise, information quantities are assumed finite and measured in nats; integrals replace sums in continuous spaces. We write $\mathbb{E}_{S,A}D_{\mathrm{KL}}(\cdot\|\cdot)$ for the conditional Kullback Leibler divergence (Cover & Thomas, 2006).

State propagation, action influence, and intervention. Three quantities describe different aspects of a controlled transition:

$$\mathcal{M}_\pi \doteq I(S; S'|A), \quad \mathcal{J}_\pi \doteq I(A; S'|S), \quad \mathcal{T}_\pi \doteq \mathbb{E}_{S,A} D_{\mathrm{KL}}(P(\cdot|S,A) \| P_0(\cdot|S)). \tag{1}$$

Morphological computation, $\mathcal{M}_\pi$, measures how much the current state helps predict the successor once the action is known (Zahedi & Ay, 2013). *Realized action information*, $\mathcal{J}_\pi$, exchanges those roles: at a given state, how much does knowing the action help predict what happens next? *Intervention*, $\mathcal{T}_\pi$, measures how far the executed actions move the transition law from passive dynamics (Todorov, 2006; 2004; 2009).

Empowerment asks how much influence the action channel could provide. It is the capacity $\mathcal{E}(s; \mathcal{V}) \doteq \sup_{\nu \in \mathcal{V}} I_\nu(A; S'|S = s)$, where $\mathcal{V}$ specifies admissible action distributions, for example through a power constraint (Klyubin et al., 2005; Salge et al., 2013; 2014). A strong motor makes many futures available even when the controller never uses it, and empowerment records that capability. What $\mathcal{J}_\pi$ and $\mathcal{T}_\pi$ record is narrower: they describe the policy actually deployed.

The regularization choices. Write $\mathcal{R}_\pi \doteq I(S; A)$ for the state information encoded in an action. We compare three objectives to be maximized:

$$\underbrace{J_{\mathrm{ext}} + \beta \mathcal{M}_\pi}_{\text{MC}}, \qquad \underbrace{J_{\mathrm{ext}} - \beta \mathcal{R}_\pi}_{\text{Compression}}, \qquad \underbrace{J_{\mathrm{ext}} - \beta \mathcal{T}_\pi}_{\text{Direct}}. \tag{2}$$

In (2), MC rewards conditional state propagation, Compression penalizes state encoding, and Direct penalizes intervention. Equal numerical coefficients $\beta \geq 0$ need not impose comparable regularization, so we compare the selected controllers at matched task performance.

To connect the quantities in (1), consider approximating $P(s'|s,a)$ by a reference that can see only the action, or only the state. The best such approximation is a ***KL projection*** of P onto the restricted family (Csiszár, 1975; Amari, 2016; Ay et al., 2017).

Proposition 1 (Complementary KL projections)**.** *Let $Q_A(\cdot|a)$ and $Q_S(\cdot|s)$ range over action-only and state-only successor reference kernels. Then*

$$\mathcal{M}_\pi = \min_{Q_A} \mathbb{E}_{S,A} D_{\mathrm{KL}}(P(\cdot|S,A) \| Q_A(\cdot|A)), \quad \mathcal{J}_\pi = \min_{Q_S} \mathbb{E}_{S,A} D_{\mathrm{KL}}(P(\cdot|S,A) \| Q_S(\cdot|S)).$$

The minimizers are the conditional successor laws $P_\pi(s'|a)$ and $\overline{P}_\pi(s'|s) \doteq \int P(s'|s,a)\pi(a|s)da$, respectively. Fixing the state-only reference to P_0 gives

$$\mathcal{T}_\pi = \mathcal{J}_\pi + \mathbb{E}_S D_{\mathrm{KL}}(\overline{P}_\pi(\cdot|S) \| P_0(\cdot|S)). \tag{3}$$

The two projections exchange the roles of state and action. The passive kernel already lies in the state-only family, so intervention needs no third projection: (3) splits it into the action influence $\mathcal{J}_\pi$ and a second term measuring how far the action-averaged dynamics have moved from passive motion. Only the first is a projection residual, which is why intervention is not generally a mutual information. Propositions 6 and 7 give the full form and proof. The counterexamples in Appendix A.1.2 show why these three readings are not interchangeable. To understand what an MC bonus does, we must now separate what an action reveals from what it physically does.

3 THE BODY GATE

An action chosen from the state carries a trace of that state, so knowing the action already tells us something about where the system was. Conditioning on it can therefore lower MC even when the mechanics have not changed at all. The chain rule makes this dependence explicit:

$$\mathcal{M}_\pi = \underbrace{I(S; S')}_{\text{closed-loop prediction}} - \underbrace{I(S; A)}_{\text{controller encoding}} + \underbrace{I(S; A|S')}_{\text{residual}} \geq I(S; S') - I(S; A). \tag{4}$$

The first term describes predictability across the controlled transition (Zahedi et al., 2010; Martius et al., 2013); the second measures how much state information the policy expresses in its actions. Their relationship links external predictability with internal controller complexity (Jost, 2004). A controller that holds prediction fixed while encoding less raises the lower bound, though not necessarily MC itself, since the residual is free to move as well. So (4) hints at a compression reading without saying which part of the controller's information an MC bonus actually acts on. Answering that requires separating what a command reveals from what it physically does.

The counterfactual passive successor and controller overlap. Consider a pendulum in motion. Gravity, momentum, and friction shape its successor state, while the motor applies a torque. The *counterfactual passive successor* is where it would arrive with the torque withheld and everything else unchanged. Write the transition as a structural causal model $S' = f(S, A, N)$, where N is independent of the current state and controller randomness. For zero actuation a_0, define $Y \doteq S'_{A \leftarrow a_0} = f(S, a_0, N)$. The replacement $A \leftarrow a_0$ changes only the action in the transition equation (Pearl, 2009); A still denotes the command actually chosen.

Given S, the passive successor depends only on the independent disturbance, so $I(A; Y|S) = 0$, also written $A - S - Y$. This is exactly what makes the controller's state dependence split in two. Expanding $I(A; S, Y)$ in two orders and using $A - S - Y$ gives

$$I(S; A) = \underbrace{I(A; Y)}_{\text{body-visible controller information}} + \underbrace{I(A; S|Y)}_{\text{complementary controller information}}. \tag{5}$$

The first term is what the command A shares with the passive successor Y, the overlap between controller and body; the second is what A still reveals about S once Y is known. A single velocity reading contributes to both: for a coasting mass most of what A carries about that velocity already lies in Y, so the overlap is large, while for a heavily damped mass almost none of it does. The overlap therefore measures how much of the same velocity information is already carried by passive motion. For a fixed controller and state ensemble, the body determines the overlap, but not whether the task still needs the reading: a mass coasting predictably toward its target must still be braked before it arrives, and braking needs the velocity.

What the MC objective rewards. MC uses the actual successor S'. To connect it to Y, suppose that the known action lets us undo its effect: $S' = T_A(Y)$, with each T_a invertible. Additive actuation, $S' = Y + g(A)$ with $g(a_0) = 0$, satisfies this condition because subtracting $g(A)$ recovers Y. This recoverability is separate from noise independence; it is not guaranteed for every pendulum model or nonlinear transition. Under it, S' and Y contain the same state information conditional on the action, and the chain rule gives

$$\mathcal{M}_\pi = I(S; Y|A) = I(S; Y) - I(S; A) + I(A; S|Y) = I(S; Y) - I(A; Y). \tag{6}$$

The complementary information cancels. MC is passive prediction less the part already revealed by the command, so the regularized return becomes

$$J_{\text{ext}}(\pi) + \beta \mathcal{M}_\pi = J_{\text{ext}}(\pi) + \beta I(S; Y) - \beta I(A; Y). \tag{7}$$

At a fixed state distribution and passive channel, $I(S; Y)$ is constant across policies. Equation (7) therefore penalizes the body-visible component alone in (5), whereas generic compression penalizes all of $I(S; A)$. Under on-policy evaluation, the policy also changes which states it visits: it can raise MC by reducing $I(A; Y)$ or by visiting states whose passive motion is more predictable. Holding the state ensemble fixed isolates the first effect; Appendix A.2.2 treats the general case. If the passive successor forgets the state entirely, $I(A; Y) = 0$ and the penalty vanishes. In the ideal limit where it reveals the state completely, all controller information is body-visible and the penalty becomes ordinary compression. Between these extremes, its strength can differ across state directions.

How reliably does the body carry a direction? Consider the linear Gaussian model $S' = FS + GA + N$ with controller $A = KS + \epsilon$, where F describes passive propagation, G actuation, and K feedback. The state S, process noise N, and action noise ϵ are independent centered Gaussians with positive definite covariances Σ_S, Σ_N, and Σ_ϵ. Hold the mechanics and these covariances fixed while varying K; the fixed state law $\mathcal{N}(0, \Sigma_S)$ is the reference ensemble d_{ref}. The passive successor $Y = FS + N$ is then a noisy record of the present state carried forward by the body. Some features of the starting state survive in that record; others are obscured by damping or disturbance. Find the pair of directions in S and Y with the greatest correlation, then repeat using directions uncorrelated with those already chosen. This splits the passive channel into independent scalar channels, the *passive modes* (Hotelling, 1936), whose informations add. Mode i has *reliability*

$$b_i \doteq \rho_i^2 = 1 - e^{-2\mathcal{M}_i} \in [0, 1), \qquad \sum_i \mathcal{M}_i = I(S; Y), \tag{8}$$

where ρ_i is the correlation within the ith pair and $\mathcal{M}_i$ the information it carries on the reference ensemble. Thus b_i is the fraction of that mode's variance the passive successor explains. Near one,

the body carries the direction almost intact; near zero, it forgets it within a step. Reliability is not a property of the mechanics alone: how much the state varies, how large the disturbance is, and how long we wait all enter. We can now describe the penalty in the regime of *weak state encoding*, where the state-dependent part of the action is small relative to the action noise:

Theorem 2 (Gaussian body-gated compression, informal)**.** *In the Gaussian model above, an MC bonus applies, to leading order in weak state encoding, a compression penalty weighted by each passive mode's reliability b_i. Ordinary policy compression weights every mode equally.*

The coefficient β in (2) sets how hard the penalty pushes, and the reliabilities in (8) decide where that push lands. Since b_i is evaluated on d_{ref}, it describes the body rather than the feedback (Proposition 12). The identities (5) and (6) are exact; their interpretation as compression with fixed modal weights is a weak encoding approximation. Theorems 10 and 14 give the formal results.

Example 1 (Adding correction to a passive walker)**.** A strictly unactuated walker has a constant action, so both terms on the right of (5) vanish: there is no controller channel to compress. Add a corrective controller and consider a gait deviation measured at successive heel strikes. The passive successor explains 90% of its variance under light gearing and 20% under heavy gearing, so at the same weak encoding the MC penalty is weaker under heavy gearing (see Example 12). Stronger erasure can also stabilize a gait, so a smaller weight does not imply worse walking.

Remark 1 (Connection to information bottleneck)**.** Along $A - S - Y$, the action is a representation of the state and the passive successor Y is a relevance variable. Classical information bottleneck compresses the representation while preserving information about relevance (Tishby et al., 1999). At fixed P_{SY}, the MC objective (7) instead penalizes $I(A;Y)$, while the external task determines what the action must retain. The passive modes are the canonical directions of Gaussian information bottleneck (Chechik et al., 2005); for Gaussian encoding of one mode, b_i is the initial slope of controller overlap $I(A;Y)$ against controller information $I(A;S)$ (Corollary 18). The largest weight also bounds arbitrary encoders: any controller satisfying $A - S - Y$ can share at most a fraction $\max_i b_i$ of its state information with this passive successor. This is a *strong data processing inequality* (Polyanskiy & Wu, 2015; Ngampruetikorn & Schwab, 2021) (see Remark 3 and (23)).

The action-side counterpart. The Gaussian action channel $A \to S'$, conditional on S, has a similar geometry. With physical noise fixed, more action randomness makes commands easier to distinguish by their effects, but each additional unit of intervention yields less information within each mode. Appendices A.3.2 and A.3.3 give the derivation and connection to empowerment.

4 Closed-loop body gating

The single-step split (5) compares the information in a command with what the next passive motion carries. Across successive decisions, previous commands matter to this comparison. Consider a cart that receives a push, coasts, and then receives a corrective push. Its later velocity partly reflects the first push, so a correlation between that command and the later state need not mean that the controller read the state. The first command may also have revealed information that remains relevant at the second decision. We should measure what the second command adds beyond what the first already disclosed. *Directed information* respects this order. It measures what each command reveals about states observed by the time it is chosen, conditional on the earlier commands (Massey, 1990; Raginsky, 2011). Write $S^{0:t} \doteq (S_0, \dots, S_t)$, and similarly for actions. Over T decisions, the *directed controller rate* is

$$\mathcal{R}_T^{\rightarrow} \doteq I(S^{0:T-1} \to A^{0:T-1}) \doteq \sum_{t=0}^{T-1} I(S^{0:t}; A_t | A^{0:t-1}). \tag{9}$$

Only states available when the command is chosen enter each term in (9). Conditioning on earlier commands accounts for information they already reveal; the history is empty at $t = 0$. Trajectory mutual information also includes the influence of earlier actions on later states, as Example 2 shows; see Appendix A.4.1 for the decomposition (Massey & Massey, 2005).

Example 2 (An action creates its own future observation)**.** Let $S_0 = 0$, choose A_0 as an independent fair bit, and set $S_1 = A_0$, $A_1 = 0$. The trajectories share $\log 2$ nats because the later state records the first command. The directed controller rate is nevertheless zero, since neither command conveys a state observation. This is the simplest version of pushing the cart at random and observing the motion that the push created.

We now ask how much of this directed information overlaps the passive successor. For the cart, this is its next motion if the current push is withheld, with earlier pushes unchanged. Formally, let $S_{t+1} = f_t(S_t, A_t, N_t)$ and define $Y_{t+1} \doteq f_t(S_t, a_0, N_t)$. A fresh disturbance independent of the available history and current controller randomness gives $I(A_t; Y_{t+1}|S_t, A^{0:t-1}) = 0$. Suppose also that, given the current state and past actions, earlier states do not change the action distribution. A Markov policy satisfies this condition because it chooses actions from the current state alone. Each directed information increment in (9) then reduces to $I(S_t; A_t|A^{0:t-1})$. Applying the single-step split (5) conditional on past commands and summing gives the sequential body gate.

Theorem 3 (Sequential body gating)**.** *Under the two conditions above,*

$$\mathcal{R}_T^{\rightarrow} = \underbrace{\sum_{t=0}^{T-1} I(A_t; Y_{t+1}|A^{0:t-1})}_{\textit{body-visible}} + \underbrace{\sum_{t=0}^{T-1} I(A_t; S_t|Y_{t+1}, A^{0:t-1})}_{\textit{complementary}}. \tag{10}$$

Each term is conditioned on the earlier commands, so information a previous command already expressed is not counted again.

The first sum in (10) measures information shared by each command and the next passive motion, beyond what earlier commands already reveal. To connect this overlap to MC on actual successors, suppose the current action allows recovery of the passive successor: $S_{t+1} = T_{t,A_t}(Y_{t+1})$, with each $T_{t,a}$ invertible. The corresponding history-conditioned MC in (32) is $\sum_{t=0}^{T-1} I(S_t; S_{t+1}|A^{0:t}) = \sum_{t=0}^{T-1} I(S_t; Y_{t+1}|A^{0:t-1}) - \sum_{t=0}^{T-1} I(A_t; Y_{t+1}|A^{0:t-1})$. Thus an MC bonus penalizes accumulated body-visible information when the passive predictive sum is fixed. In a learned loop, that sum can also change as the policy changes the states visited and what earlier commands reveal. The exact decomposition still holds; Theorem 23 and Corollary 24 give the full statements.

The gate can change along a trajectory. Earlier commands alter the uncertainty that dynamics and disturbance carry forward. In the Gaussian model, reliability therefore evolves through a Kalman-like recursion along the trajectory (see Proposition 26 and Example 14). Whether reducing the body-visible component lowers intervention depends on what the task permits.

5 When information savings become physical savings

We now know which controller information an MC bonus acts on. Whether reducing it yields physical savings is a separate question. For example, a fixed strong push reveals nothing about the state, whereas a small corrective force can reveal whether the system is left or right of its target. We therefore compare intervention and controller information at a common task requirement.

Definition 1 (Cheap control and informational control frontiers)**.** Let $C_{\mathrm{task}}(\pi) \doteq -J_{\mathrm{ext}}(\pi)$ and let Π be the admissible policy class. For a task cost bound c, define

$$\mathcal{C}^\star(c; \Pi) \doteq \inf_{\pi \in \Pi: C_{\mathrm{task}}(\pi) \le c} \mathcal{T}_\pi, \qquad \mathcal{R}^\star(c; \Pi) \doteq \inf_{\pi \in \Pi: C_{\mathrm{task}}(\pi) \le c} I_\pi(S; A). \tag{11}$$

Both use the on-policy law unless a fixed reference is declared.

The frontiers in (11) compare the least intervention and controller information needed to meet the task bound. In the Gaussian system, intervention is the expected quadratic action cost $\mathcal{T}_\pi = \frac{1}{2}\mathbb{E}[A^\top R_{\mathrm{T}} A]$, with $R_{\mathrm{T}} \doteq G^\top \Sigma_N^{-1} G$ the action space Fisher matrix of the transition law; see Proposition 29 and Remark 8 in Appendix A.5.3. This matrix weights action effects relative to process noise and gives zero weight to actions that leave the transition unchanged. One might hope that low controller information guarantees low intervention. It does not.

Theorem 4 (Information alone does not guarantee cheap control)**.** *Consider $S' = FS + GA + N$ with independent Gaussian noise of covariance $\Sigma_N \succ 0$ and passive reference $\mathcal{N}(Fs, \Sigma_N)$. Fix the state law for one transition and a centered action Z with finite second moment. For an admissible constant shift u, set $A_u = Z + u$. Then $I(S; A_u)$, $I(A_u; Y)$, and $I(A_u; S|Y)$ are unchanged, whereas $\mathcal{T}(u) = \mathcal{T}(0) + \frac{1}{2} u^\top R_{\mathrm{T}} u$. If arbitrarily large shifts are admissible along v with $Gv \neq 0$, intervention is unbounded at the same information values.*

A constant shift leaves informational dependence unchanged, while the actuator responds to the action's magnitude. The shifted policies need not meet the same task requirement, so the result rules

out a guarantee from information alone, rather than comparing equally good controllers. The proof appears in Appendix A.5.2.

When the task permits less intervention. Near a good controller some parameter changes barely affect performance while others spoil the task at once, and that difference decides what a regularizer can exploit. Let θ parameterize a smooth finite-dimensional policy class, such as the feedback gains in our Gaussian experiments. Assume a strict local minimum $\theta^\star$ of the task cost C with positive definite Hessian $H \doteq \nabla^2 C(\theta^\star)$ measuring this sensitivity. For a smooth regularizer $R(\theta)$, write $g_R \doteq \nabla R(\theta^\star)$ and $g_T \doteq \nabla \mathcal{T}(\theta^\star)$. A small penalty βR then moves the optimum by roughly $-\beta H^{-1} g_R$, the inverse Hessian permitting a larger excursion where performance falls away slowly. Whether that excursion also lowers intervention depends on how the two gradients align once the task's tolerance is taken into account.

Definition 2 (Alignment of task, body, and action)**.** For nonzero g_R, g_T, define their alignment in the metric given by the inverse Hessian as

$$\chi_R \doteq \frac{g_T^\top H^{-1} g_R}{\sqrt{(g_T^\top H^{-1} g_T)(g_R^\top H^{-1} g_R)}} \in [-1, 1]. \tag{12}$$

For $R = I(A;Y)$, the passive dynamics determine which state dependence the penalty targets. The inverse Hessian accounts for task sensitivity, so χ_R measures whether the resulting change also reduces intervention.

Theorem 5 (Local alignment of task, body, and action)**.** *Suppose C is three times and $R, \mathcal{T}$ twice continuously differentiable near $\theta^\star$. Let θ_β be the local minimizer of $C + \beta R$ through $\theta^\star$ for small $\beta > 0$. At a sufficiently small task cost increase $\delta \doteq C(\theta_\beta) - C(\theta^\star) > 0$,*

$$\mathcal{T}(\theta_\beta) - \mathcal{T}(\theta^\star) = -\sqrt{2\delta}\chi_R \sqrt{g_T^\top H^{-1} g_T} + O(\delta). \tag{13}$$

Positive alignment lowers intervention to leading order; negative alignment raises it. Direct, $R = \mathcal{T}$, attains $\chi_R = 1$, giving the largest leading reduction at fixed small task loss.

Equation (13) compares local optima independently of the optimization algorithm. At $\chi_R = 0$, intervention has no first-order change along the regularization path; higher-order terms decide its sign. Rankings can change farther from the optimum. By (7), raw on-policy MC uses $R = I_\pi(A;Y) - I_\pi(S;Y)$, so its alignment reflects visitation as well as encoding. The body gate identifies what information is penalized, and alignment determines whether that penalty encourages physical savings. Theorem 30 and its corollaries (Appendix A.5.4) give the proof and consequences.

Remark 2 (Why not optimize control intervention directly?)**.** When the desired intervention objective is available, Direct is the appropriate baseline. The KL intervention measure requires passive reference dynamics and a noise model, and can become singular as noise vanishes (Stein et al., 2026). A squared action norm avoids a transition model but depends on action coordinates and units. Appendix A.5.3 develops these distinctions. Appendix A.5.6 gives the mechanical energy balance (Proposition 36) and explains why intervention differs from actuator work (Remark 9).

6 Experiments

We test whether the predicted selectivity reduces intervention at comparable task performance, first in an LQG system, and then in an unmodified benchmark.

Push and Coast. We construct Push and Coast to isolate the situation the body gate describes (Appendix B.1 gives the dynamics). A mass travels toward a target while a separate alignment error must be held small. A push creates momentum that carries the mass onward after the drive is reduced, but it does not correct alignment. The controller can therefore ease the drive while retaining alignment feedback, although it must still brake before overshooting. A parameter α controls how much velocity survives each step. Reducing α weakens coasting without changing the task or alignment dynamics, so we can test whether the benefit depends on the body. The task has a quadratic state cost with no action penalty, so reduced intervention is not already rewarded by the task itself. Using the known linear Gaussian model, we optimize each objective at common task cost limits.

At 10% allowed task degradation, both Compression and MC weaken feedback relative to the controller optimized without regularization, but treat alignment differently: Compression retains 51% of

drive and 70% of alignment feedback, while MC retains 40% and 98%, respectively. Since coasting sustains the mass's motion, easing the drive costs little in task performance. MC retains alignment correction and uses 62.6% less intervention at the same task cost (Figure 2a,b). Reducing α from 0.95 to 0.50 shrinks this advantage to 12.2%, so the benefit tracks how well the body sustains motion. The local alignments, 0.896 for MC and 0.654 for Compression, also predict this ordering near the task optimum (Theorem 5). Learning with an on-policy MC bonus reproduces this selectivity. Across ten policy gradient training seeds, MC uses 3.6% less intervention than Compression at comparable task costs. Both policies remain within 0.2% of the task optimum, compared with the 10% allowance above. Body-visible information in (5) is lower by 0.101 nats, with no resolved change in complementary information, consistent with (7). The sequential decomposition in Theorem 3 reveals the same empirical pattern across successive decisions (Appendix B.2).

Table 1: **Selective compression need not reduce intervention.** Differences from Compression: MC for learned Push and BODY for Ball in Cup. Brackets give 95% bootstrap intervals. Physical cost is intervention $\mathcal{T}_\pi$ for Push and squared action for Ball in Cup. Intervals include training variability for Push and evaluation variability for fixed Ball in Cup policies (Appendices B.2 and B.3).

Quantity	Learned Push	Ball in Cup
Task cost / return	$\Delta C = -0.000028\ [-0.000254, 0.000189]$	$\Delta J_{\text{ext}} = +13.4\ [-101.1, 130.0]$
Body-visible information	$-0.101\ [-0.135, -0.076]$	$-0.01151\ [-0.01221, -0.01071]$
Complementary information	$-0.020\ [-0.139, 0.109]$	$+0.02098\ [0.02034, 0.02158]$
Physical cost	$\Delta\mathcal{T} = -70.74\ [-87.84, -53.96]$	$\Delta \sum_t \|A_t\|^2 = +17.73\ [-23.14, 57.66]$

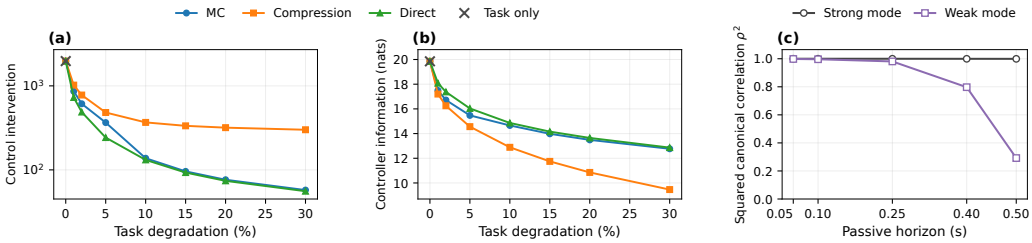

Figure 2: **Selective compression and persistent body memory.** **(a,b)** Push and Coast with the coasting body. Direct and Compression target the cheap control and information frontiers within the same policy class. Crosses mark the optimum without regularization. MC uses less intervention than Compression while retaining more information, approaching Direct at larger task allowances. Thus, stronger overall compression need not make control cheaper. **(c)** Nonlinear pendulum under zero torque. Passive directions separate over longer intervals, revealing uneven persistence of body memory: one squared canonical correlation remains near one through 0.50 s while the other falls to about 0.29. Points show interquartile means over 20 seeds with 95% bootstrap intervals.

Ball in Cup. In DMC Ball in Cup Catch (Tassa et al., 2018), the controller moves a cup to catch an unactuated ball attached by a tether. Moving the cup sets the ball swinging. Gravity, inertia, and the tether carry its motion between commands, but the controller must still place the cup beneath the returning ball at the right time. A small correction can decide the catch. The reward records the ball being inside the cup and contains no action penalty. Starting from a common policy, we optimize corrections within the same policy class. To isolate the body-visible penalty in (7), we minimize $I(A; Y_\ell)$, where Y_ℓ extends Y to a passive future at horizon ℓ. We call this objective BODY and compare it with Compression. Both use Gaussian covariance estimates on the same fixed state ensemble, isolating changes in controller encoding from changes in the states visited (Corollary 34).

Compared with Compression, BODY retains 0.0210 nats more complementary information while sharing 0.0115 nats less with the passive future (Table 1). It therefore retains more controller information overall, including more of the complementary component that its objective does not directly reward. These covariance diagnostics support the predicted selectivity in a nonlinear system. Both policies catch on 86.7% of evaluation episodes, but the intervals for differences in return and squared action include zero. The comparison therefore establishes neither a physical advantage nor equivalence, consistent with Section 5: selective compression alone does not guarantee physical savings.

Why the horizon matters. Measuring this selectivity requires a physically meaningful horizon. Over a very short interval, the passive future can closely resemble the starting state (Theorem 38). Longer horizons reveal which features the body continues to carry, exposing uneven persistence of body memory (Figure 2c). In the Gaussian model, the squared canonical correlations are the body gate reliabilities b_i, whose differences determine selectivity across directions. For Ball in Cup, we fix $\ell = 8$ steps (0.16 s), where passive directions separate, before any regularized optimization (Appendix B.3). Further horizon studies, including MountainCar, appear in Appendices B.5 and B.6.

7 Related work

Morphology and cheap control. Passive walking and embodied design show how useful behavior can emerge from mechanics with little active correction (McGeer, 1990; Collins et al., 2005; Paul, 2006; Pfeifer & Bongard, 2006). Information-theoretic MC measures describe state propagation conditional on action (Zahedi & Ay, 2013), while cheap control theory studies how embodiment changes the controller complexity needed to realize behavior (Montúfar et al., 2015). Work on hopping, synergy, and embodied learning connects these ideas to concrete behavior (Ghazi-Zahedi et al., 2016; 2017; Montúfar et al., 2016; Langer & Ay, 2021). Following the distinction between computation and the facilitation of control (Müller & Hoffmann, 2017), we ask which controller information overlaps passive dynamics and when reducing that overlap has a physical benefit.

Information-limited control. Rate distortion and bounded rationality formalize tradeoffs between decision quality and information (Sims, 2006; Shafieepoorfard et al., 2013; Ortega & Braun, 2013; Ortega et al., 2015; Genewein et al., 2015; Kostina & Hassibi, 2019; Arumugam & Van Roy, 2021). Feedback-control formulations extend these tradeoffs through memory and directed information (Tatikonda et al., 2004; Fox & Tishby, 2016a;b; Tanaka et al., 2018), including the role of mechanical compliance (Haninger, 2021). Information penalties also shape learned policies (Goyal et al., 2019; Leibfried & Grau-Moya, 2020; Malloy et al., 2020; Furuta et al., 2021). Frugal inference likewise separates inference information from physical motion effort (Olivos-Castillo et al., 2025). We identify the component of controller information selected by the passive body. Its reduction is distinct from lowering the whole information rate, and neither alone guarantees less intervention.

Empowerment and control intervention. Empowerment measures available action influence as channel capacity (Klyubin et al., 2005; Salge et al., 2013); variational and skill-discovery objectives use related information criteria for exploration and pretraining (Mohamed & Rezende, 2015; Gregor et al., 2016; Eysenbach et al., 2019; Sharma et al., 2020; Eysenbach et al., 2022). KL control instead penalizes departure from a reference transition law (Kappen, 2005; Todorov, 2004; 2006; 2009; Theodorou & Todorov, 2012), with connections to probabilistic inference (Levine, 2018) and limitations at vanishing noise (Datar & Ay, 2026; Stein et al., 2026). These are established ingredients. The common geometry makes their differences explicit; our body-gating, no-go, alignment, and sequential results explain when controller simplification agrees with reduced intervention.

8 Discussion and limitations

Embodiment shapes which state information persists through the system's passive dynamics. Our results identify the controller information shared with that motion and track it across successive decisions. Under the stated conditions, an MC objective selectively penalizes this overlap. Whether that selectivity makes control cheaper depends on the task: information about predictable motion can still be essential for correcting it. The alignment result identifies when reducing the targeted information also reduces intervention near the task optimum. This gives a concrete account of when the body's contribution can become a saving in control.

The scope of these conclusions is precise. The exact MC interpretation requires independent current disturbances and recovery of the passive successor from the actual successor and action. Gaussian weights describe a fixed state ensemble, and the alignment guarantee is local to an optimum with positive definite curvature. Intervention is measured relative to a specified passive reference and noise model. Extending these results to learned dynamics and larger policy changes would help turn the theory into a practical guide for deciding where feedback can safely be relaxed. The promise of embodiment is realized when useful passive motion lets the controller relinquish costly corrections without sacrificing the task.

AI USE STATEMENT

Generative AI tools were used in four ways. First, to polish and refine the manuscript text: the authors specified all content, structure, and claims, using AI assistance to improve the prose of the introduction and expository passages. Second, to generate implementation code for the small-scale reinforcement learning experiments. Third, to independently verify the proofs of three of the results; these proofs were written by the authors and checked by them, with AI used as an additional check rather than as a source of the arguments. Fourth, for routine editorial tasks such as reference formatting and consistency checking. All theoretical results, experimental designs, and scientific claims originate with the authors, who verified every AI-assisted output and take full responsibility for the submitted content.

REPRODUCIBILITY STATEMENT

Appendix A provides the assumptions and proofs; Appendix B describes the experiments and uncertainty estimates. Minimal code supporting the theory is available at `https://anonymous.4open.science/r/mclr-code-2F23`.

REFERENCES

Shun-ichi Amari. *Information Geometry and Its Applications*. Springer, 2016.

Dilip Arumugam and Benjamin Van Roy. Deciding what to learn: A rate-distortion approach. In *International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pp. 373–382, 2021.

Nihat Ay. Geometric design principles for brains of embodied agents. *KI - Künstliche Intelligenz*, 29(4):389–399, 2015.

Nihat Ay, Jürgen Jost, Hông Vân Lê, and Lorenz Schwachhöfer. *Information geometry*, volume 64. Springer, 2017.

Gal Chechik, Amir Globerson, Naftali Tishby, and Yair Weiss. Information bottleneck for Gaussian variables. *Journal of Machine Learning Research*, 6:165–188, 2005.

Steve Collins, Andy Ruina, Russ Tedrake, and Martijn Wisse. Efficient bipedal robots based on passive-dynamic walkers. *Science*, 307(5712):1082–1085, 2005.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2nd edition, 2006.

Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

Adwait Datar and Nihat Ay. Wasserstein KL-divergence for Gaussian distributions. In *Geometric Science of Information (GSI)*, volume 16034 of *Lecture Notes in Computer Science*, pp. 91–101. Springer, 2026.

Milan S Derpich, Eduardo I Silva, and Jan Østergaard. Fundamental inequalities and identities involving mutual and directed informations in closed-loop systems. *arXiv preprint arXiv:1301.6427*, 2013.

Benjamin Eysenbach, Abhishek Gupta, Julian Ibarz, and Sergey Levine. Diversity is all you need: Learning skills without a reward function. In *International Conference on Learning Representations (ICLR)*, 2019.

Benjamin Eysenbach, Ruslan Salakhutdinov, and Sergey Levine. The information geometry of unsupervised reinforcement learning. In *International Conference on Learning Representations (ICLR)*, 2022.

Roy Fox and Naftali Tishby. Minimum-information LQG control, part I: Memoryless controllers. In *IEEE Conference on Decision and Control (CDC)*, pp. 5610–5616, 2016a.

Roy Fox and Naftali Tishby. Minimum-information LQG control, part II: Retentive controllers. In *IEEE Conference on Decision and Control (CDC)*, pp. 5603–5609, 2016b.

Hiroki Furuta, Tatsuya Matsushima, Tadashi Kozuno, Yutaka Matsuo, Sergey Levine, Ofir Nachum, and Shixiang Shane Gu. Policy information capacity: Information-theoretic measure for task complexity in deep reinforcement learning. In *International Conference on Machine Learning (ICML)*, pp. 3541–3552, 2021.

Mariano Garcia, Anindya Chatterjee, Andy Ruina, and Michael Coleman. The simplest walking model: Stability, complexity, and scaling. *Journal of Biomechanical Engineering*, 120(2):281–288, 1998.

Tim Genewein, Felix Leibfried, Jordi Grau-Moya, and Daniel A. Braun. Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle. *Frontiers in Robotics and AI*, 2:27, 2015.

Keyan Ghazi-Zahedi. *Morphological Intelligence: Measuring the Body's Contribution to Intelligence*. Springer, 2019.

Keyan Ghazi-Zahedi, Daniel F. B. Haeufle, Guido Montúfar, Syn Schmitt, and Nihat Ay. Evaluating morphological computation in muscle and DC-motor driven models of hopping movements. *Frontiers in Robotics and AI*, 3:42, 2016.

Keyan Ghazi-Zahedi, Carlotta Langer, and Nihat Ay. Morphological computation: Synergy of body and brain. *Entropy*, 19(9):456, 2017.

Anirudh Goyal, Riashat Islam, Daniel Strouse, Zafarali Ahmed, Matthew Botvinick, Hugo Larochelle, Yoshua Bengio, and Sergey Levine. InfoBot: Transfer and exploration via the information bottleneck. In *International Conference on Learning Representations (ICLR)*, 2019.

Karol Gregor, Danilo Jimenez Rezende, and Daan Wierstra. Variational intrinsic control. *arXiv preprint arXiv:1611.07507*, 2016.

Kevin Haninger. Minimum directed information: A design principle for compliant robots. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11953–11960, 2021.

Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3–4):321–377, 1936.

Jürgen Jost. External and internal complexity of complex adaptive systems. *Theory in Biosciences*, 123(1):69–88, 2004.

Hilbert J. Kappen. Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11):P11011, 2005.

Alexander S. Klyubin, Daniel Polani, and Chrystopher L. Nehaniv. Empowerment: A universal agent-centric measure of control. In *IEEE Congress on Evolutionary Computation*, pp. 128–135, 2005.

Victoria Kostina and Babak Hassibi. Rate-cost tradeoffs in control. *IEEE Transactions on Automatic Control*, 64(11):4525–4540, 2019.

Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, 2004.

Carlotta Langer and Nihat Ay. How morphological computation shapes integrated information in embodied agents. *Frontiers in Psychology*, 12:716433, 2021.

Felix Leibfried and Jordi Grau-Moya. Mutual-information regularization in Markov decision processes and actor-critic learning. In *Conference on Robot Learning (CoRL)*, pp. 360–373, 2020.

Sergey Levine. Reinforcement learning and control as probabilistic inference: Tutorial and review. *arXiv preprint arXiv:1805.00909*, 2018.

Tailia Malloy, Chris R. Sims, Tim Klinger, Miao Liu, Matthew Riemer, and Gerald Tesauro. Deep RL with information constrained policies: Generalization in continuous control. *arXiv preprint arXiv:2010.04646*, 2020.

Georg Martius, Ralf Der, and Nihat Ay. Information driven self-organization of complex robotic behaviors. *PLOS ONE*, 8(5):e63400, 2013.

James L. Massey. Causality, feedback and directed information. In *International Symposium on Information Theory and Its Applications (ISITA)*, pp. 303–305, 1990.

James L. Massey and P. C. Massey. Conservation of mutual and directed information. In *IEEE International Symposium on Information Theory (ISIT)*, pp. 157–158, 2005.

Tad McGeer. Passive dynamic walking. *The International Journal of Robotics Research*, 9(2): 62–82, 1990.

Shakir Mohamed and Danilo Jimenez Rezende. Variational information maximisation for intrinsically motivated reinforcement learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 28, 2015.

Guido Montúfar, Keyan Ghazi-Zahedi, and Nihat Ay. A theory of cheap control in embodied systems. *PLOS Computational Biology*, 11(9):e1004427, 2015.

Guido Montúfar, Keyan Ghazi-Zahedi, and Nihat Ay. Information theoretically aided reinforcement learning for embodied agents. *arXiv preprint arXiv:1605.09735*, 2016.

Vincent C. Müller and Matej Hoffmann. What is morphological computation? on how the body contributes to cognition and control. *Artificial Life*, 23(1):1–24, 2017.

Vudtiwat Ngampruetikorn and David J. Schwab. Perturbation theory for the information bottleneck. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 21008–21018, 2021.

Frank Nielsen. An elementary introduction to information geometry. *Entropy*, 22(10):1100, 2020.

Itzel Olivos-Castillo, Paul Schrater, and Xaq Pitkow. Frugal inference for control. *arXiv preprint arXiv:2406.14427v3*, 2025.

Pedro A. Ortega and Daniel A. Braun. Thermodynamics as a theory of decision-making with information-processing costs. *Proceedings of the Royal Society A*, 469(2153):20120683, 2013.

Pedro A Ortega, Daniel A Braun, Justin Dyer, Kee-Eung Kim, and Naftali Tishby. Information-theoretic bounded rationality. *arXiv preprint arXiv:1512.06789*, 2015.

Chandana Paul. Morphological computation: A basis for the analysis of morphology and control requirements. *Robotics and Autonomous Systems*, 54(8):619–630, 2006.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009.

Gabriel Peyré and Marco Cuturi. Computational optimal transport with applications to data sciences. *Foundations and trends® in machine learning*, 11(5-6):355–607, 2019.

Rolf Pfeifer and Josh C. Bongard. *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press, 2006.

Yury Polyanskiy and Yihong Wu. Strong data-processing inequalities for channels and Bayesian networks. *arXiv preprint arXiv:1508.06025*, 2015.

Yury Polyanskiy and Yihong Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2025.

Maxim Raginsky. Directed information and Pearl's causal calculus. In *2011 49th Annual Allerton Conference on Communication, Control, and Computing*, pp. 958–965, 2011.

Jonathan Rubin, Ohad Shamir, and Naftali Tishby. Trading value and information in MDPs. In *Decision Making with Imperfect Decision Makers*, pp. 57–74. Springer, 2012.

Christoph Salge, Cornelius Glackin, and Daniel Polani. Approximation of empowerment in the continuous domain. *Advances in Complex Systems*, 16(2–3):1250079, 2013.

Christoph Salge, Cornelius Glackin, and Daniel Polani. Empowerment: An introduction. In *Guided Self-Organization: Inception*, pp. 67–114. Springer, 2014.

Ehsan Shafieepoorfard, Maxim Raginsky, and Sean P Meyn. Rational inattention in controlled Markov processes. In *2013 American Control Conference*, pp. 6790–6797. IEEE, 2013.

Ehsan Shafieepoorfard, Maxim Raginsky, and Sean P. Meyn. Rationally inattentive control of Markov processes. *SIAM Journal on Control and Optimization*, 54(2):987–1016, 2016.

Archit Sharma, Shixiang Gu, Sergey Levine, Vikash Kumar, and Karol Hausman. Dynamics-aware unsupervised discovery of skills. In *International Conference on Learning Representations (ICLR)*, 2020.

Christopher A Sims. Rational inattention: Beyond the linear-quadratic case. *American Economic Review*, 96(2):158–163, 2006.

Mark W. Spong, Seth Hutchinson, and M. Vidyasagar. *Robot Modeling and Control*. Wiley, 2 edition, 2020.

Viktor Stein, Adwait Datar, and Nihat Ay. Well-posed KL-regularized control via Wasserstein and Kalman-Wasserstein KL divergences. In *International Conference on Machine Learning (ICML)*, 2026.

Takashi Tanaka, Peyman Mohajerin Esfahani, and Sanjoy K. Mitter. LQG control with minimum directed information: Semidefinite programming approach. *IEEE Transactions on Automatic Control*, 63(1):37–52, 2018.

Takashi Tanaka, Henrik Sandberg, and Mikael Skoglund. Transfer-entropy-regularized Markov decision processes. *IEEE Transactions on Automatic Control*, 67(4):1944–1951, 2022.

Yuval Tassa, Yotam Doron, Alistair Muldal, Tom Erez, Yazhe Li, Diego de Las Casas, David Budden, Abbas Abdolmaleki, Josh Merel, Andrew Lefrancq, Timothy Lillicrap, and Martin Riedmiller. DeepMind Control Suite. *arXiv preprint arXiv:1801.00690*, 2018.

Sekhar Tatikonda and Sanjoy Mitter. The capacity of channels with feedback. *IEEE Transactions on Information Theory*, 55(1):323–349, 2009.

Sekhar Tatikonda, Anant Sahai, and Sanjoy Mitter. Stochastic linear control over a communication channel. *IEEE Transactions on Automatic Control*, 49(9):1549–1561, 2004.

Evangelos A. Theodorou and Emanuel Todorov. Relative entropy and free energy dualities: Connections to path integral and KL control. In *IEEE Conference on Decision and Control (CDC)*, pp. 1466–1473, 2012.

Stas Tiomkin and Naftali Tishby. A unified Bellman equation for causal information and value in Markov decision processes. arXiv preprint arXiv:1703.01585, 2017.

Naftali Tishby and Daniel Polani. Information theory of decisions and actions. In *Perception-Action Cycle: Models, Architectures, and Hardware*, pp. 601–636. Springer, 2011.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. In *Allerton Conference on Communication, Control, and Computing*, pp. 368–377, 1999.

Emanuel Todorov. Optimality principles in sensorimotor control. *Nature Neuroscience*, 7(9):907–915, 2004.

Emanuel Todorov. Linearly-solvable Markov decision problems. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 19, 2006.

Emanuel Todorov. Efficient computation of optimal actions. *Proceedings of the National Academy of Sciences*, 106(28):11478–11483, 2009.

Dor Tsur, Oron Sabag, Navin Kashyap, Haim Permuter, and Gerhard Kramer. Directed information: Estimation, optimization and applications in communications and causality. *arXiv preprint arXiv:2602.09711*, 2026.

Keyan Zahedi and Nihat Ay. Quantifying morphological computation. *Entropy*, 15(5):1887–1915, 2013.

Keyan Zahedi, Nihat Ay, and Ralf Der. Higher coordination with less control—A result of information maximization in the sensorimotor loop. *Adaptive Behavior*, 18(3-4):338–355, 2010.

A Theory and proofs

This appendix provides the formal statements and proofs behind the main argument, together with supporting results. Throughout, regular conditional distributions are assumed to exist, and information quantities are finite. Integrals replace sums for continuous variables. The proofs follow the main argument, from transition geometry and the exact body gate to Gaussian channels, feedback, and physical savings. The final subsection treats nonlinear approximations and the choice of a physical horizon.

A.1 Transition geometry and separations

A.1.1 Dual transition geometry

The classical *Golden Formula* separates expected KL divergence into information and marginal mismatch (Polyanskiy & Wu, 2025, Theorem 4.1). For any input U, output V, and reference law Q_V for which the terms are finite,

$$\mathbb{E}_U D_{\mathrm{KL}}(P_{V|U}(\cdot|U)\|Q_V) = I(U;V) + D_{\mathrm{KL}}(P_V\|Q_V). \tag{14}$$

Conditioning (14) on the action or on the state gives the full projection identities underlying Proposition 1.

Proposition 6 (Dual KL projection)**.** *Fix a joint law $d(s)\pi(a|s)P(s'|s,a)$. Let $P_\pi(\cdot|a)$ and $\overline{P}_\pi(\cdot|s)$ be its conditional successor laws as defined in Proposition 1. For any reference kernels with finite expected divergences,*

$$\mathbb{E}_{S,A} D_{\mathrm{KL}}(P(\cdot|S,A)\|Q_A(\cdot|A)) = I(S;S'|A) + \mathbb{E}_A D_{\mathrm{KL}}(P_\pi(\cdot|A)\|Q_A(\cdot|A)),$$
$$\mathbb{E}_{S,A} D_{\mathrm{KL}}(P(\cdot|S,A)\|Q_S(\cdot|S)) = I(A;S'|S) + \mathbb{E}_S D_{\mathrm{KL}}(\overline{P}_\pi(\cdot|S)\|Q_S(\cdot|S)).$$

Consequently, the corresponding infima over Q_A and Q_S are $\mathcal{M}_\pi$ and $\mathcal{J}_\pi$, attained at the displayed conditional marginals. Taking $Q_S = P_0$ gives the control intervention decomposition in Proposition 1.

Proof. Apply (14) with $U = S$, $V = S'$, and reference $Q_A(\cdot|a)$ conditionally on $A = a$, then average over A. Exchanging S and A gives the second identity. Each remaining KL term is nonnegative and vanishes at the corresponding conditional successor law, proving the minima. Setting $Q_S = P_0$ yields (3). □

For an information geometric interpretation, take finite alphabets and strictly positive $\mu(s,a) \doteq d(s)\pi(a|s)$ and kernels. Write $\mathcal{Q}_A$ and $\mathcal{Q}_S$ for the action-only and state-only families, and $P_A^*(s'|s,a) \doteq P_\pi(s'|a)$ and $P_S^*(s'|s,a) \doteq \overline{P}_\pi(s'|s)$ for their projections; see Figure 3. We work in the product of conditional probability simplexes. Both reference families are affine in row probabilities and row log odds. The minimization varies the second KL argument, also called an *m-projection* or *reverse I-projection*; both projections have this orientation (Csiszár, 1975; Amari, 2016; Nielsen, 2020; Ay et al., 2017). Their complementary roles exchange state and action. At a positive kernel Q,

the Fisher inner product of tangent vectors U and V, each summing to zero within every conditional row, is

$$g_Q(U,V) \doteq \sum_{s,a,s'} \mu(s,a)\frac{U(s'|s,a)V(s'|s,a)}{Q(s'|s,a)}.$$

At P_A^*, let $U = P - P_A^*$ and take V tangent to $\mathcal{Q}_A$. Since P_A^* averages the transition law over states conditional on the action, the residual U satisfies $\mathbb{E}[U(s'|S,a)|A=a] = P_\pi(s'|a) - P_\pi(s'|a) = 0$. Both V and P_A^* depend only on (a,s'), so averaging over S in the Fisher inner product gives $g_{P_A^*}(U,V) = 0$. Exchanging state and action proves the other orthogonality. These are the square corners in Figure 3; the algebraic Pythagorean identities themselves hold beyond this finite positive setting.

The projection selects the best restricted predictor. Fixing the state-only reference to the passive kernel instead measures intervention.

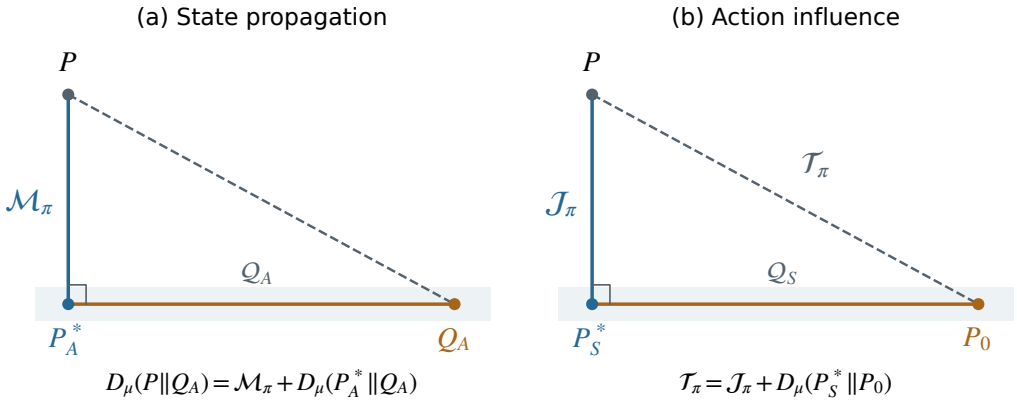

Figure 3: Complementary KL projections of the controlled kernel P. Here D_μ averages KL under $\mu(s,a) \doteq d(s)\pi(a|s)$. The strips denote action-only and state-only kernel families, with projections P_A^* and P_S^*; Q_A is an arbitrary action-only reference. Square corners indicate Fisher orthogonality at each projection. Projection residuals are MC (left) and realized action information (right). The passive reference lies within the state-only family, so intervention decomposes through that projection. Lengths are schematic, not KL values; the two kernel families are not asserted to be mutually orthogonal.

Definition 3 (Control intervention)**.** The transition KL cost relative to the declared passive kernel P_0 is

$$\tau(s,a) \doteq D_{\mathrm{KL}}(P(\cdot|s,a)\|P_0(\cdot|s)).$$

The intervention of a policy is its average transition KL cost, $\mathcal{T}_\pi = \mathbb{E}_{S,A}[\tau(S,A)]$.

This definition depends on the executed actions, including any common displacement they produce. The next consequence of Proposition 6 separates that displacement from action distinguishability.

Proposition 7 (Control intervention decomposition)**.** *For the joint law $\mathbb{P}_\pi^d$, let P_0 be the declared reference kernel and let $\overline{P}_\pi(s'|s) \doteq \int P(s'|s,a)\pi(a|s)da$. If the expected transition KL is finite, then*

$$\mathcal{T}_\pi = I(S';A|S) + \mathbb{E}_S D_{\mathrm{KL}}(\overline{P}_\pi(\cdot|S)\|P_0(\cdot|S)).$$

Proof. Take $Q_S = P_0$ in the second identity of Proposition 6. □

Intervention includes both action distinguishability and displacement of the average successor. This explains why action information alone is insufficient, and motivates the following examples separating the three information views.

A.1.2 SEPARATING MORPHOLOGICAL COMPUTATION, EMPOWERMENT, AND INTERVENTION

The following examples show why morphological computation, empowerment, and intervention cannot substitute for one another. Each measures a different aspect of the interaction between controller and body.

Theorem 8 (Coexistence of MC and empowerment without intervention)**.** *Let $\mathcal{S} = \mathcal{A} = \mathbb{Z}_n = \{0, \dots, n-1\}$, let S be uniform, and let $S' = (S + A + N) \bmod n$, with independent full support noise N. Deploy $A = 0$ and declare $P_0(\cdot|s) = P(\cdot|s, 0)$. With unrestricted action sources for empowerment,*

$$\mathcal{M}_\pi = \mathcal{E}(s) = \log n - H(N), \qquad \mathcal{J}_\pi = \mathcal{T}_\pi = 0.$$

Proof. Under the deployed policy, $S' = (S + N) \bmod n$ is uniform. Thus $H(S'|A) = \log n$ and $H(S'|S, A) = H(N)$, giving $\mathcal{M}_\pi = \log n - H(N)$. At fixed s, every action source has output entropy at most $\log n$, while the conditional output entropy given the action is $H(N)$. A uniform action source makes the output uniform and therefore achieves the capacity $\log n - H(N)$. The deployed action is constant, giving zero realized action information. Its transition equals the passive kernel, so every realized intervention divergence is zero. □

The equality of MC and empowerment in Theorem 8 reflects the symmetry of state and action in this construction. MC remains $\log n - H(N)$ under any action law independent of S, including randomized commands.

Example 3 (High MC and empowerment with zero intervention)**.** In Theorem 8, take noise concentrated near zero while retaining positive mass everywhere. Empowerment and MC approach $\log n$, but the passive policy has zero realized action information and zero intervention. The body strongly propagates state information and offers action influence that the controller does not use.

For the walker in Figure 1, this distinction is between actuators left idle during passive motion and actuators absent altogether. Idle actuators can retain positive empowerment while incurring zero intervention. An unactuated walker has zero empowerment, even when its mechanics propagate state information (Remark 5).

Dependence between state and action introduces a separate issue. MC measures what the state predicts about the successor once the action is known. If the command already reveals the state, conditioning on it can reduce MC even when the physical dynamics are unchanged.

Example 4 (Different MC under identical dynamics)**.** Let S be uniform on n states and let every action leave it unchanged, $S' = S$. Suppose the action set permits an injective encoding of S. Then

$$\mathcal{M}_\pi = I(S; S'|A) = H(S|A).$$

Under a constant policy, observing the action reveals nothing about the state, so $\mathcal{M}_\pi = \log n$. Under an injective policy $A = f(S)$, observing the action recovers the state, so $\mathcal{M}_\pi = 0$. The dynamics are identical in both cases. MC changes solely because the command reveals different amounts of state information, even though it has no physical effect.

The distinction also works in the other direction: available action influence need not require any passive state memory.

Example 5 (High empowerment with zero MC)**.** Let $S' = A + N$, where $N \sim \mathcal{N}(0, \sigma_N^2)$ is independent of (S, A) and $\sigma_N^2 > 0$. Impose an action power bound $P_{\max}$. The successor has no state dependence once A is known, so $\mathcal{M}_\pi = 0$ for every policy. Empowerment is $\frac{1}{2}\log(1 + P_{\max}/\sigma_N^2)$ and can be large. Thus available action influence need not come with passive propagation.

Even when state propagation is held fixed, intervention can vary. The next example complements the controller information obstruction in Theorem 4 by holding MC fixed while changing action variance.

Example 6 (Changing intervention at fixed MC)**.** Let $S' = \alpha S + bA + N$, with $b \neq 0$ and centered independent Gaussian S, A, N of variances $\sigma_S^2, v, \sigma_N^2$, where $\sigma_N^2 > 0$. Relative to zero actuation,

$$\mathcal{M}_\pi = \tfrac{1}{2}\log(1 + \alpha^2\sigma_S^2/\sigma_N^2), \quad \mathcal{J}_\pi = \tfrac{1}{2}\log(1 + b^2 v/\sigma_N^2), \quad \mathcal{T}_\pi = b^2 v/(2\sigma_N^2).$$

Conditioning on A removes the known shift bA, leaving MC independent of v. Increasing $v = \mathbb{E}[A^2]$ therefore increases squared action effort and intervention without changing MC.

A large mismatch can also come from the choice of reference. The following example requires a nominal reference outside the available action family.

Example 7 (Large nominal intervention with zero MC and empowerment)**.** Suppose every available action produces $\mathcal{N}(m, 1)$ independently of the state, but a declared nominal reference produces $\mathcal{N}(0, 1)$. Then MC, realized action information, and empowerment are zero, while nominal intervention is $m^2/2$. This construction does not use $P_0 = P(\cdot|s, a_0)$ with an available passive action: if all available actions have the same kernel and the reference is one of them, intervention must also be zero.

These distinctions also matter for exploration in reinforcement learning. High MC need not imply that an agent visits new states, and broad state visitation need not reflect any influence from its actions (Examples 8 and 9).

Example 8 (High MC without exploration)**.** Return to the identity dynamics $S' = S$ in Example 4, with a uniform initial state and a constant policy. MC equals $\log n$, yet each trajectory remains at its starting state and visits nothing new. The passive kernel is $P_0(\cdot|s) = \delta_s$, so intervention is zero. Since actions have no effect, empowerment is also zero. High MC can therefore reflect persistent state memory without exploration.

Example 9 (Broad visitation with zero MC, empowerment, and intervention)**.** If $P_0(s'|s) = Q(s')$ is independent of s and every action leaves this law unchanged, a trajectory may visit many states purely through passive randomness. Yet S' is independent of both S and A, so MC, realized action information, and empowerment vanish. Intervention is also zero because every action produces the passive law. Broad state visitation alone does not imply that any of these information quantities is positive.

Examples 3 and 5 separate available influence from deployed intervention and state memory. Example 4 shows how action conditioning changes MC even at fixed dynamics, while Example 6 changes intervention at fixed MC. Example 7 isolates the role of the reference. None provides a task performance guarantee: whether feedback can be reduced while preserving performance remains a separate question. Applying the body gate to nonlinear mechanics also requires assessing approximation error and the observation horizon, as developed in Appendices A.6.1–A.6.3.

A.2 The exact body gate and the MC objective

A.2.1 The distribution-free body gate

Defining a passive kernel alone does not identify the factual/counterfactual same-noise coupling from arbitrary observational data. We use the declared structural causal model (SCM) (Pearl, 2009) to specify that coupling.

The next proposition shows when the action and passive successor are independent given the current state. It also specifies when the passive successor can be recovered from the actual successor and the chosen action, allowing us to connect the body gate to morphological computation.

Proposition 9 (SCM realization of the body gate hypotheses)**.** *Let $S' = f(S, A, N)$ and $A = \phi_\pi(S, U)$, where ϕ_π is a measurable function implementing policy π using state S and controller randomness U, and N is independent of (S, U). Set $Y = f(S, a_0, N)$. Then $A - S - Y$. If measurable bijections T_a, with measurable inverses, satisfy $f(s, a, n) = T_a(f(s, a_0, n))$ on the relevant support, then $S' = T_A(Y)$. Additive actuation is the case $T_a(y) = y+g(a)$, with $g(a_0) = 0$.*

Proof. Conditional on $S = s$, the variables $\phi_\pi(s, U)$ and $f(s, a_0, N)$ depend on independent innovations. Their conditional joint law factors, proving $I(A; Y|S) = 0$. Substituting the definition of Y into the assumed transformation gives $S' = T_A(Y)$. At each fixed action, the measurable inverse then recovers Y from S'. □

The Markov relation follows from exogeneity; the MC identification additionally needs the transformation property. With these roles separated, the next result can state the information split without restricting the distributions to be Gaussian.

Theorem 10 (Distribution-free body-gated compression)**.** *Let (A, S, Y) satisfy $I(A; Y|S) = 0$. Define the passive predictive information $\mathcal{M}_{\text{std}} \doteq I(S; Y)$, conditional passive propagation $\mathcal{M}_A \doteq$*

$I(S;Y|A)$*, controller information* $\mathcal{R}_\pi \doteq I(S;A)$*, body-visible information* $\mathcal{R}_{\mathrm{red}} \doteq I(A;Y)$*, and body-complementary information* $\mathcal{R}_{\mathrm{comp}} \doteq I(A;S|Y)$*. Then*

$$\mathcal{M}_{\mathrm{std}} - \mathcal{M}_A = \mathcal{R}_{\mathrm{red}}, \qquad \mathcal{R}_\pi = \mathcal{R}_{\mathrm{red}} + \mathcal{R}_{\mathrm{comp}}, \qquad 0 \le \mathcal{R}_{\mathrm{red}} \le \mathcal{R}_\pi.$$

In particular,

$$\mathcal{M}_A = I(S;Y) - I(S;A) + I(A;S|Y).$$

If $S' = T_A(Y)$ *with each* T_a *a measurable bijection with measurable inverse, then* $\mathcal{M}_\pi = I(S;S'|A) = \mathcal{M}_A$*. At fixed* P_{SY}*, maximizing* $J_{\mathrm{ext}}(\pi) + \beta\mathcal{M}_\pi$ *is therefore equivalent to maximizing* $J_{\mathrm{ext}}(\pi) - \beta I(A;Y)$ *for any fixed* $\beta \ge 0$*.*

Proof. Expand the information between (S, A) and Y in the two possible orders. Conditional independence eliminates one term:

$$I(S, A; Y) = I(S;Y) + I(A;Y|S) = I(S;Y) = I(A;Y) + I(S;Y|A).$$

Rearranging gives the passive information identity. Next expand the information between A and (S, Y):

$$I(A; S, Y) = I(A;S) + I(A;Y|S) = I(A;S) = I(A;Y) + I(A;S|Y).$$

This proves the controller information split and the upper bound by nonnegativity of conditional mutual information. Nonnegativity also gives the lower bound. For the MC interpretation, condition on $A = a$. The bijection $Y \leftrightarrow T_a(Y)$ preserves conditional mutual information, so $I(S;S'|A) = I(S;Y|A)$. Finally, fixing P_{SY} makes $I(S;Y)$ constant across policies. Substitution into the regularized return gives the claimed equivalence. □

The identity distinguishes information shared with passive motion from the information that motion does not reveal. Example 10 makes the two extremes explicit before we turn to partial retention in noisy mechanical systems.

Example 10 (A body that retains one coordinate)**.** Let $S = (S_1, S_2)$ contain independent fair bits and let $Y = S_1$. A controller choosing $A = S_1$ uses $\log 2$ nats, all body-visible; one choosing $A = S_2$ uses the same amount, all complementary. Generic compression treats them equally, while a penalty on $I(A;Y)$ targets only the first. Which controller can be simplified depends on whether the task needs the retained coordinate, the forgotten one, or both.

The two controllers encode an equal amount of information, but only one overlaps what the body retains. Physical systems usually lie between these extremes. The Gaussian specialization in Appendix A.3.1 describes that partial retention direction by direction.

A.2.2 From a measure to a learning objective

How the three objectives enter a learner. To use the objectives in (2) for learning, we must distinguish a fixed control cost from an information penalty that changes with the policy. Intervention has the form

$$\mathcal{T}_\pi = \mathbb{E}_{S,A}[\tau(S, A)],$$

where $\tau(s, a) = D_{\mathrm{KL}}(P(\cdot|s, a)\|P_0(\cdot|s))$ is fixed by the plant and passive reference. Direct therefore subtracts a fixed cost from the task reward. Compression and MC instead use logarithmic terms that depend on distributions induced by the current policy:

$$\mathcal{R}_\pi = \mathbb{E}\Big[\log \tfrac{\pi(A|S)}{\overline{\pi}(A)}\Big], \qquad \mathcal{M}_\pi = \mathbb{E}\Big[\log \tfrac{P(S'|S,A)}{P_\pi(S'|A)}\Big], \tag{15}$$

where $\overline{\pi}$ is the marginal action law and $P_\pi(\cdot|a)$ is the successor law conditional on the action.

This policy dependence might appear to require differentiating the induced reference laws as well as the sampling distribution. At the population level, however, the explicit derivatives inside the logarithms have zero expectation. For example, under regularity conditions permitting differentiation through the integrals,

$$\mathbb{E}\big[\nabla_\theta \log \overline{\pi}_\theta(A)\big] = \int \overline{\pi}_\theta(a)\nabla_\theta \log \overline{\pi}_\theta(a)\, da = \int \nabla_\theta \overline{\pi}_\theta(a)\, da = 0.$$

The same normalization argument applies to $P_\pi(\cdot|a)$ and $\pi(\cdot|s)$. Thus trajectory policy gradients can use the current logarithmic terms in (15) as rewards, holding them fixed during each gradient calculation and updating them as the policy changes. For the normalized discounted state law, the per-step regularization terms include a factor $1-\gamma$, which can be absorbed into the regularization weight. This gives the exact population gradient; estimating the distributions from finite data can introduce error.

The resulting gradient includes changes in state visitation. This matters for interpreting an MC bonus: the Gaussian body gate describes selective compression at a fixed state ensemble, whereas learning can also change the ensemble itself. We next separate these effects to identify what MC rewards when the visited states are allowed to change.

Theorem 11 (Exact MC-regularized control decomposition)**.** *For every policy, use its declared state law and a common passive channel $P_Y(\cdot|s)$. Suppose $A - S - Y$ and $S' = T_A(Y)$ with the invertibility conditions of Proposition 9. Define $\mathcal{L}_\beta(\pi) \doteq J_{\mathrm{ext}}(\pi) + \beta\mathcal{M}_\pi$ for $\beta \geq 0$. Then*

$$\mathcal{L}_\beta(\pi) = J_{\mathrm{ext}}(\pi) + \beta I_\pi(S;Y) - \beta I_\pi(A;Y). \tag{16}$$

If the state law is fixed across policies, maximizing this objective is equivalent to maximizing $J_{\mathrm{ext}}(\pi) - \beta I_\pi(A;Y)$. With policy-dependent occupancy, both information terms can vary.

Proof. Apply Theorem 10 separately to each policy's joint law and substitute its MC identity into $\mathcal{L}_\beta$. Holding both P_S and $P_Y(\cdot|s)$ fixed holds the joint law P_{SY} fixed, so $I(S;Y)$ is an additive constant in that comparison. □

To distinguish the two information terms in (16), we compare controller encoding and visited states separately.

Proposition 12 (Behavioral, occupancy-matched, and fully standardized MC)**.** *Assume the same passive channel and invertible transformations. Let q_A be a fixed action law independent of the state. Behavioral MC, $\mathcal{M}_{\mathrm{beh}}(\pi)$, uses $d_\pi(s)\pi(a|s)P_Y(y|s)$. The occupancy-matched probe, $\mathcal{M}_{\mathrm{occ}}(\pi)$, replaces $\pi(a|s)$ by $q_A(a)$. Fully standardized MC, $\mathcal{M}_0$, additionally replaces d_π by d_{ref}. In each case, compute $I(S';S|A)$ with $S' = T_A(Y)$. Then*

$$\begin{aligned} \mathcal{M}_{\mathrm{occ}}(\pi) &= I_{d_\pi}(S;Y), & \mathcal{M}_0 &= I_{d_{\mathrm{ref}}}(S;Y), \\ \mathcal{M}_{\mathrm{occ}}(\pi) - \mathcal{M}_{\mathrm{beh}}(\pi) &= I_\pi(A;Y), & \mathcal{M}_{\mathrm{occ}}(\pi) - \mathcal{M}_0 &= I_{d_\pi}(S;Y) - I_{d_{\mathrm{ref}}}(S;Y). \end{aligned} \tag{17}$$

Consequently, $\mathcal{M}_0$ is independent of the policy when the reference laws and plant are fixed.

Proof. The independent probe gives $A \perp (S,Y)$. Invertibility at fixed A then yields $I(S';S|A) = I(Y;S|A) = I(Y;S)$, proving the first two formulas. Under the behavioral law, Theorem 10 instead gives $\mathcal{M}_{\mathrm{beh}} = I_{d_\pi}(S;Y) - I_\pi(A;Y)$. Subtraction proves the remaining identities. No policy parameter remains when both reference laws are fixed. □

Equations (17) describe statistical estimands. The next consequence of Proposition 9 states which change corresponds to an actual intervention on the controller.

Proposition 13 (Stochastic policy intervention semantics of the action probe)**.** *Under the SCM assumptions of Proposition 9, replace the behavioral action mechanism by an exogenous probe $A \sim q_A$ independent of S. For a declared state law $d(s)$, the resulting interventional joint law is*

$$P^{\mathrm{do}(A\sim q_A)}(s,a,y) = d(s)q_A(a)P_Y(y|s).$$

Consequently the occupancy-matched probe in (17) *can be interpreted as a stochastic policy intervention that severs the endogenous edge $S \to A$ while leaving the passive body mechanism unchanged. By contrast, replacing the state law by an arbitrary fixed reference d_0 is a standardization operation unless the experiment can physically reset or randomize states according to d_0.*

Proof. A stochastic intervention replaces the conditional action mechanism by the declared exogenous kernel q_A and leaves the other structural mechanisms unchanged (Pearl, 2009). We use the independence assumption of Proposition 9: the current disturbance N is independent of the state and controller randomness. The passive successor $Y = f(S,a_0,N)$ depends on S and N, but not on the action chosen by the controller. Replacing the action rule therefore leaves $P_Y(y|s)$ unchanged. With the declared state law d and the independent action law q_A, their product gives the stated factorization. □

The action probe removes the state dependence of the action mechanism. Replacing the state ensemble is a different operation and requires the ability to reset or sample the physical state from the reference law. The distinction matters whenever a statistical comparison is interpreted as a change that a controller could actually make. Fixing both state and action laws therefore produces a plant diagnostic, whereas fixing only the state law still permits a trainable controller redundancy objective. Appendix A.4 extends this distinction through a causal sequence of decisions.

A.3 Gaussian state and action channels

A.3.1 Gaussian directional geometry

The Gaussian specialization identifies the directions weighted by the exact overlap. The normalized matrices in (18) determine the weak encoding penalties in (19). This comparison uses a common state ensemble; it does not require different policies to induce the same stationary covariance.

Theorem 14 (Gaussian body-gated policy compression)**.** *Let $S \sim \mathcal{N}(0, \Sigma_S)$, $N \sim \mathcal{N}(0, \Sigma_N)$, and $\epsilon \sim \mathcal{N}(0, \Sigma_\epsilon)$ be mutually independent with positive definite covariances. Let $A = KS + \epsilon$, $S' = FS + GA + N$, and $Y = FS + N$. Hold Σ_S, F, G, Σ_N, and Σ_ϵ fixed when varying K. Define the normalized passive signal matrix M_S, body gate B_S, and controller precision X_K by*

$$M_S \doteq \Sigma_S^{1/2} F^\top \Sigma_N^{-1} F \Sigma_S^{1/2}, \quad B_S \doteq M_S (I + M_S)^{-1}, \quad X_K \doteq \Sigma_S^{1/2} K^\top \Sigma_\epsilon^{-1} K \Sigma_S^{1/2}. \tag{18}$$

Here I denotes an identity matrix of the appropriate size. Also define

$$\Pi_K \doteq K^\top \Sigma_\epsilon^{-1} K, \quad Q \doteq (\Sigma_S^{-1} + \Pi_K)^{-1}, \quad \Theta \doteq \Sigma_S F^\top (\Sigma_N + F \Sigma_S F^\top)^{-1} F \Sigma_S.$$

Then the following statements hold:

(i) Exact information formulas.

$$\mathcal{M}_\pi = \tfrac{1}{2} \log\det(I + \Sigma_N^{-1} F Q F^\top),$$
$$I(S;Y) = \tfrac{1}{2} \log\det(I + M_S), \qquad I(S;A) = \tfrac{1}{2} \log\det(I + X_K).$$

(ii) Exact overlap. $I(S;Y) - \mathcal{M}_\pi = I(A;Y)$, *with* $0 \le I(A;Y) \le I(S;A)$.

(iii) Factorization. $\Theta = \Sigma_S^{1/2} B_S \Sigma_S^{1/2}$.

(iv) Spectrum. *B_S is symmetric with $0 \preceq B_S \prec I$. Its eigenvalues are the squared canonical correlations between S and Y. If m_i are the eigenvalues of M_S and $\mathcal{M}_i \doteq \frac{1}{2}\log(1 + m_i)$, then*

$$b_i = \frac{m_i}{1 + m_i} = 1 - e^{-2\mathcal{M}_i}, \qquad \sum_i \mathcal{M}_i = I(S;Y), \qquad \det(I - B_S) = e^{-2I(S;Y)}.$$

(v) Local compression. *As $\Pi_K \to 0$,*

$$I(A;Y) = \tfrac{1}{2} \operatorname{tr}(\Theta \Pi_K) + O(\|\Pi_K\|^2), \qquad I(S;A) = \tfrac{1}{2} \operatorname{tr}(\Sigma_S \Pi_K) + O(\|\Pi_K\|^2).$$

Equivalently, as $X_K \to 0$,

$$I(A;Y) = \tfrac{1}{2} \operatorname{tr}(B_S X_K) + O(\|X_K\|^2), \qquad I(S;A) = \tfrac{1}{2} \operatorname{tr}(X_K) + O(\|X_K\|^2). \tag{19}$$

In an orthonormal eigenbasis of B_S, write x_i for the diagonal entries of X_K. The leading penalties are therefore $\frac{1}{2}\sum_i b_i x_i$ and $\frac{1}{2}\sum_i x_i$, respectively. This is the mode-weighted compression stated informally in Theorem 2; it does not require the encoder to be diagonal in that basis.

(vi) Vanishing and noiseless modes. *The null space of B_S is* $\ker(\Sigma_N^{-1/2} F \Sigma_S^{1/2})$. *For $\Sigma_N = \eta \Sigma_{N,0}$ with $\eta \downarrow 0$, the gate converges to the orthogonal projector onto* $\operatorname{range}(\Sigma_S^{1/2} F^\top \Sigma_{N,0}^{-1/2})$. *If F has full column rank, the limit is I.*

Proof. We first compute the information quantities, then identify their common spectral structure. Gaussian conditioning gives $\operatorname{Cov}(S|A) = Q$. Given A, the action effect GA is a known translation, so the conditional successor has signal covariance $FQF^\top$ and independent noise covariance Σ_N. The Gaussian channel formula gives the first expression in (i). Applying the same formula to $Y = FS + N$ and $A = KS + \epsilon$, and using $\det(I + UV) = \det(I + VU)$ for compatible matrices, gives

the other two. The independence assumptions imply $A - S - Y$, and additive actuation is invertible at fixed action. Part (ii) therefore follows from Theorem 10.

For the factorization, set $W \doteq \Sigma_N^{-1/2} F \Sigma_S^{1/2}$. Then $M_S = W^\top W$, and

$$\Sigma_S^{-1/2} \Theta \Sigma_S^{-1/2} = W^\top (I + WW^\top)^{-1} W = M_S (I + M_S)^{-1} = B_S.$$

The middle equality follows by multiplying the identity $(I + W^\top W) W^\top = W^\top (I + WW^\top)$ by the corresponding inverses. This proves (iii). Diagonalizing the symmetric positive semidefinite matrix M_S now gives $b_i = m_i/(1 + m_i)$ and all the scalar and determinant identities in (iv). Moreover,

$$B_S = \Sigma_S^{1/2} F^\top (F \Sigma_S F^\top + \Sigma_N)^{-1} F \Sigma_S^{1/2}.$$

Since $\mathrm{Cov}(S, Y) = \Sigma_S F^\top$, this is the normalized regression matrix whose eigenvalues are the squared canonical correlations (Chechik et al., 2005). This proves the remaining spectral statement.

For (v), expand the posterior covariance near zero feedback precision:

$$Q = \Sigma_S - \Sigma_S \Pi_K \Sigma_S + O(\|\Pi_K\|^2).$$

Substitution into the log determinant in (i), using $d \log\det V = \mathrm{tr}(V^{-1} dV)$ for a nonsingular matrix V, yields

$$\mathcal{M}_\pi - I(S; Y) = -\tfrac{1}{2} \mathrm{tr}(\Theta \Pi_K) + O(\|\Pi_K\|^2).$$

By the identity in part (ii), $\mathcal{M}_\pi - I(S; Y) = -I(A; Y)$, so changing the sign gives the expansion of the body-visible information. Expanding $\log\det(I + X_K)$ gives the controller information formula. Since Σ_S is fixed and nonsingular, $X_K = \Sigma_S^{1/2} \Pi_K \Sigma_S^{1/2}$ has an equivalent small matrix limit and remainder order.

Finally, $M_S = W^\top W$ and B_S have the same null space. Under the noise scaling in (vi), each positive eigenvalue of M_S scales as $1/\eta$, so its gate weight converges to one, while zero eigenvalues remain zero. The limit is the orthogonal projector onto the nonzero eigenspace. This proves (vi). □

The operator weights feedback by the reliability of passive propagation, rather than by state dependence alone. We next examine what these weights mean statistically and how they simplify in one dimension.

What the operator measures. Theorem 14 identifies Θ algebraically. Its statistical meaning is equally simple: how much state uncertainty an observation of the passive successor removes.

Corollary 15 (Interpretation of the passive propagation matrix). Under Theorem 14, let $Y = FS + N$ and $\Theta \doteq \Sigma_S^{1/2} B_S \Sigma_S^{1/2}$. Then

$$\Theta = \Sigma_S - \mathrm{Cov}(S|Y), \quad 0 \preceq \Theta \preceq \Sigma_S.$$

Thus Θ is exactly the reduction in state covariance obtained by observing the passive successor.

Proof. Gaussian conditioning gives $\mathrm{Cov}(S|Y) = \Sigma_S - \Sigma_S F^\top (F \Sigma_S F^\top + \Sigma_N)^{-1} F \Sigma_S$. The second term is Θ by Theorem 14. The covariance reduction is positive semidefinite and cannot exceed the prior covariance. □

Corollary 15 shows that a large gate weight means that the passive successor reveals a large fraction of a normalized state direction's variance. This interpretation does not depend on what the task values. For computing all three transition quantities, we also need the action channel formulas.

Corollary 16 (Linear Gaussian closed forms). Under the positive definite covariance assumptions of Theorem 14, for $S \sim \mathcal{N}(0, \Sigma_S)$, $A = KS + \epsilon$, $\epsilon \sim \mathcal{N}(0, \Sigma_\epsilon)$, $S' = FS + GA + N$, and $N \sim \mathcal{N}(0, \Sigma_N)$ independent,

$$\Sigma_{S|A} = (\Sigma_S^{-1} + K^\top \Sigma_\epsilon^{-1} K)^{-1}, \quad \mathcal{M}_\pi = \frac{1}{2} \log\det(I + \Sigma_N^{-1} F \Sigma_{S|A} F^\top),$$

$$\mathcal{J}_\pi = \frac{1}{2} \log\det(I + \Sigma_N^{-1} G \Sigma_\epsilon G^\top), \quad I(S; A) = \frac{1}{2} \log\det(I + \Sigma_S K^\top \Sigma_\epsilon^{-1} K),$$

$$\mathcal{T}_\pi = \frac{1}{2} \mathrm{tr}[G^\top \Sigma_N^{-1} G (K \Sigma_S K^\top + \Sigma_\epsilon)].$$

Proof. The state-side formulas are contained in Theorem 14. Conditioning on S leaves $G\epsilon + N$, so the Gaussian action channel has signal covariance $G\Sigma_\epsilon G^\top$ and noise covariance Σ_N, yielding $\mathcal{J}_\pi$. The intervention formula follows from the equal covariance Gaussian KL $\frac{1}{2}a^\top G^\top \Sigma_N^{-1} Ga$ averaged over A, whose covariance is $K\Sigma_S K^\top + \Sigma_\epsilon$. □

Corollary 16 shows why the three quantities respond differently to feedback. Once the state is known, the mean action KS is known too. Only the random action component ϵ contributes to $\mathcal{J}_\pi$, although both components contribute to intervention. The scalar case makes the connection to policy compression explicit: when passive motion carries state information, increasing feedback raises controller information and lowers MC.

Corollary 17 (Exact scalar policy compression form). In the scalar specialization of Theorem 14, let $S \sim \mathcal{N}(0, \sigma_S^2)$, $A = kS + \epsilon_A$ with $\epsilon_A \sim \mathcal{N}(0, \sigma_A^2)$, and $Y = \alpha S + N$ with $N \sim \mathcal{N}(0, \sigma_N^2)$ independent. Assume $\sigma_S^2, \sigma_A^2, \sigma_N^2 > 0$ and define $x \doteq \alpha^2 \sigma_S^2 / \sigma_N^2$. Then

$$I(S; A) = \tfrac{1}{2} \log(1 + \tfrac{k^2 \sigma_S^2}{\sigma_A^2}), \quad \mathcal{M}_\pi = \tfrac{1}{2} \log[1 + x \exp(-2I(S; A))].$$

For fixed x and σ_A^2, $\mathcal{M}_\pi$ is nonincreasing in k^2, and is strictly decreasing when $x > 0$ and $\sigma_S^2 > 0$.

Proof. Gaussian conditioning gives $\mathrm{Var}(S|A) = \sigma_S^2/(1+u)$ with $u = k^2\sigma_S^2/\sigma_A^2$. Therefore $\mathcal{M}_\pi = \frac{1}{2}\log[1 + x/(1+u)]$. Since $\exp[2I(S;A)] = 1 + u$, substitution gives the stated identity. □

In one dimension there is no choice of which mode to compress: for $x > 0$, MC and controller information order feedback strengths oppositely. This monotone relation does not make their regularized objectives identical at the same coefficient, and a changed occupancy can change x. In several dimensions, different propagation strengths give the distinct weights in Theorem 14.

Constructing the passive modes. Normalize the state by $\Sigma_S^{-1/2}$ and the disturbance by $\Sigma_N^{-1/2}$. Let $W \doteq \Sigma_N^{-1/2} F \Sigma_S^{1/2}$, and take a singular value decomposition of W. Rotating the normalized state and successor by its right and left singular vectors yields independent informative pairs of the form $\widetilde{Y}_i = \sqrt{m_i}\widetilde{S}_i + \eta_i$, where $\widetilde{S}_i$ and η_i are independent standard Gaussians. Output coordinates containing only noise carry no state information; untransmitted state directions contribute zero. Invertible coordinate changes preserve information, and independence makes it additive, giving $I(S;Y) = \sum_i \log(1 + m_i)/2$.

For each informative pair, $\mathrm{Var}(\widetilde{S}_i|\widetilde{Y}_i) = 1/(1+m_i)$. The fraction of variance explained is therefore $m_i/(1 + m_i) = b_i$. Equivalently, $\mathrm{Corr}(\widetilde{S}_i, \widetilde{Y}_i)^2 = b_i$. Thus reliability can change with state variation, noise, or observation interval even when the mechanics are unchanged.

Overlap at finite encoding. To see exactly how reliability controls overlap, consider a single jointly Gaussian triple. The key fact is that the action and passive successor can be correlated only through their common state; Corollary 18 gives the resulting relation between overlap and encoding.

Corollary 18 (Gaussian overlap curve). Let scalar (A, S, Y) be jointly Gaussian with positive marginal variances, $A - S - Y$, and finite $R \doteq I(S; A)$. Set $b \doteq \mathrm{Corr}(S, Y)^2 < 1$. Then

$$I(A; Y) = -\tfrac{1}{2} \log(1 - b + be^{-2R}), \qquad I(A; Y) = bR + O(R^2) \text{ as } R \downarrow 0. \tag{20}$$

For these Gaussian encoders, $I(A;Y) \le bI(S;A)$, and the ratio $I(A;Y)/I(S;A)$ tends to b as $I(S;A) \downarrow 0$ through positive values.

Proof. For jointly Gaussian variables, conditional independence makes the conditional covariance vanish:

$$0 = \mathrm{Cov}(A, Y|S) = \mathrm{Cov}(A, Y) - \frac{\mathrm{Cov}(A, S)\,\mathrm{Cov}(S, Y)}{\mathrm{Var}(S)}.$$

Dividing by the marginal standard deviations gives $\rho_{AY} = \rho_{AS}\rho_{SY}$. Intuitively, the correlation shared by the two observations is the product of how strongly each observes the common source. A scalar Gaussian pair has information $-\log(1 - \rho^2)/2$. Therefore $\rho_{AS}^2 = 1 - e^{-2R}$ and $\rho_{AY}^2 = b(1 - e^{-2R})$, proving the exact curve. Its derivative at zero is b. Its second derivative is nonpositive, since its first derivative is $be^{-2R}/(1 - b + be^{-2R})$. A concave curve through the origin lies below its tangent there, proving the bound and limiting ratio. □

Equation (20) is exact for Gaussian encoders. The multivariate weighted formula in Theorem 14(v) is a local expansion that also allows encoders to mix passive modes.

Remark 3 (Information bottleneck and the MC objective)**.** We expand the connection in Remark 1. In the Markov chain $A - S - Y$, the action represents the state and the passive successor supplies a relevance variable. Classical information bottleneck compresses the source while preserving information about relevance (Tishby et al., 1999). In our variables, it minimizes $I(S;A) - \lambda I(A;Y)$ for $\lambda > 0$ at fixed P_{SY}. Under the recoverability conditions of (7), an MC bonus instead gives $J_{\text{ext}} - \beta I(A;Y)$ up to a constant when P_{SY} is fixed (Theorem 10). It penalizes overlap with the passive successor, while the task determines which state dependence must remain. Thus the objectives share an information geometry but use the overlap differently.

For jointly Gaussian (S, Y), their canonical directions are the relevance directions of Gaussian information bottleneck (Chechik et al., 2005). The squared canonical correlations b_i are our body gate reliabilities. Corollary 18 shows that b_i is the initial overlap gained per unit of weak Gaussian encoding in mode i. The largest reliability also limits arbitrary encoders, as follows.

Information retained through the passive channel. Fix the state law and passive channel, hence P_{SY}, and vary the stochastic encoder $P_{A|S}$, with joint law $P_{SAY} = P_{SY}P_{A|S}$. The *information contraction coefficient* for this fixed input law is (Polyanskiy & Wu, 2015; Ngampruetikorn & Schwab, 2021)

$$\eta(P_{SY}) \doteq \sup_{P_{A|S}\colon 0<I(A;S)<\infty} \frac{I(A;Y)}{I(A;S)}. \tag{21}$$

The supremum ranges over all stochastic encoders, not just a chosen physical policy class. It is the largest fraction of controller information that can also be visible in the passive successor. The corresponding strong data processing inequality and the split in (5) give

$$I(A;Y) \le \eta(P_{SY})I(A;S), \quad I(A;S|Y) \ge [1-\eta(P_{SY})]I(A;S). \tag{22}$$

For jointly Gaussian (S, Y),

$$\eta(P_{SY}) = \max_i b_i, \tag{23}$$

giving the coefficient in (22). To establish the upper bound, first take a scalar standard Gaussian pair S, Y with squared correlation $b < 1$. Applying the conditional entropy power inequality to $Y = \sqrt{b}S + \sqrt{1-b}N$, with independent standard Gaussian N (Polyanskiy & Wu, 2025) gives, for any encoder $A - S - Y$ and $R = I(A;S) < \infty$,

$$I(A;Y) \le -\tfrac{1}{2}\log(1 - b + be^{-2R}) \le bR. \tag{24}$$

The second inequality follows from concavity of the displayed function of R, whose value at zero is zero and whose initial slope is b. Unlike the equality in Corollary 18, this upper bound does not require a Gaussian encoder.

For vectors, canonical coordinates give independent Gaussian pairs $(\widetilde{S}_i, \widetilde{Y}_i)$ with squared correlations b_i; untransmitted state directions can be paired with independent noise, and noise-only output directions can be discarded. Write $\widetilde{S}^{<i} = (\widetilde{S}_1, \dots, \widetilde{S}_{i-1})$, and similarly for $\widetilde{Y}^{<i}$. Applying the scalar bound with representation $(A, \widetilde{Y}^{<i})$ yields

$$\begin{aligned} I(A;Y) &= \sum_i I(A;\widetilde{Y}_i|\widetilde{Y}^{<i}) \le \sum_i b_i I(A;\widetilde{S}_i|\widetilde{Y}^{<i}) \\ &\le (\max_i b_i)\sum_i I(A;\widetilde{S}_i|\widetilde{S}^{<i}) = (\max_i b_i)I(A;S). \end{aligned} \tag{25}$$

For the second line, earlier passive coordinates are noisy observations of earlier state coordinates. Data processing gives $I(\widetilde{S}_i; A, \widetilde{Y}^{<i}) \le I(\widetilde{S}_i; A, \widetilde{S}^{<i})$, and independence of the state modes converts these to the conditional terms above. This proves the bound for arbitrary encoders, including those that mix modes.

To show that no smaller coefficient works for all encoders, choose a standardized state mode $\widetilde{S}_i$ with maximal reliability $b_i = \max_j b_j$ and encode it as $A_\varepsilon = \sqrt{\varepsilon}\,\widetilde{S}_i + Z$, where Z is independent standard Gaussian noise and $\varepsilon > 0$. Then

$$I(A_\varepsilon;S) = \tfrac{1}{2}\log(1+\varepsilon), \quad I(A_\varepsilon;Y) = \tfrac{1}{2}\log\frac{1+\varepsilon}{1+(1-b_i)\varepsilon}.$$

Their ratio tends to $b_i = \max_j b_j$ as $\varepsilon \downarrow 0$, proving sharpness. The scalar coefficient captures only the maximal reliability. The body gate operator retains the reliability of every mode and its orientation, giving a more detailed account of selectivity. For non-Gaussian channels, (21) still applies, but its value need not equal the largest squared canonical correlation.

Remark 4 (Overlap does not make feedback unnecessary)**.** A mass moving away from its target has a predictable passive successor. Its velocity can receive a strong body gate weight even though braking requires velocity feedback. The gate identifies what the penalty targets; the task determines whether reducing it is useful.

Remark 4 separates passive predictability from usefulness to the task. Example 11 changes damping in a mass, while Example 12 changes the passive return map around a gait. Both hold the state ensemble fixed to isolate how the mechanics change passive predictability.

Example 11 (Sliding mass: the same controller, different damping)**.** Consider a unit mass with state $S = (x, v)^\top$, with position and velocity expressed in fixed reference units. Its deterministic equations are $\dot{x} = v$ and $\dot{v} = -\kappa v + a$, where κ is damping and a is the applied force. A forward Euler step of length $h = 0.1$ gives $x' = x + hv$ and $v' = (1 - h\kappa)v + ha$. For this comparison, choose a centered Gaussian state ensemble with covariance $\Sigma_S = I$ and an independent Gaussian disturbance with covariance $\Sigma_N = 0.4I$. The discretized dynamics then have the form $S' = F_\kappa S + GA + N$, with

$$F_\kappa \doteq \begin{bmatrix} 1 & 0.1 \\ 0 & 1 - 0.1\kappa \end{bmatrix}, \qquad G \doteq \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}, \qquad \Sigma_S = I, \qquad \Sigma_N = 0.4I.$$

Use $A = KS + \epsilon$ with $K = [0.4 \;\; 0.8]$ and independent $\epsilon \sim \mathcal{N}(0, 1)$. Controller information is 0.2939 nats in both bodies. For heavy damping, $\kappa = 8$, the body gate is

$$B_S = \begin{bmatrix} 0.7097 & 0.0645 \\ 0.0645 & 0.0968 \end{bmatrix},$$

with eigenvalues 0.0901 and 0.7164. For light damping, $\kappa = 0.5$, it is

$$B_S = \begin{bmatrix} 0.7127 & 0.0219 \\ 0.0219 & 0.6936 \end{bmatrix},$$

with eigenvalues 0.6793 and 0.7270. Substituting in the exact Gaussian formulas gives $I(A; Y) = 0.0642$ nats under heavy damping and 0.1912 under light damping. The controller encodes the same state information, but more of it overlaps the passive successor when momentum persists. This calculation holds the state law fixed; it does not assert that the two bodies have the same on-policy occupancy.

Walker return maps and derived weights. Example 1 uses a Poincaré return map, which records the state at successive heel strikes. A periodic gait becomes a fixed point, and local deviations propagate through the return map's linearization (McGeer, 1990; Garcia et al., 1998). The specified maps compare canonical, lightly geared, and heavily geared variants. Negative controller gains make the commands corrective; changing their signs does not change the one-step information or KL values at the fixed centered ensemble.

Example 12 (A stylized walker and geared variants)**.** Take $S' = F_j S + A + N$, with $\Sigma_S = I$, $\Sigma_N = 0.01I$, and

$$F_{\text{can}} \doteq \operatorname{diag}(0.85, 0.55), \quad F_{\text{light}} \doteq \operatorname{diag}(0.45, 0.30), \quad F_{\text{heavy}} \doteq \operatorname{diag}(0.15, 0.05).$$

These illustrative maps model progressively stronger contraction of normalized gait deviations through transmission damping or friction. They are not estimates from particular physical walkers. Use $A = KS + \epsilon$, with $K = -\operatorname{diag}(0.3, 0.5)$ and $\Sigma_\epsilon = I$. For coordinate i, write λ_i for its passive factor and k_i for its feedback magnitude. Its passive signal-to-noise ratio is $m_i = \lambda_i^2/0.01$, its reliability is $b_i = m_i/(1 + m_i)$, and its controller information is $R_i = \log(1 + k_i^2)/2$. The posterior state variance is $1/(1 + k_i^2)$, giving

$$I(A; Y) = \tfrac{1}{2} \sum_{i=1}^{2} \log \frac{1 + m_i}{1 + m_i/(1 + k_i^2)}, \qquad \mathcal{M}_\pi = \tfrac{1}{2} \sum_{i=1}^{2} \log \left(1 + \frac{m_i}{1 + k_i^2}\right).$$

Consequently, the exact Gaussian calculation gives the following values in nats:

Variant	Gate weights	$I(S;Y)$	$I(A;Y)$	$\mathcal{M}_\pi$
Canonical	$(0.9863, 0.9680)$	3.8679	0.1501	3.7179
Lightly geared	$(0.9529, 0.9000)$	2.6795	0.1402	2.5393
Heavily geared	$(0.6923, 0.2000)$	0.7009	0.0498	0.6511

The controller information is 0.1547 nats in every body. Subtracting the overlap gives complementary information of approximately 0.0046, 0.0145, and 0.1048 nats, respectively. These values follow from the exact Gaussian formulas, with all bodies evaluated on the same initial state ensemble.

The action channel is the same in all three models: $G = I$, $\Sigma_\epsilon = I$, and $\Sigma_N = 0.01I$. Its two signal-to-noise ratios are 100, so each action reliability is $100/101$. Total realized action information is 4.6151 nats. By Corollary 16, intervention is $\frac{1}{2}\operatorname{tr}[\Sigma_N^{-1}(K\Sigma_S K^\top + \Sigma_\epsilon)]$. The mean feedback contributes 17 and the independent action randomness contributes 100, giving total intervention 117 in every body.

Changing the passive return map changes the state channel without changing this example's actuation channel. A physical change of gearing could change both F and G; holding G fixed here isolates the former effect. Lower state reliability also need not mean worse walking: strong contraction can suppress unwanted gait deviations. Appendix A.4.2 continues these same maps through successive corrections with fresh independent disturbances.

Remark 5 (The truly unactuated walker)**.** In a walker with no actuator, the action is constant and $S' = Y$. Both controller information components are zero, while $I(S;Y)$ can be positive. Such a machine motivates cheap control, but there is no controller channel for an MC bonus to compress. The preceding calculation concerns an actuated extension around a passive gait.

These examples separate what the body carries from what the controller encodes. Changing the mechanics leaves controller information unchanged but alters how much of it overlaps the passive successor. The unactuated walker in Remark 5 makes the distinction especially clear: MC can be positive even when no controller is present. The useful question is therefore not simply how large MC is, but which controller information the body already carries and whether the task permits that feedback to be weakened. The body gate answers the first question; task performance must answer the second. These comparisons hold the state ensemble fixed, as discussed in Appendix A.2.2.

A.3.2 The Gaussian action channel

The Gaussian model lets us identify which part of intervention is visible as action information. We use the action metric R_{T} introduced in Section 5; Appendix A.5.3 derives its Fisher interpretation. Write $A = KS + \epsilon$, where KS is the mean action at state S and ϵ is the fresh random component, also called the action innovation. Once S is known, KS is fixed. Only ϵ therefore contributes to the action information $I(A; S'|S)$, although both components contribute to intervention.

At a fixed state $S = s$, the successor is $S' = Fs + GKs + G\epsilon + N$. The varying action effect $G\epsilon$ supplies the signal, while the independent physical disturbance N obscures it. With the disturbance distribution held fixed, spreading the action effects farther apart makes the next state more informative about which command was applied. This gain has diminishing returns: additional intervention buys progressively less action information. Theorem 19 quantifies this relation by separating the action's effects into independent Gaussian modes and computing each mode's contribution to information and intervention.

Theorem 19 (Action channel saturation and intervention decomposition)**.** *Under the centered Gaussian model of Theorem 14, define $M_A \doteq \Sigma_\epsilon^{1/2} R_{\mathrm{T}} \Sigma_\epsilon^{1/2}$ and $B_A \doteq M_A(I + M_A)^{-1}$. If $m_{A,i}$ are the eigenvalues of M_A, define $\mathcal{J}_i \doteq \frac{1}{2}\log(1 + m_{A,i})$ and $b_{A,i} \doteq m_{A,i}/(1 + m_{A,i})$. Then*

$$\mathcal{J}_\pi = \tfrac{1}{2}\log\det(I + M_A), \qquad b_{A,i} = 1 - e^{-2\mathcal{J}_i}, \qquad \det(I - B_A) = e^{-2\mathcal{J}_\pi}.$$

Writing the intervention due to action randomness as $\mathcal{T}_{\mathrm{innov}} \doteq \frac{1}{2}\operatorname{tr}(M_A)$ and its contribution in mode i as $\mathcal{T}_{\mathrm{innov},i} \doteq m_{A,i}/2$, we have

$$\frac{d\mathcal{J}_i}{d\mathcal{T}_{\mathrm{innov},i}} = \frac{1}{1 + m_{A,i}} = 1 - b_{A,i}. \tag{26}$$

Total intervention is $\mathcal{T}_\pi = \mathcal{T}_{\text{fb}} + \mathcal{T}_{\text{innov}}$, *where* $\mathcal{T}_{\text{fb}} = \frac{1}{2}\operatorname{tr}(R_{\text{T}} K \Sigma_S K^\top)$. *The modal derivative concerns innovation intervention only.*

Proof. Conditioning on S leaves $G\epsilon+N$ as the random successor component. Subtracting the Gaussian noise entropy and applying Sylvester's determinant identity gives $\mathcal{J}_\pi = \frac{1}{2}\log\det(I + M_A)$. Diagonalizing M_A proves the two spectral identities. Equal covariance Gaussian KL gives $\tau(s,a) = \frac{1}{2}a^\top R_{\text{T}} a$. Substituting $A = KS + \epsilon$ and taking expectations eliminates the cross term by independence and centering, leaving the feedback and innovation terms. In each mode, differentiating $\log(1+m)/2$ with respect to $m/2$ gives $1/(1+m)$. □

Equation (26) expresses diminishing information gain: action information grows logarithmically with the random action variance, while the intervention due to that randomness grows linearly. Feedback has a different effect. Once the state is given, changing the mean command KS only translates the conditional action channel, leaving its information unchanged. The next corollary shows that this same change can still alter intervention. Thus realized action information cannot measure the physical cost of feedback.

Corollary 20 (Gain blindness of realized action information). Under Theorem 19, at fixed Σ_ϵ, realized action information $\mathcal{J}_\pi$ has no direct dependence on K. With the same R_{T} and M_A,

$$\mathcal{T}_\pi = \underbrace{\frac{1}{2}\operatorname{tr}(R_{\text{T}} K \Sigma_S K^\top)}_{\mathcal{T}_{\text{fb}}} + \underbrace{\frac{1}{2}\operatorname{tr}(M_A)}_{\mathcal{T}_{\text{innov}}}, \quad \mathcal{J}_\pi = \frac{1}{2}\log\det(I + M_A).$$

Consequently

$$\Delta_\pi \doteq \mathcal{T}_\pi - \mathcal{J}_\pi = \mathcal{T}_{\text{fb}} + \frac{1}{2}[\operatorname{tr}(M_A) - \log\det(I + M_A)].$$

Proof. The expression for $\mathcal{J}_\pi$ in Corollary 16 contains Σ_ϵ but not K. Splitting the trace in $\mathcal{T}_\pi$ into the deterministic mean action covariance and stochastic innovation covariance gives the decomposition. Subtracting $\mathcal{J}_\pi$ gives Δ_π. □

Corollary 20 shows that changing feedback can change intervention without changing realized action information. The gap Δ_π is the Gaussian form of the divergence from the passive reference in Proposition 7: it accounts for both the mean displacement caused by feedback and the additional spread caused by action randomness. We next turn from the information carried by a given action source to empowerment, which optimizes that source under an action constraint.

A.3.3 From realized action information to empowerment

Theorem 19 measures action information for a chosen Gaussian source. Empowerment optimizes that source under an action constraint. The resulting capacity and water-filling solution are standard (Cover & Thomas, 2006). Their significance here is that the optimal action directions are determined by the same metric R_{T} that weights intervention in Section 5; see Remark 8. Directions favored by empowerment therefore need *not* be cheap to use.

Consider $S' = Fs+GA+N$ at fixed s, where $N \sim \mathcal{N}(0, \Sigma_N)$ is independent of A and $\Sigma_N \succ 0$. We constrain the action source by $\mathbb{E}\|A\|^2 \le P_{\max} < \infty$ and write Σ_U for its covariance. An optimal source can be chosen centered, so the power constraint becomes $\operatorname{tr}\Sigma_U \le P_{\max}$. The term Fs translates every successor distribution by the same amount, so it does not affect how distinguishable the actions are. The corollary below shows that a Gaussian source suffices.

Corollary 21 (Gaussian empowerment and its independence of state). For the channel and power constraint above,

$$\mathcal{E}(s; P_{\max}) = \max_{\Sigma_U \succeq 0,\ \operatorname{tr}\Sigma_U \le P_{\max}} \frac{1}{2}\log\det(I + \Sigma_N^{-1} G \Sigma_U G^\top). \tag{27}$$

This value is independent of s. Let $\lambda_1, \dots, \lambda_r > 0$ be the positive eigenvalues of R_{T}, with orthonormal eigenvectors $u_1, \dots, u_r$. If $r > 0$, an optimal source is $A \sim \mathcal{N}(0, \Sigma_U^*)$, where

$$\Sigma_U^* = \sum_{i=1}^{r} p_i^* u_i u_i^\top, \quad p_i^* = \max\{0, \mu - \tfrac{1}{\lambda_i}\}, \quad \sum_{i=1}^{r} p_i^* = P_{\max},$$
$$\mathcal{E}(s; P_{\max}) = \tfrac{1}{2} \sum_{i=1}^{r} \log(1 + \lambda_i p_i^*). \tag{28}$$

The parameter μ is chosen to satisfy the power constraint. If $G = 0$, empowerment is zero.

Proof. Translation invariance removes Fs. For an arbitrary source with covariance Σ_U, independence gives $I(A; S'|S = s) = h(GA + N) - h(N)$. Gaussian entropy maximization bounds this by the objective in (27). A centered Gaussian source attains the bound, as also follows from Theorem 19 and Sylvester's determinant identity. The formula extends to singular source covariances by continuity. Centering loses no information and avoids spending power on a mean, since $\mathbb{E}\|A\|^2 = \operatorname{tr} \Sigma_U + \|\mathbb{E}A\|^2$. The feasible covariance set is compact and the objective is continuous, so the maximum exists.

To obtain the optimizer, normalize the output noise by $\Sigma_N^{-1/2}$ and use a singular value decomposition of $\Sigma_N^{-1/2} G$. The informative scalar channels have gains $\sqrt{\lambda_i}$. Hadamard's determinant inequality shows that independent Gaussian inputs in these directions maximize information for any assigned modal variances. The remaining problem is therefore to maximize $\frac{1}{2} \sum_i \log(1 + \lambda_i p_i)$ over $p_i \geq 0$ with $\sum_i p_i \leq P_{\max}$. For positive power, optimality equalizes the marginal gains $\lambda_i / [2(1 + \lambda_i p_i)]$ across directions receiving power. Power goes first to directions with larger λ_i. As the budget grows, weaker directions can also receive power. Solving these conditions gives the water-filling allocation in (28) (Cover & Thomas, 2006). Zero power gives zero capacity. □

Equation (28) makes the connection to intervention concrete. Allocating action variance p_i along u_i produces signal-to-noise ratio $\lambda_i p_i$ and intervention $\lambda_i p_i / 2$. For the same action variance, a larger λ_i yields more action information but also more intervention. The power budget limits command magnitude, whereas intervention also accounts for how strongly each command changes the passive dynamics. Without any bound on action power, empowerment diverges whenever $G \neq 0$.

A finite horizon in an LTI system. When F and G remain constant, the system is linear time invariant (LTI). Consider $S_{t+1} = FS_t + GA_t + N_t$ with independent disturbances $N_t \sim \mathcal{N}(0, \Sigma_N)$ and fixed $S_0 = s$. Choose the entire action sequence $U_H = (A_0^\top, \dots, A_{H-1}^\top)^\top$ independently of these disturbances, so no action depends on intermediate observations. The terminal state is $S_H = F^H s + G_H U_H + N_H$, where

$$G_H \doteq [F^{H-1} G \ \cdots \ G], \qquad N_H \doteq \sum_{t=0}^{H-1} F^{H-1-t} N_t, \qquad \Sigma_H \doteq \sum_{j=0}^{H-1} F^j \Sigma_N (F^j)^\top.$$

Here G_H is the controllability matrix mapping the action sequence to the terminal state, and Σ_H is the accumulated noise covariance. Applying Corollary 21 with G_H and Σ_H gives the maximum of $I(U_H; S_H | S_0 = s)$ under a total power constraint $\mathbb{E}\|U_H\|^2 \leq P_{\max}$. The translation $F^H s$ again leaves information unchanged. This empowerment can depend on the propagation matrix F, actuation, noise, and horizon even though it is constant across starting states.

Capacity and the passive reference. The Golden Formula (14) gives another connection between empowerment and intervention. Each action produces a successor distribution. For a finite channel with unrestricted action sources, empowerment is the smallest KL radius containing all these distributions (Polyanskiy & Wu, 2025). Its center is optimized, whereas intervention measures divergence from the declared passive reference.

Proposition 22 (Empowerment as an information radius)**.** *Fix a state s in a channel with finite action and successor alphabets, and allow every action source distribution. Then*

$$\mathcal{E}(s) = \min_Q \max_a D_{\mathrm{KL}}(P(\cdot|s,a) \| Q). \tag{29}$$

For a passive reference with finite divergences,

$$\mathcal{E}(s) \le \max_a D_{\mathrm{KL}}(P(\cdot|s,a)\|P_0(\cdot|s)) = \max_a \tau(s,a). \tag{30}$$

Proof. Write $P_a \doteq P(\cdot|s,a)$ and let $P_\nu = \sum_a \nu(a)P_a$ be the output law under action source ν. For any reference Q with finite divergences, the Golden Formula gives

$$I_\nu(A;S') = \sum_a \nu(a) D_{\mathrm{KL}}(P_a\|Q) - D_{\mathrm{KL}}(P_\nu\|Q) \le \max_a D_{\mathrm{KL}}(P_a\|Q).$$

References with an infinite maximum give a trivial upper bound. Maximizing over ν therefore shows that every candidate center bounds empowerment from above.

For equality, let ν^* maximize mutual information and set $Q^* = P_{\nu^*}$. Such a maximizer exists because the input simplex is compact and finite channel mutual information is continuous. The law Q^* assigns positive mass to every reachable output: otherwise, mixing ν^* with an action reaching a missing output would have an infinite positive directional derivative, contradicting optimality. For each action a, define $\nu_t = (1-t)\nu^* + t\delta_a$. Differentiating at $t = 0$ gives

$$0 \ge \left.\frac{d}{dt} I_{\nu_t}(A;S')\right|_{t=0+} = D_{\mathrm{KL}}(P_a\|Q^*) - \mathcal{E}(s).$$

Thus every action law lies within KL radius $\mathcal{E}(s)$ of Q^*. Their divergences averaged under ν^* equal $\mathcal{E}(s)$, so their maximum equals it. This proves (29). Taking $Q = P_0(\cdot|s)$ gives (30). □

The same radius argument applies with state as the input and action held fixed, provided the finite starting state source is unrestricted.

Proposition 22 bounds empowerment by the largest available intervention at a state, not by the intervention of a deployed policy. A controller may leave that capacity unused: in Theorem 8, the passive policy has zero intervention while empowerment can remain positive. Capacity describes what the channel affords under the declared input constraint. Realized intervention depends on which actions the controller actually uses. The examples in Appendix A.1.2 illustrate why these quantities must remain distinct.

A.4 Body gating through feedback

This section proves the sequential body gate in Theorem 3, showing how the information split (5) extends to successive decisions in a feedback loop.

A.4.1 Sequential body gating

Mutual information and its two chronological components. Directed information separates information used to choose commands from information that earlier commands create in later observations (Massey, 1990; Tatikonda & Mitter, 2009; Derpich et al., 2013; Tsur et al., 2026). Here the state is observed before the action at that time is chosen. For a finite horizon, the joint law therefore factors chronologically as

$$p(s^{0:T-1}, a^{0:T-1}) = \prod_{t=0}^{T-1} p(s_t|s^{0:t-1}, a^{0:t-1}) p(a_t|s^{0:t}, a^{0:t-1}).$$

Factor the state and action marginal laws in their own chronological orders. The logarithm of the joint law divided by the product of these marginals separates into the sum of two log ratios:

$$\log \frac{p(s^{0:T-1}, a^{0:T-1})}{p(s^{0:T-1})p(a^{0:T-1})} = \sum_t \log \frac{p(a_t|s^{0:t}, a^{0:t-1})}{p(a_t|a^{0:t-1})} + \sum_t \log \frac{p(s_t|s^{0:t-1}, a^{0:t-1})}{p(s_t|s^{0:t-1})}.$$

Taking expectations gives

$$I(S^{0:T-1}; A^{0:T-1}) = \mathcal{R}_T^{\rightarrow} + \sum_{t=0}^{T-1} I(A^{0:t-1}; S_t|S^{0:t-1}). \tag{31}$$

Equation (31) gives the chronological decomposition used in Section 4. The same argument uses Radon–Nikodym derivatives when densities are not available. The forward sum includes the current state; the reverse sum includes only earlier actions. Including A_t in the reverse term would destroy this chronology and would not give the stated identity (Massey & Massey, 2005).

Across several decisions, the causal ordering matters. In Example 2, all forward increments in (9) vanish, while $I(A_0; S_1|S_0) = \log 2$ is the sole reverse increment in (31). The state records an earlier action, although no command conveys a state observation.

Conditional one-step identities. We now focus on the controller directed information $\mathcal{R}^{\to}_T$ in (31) and ask how much of it is shared with passive motion. At each decision, the same noise counterfactual holds the realized past fixed. The passive successor is $Y_{t+1} = f_t(S_t, a_0, N_t)$, where the fresh disturbance N_t is independent of the history and controller randomness. Once S_t is given, the action cannot reveal anything more about Y_{t+1}, even if the controller remembers earlier states. Thus $I(A_t; Y_{t+1}|S_t, A^{0:t-1}) = 0$. The following definitions keep the conditioning explicit.

Definition 4 (Sequential body-gated quantities)**.** For $t = 0, \dots, T-1$, suppose $A_t - S_t - Y_{t+1}$ conditional on $A^{0:t-1}$. Define

$$\mathcal{M}^{\to}_{\mathrm{std},T} \doteq \sum_{t=0}^{T-1} I(S_t; Y_{t+1}|A^{0:t-1}), \qquad \mathcal{M}^{\to}_{\mathrm{beh},T} \doteq \sum_{t=0}^{T-1} I(S_t; Y_{t+1}|A^{0:t}),$$
$$\mathcal{R}^{\to}_{\mathrm{red},T} \doteq \sum_{t=0}^{T-1} I(A_t; Y_{t+1}|A^{0:t-1}), \qquad \mathcal{R}^{\to}_{\mathrm{comp},T} \doteq \sum_{t=0}^{T-1} I(A_t; S_t|Y_{t+1}, A^{0:t-1}).$$

The first two sums compare passive prediction before and after the current command is known. The last two separate the current state information in each command into information shared with passive motion and information that passive motion does not reveal. The theorem below gives this split and specifies when it also accounts for the full controller directed information.

Theorem 23 (Sequential body gate and controller information)**.** *Under Definition 4,*

$$\mathcal{M}^{\to}_{\mathrm{std},T} - \mathcal{M}^{\to}_{\mathrm{beh},T} = \mathcal{R}^{\to}_{\mathrm{red},T}, \qquad \sum_{t=0}^{T-1} I(S_t; A_t|A^{0:t-1}) = \mathcal{R}^{\to}_{\mathrm{red},T} + \mathcal{R}^{\to}_{\mathrm{comp},T}.$$

If the controller uses only the current state and past actions, so that A_t is conditionally independent of $S^{0:t-1}$ given $(S_t, A^{0:t-1})$, the second identity gives the directed information split (10) *in Theorem 3. A Markov policy satisfies this condition. If also $S_{t+1} = T_{t,A_t}(Y_{t+1})$ with each $T_{t,a}$ invertible, then*

$$\mathcal{M}^{\to}_{\mathrm{beh},T} = \sum_{t=0}^{T-1} I(S_t; S_{t+1}|A^{0:t}).$$

Proof. At each time, the chain rule and the conditional independence $I(A_t; Y_{t+1}|S_t, A^{0:t-1}) = 0$ from Definition 4 give

$$I(S_t; Y_{t+1}|A^{0:t-1}) - I(S_t; Y_{t+1}|A^{0:t}) = I(A_t; Y_{t+1}|A^{0:t-1}),$$
$$I(S_t; A_t|A^{0:t-1}) = I(A_t; Y_{t+1}|A^{0:t-1}) + I(A_t; S_t|Y_{t+1}, A^{0:t-1}).$$

These are the conditional versions of the identities in Theorem 10. Summing these identities over time and substituting Definition 4 gives the two decompositions in the theorem. For the directed information identification, expand

$$I(S^{0:t}; A_t|A^{0:t-1}) = I(S_t; A_t|A^{0:t-1}) + I(S^{0:t-1}; A_t|S_t, A^{0:t-1}).$$

The assumed conditional independence of the current action from earlier states makes the final term zero. Thus the sum of current state contributions equals the controller directed information.

Finally, conditioning on $A^{0:t}$ fixes the current action and therefore the invertible transformation between Y_{t+1} and S_{t+1}. Invariance of mutual information under this transformation gives $I(S_t; Y_{t+1}|A^{0:t}) = I(S_t; S_{t+1}|A^{0:t})$. Summing proves the last equality. □

If $S_{t+1} = T_{t,A_t}(Y_{t+1})$ with each $T_{t,a}$ invertible, knowing the current action allows recovery of the passive successor. Combining the identities in Theorem 23 then gives

$$\sum_{t=0}^{T-1} I(S_t; S_{t+1}|A^{0:t}) = \sum_{t=0}^{T-1} I(S_t; Y_{t+1}|A^{0:t-1}) - \sum_{t=0}^{T-1} I(A_t; Y_{t+1}|A^{0:t-1}). \tag{32}$$

Equation (32) is the MC identity stated in Section 4. Its left side conditions on all commands through the current decision, so it generally differs from $\sum_t I(S_t; S_{t+1}|A_t)$, which conditions only on the current command.

Under the controller assumption in Theorem 23, earlier states provide no additional information about the current action once the current state and past actions are given. The directed controller rate defined in (9) therefore reduces to the following form.

Corollary 24 (Directed controller rate). Under Definition 4, if A_t is conditionally independent of $S^{0:t-1}$ given $(S_t, A^{0:t-1})$, then

$$\mathcal{R}^{\rightarrow}_T = \sum_{t=0}^{T-1} I(S_t; A_t|A^{0:t-1}) = \mathcal{R}^{\rightarrow}_{\mathrm{red},T} + \mathcal{R}^{\rightarrow}_{\mathrm{comp},T}.$$

Proof. The controller assumption removes the contribution from earlier states once the current state is given, as shown in the proof of Theorem 23. Applying its controller information decomposition gives the result. □

At $T = 1$, the action history is empty and the split reduces to (5).

Conditioning on past commands accounts for information already revealed through actions. A controller may, however, observe a state feature long before expressing it in a command. Conditioning additionally on past states would exclude that information even when a later command reveals it for the first time. With this conditioning, we measure only the information the command carries about the current state beyond the full past. We call the resulting sum the *state innovation rate*. This differs from the action innovation in Theorem 19, which is fresh randomness added to the command. Corollary 25 separates these contributions to the directed rate, and Example 13 illustrates why the distinction matters.

Corollary 25 (State innovation given the full history). For an arbitrary policy using the observed history, let $H_t \doteq (S^{0:t-1}, A^{0:t-1})$ and assume $I(A_t; Y_{t+1}|S_t, H_t) = 0$. Then

$$\mathcal{R}^{\rightarrow}_{\mathrm{innov},T} \doteq \sum_{t=0}^{T-1} I(S_t; A_t|H_t) = \sum_{t=0}^{T-1} I(A_t; Y_{t+1}|H_t) + \sum_{t=0}^{T-1} I(A_t; S_t|Y_{t+1}, H_t). \tag{33}$$

The full directed controller rate also includes information about earlier states:

$$\mathcal{R}^{\rightarrow}_T = \sum_{t=0}^{T-1} I(S^{0:t-1}; A_t|A^{0:t-1}) + \mathcal{R}^{\rightarrow}_{\mathrm{innov},T}. \tag{34}$$

Proof. Apply the conditional chain rule argument from Theorem 23 given H_t. The assumed conditional independence gives the split in (33) and the corresponding passive prediction identity:

$$\sum_{t=0}^{T-1} I(S_t; Y_{t+1}|H_t) - \sum_{t=0}^{T-1} I(S_t; Y_{t+1}|H_t, A_t) = \sum_{t=0}^{T-1} I(A_t; Y_{t+1}|H_t).$$

For the full directed rate, expand each increment by separating earlier states from the current state:

$$I(S^{0:t}; A_t|A^{0:t-1}) = I(S^{0:t-1}; A_t|A^{0:t-1}) + I(S_t; A_t|H_t).$$

Summing gives (34). □

The earlier state contribution in (34) can be positive even for a Markov policy. Such a policy need not consult earlier observations, but its current state may still contain them. The following example shows why conditioning on the full past can remove information that a command is only now revealing.

Example 13 (An old observation first appears in a later command)**.** Let S_0 be a fair bit that persists as $S_1 = S_0$. Choose $A_0 = 0$ and $A_1 = S_1$. This is a time-dependent Markov policy. The controller observes the bit at time zero, but its first command reveals nothing about it. The second command reveals the bit completely, giving

$$\mathcal{R}_2^{\rightarrow} = I(S_0; A_0) + I(S_0, S_1; A_1|A_0) = 0 + \log 2.$$

The state innovation rate is nevertheless zero:

$$\mathcal{R}_{\mathrm{innov},2}^{\rightarrow} = I(S_0; A_0) + I(S_1; A_1|S_0, A_0) = 0.$$

The second observation supplies no new information because S_1 is already determined by S_0. All of the directed rate comes from the earlier state contribution $I(S_0; A_1|A_0) = \log 2$. The bit is old to the controller but newly revealed through its command.

Example 13 explains why the main body gate conditions on past commands rather than the full state history. It counts state information when commands first reveal it, even if the controller observed it earlier. The Gaussian recursion in the next section uses the same conditioning to track what successive commands reveal.

A.4.2 The evolving Gaussian gate

The sequential body gate depends on what earlier commands have already revealed about the state. Viewed through the action record, each new command is a noisy observation, and the dynamics propagate the uncertainty remaining after that observation. This gives the Kalman covariance recursion described in Section 4. The proposition below uses it to compute the evolving gate and the information revealed at each decision.

Proposition 26 (Causal Gaussian body gain recursion)**.** *Let $S_{t+1} = F_t S_t + G_t A_t + N_t$ and $A_t = K_t S_t + \epsilon_t$, with deterministic matrices, a Gaussian initial state, and mutually independent Gaussian innovations independent of that state. Their covariances are $\Sigma_{N,t} \succ 0$ and $\Sigma_{\epsilon,t} \succ 0$. Define $\Sigma_t^- \doteq \mathrm{Cov}(S_t|A^{0:t-1}) \succ 0$ and $\Sigma_t^+ \doteq \mathrm{Cov}(S_t|A^{0:t})$. These covariances do not depend on the realized commands and satisfy*

$$\Sigma_t^+ = ((\Sigma_t^-)^{-1} + K_t^\top \Sigma_{\epsilon,t}^{-1} K_t)^{-1}, \quad \Sigma_{t+1}^- = F_t \Sigma_t^+ F_t^\top + \Sigma_{N,t}. \tag{35}$$

The controller information increment is

$$I(S_t; A_t|A^{0:t-1}) = \tfrac{1}{2} \log \frac{\det \Sigma_t^-}{\det \Sigma_t^+}.$$

For the passive successor $Y_{t+1} = F_t S_t + N_t$, the predictive information before and after observing the command is

$$I(S_t; Y_{t+1}|A^{0:t-1}) = \tfrac{1}{2} \log\det(I + \Sigma_{N,t}^{-1} F_t \Sigma_t^- F_t^\top),$$
$$I(S_t; Y_{t+1}|A^{0:t}) = \tfrac{1}{2} \log\det(I + \Sigma_{N,t}^{-1} F_t \Sigma_t^+ F_t^\top).$$

The conditional body gate is $B_{S,t} = M_{S,t}(I + M_{S,t})^{-1}$, where $M_{S,t} \doteq (\Sigma_t^-)^{1/2} F_t^\top \Sigma_{N,t}^{-1} F_t (\Sigma_t^-)^{1/2}$. Its eigenvalues are the squared conditional canonical correlations of S_t and Y_{t+1} given $A^{0:t-1}$.

Proof. Treat A_t as a Gaussian observation of S_t. Completing the square in the conditional density adds $K_t^\top \Sigma_{\epsilon,t}^{-1} K_t$ to the prior precision, giving the first equation in (35). The reduction in Gaussian state entropy gives the controller information increment.

Conditioning on the commands through time t makes $G_t A_t$ a known translation. The remaining uncertainty comes from $F_t S_t + N_t$, giving the second equation in (35). Starting from the fixed initial covariance, these updates depend only on the deterministic matrices and noise covariances, so they are independent of the realized commands. Applying the exact Gaussian information formulas in Theorem 14(i) with the conditional state covariances Σ_t^- and Σ_t^+ gives the predictive information before and after observing A_t. Finally, Theorem 14 applied to the conditional prior Σ_t^- gives the gate and its canonical correlation interpretation. Conditional means do not affect these quantities. □

The first example isolates how the action record changes the gate while both the mechanics and feedback remain fixed.

Example 14 (Two decisions for a coasting velocity)**.** Let S_t be a scalar velocity deviation with dynamics $S_{t+1} = 0.8S_t + A_t + N_t$ and corrective feedback $A_t = -S_t + \epsilon_t$. Take independent centered Gaussian S_0, N_t, and ϵ_t with variances 1, 0.36, and 1, respectively. The initial prior variance is 1. Observing A_0 reduces it to $1/2$, and (35) gives the next prior variance as $0.8^2(1/2) + 0.36 = 0.68$.

The exact Gaussian formulas in Proposition 26, together with the sequential split, give the following information quantities in nats:

Decision	Prior variance	Gate weight	Body-visible	Complementary
$t = 0$	1.00	0.6400	0.1928	0.1537
$t = 1$	0.68	0.5473	0.1252	0.1342

The two information columns sum to controller increments of 0.3466 and 0.2594 nats. Across both decisions, body-visible information totals 0.3180 nats and complementary information totals 0.2879 nats, giving a directed rate of 0.6060 nats. The second decision starts with less state uncertainty given the action record, so the same mechanics receive a lower gate weight.

Example 15 applies the recursion across walking strides, showing how passive contraction changes the information available at later decisions.

Example 15 (The walker across successive strides)**.** Return to a gait mode from Example 12, with passive factor λ, corrective gain magnitude k, action noise variance 1, and transition noise variance 0.01. Write $p_t \doteq \mathrm{Var}(S_t|A^{0:t-1})$ and $q_t \doteq \mathrm{Var}(S_t|A^{0:t})$ for the prior and posterior variances given the action record. Starting from $p_0 = 1$, Proposition 26 gives

$$q_t = \frac{p_t}{1 + k^2 p_t}, \qquad p_{t+1} = \lambda^2 q_t + 0.01, \qquad b_t = \frac{\lambda^2 p_t}{0.01 + \lambda^2 p_t}.$$

Here b_t is the gate weight at stride t. The controller information increment $r_t \doteq I(S_t; A_t|A^{0:t-1})$ and body-visible increment $\ell_t \doteq I(A_t; Y_{t+1}|A^{0:t-1})$ are

$$r_t = \tfrac{1}{2}\log(1 + k^2 p_t), \qquad \ell_t = \tfrac{1}{2}\log\frac{0.01 + \lambda^2 p_t}{0.01 + \lambda^2 q_t}.$$

Their difference $r_t - \ell_t$ is the complementary increment. Summing over coordinates and strides gives the two components of the directed controller rate.

For the second coordinate, $k = 0.5$ gives $q_0 = 0.8$. The canonical map, with $\lambda = 0.55$, gives $p_1 = 0.252$ and $b_1 \approx 0.8840$. The heavily geared map, with $\lambda = 0.05$, gives $p_1 = 0.012$ and $b_1 \approx 0.0030$. Strong contraction suppresses the earlier gait deviation, leaving fresh disturbances to account for most of the remaining uncertainty. Little of that uncertainty is then carried into the next passive successor.

These examples show how fixed mechanics can produce changing gate weights across decisions. Equation (35) tracks both the information disclosed by commands and the uncertainty propagated and added by the dynamics, providing the conditional state ensemble needed for the sequential body gate.

A.5 From information to physical savings

A.5.1 Properties of the cheap control and information frontiers

Definition 1 compares physical and informational optima under the same task constraint. For the body-visible surrogate, one can similarly define $\mathcal{B}^\star(c; \Pi) \doteq \inf_{\pi \in \Pi : C_{\mathrm{task}}(\pi) \le c} I_\pi(A; Y)$ under a declared evaluation law. The three frontiers share the following monotonicity properties.

The intervention objective is Definition 3. The frontier definitions use a task cost upper bound c. Increasing c enlarges the feasible set, so all three frontiers are nonincreasing, with value $+\infty$ when infeasible. This monotonicity holds for any fixed policy class and does not require convexity. Convexity requires additional structure (Remarks 6 and 7).

Remark 6 (Convexity of the physical frontiers)**.** For a finite discounted MDP, normalized state-action occupancies $\mu(s,a)$ of stationary randomized policies satisfy linear balance equations for state visitation, and therefore form a convex set. Suppose task cost and intervention are fixed linear functionals, $C_{\mathrm{task}} = \langle \mu, c_{\mathrm{task}} \rangle$ and $\mathcal{T} = \langle \mu, \tau \rangle$, where $c_{\mathrm{task}}(s,a)$ is the one-stage task cost with any discount normalization absorbed into it and $\tau(s,a)$ is the conditional KL departure. Take two feasible occupancies with task bounds c_1 and c_2. Their mixture with weight $\lambda \in [0,1]$ is also a valid occupancy. Its task cost is at most $\lambda c_1 + (1-\lambda)c_2$, and its intervention is the same weighted average of the two interventions. Applying this construction to optimal occupancies, or to a sequence approaching each optimum, gives

$$\mathcal{C}^\star(\lambda c_1 + (1-\lambda)c_2; \Pi) \le \lambda \mathcal{C}^\star(c_1; \Pi) + (1-\lambda)\mathcal{C}^\star(c_2; \Pi).$$

Thus the physical frontier is convex under these assumptions. The same proof applies when intervention is replaced by squared action cost $\|a\|^2$ at each decision. A restricted parametric policy family need not realize mixtures of its occupancies, so this convexity argument does not automatically apply to the plotted restricted class curves in Figure 2.

Remark 7 (Convexity of information frontiers)**.** At fixed P_S, $I(S;A)$ is convex in the encoder $P_{A|S}$. At fixed P_{SY} with $A - S - Y$, $I(A;Y)$ is also convex in that encoder because the induced channel $P_{A|Y}$ depends linearly on it while P_Y remains fixed. Suppose the admissible encoder family is convex and task cost is affine in the encoder, so mixing encoders mixes their task costs in the same proportions. The argument above then establishes convexity of the information and body-visible frontiers. Fixing the state ensemble for information evaluation alone is insufficient. A policy also changes future state visitation, so a general multistep task cost is not affine in the policy kernel.

Policy restrictions and regularization paths. The curves in Figure 2 compare controllers within the chosen policy class at common task cost limits. They need not attain the frontiers available to unrestricted policies. Direct provides the intervention reference within this class, against which MC and Compression are compared. A separate issue arises when controllers are obtained by varying a regularization coefficient instead of imposing task cost limits. Even with exact optimization, such a regularization path can skip parts of a nonconvex frontier because some constrained optima minimize no weighted objective.

A.5.2 Information-only no-go result

The constant shift construction in Theorem 4 holds the state law fixed and changes the action marginal without changing the information carried by the action. Theorem 27 below proves the general formula, including the centered case used in the main text.

Theorem 27 (Information-only regularization cannot guarantee physical cheapness)**.** *Let (S,Z) have any joint law with $\mathbb{E}\|Z\|^2 < \infty$, and let $N \sim \mathcal{N}(0,\Sigma_N)$ with $\Sigma_N \succ 0$ be independent of (S,Z). Define $Y = FS + N$, $A_u = Z + u$, and $S'_u = FS + GA_u + N$ for a constant vector u. Use the passive reference $P_0(\cdot|s) = \mathcal{N}(Fs, \Sigma_N)$. Then*

$$I(S;A_u) = I(S;Z), \qquad I(A_u;Y) = I(Z;Y), \qquad I(A_u;S|Y) = I(Z;S|Y).$$

With $R_{\mathrm{T}} = G^\top \Sigma_N^{-1} G$,

$$\mathcal{T}(u) = \tfrac{1}{2}\mathbb{E}[(Z+u)^\top R_{\mathrm{T}}(Z+u)].$$

If $\mathbb{E}Z = 0$, then $\mathcal{T}(u) = \mathcal{T}(0) + \frac{1}{2}u^\top R_{\mathrm{T}} u$. If shifts of arbitrarily large magnitude are admissible along a vector v with $Gv \neq 0$, intervention is unbounded while all displayed information quantities remain fixed. Thus no function of these information quantities alone gives a universal finite intervention bound over this class.

Proof. Translation by u is a measurable bijection, so mutual information is invariant under $Z \mapsto Z + u$, including when conditioned on Y. This proves the information equalities. At a fixed state and action, the passive and controlled kernels have equal covariance and means separated by Ga. Their KL divergence is

$$D_{\mathrm{KL}}(\mathcal{N}(Fs + Ga, \Sigma_N) \| \mathcal{N}(Fs, \Sigma_N)) = \tfrac{1}{2} a^\top G^\top \Sigma_N^{-1} G a.$$

Averaging over A_u proves the intervention formula. Expanding the quadratic form produces a cross term $u^\top R_{\mathrm{T}} \mathbb{E}Z$, which vanishes under centering. For $u = tv$ with $Gv \neq 0$, positive definiteness of Σ_N gives $v^\top R_{\mathrm{T}} v > 0$, so intervention grows quadratically in t. Since the information arguments are fixed, any finite bound depending only on them is eventually exceeded. □

The translation example rules out a universal implication. A restricted comparison is still possible if action means are controlled and the policy noise geometry is compatible with the intervention metric.

Proposition 28 (Information bounds for feedback intervention)**.** *In the centered Gaussian model of Theorem 14, define an action space encoding matrix by*

$$X_A \doteq \Sigma_\epsilon^{-1/2} K \Sigma_S K^\top \Sigma_\epsilon^{-1/2}.$$

The feedback part of intervention is $\mathcal{T}_{\mathrm{fb}} \doteq \frac{1}{2}\operatorname{tr}(R_{\mathrm{T}} K \Sigma_S K^\top)$*. With* $M_A \doteq \Sigma_\epsilon^{1/2} R_{\mathrm{T}} \Sigma_\epsilon^{1/2}$*,*

$$\mathcal{T}_{\mathrm{fb}} = \tfrac{1}{2}\operatorname{tr}(M_A X_A), \qquad I(S;A) = \tfrac{1}{2}\log\det(I + X_A).$$

If the operator norm satisfies $\|X_A\|_2 \le \varepsilon_{\mathrm{enc}}$ *for a finite bound* $\varepsilon_{\mathrm{enc}} \ge 0$*, then*

$$\lambda_{\min}(M_A) I(S;A) \le \mathcal{T}_{\mathrm{fb}} \le (1+\varepsilon_{\mathrm{enc}})\lambda_{\max}(M_A) I(S;A).$$

Proof. Cyclicity of trace proves the intervention identity, and the Gaussian channel formula gives the information identity. For $0 \le x \le \varepsilon_{\mathrm{enc}}$, integration of $1/(1+x)$ gives $x/(1+\varepsilon_{\mathrm{enc}}) \le \log(1+x) \le x$. Apply these inequalities to the eigenvalues of X_A, then use

$$\lambda_{\min}(M_A)\operatorname{tr}(X_A) \le \operatorname{tr}(M_A X_A) \le \lambda_{\max}(M_A)\operatorname{tr}(X_A).$$

Combining the inequalities proves both bounds. □

Proposition 28 concerns feedback intervention; the random action component adds a constant only when policy noise is fixed. The bounds are tightest when encoding is weak and the eigenvalues of M_A are comparable and nonzero. Even then, it does not tell us which feedback changes preserve the task. That requires the task curvature geometry in Appendix A.5.4.

A.5.3 The Fisher geometry and choice of intervention

As discussed in Remark 2, the meaning of cheap control depends on how intervention is measured. We identify the local Fisher geometry of the transition KL cost $\tau(s,a)$ from Definition 3 and explain how the passive reference and process noise shape this measure.

The following standard expansion identifies Fisher information as the local quadratic cost of intervention (Polyanskiy & Wu, 2025).

Proposition 29 (Local quadratic form of intervention)**.** *Fix a state* s *and let* p_a *be a twice continuously differentiable family of transition densities with common support near* $a = 0$*, where* p_0 *is the passive density. Assume that the normalization and KL integrals can be differentiated twice under the integral sign and that the KL Hessian is continuous. With score* $\zeta(s') \doteq \nabla_a \log p_a(s')|_{a=0}$ *and Fisher matrix* $\mathcal{I}_F(s) \doteq \mathbb{E}_{p_0}[\zeta\zeta^\top]$*,*

$$\tau(s,a) = D_{\mathrm{KL}}(p_a \| p_0) = \tfrac{1}{2} a^\top \mathcal{I}_F(s) a + o(\|a\|^2).$$

Proof. With the state s fixed, integrate over the successor s' and define $D(a) \doteq \int p_a(s') \log \frac{p_a(s')}{p_0(s')}\, ds'$. The action a is the parameter of the density, not the integration variable. Normalization gives $\int \nabla_a p_a(s')\, ds' = 0$, so differentiating under the integral yields $\nabla_a D(a) = \int \nabla_a p_a(s') \log \frac{p_a(s')}{p_0(s')}\, ds'$. At $a = 0$, the logarithm vanishes and $\nabla_a D(0) = 0$. Differentiating again produces one term containing this logarithm and a second from its derivative. The first vanishes at zero, leaving

$$\nabla_a^2 D(0) = \int \frac{\nabla_a p_a(s')|_{a=0}\, \nabla_a p_a(s')|_{a=0}^\top}{p_0(s')}\, ds' = \mathbb{E}_{p_0}[\zeta\zeta^\top] = \mathcal{I}_F(s).$$

Since $D(0) = 0$, Taylor expansion with the assumed continuous Hessian gives $D(a) = \frac{1}{2} a^\top \mathcal{I}_F(s) a + o(\|a\|^2)$. □

Remark 8 (Three readings of the action metric R_{T})**.** For the Gaussian transition $P(\cdot|s,a) = \mathcal{N}(Fs + Ga, \Sigma_N)$, the action metric R_{T} introduced in Section 5 has three related interpretations.

1. **Fisher information.** The score with respect to the action at zero action is $G^\top \Sigma_N^{-1}(s' - Fs)$, whose covariance under the passive law is R_{T}.
2. **Intervention cost.** The exact cost is $\tau(s, a) = \frac{1}{2} a^\top R_{\mathrm{T}} a$, so Proposition 29 has no remainder in this model.
3. **Action channel geometry.** Since

$$R_{\mathrm{T}} = (\Sigma_N^{-1/2} G)^\top (\Sigma_N^{-1/2} G),$$

its eigenvalues measure the strength of action effects after normalizing the process noise. Together with their eigenvectors, they determine the optimal variance allocation for empowerment in Corollary 21.

The *same* matrix therefore determines how distinguishable an action is, how much intervention it incurs, and which directions empowerment favors under a squared action budget.

The action metric also weights the intervention generated by state feedback: for $A = KS + \epsilon$ in the centered Gaussian model, $\mathcal{T}_{\mathrm{fb}} = \frac{1}{2} \operatorname{tr}(R_{\mathrm{T}} K \Sigma_S K^\top)$, as in Theorem 19. It is distinct from the passive body gate B_S, which describes what state information passive motion retains. It also differs from the task Hessian H, which measures the cost of changing policy parameters rather than actions.

Coordinate invariance is not independence of the physical model. KL is unchanged when both successor distributions are pushed forward by the same invertible coordinate transformation. Relabeling actions consistently also preserves the transition divergences. By contrast, an unweighted norm $\|a\|^2$ changes with action units. Nevertheless, transition KL depends on the chosen passive reference and noise model. For a fixed positive definite covariance $\Sigma_{N,0}$, write $\Sigma_N = \eta \Sigma_{N,0}$ with noise scale $\eta > 0$. The transition KL is $a^\top G^\top \Sigma_{N,0}^{-1} G a / (2\eta)$, so it diverges as $\eta \downarrow 0$ whenever $Ga \neq 0$. In a deterministic model, distinct controlled and passive point masses have infinite KL. A normalized mean-shift diagnostic is therefore not an empirical KL unless a noise model giving that interpretation is specified.

For Gaussian successor laws with equal covariance, the squared 2-Wasserstein distance with Euclidean ground distance is $\|Ga\|^2$ (Peyré & Cuturi, 2019). It measures displacement of the successor mean and weights actions by $G^\top G$, without the inverse noise covariance in transition KL. For these same laws, the Wasserstein KL divergence of Datar & Ay (2026) equals $\frac{1}{2}\|Ga\|^2$. Although the two measures differ in general, both assess mean displacement without inverse noise weighting in this comparison. Related control formulations also use Wasserstein and Kalman–Wasserstein KL divergences (Stein et al., 2026).

We use KL because it measures departure from passive dynamics relative to process noise and supports the exact decomposition into action information and transition mismatch in Proposition 7. The choice of geometry determines what intervention measures. Neither statistical distinguishability nor action magnitude is automatically mechanical work (Appendix A.5.6).

A.5.4 Local task–body–action alignment

Theorem 5 compares regularizers near the same task optimum. The alignment coefficient is defined in Definition 2. Explicit smoothness assumptions below justify the Taylor remainder orders; they also distinguish this local theorem from a finite loss comparison between arbitrary trained policies.

Theorem 30 (Local cheap control alignment at matched task loss)**.** *Let $\theta^\star$ lie in the interior of an open parameter domain. Suppose C is three times continuously differentiable and has $\nabla C(\theta^\star) = 0$ and $H = \nabla^2 C(\theta^\star) \succ 0$. Let R and $\mathcal{T}$ be twice continuously differentiable nearby, with nonzero gradients $g_R = \nabla R(\theta^\star)$ and $g_T = \nabla \mathcal{T}(\theta^\star)$. For sufficiently small $\beta \geq 0$, the local stationary branch θ_β of $C + \beta R$ through $\theta^\star$ consists of strict local minima and satisfies*

$$\begin{aligned} \theta_\beta &= \theta^\star - \beta H^{-1} g_R + O(\beta^2), \\ C(\theta_\beta) - C(\theta^\star) &= \tfrac{1}{2}\beta^2 g_R^\top H^{-1} g_R + O(\beta^3), \\ \mathcal{T}(\theta_\beta) - \mathcal{T}(\theta^\star) &= -\beta g_T^\top H^{-1} g_R + O(\beta^2). \end{aligned}$$

For a sufficiently small positive task cost increase δ, the branch therefore obeys

$$\mathcal{T}(\theta_\beta) - \mathcal{T}(\theta^\star) = -\sqrt{2\delta}\frac{g_T^\top H^{-1} g_R}{\sqrt{g_R^\top H^{-1} g_R}} + O(\delta). \tag{36}$$

Equivalently, the leading coefficient is $-\chi_R\sqrt{2 g_T^\top H^{-1} g_T}$, with χ_R as defined in Section 5. Among such regularizers, direct intervention $R = \mathcal{T}$ maximizes the leading decrease at matched δ.

Proof. Apply the implicit function theorem to $\nabla C(\theta) + \beta \nabla R(\theta) = 0$. Its derivative with respect to θ at $(\theta^\star, 0)$ is the invertible matrix H. A unique differentiable local stationary branch exists, and continuity of the Hessian makes it a branch of strict local minima for small β. Differentiating stationarity at zero gives $\theta_0' = -H^{-1} g_R$.

Write $h_\beta = \theta_\beta - \theta^\star$. The smoothness assumptions yield

$$0 = H h_\beta + \beta g_R + O(\|h_\beta\|^2 + \beta \|h_\beta\|).$$

Since $h_\beta = O(\beta)$, substitution gives $h_\beta = -\beta H^{-1} g_R + O(\beta^2)$. Taylor expansion of C and $\mathcal{T}$ at $\theta^\star$ gives the stated task cost and intervention expansions.

Set $\delta(\beta) \doteq C(\theta_\beta) - C(\theta^\star)$ and $H_\beta \doteq \nabla^2 (C + \beta R)(\theta_\beta)$. Differentiating the stationarity equation $\nabla C(\theta_\beta) + \beta \nabla R(\theta_\beta) = 0$ with respect to β gives $\theta_\beta' = -H_\beta^{-1} \nabla R(\theta_\beta)$. Using $\nabla C(\theta_\beta) = -\beta \nabla R(\theta_\beta)$, we obtain

$$\delta'(\beta) = \beta\, \nabla R(\theta_\beta)^\top H_\beta^{-1} \nabla R(\theta_\beta) > 0$$

for sufficiently small $\beta > 0$, because $H_\beta \succ 0$ and $\nabla R(\theta_\beta) \neq 0$. Thus each sufficiently small positive task increase δ corresponds to a unique point on the positive branch. Inverting the task cost expansion gives

$$\beta = \sqrt{\frac{2\delta}{g_R^\top H^{-1} g_R}} + O(\delta).$$

Substitution proves (36). Finally, Cauchy–Schwarz in the positive definite inner product $\langle u, v \rangle_{H^{-1}} \doteq u^\top H^{-1} v$ gives

$$\frac{g_T^\top H^{-1} g_R}{\sqrt{g_R^\top H^{-1} g_R}} \leq \sqrt{g_T^\top H^{-1} g_T}.$$

Equality holds exactly when g_R is positively proportional to g_T. Choosing $R = \mathcal{T}$ attains this bound and therefore maximizes the leading intervention decrease at matched task loss. □

The alignment coefficient compares a surrogate direction with the physical objective itself. Its upper bound identifies the natural local reference for all such comparisons.

Corollary 31 (Direct intervention is the local oracle). Under Theorem 30, $R = \mathcal{T}$ gives $\chi_R = 1$ and the largest leading intervention decrease at matched infinitesimal task loss. Equality for another admissible regularizer requires g_R to be positively proportional to g_T.

Proof. Theorem 30 bounds the leading intervention decrease by its value at $\chi_R = 1$, attained exactly when g_R is positively proportional to g_T. Choosing $R = \mathcal{T}$ attains this bound. □

Cheap control here means minimizing the control intervention $\mathcal{T}_\pi$ subject to a task bound, as formalized by the physical frontier in (11), with Direct providing the corresponding local benchmark.

To compare two information surrogates, the common task loss normalization removes their arbitrary coefficient scales.

Corollary 32 (Surrogates at matched task loss). Under Theorem 30, let R_1, R_2 have nonzero gradients and local branches $\theta_1(\delta), \theta_2(\delta)$ with the same sufficiently small positive task loss δ. Then

$$\mathcal{T}(\theta_1(\delta)) - \mathcal{T}(\theta_2(\delta)) = -\sqrt{2\delta}(\chi_{R_1} - \chi_{R_2})\sqrt{g_T^\top H^{-1} g_T} + O(\delta).$$

If $\chi_{R_1} > \chi_{R_2}$, the first branch has smaller intervention for all sufficiently small positive δ. Equal alignment leaves the ordering undecided at this order.

Proof. Subtract the two matched loss expansions in Theorem 30. The common task optimum fixes the same g_T and H for both branches. When the alignment difference is positive, its nonzero $\sqrt{\delta}$ term dominates the $O(\delta)$ remainder for sufficiently small δ. □

Corollary 32 makes the comparison independent of the arbitrary scale of each regularization coefficient. It orders nearby policies at the same task loss; farther away, the remainder terms can change that ordering. The next scalar consequence explains why a noticeable change in feedback can initially cause a much smaller change in task return.

Corollary 33 (Scalar parameter response and task return loss). Let the scalar setting of Corollary 17 have fixed $x > 0$, $q \doteq \sigma_S^2/\sigma_A^2 > 0$, and a three times continuously differentiable task return with strict local maximum $k^\star$ and curvature $\kappa_R \doteq -J''_{\mathrm{ext}}(k^\star) > 0$. Define

$$c_\star \doteq \frac{qx}{(1+q(k^\star)^2)(1+q(k^\star)^2+x)}.$$

The local maximizer k_β of $J_{\mathrm{ext}} + \beta\mathcal{M}$ satisfies

$$k^\star - k_\beta = \frac{c_\star k^\star}{\kappa_R}\beta + O(\beta^2), \qquad J_{\mathrm{ext}}(k^\star) - J_{\mathrm{ext}}(k_\beta) = \frac{c_\star^2 (k^\star)^2}{2\kappa_R}\beta^2 + O(\beta^3).$$

Proof. The exact scalar formula gives $\mathcal{M}(k) = \frac{1}{2}\log[(1+qk^2+x)/(1+qk^2)]$, hence $\mathcal{M}'(k^\star) = -c_\star k^\star$. Apply the parameter expansion in Theorem 30 with $C = -J_{\mathrm{ext}}$ and $R = -\mathcal{M}$, then expand the task return at its stationary point. When $k^\star = 0$, it remains stationary for every small β, and the displayed leading coefficients vanish. □

For weak feedback, $c_\star$ approaches $qx/(1+x)$, recovering the small-gain quadratic approximation. Corollary 33 describes the change between neighboring optimized policies as the regularization weight varies.

Corollary 31 is an oracle for the declared intervention objective near a common optimum, not an algorithm ranking at finite task loss. Applying it to raw MC also requires accounting for the policy's effect on occupancy.

Corollary 34 (MC alignment separates redundancy and occupancy). Under the assumptions of Theorems 11 and 30, set $R_{\mathrm{MC}} \doteq I_\pi(A;Y) - I_\pi(S;Y)$. At $\theta^\star$, let $g_{\mathrm{red}} \doteq \nabla I_\pi(A;Y)$ and $g_{\mathrm{occ}} \doteq \nabla I_\pi(S;Y)$. The coefficient of the first-order intervention change is negative exactly when

$$g_T^\top H^{-1} g_{\mathrm{red}} > g_T^\top H^{-1} g_{\mathrm{occ}}.$$

At fixed state law and passive channel, $g_{\mathrm{occ}} = 0$.

Proof. Differentiating the exact objective identity gives $g_R = g_{\mathrm{red}} - g_{\mathrm{occ}}$. The leading intervention change is $-\beta g_T^\top H^{-1} g_R$. Substitution gives the criterion. If the two inner products are equal, this coefficient vanishes and higher-order terms determine the sign. Fixing P_{SY} makes its information gradient zero. □

Corollary 34 shows how raw MC can change intervention through both encoding and visitation. The BODY experiment in Section 6 isolates the encoding contribution by optimizing the body-visible penalty on a fixed state ensemble, with the passive channel held unchanged.

The following example holds the geometry simple enough to see separately why task tolerance and physical improvement must agree.

Example 16 (Task tolerance alone does not ensure physical savings)**.** Let $\theta = (\theta_1, \theta_2)$, $C(\theta) = \frac{1}{2}(100\theta_1^2 + \theta_2^2)$, and $\mathcal{T}(\theta) = 1 + \theta_1$ near $\theta^\star = 0$. The task permits ten times as much displacement along θ_2 as along θ_1 at the same small cost increase. Direct intervention selects the first direction, giving $\chi_{\mathrm{T}} = 1$ and a leading reduction $\sqrt{2\delta}/10$. A regularizer $R(\theta) = \theta_2$ instead moves only in the task's tolerant direction. It has $\chi_R = 0$ and does not change intervention at all in this example. A regularizer $R(\theta) = -\theta_1$ has $\chi_R = -1$ and increases intervention by $\sqrt{2\delta}/10$. Thus task tolerance alone is insufficient: the selected direction must also change the physical objective in the right way.

Example 16 isolates alignment in policy parameter space. Appendix A.5.5 develops the physical counterpart of sequential information accounting: intervention must also be accumulated along the trajectory, rather than inferred solely from its endpoint.

A.5.5 INTERVENTION OVER TRAJECTORIES

Actions whose effects cancel at the endpoint can still incur intervention along the trajectory. A path comparison records each departure from passive dynamics and makes this distinction explicit.

Proposition 35 (Trajectory divergence equals cumulative intervention)**.** *For a horizon T, let $\mathbb{P}^{\pi}_{T}$ be the joint state-action trajectory law with initial law ρ_0, causal policy kernels, and controlled transitions P. Let $\mathbb{P}^{\pi}_{0,T}$ use the same initial law and policy kernels, replacing only P by P_0. Assume absolute continuity on visited transitions and define*

$$\mathcal{T}_T^{\mathrm{path}} \doteq \mathbb{E}_{\pi} \sum_{t=0}^{T-1} D_{\mathrm{KL}}(P(\cdot|S_t, A_t) \| P_0(\cdot|S_t)).$$

Then $D_{\mathrm{KL}}(\mathbb{P}^{\pi}_{T} \| \mathbb{P}^{\pi}_{0,T}) = \mathcal{T}_T^{\mathrm{path}}$. The KL between their terminal-state marginals is at most this value.

Proof. Factor both joint laws into their initial, policy, and transition kernels. Along any fixed trajectory the initial and policy factors are identical, so their likelihood ratio contributions cancel. The remaining log ratio is the sum of transition log ratios. Taking its expectation under the controlled law gives the conditional KL sum, by the KL divergence chain rule (Cover & Thomas, 2006). Projecting the trajectory onto its terminal state cannot increase KL, proving the inequality. □

In Proposition 35, the reference retains the same action rule but changes the state dynamics, so its action distribution need not match the controlled process after averaging over states. Example 17 shows why this construction measures accumulated intervention rather than endpoint displacement.

Example 17 (Endpoint cancellation)**.** Let $S_{t+1} = S_t + A_t + N_t$, with independent Gaussian noise of variance $\sigma_N^2 > 0$ and fixed $S_0 = 0$. Apply the deterministic sequence $A_0 = u$, $A_1 = -u$. The terminal law is the same as under zero actuation: $S_2 = N_0 + N_1$. Its divergence from the passive terminal law is zero, but the two transitions contribute a total intervention of u^2/σ_N^2. Looking only at the endpoint would miss both interventions. This is the physical counterpart of the need to account for the direction and timing of controller information.

A.5.6 MECHANICAL ENERGY, ACTUATOR WORK, AND CONTROL INTERVENTION

Stored kinetic and potential energy can sustain motion even without actuation. The mechanical energy balance separates this passive contribution from actuator work and makes clear why neither is measured by KL intervention.

Let $q(t) \in \mathbb{R}^n$ be a robot's generalized position and $u(t) \in \mathbb{R}^p$ its actuator input. Consider a twice differentiable trajectory satisfying the standard manipulator equations with viscous damping (Spong et al., 2020):

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + \nabla V(q) + D\dot{q} = B(q)u. \tag{37}$$

Here $M(q)$ is a continuously differentiable symmetric positive definite inertia matrix, $V(q)$ is a continuously differentiable potential energy, D is a symmetric positive semidefinite damping matrix, and $B(q)$ maps actuator inputs to generalized forces. The matrix $C(q, \dot{q})$ represents the inertial terms arising from motion and is chosen so that $\dot{M}(q) - 2C(q, \dot{q})$ is skew symmetric, where $\dot{M}$ is the derivative of $M(q(t))$ along the trajectory. Total mechanical energy is the sum of kinetic and potential energy,

$$E(q, \dot{q}) \doteq \tfrac{1}{2}\dot{q}^{\top} M(q)\dot{q} + V(q).$$

The following standard balance identifies how actuator work and damping change this stored energy.

Proposition 36 (Mechanical energy balance separates passive storage from active work)**.** *For the mechanical system in* (37)*, the rate of change of energy is*

$$\dot{E} = \underbrace{\dot{q}^{\top} B(q)u}_{\textit{actuator power}} - \underbrace{\dot{q}^{\top} D\dot{q}}_{\textit{dissipation rate}}. \tag{38}$$

Consequently, over any interval $[t_0, t_1]$ on which the terms are integrable,

$$E(t_1) - E(t_0) = \int_{t_0}^{t_1} \dot{q}^{\top} B(q)u \, dt - \int_{t_0}^{t_1} \dot{q}^{\top} D\dot{q} \, dt. \tag{39}$$

Here $E(t)$ abbreviates $E(q(t), \dot{q}(t))$. The first integral is signed actuator work. With $u = 0$, energy is conserved when $D = 0$ and is nonincreasing when $D \succeq 0$.

Proof. Write $v \doteq \dot{q}$ temporarily. Kinetic energy depends on time through both velocity and the inertia matrix evaluated at $q(t)$. The product rule gives

$$\begin{aligned}\frac{d}{dt}(\tfrac{1}{2}v^\top M(q)v) &= \tfrac{1}{2}\dot{v}^\top Mv + \tfrac{1}{2}v^\top \dot{M}v + \tfrac{1}{2}v^\top M\dot{v} \\ &= v^\top M\dot{v} + \tfrac{1}{2}v^\top \dot{M}v.\end{aligned}$$

The first and third terms are equal because M is symmetric. The chain rule gives $dV(q(t))/dt = \nabla V(q)^\top v$, so

$$\dot{E} = v^\top M\dot{v} + \tfrac{1}{2}v^\top \dot{M}v + \nabla V(q)^\top v.$$

From (37), $M\dot{v} = Bu - Cv - \nabla V - Dv$. Substitution yields

$$\begin{aligned}\dot{E} &= v^\top Bu - v^\top Cv - v^\top \nabla V - v^\top Dv + \tfrac{1}{2}v^\top \dot{M}v + \nabla V^\top v \\ &= v^\top Bu - v^\top Dv + \tfrac{1}{2}v^\top(\dot{M} - 2C)v.\end{aligned}$$

The potential terms cancel because they are the same scalar with opposite signs. They describe energy exchanged between kinetic and potential forms, which does not change their sum.

For any real skew symmetric matrix Z, transposing the scalar $v^\top Zv$ gives $v^\top Z^\top v = -v^\top Zv$, so $v^\top Zv = 0$. Applying this identity to $Z = \dot{M} - 2C$ proves (38). Integration gives (39). Since $D \succeq 0$, the dissipation rate is nonnegative, establishing the passive conclusions. □

An ideal unactuated pendulum exchanges kinetic and potential energy while their sum remains constant. With damping, stored energy decreases even though actuator work remains zero. An actuator can supply energy or remove it through braking, so its signed work can be positive or negative. These physical quantities differ from statistical departure from a passive transition law.

Remark 9 (KL intervention does not determine mechanical work)**.** The work balance in (39) remains meaningful without process noise, whereas KL between distinct deterministic successor laws is infinite. For a finite comparison, consider a mass $m > 0$ with velocity v and a constant force u applied over a step of duration $h > 0$. Let $v' = v + hu/m + N$, where $N \sim \mathcal{N}(0, \sigma^2)$ and $\sigma > 0$, and use the same model with $u = 0$ as the passive reference. The transition KL is

$$D_{\mathrm{KL}}(P(\cdot|v, u)\|P_0(\cdot|v)) = \frac{h^2u^2}{2m^2\sigma^2}.$$

For kinetic energy $K_{\mathrm{mech}}(v) \doteq mv^2/2$, the expected change is

$$\mathbb{E}[K_{\mathrm{mech}}(v') - K_{\mathrm{mech}}(v)|v, u] = huv + \frac{h^2u^2}{2m} + \frac{m\sigma^2}{2}.$$

The last term is also present under passive motion. Subtracting that passive expectation leaves $huv + h^2u^2/(2m)$, which equals the work of the force along the corresponding noiseless motion. Opposite forces of equal magnitude therefore have the same transition KL but can have opposite signs of work. For $m = h = v = 1$ and $u = \pm 1$, both interventions equal $1/(2\sigma^2)$, while the energy changes relative to passive motion are $3/2$ and $-1/2$, respectively.

Thus KL intervention measures dimensionless statistical departure, reported in nats, rather than stored energy or signed work. An energetic interpretation requires an additional physical model that specifies the cost of interest, such as actuator energy consumption, dissipation, or energy recovered during braking.

A.6 Nonlinear channels and physical horizons

Applying the Gaussian body gate to nonlinear mechanics raises two questions: whether a local linear approximation describes the states being evaluated, and whether the observation horizon reveals differences in how reliably passive motion retains state information.

A.6.1 What local covariance geometry can establish

Linearizing the passive dynamics at one state need not capture the covariance geometry between the evaluated states and their passive successors. The approximation depends both on the curvature of the dynamics and on how widely those states are spread around the linearization point. Proposition 37 bounds the resulting covariance errors. We use $\|\cdot\|_F$ for the Frobenius norm and $\|\cdot\|_2$ for the operator norm.

Proposition 37 (Moment bound for a nonlinear passive channel)**.** *Let $S = s_0 + X$, with $\mathbb{E}X = 0$, covariance Σ_X, and finite fourth moment. Let $Y = f(S) + N$, where N is independent of X, centered, and has covariance Σ_N. Suppose f is twice continuously differentiable along the segments from s_0 to the support of S, with bilinear Hessian operator norm at most L_f. Write $F_0 \doteq Df(s_0)$. Then*

$$\|\operatorname{Cov}(X,Y) - \Sigma_X F_0^\top\|_F \le \tfrac{1}{2} L_f \mathbb{E}\|X\|^3,$$
$$\|\operatorname{Cov}(Y) - (F_0 \Sigma_X F_0^\top + \Sigma_N)\|_F \le L_f \|F_0\|_2 \mathbb{E}\|X\|^3 + \tfrac{1}{4} L_f^2 \mathbb{E}\|X\|^4.$$

A sufficient condition for convergence of the nonlinear canonical correlation matrix to its local linear counterpart is $X = \varrho X_0$ with $\varrho \downarrow 0$, $\operatorname{Cov}(X_0) \succ 0$, finite fourth moment, and fixed $\Sigma_N \succ 0$.

Proof. Taylor's theorem gives $f(s_0 + X) = f(s_0) + F_0 X + R_f(X)$ with $\|R_f(X)\| \le L_f \|X\|^2/2$. Since X is centered and independent of N,

$$\operatorname{Cov}(X,Y) - \Sigma_X F_0^\top = \mathbb{E}[X R_f(X)^\top].$$

The Frobenius norm of a rank one matrix $uv^\top$ is $\|u\|\|v\|$, yielding the first bound. Expanding the output covariance gives two cross terms between $F_0 X$ and $R_f(X)$, together with $\operatorname{Cov}(R_f(X))$. The sum of the cross term norms is at most $L_f \|F_0\|_2 \mathbb{E}\|X\|^3$. Also,

$$\|\operatorname{Cov}(R_f(X))\|_F \le \mathbb{E}\|R_f(X) - \mathbb{E}R_f(X)\|^2 \le \mathbb{E}\|R_f(X)\|^2 \le \tfrac{1}{4} L_f^2 \mathbb{E}\|X\|^4.$$

These estimates prove the second bound.

For the final claim, define the normalized cross-covariance

$$C_\varrho \doteq \Sigma_X^{-1/2} \operatorname{Cov}(X,Y) \operatorname{Cov}(Y)^{-1/2}.$$

The canonical correlation matrix on the state coordinates is $C_\varrho C_\varrho^\top$. Under $X = \varrho X_0$, the cross-covariance error is $O(\varrho^3)$ and source whitening has norm $O(\varrho^{-1})$. Both the nonlinear and linear output covariances remain bounded below by the fixed $\Sigma_N \succ 0$. Their difference is $O(\varrho^3)$, so local smoothness of the inverse square root gives the same order for the difference of their inverse square roots. Consequently, the nonlinear and linear normalized cross-covariances differ by $O(\varrho^2)$, and their canonical correlation matrices converge to one another. □

The contraction $X = \varrho X_0$ shrinks the evaluated states toward s_0 without changing the shape of their distribution. These bounds justify a local approximation to covariance geometry. They do not make the Gaussian information formulas exact for a nonlinear channel.

A.6.2 Why a fine timestep can hide selectivity

The observation horizon is the physical time between the starting state and the passive future and may span several simulator steps. Over a very short interval, the successor nearly reproduces the starting state, so every direction can appear almost perfectly retained. This obscures the contrast that the body gate uses to weight information. Theorem 38 makes this precise for a Gaussian channel whose process noise variance scales with elapsed time, as in a linear diffusion. It motivates measuring passive memory over a physical horizon, as in Figure 2c, rather than identifying that horizon with one simulator step.

Theorem 38 (Fine timescale saturation under diffusion scaling)**.** *Let S be Gaussian with fixed covariance $\Sigma_S \succ 0$, and let $Y_\Delta = F_\Delta S + N_\Delta$, with independent centered Gaussian N_Δ. Suppose, in operator norm,*

$$F_\Delta = I + \Delta L_0 + o(\Delta), \qquad \Sigma_{N,\Delta} \doteq \operatorname{Cov}(N_\Delta) = \Delta Q_0 + o(\Delta), \qquad Q_0 \succ 0.$$

Define $M_\Delta \doteq \Sigma_S^{1/2} F_\Delta^\top \Sigma_{N,\Delta}^{-1} F_\Delta \Sigma_S^{1/2}$ *and* $B_\Delta \doteq M_\Delta (I + M_\Delta)^{-1}$. *If* S *has dimension* n_S, *and* $b_i(\Delta)$ *are the eigenvalues of* B_Δ, *then*

$$B_\Delta \to I, \qquad b_i(\Delta) = 1 - O(\Delta), \qquad I(S; Y_\Delta) = \tfrac{n_S}{2} \log(1/\Delta) + O(1).$$

Proof. The scaled noise covariance converges to $Q_0 \succ 0$, so $\Delta \Sigma_{N,\Delta}^{-1} \to Q_0^{-1}$. Since $F_\Delta \to I$,

$$\Delta M_\Delta \to \Sigma_S^{1/2} Q_0^{-1} \Sigma_S^{1/2} \succ 0.$$

Every eigenvalue $m_i(\Delta)$ of M_Δ is therefore bounded above and below by positive multiples of $1/\Delta$ for small Δ. Since $1 - b_i(\Delta) = 1/(1 + m_i(\Delta))$, each deficit is $O(\Delta)$ and $B_\Delta \to I$. The Gaussian channel formula gives

$$I(S; Y_\Delta) = \tfrac{1}{2} \log\det(I + M_\Delta) = \tfrac{n_S}{2} \log(1/\Delta) + \tfrac{1}{2} \log\det(\Delta I + \Delta M_\Delta).$$

The final determinant converges to a positive finite value, proving the asymptotic. □

The theorem assumes Gaussian noise in every state direction, with variance proportional to the sampling interval. Noise only in velocity, or observation noise whose variance stays fixed as the interval shrinks, need not give the same limit. The pendulum study in Figure 2c illustrates the motivation for a longer horizon without satisfying every assumption of this theorem. Over longer horizons, data processing limits how much information passive motion can retain about the starting state. This information need not follow the same trend as the linear correlations plotted in the experiments.

A.6.3 Information retained over longer horizons

Under passive Markov dynamics, the state S_ℓ at horizon ℓ contains all the information about the starting state S_0 available to the next transition. Thus $S_0 - S_\ell - S_{\ell+1}$ is a Markov chain, and the data processing inequality gives

$$I(S_0; S_{\ell+1}) \le I(S_0; S_\ell).$$

Passive evolution can therefore preserve or lose information about the starting state, but cannot increase it. After conditioning on the chosen actions, the same conclusion holds when an action sequence is chosen in advance, independently of the initial state and process disturbances. This excludes feedback that could supply information about the starting state through later commands. We compare successive horizons using the same initial distribution and prefixes of that sequence. Setting every action to zero gives the passive case used in the horizon studies.

Theorem 39 (Monotonicity of the standardized horizon measure)**.** *Let controlled Markov dynamics evolve under an action sequence chosen independently of the initial state and process disturbances. For horizon* $\ell \ge 1$, *define* $\mathcal{M}_\ell \doteq I(S_\ell; S_0 | A^{0:\ell-1})$ *using a common initial law and prefixes of the same action sequence. Then* $\mathcal{M}_{\ell+1} \le \mathcal{M}_\ell$.

Proof. Conditional on $A^{0:\ell}$, the dynamics give the Markov chain $S_0 - S_\ell - S_{\ell+1}$. Applying data processing under this conditioning gives

$$I(S_0; S_{\ell+1} | A^{0:\ell}) \le I(S_0; S_\ell | A^{0:\ell}).$$

The unused action A_ℓ is conditionally independent of (S_0, S_ℓ) given the earlier actions, because the entire action sequence was chosen independently of the initial state and process disturbances. It can therefore be removed from the conditioning on the right, giving

$$I(S_0; S_\ell | A^{0:\ell}) = I(S_0; S_\ell | A^{0:\ell-1}) = \mathcal{M}_\ell.$$

The left side of the inequality is $\mathcal{M}_{\ell+1}$, proving the claim. Actions may be correlated with one another; independence across times is not required. □

Theorem 39 concerns mutual information, whereas the pendulum and MountainCar horizon plots report squared canonical correlations (Figures 2c and 6). These measure linear dependence in the chosen state coordinates. Nonlinear evolution can make an existing dependence more nearly linear without adding information about the starting state. The late rise of the weak MountainCar correlation therefore does not establish a recovery of lost information and does not contradict the theorem.

B EXPERIMENTAL EVIDENCE AND ADDITIONAL RESULTS

These studies support three aspects of the theory: selective compression, the conditions for physical savings, and the dependence of information diagnostics on the state ensemble and observation horizon. Gaussian quantities use Corollary 16. The distinction between a computed regularization path and a global frontier is explained in Appendix A.5.1. A finite regularization sweep is not a recovered global frontier. The linear Gaussian frontiers are analytic population quantities, so sampling intervals do not apply. Unless stated otherwise, intervals are 95% bootstrap intervals, paired for policy comparisons; an interval containing zero establishes neither a directional effect nor equality.

B.1 PUSH AND COAST: SELECTIVE FEEDBACK AND PHYSICAL SAVINGS

A push gives a mass momentum, allowing it to coast toward a target. A separate alignment error still needs correction. This construction tests whether MC can reduce drive feedback while retaining alignment feedback, as suggested by the Gaussian body gate in Theorem 14. For position error e_t, velocity v_t, alignment error h_t, and actions (u_t, c_t), the dynamics are

$$\begin{aligned} e_{t+1} &= e_t + \Delta t v_t + \tfrac{1}{2}\Delta t^2 u_t + N_t^e, \\ v_{t+1} &= \alpha v_t + \Delta t u_t + N_t^v, \\ h_{t+1} &= 0.8 h_t + c_t + N_t^h. \end{aligned} \tag{40}$$

Here $\Delta t = 0.5$ s and the noise is Gaussian. Setting $\alpha = 0.95$ or 0.50 changes coasting without changing the alignment dynamics or task. We optimize three linear feedback gains, with fixed Gaussian action noise, under a quadratic task cost that penalizes states only. Costs and information are summed over ten decisions.

At a common task threshold, MC maximizes behavioral MC, Compression minimizes $I(S;A)$, and Direct minimizes intervention $\mathcal{T}$. Fixed overlap instead minimizes $I(A;Y)$ on a common reference ensemble, isolating the penalty in Theorem 11. Table 2 shows that MC preserves alignment feedback while reducing drive feedback. Its intervention advantage over Compression falls from 62.6% to 12.2% when coasting is weakened.

Table 2: **Selective feedback at a common task tolerance.** Each regularized policy permits 10% greater task cost than the task optimum. Intervention reduction and feedback retention are relative to that optimum. Drive retention uses $\sqrt{k_e^2 + k_v^2}$, and alignment retention uses $|k_h|$. Values are population quantities. The local coefficient χ_R is defined in (12).

Body	Method	$\mathcal{T}$	Reduction (%)	Drive retained (%)	Alignment retained (%)	χ_R
Coasting	Task only	1972.3	–	100.0	100.0	–
	MC	137.6	93.0	39.8	98.3	0.896
	Compression	367.4	81.4	50.7	70.2	0.654
	Fixed overlap	318.0	83.9	42.0	94.4	0.909
	Direct	131.4	93.3	40.6	99.5	1.000
Damped	Task only	1218.0	–	100.0	100.0	–
	MC	238.8	80.4	33.5	97.9	0.891
	Compression	272.1	77.7	44.3	70.9	0.634
	Fixed overlap	251.3	79.4	34.2	95.2	0.909
	Direct	132.1	89.2	33.3	99.2	1.000

Theorem 30 predicts the leading change near the task optimum, not the ordering at every tolerance: fixed overlap has higher χ_R than MC but greater intervention at the reported operating points. Scaling all process noise standard deviations from 0.5 to 2 times their original values preserves lower MC intervention than Compression in all ten tested body and noise combinations.

B.2 LEARNED PUSH: SELECTIVITY ACROSS DECISIONS

Does this pattern survive learning from sampled trajectories? Across ten independent training seeds, MC and Compression incur task degradations of 0.16% and 0.17%, respectively, while

MC uses 3.6% less intervention. Table 1 reports lower overlap with the passive future and no resolved change in complementary information. The difference in total controller information is -0.121 $[-0.262, 0.027]$ nats.

Conditioning on previous commands reveals the same selective pattern: sequential overlap decreases by 0.0360 nats (MC minus Compression: -0.0360 $[-0.0475, -0.0263]$), while the complementary difference is -0.0147 $[-0.1260, 0.1104]$ nats. These are paired 95% bootstrap intervals across training seeds, with information summed over ten decisions. The sequential comparison illustrates Theorem 23. These learned policies remain much closer to the task optimum than the analytic comparison at 10% tolerance.

B.3 Ball in Cup: selective compression in nonlinear mechanics

Environment and policy construction. The DMC `ball_in_cup/catch` task has horizontal and vertical cup sliders and an unactuated ball connected by a spatial tendon. Moving the cup sets the ball in motion. Subsequent commands can exploit or disrupt that motion. Actions lie in $[-1, 1]^2$. We retain the environment's default physical parameters: motor gear 5, slider damping 3, stiffness 20, and a tendon limit of 0.3 m on the distance between ball and cup. Control steps are 0.02 s and episodes last 20 s. Initial ball positions are sampled without collisions, with horizontal coordinate in $[-0.2, 0.2]$ and vertical coordinate in $[0.2, 0.5]$. Reward is one when the ball lies inside the target and zero otherwise. Sparse return is the undiscounted episode sum, with no action penalty.

Starting from a controller that already catches the ball, we use the same residual policy class and search procedure to minimize Gaussian overlap (BODY), total controller information (Compression), or squared action (Direct). The information penalties use the same reference states and passive futures, collected using the starting controller. Policy search uses one random seed. We use separate trial episodes to choose how much of each learned change to apply, then fix these choices and test the resulting controllers on 60 held-out episodes.

What the passive future retains. The fine timescale saturation in Theorem 38 motivates looking beyond very short horizons. At eight control steps (0.16 s), passive rollouts reveal clear differences in predictability across state directions. In the Gaussian model, these differences are described by the body gate reliabilities (Theorem 14). We fix this horizon before regularization. Reducing horizontal commands by 2% preserves more reward than reducing vertical commands by the same percentage: the return difference is 136.4 $[22.4, 246.8]$. Horizontal commands also have greater passive overlap (0.6403 versus 0.4620 nats). Thus the direction with greater overlap is also more tolerant of reduced actuation in this test.

Information and physical outcomes. The controller produces a mean command $U = \mu_\theta(S)$ from each state. For the information calculation, we add independent Gaussian noise of standard deviation 0.1, giving $\widetilde{A} = U + \epsilon$, and fit a Gaussian model to (U, Y_ℓ). This gives finite measures of total information, overlap with the passive future, and complementary information through the chain $\widetilde{A} - U - Y_\ell$, as in (5). Table 3 reports these Gaussian approximations using the state/action notation introduced above. During task evaluation, the controller executes its original commands; the added noise is used only to calculate information.

Table 3: **Comparing controllers on the same states reveals selective compression.** The first three columns evaluate all controllers on the same reference states and passive futures. Complement is total minus overlap, following (5). BODY reduces overlap more than Compression while retaining more total information. The on-policy columns use the states visited by each controller and show that the overlap ordering can reverse when the state ensemble changes.

	Common reference			On-policy	
Method	$I(S;A)$	$I(A;Y_\ell)$	Complement	$I(S;A)$	$I(A;Y_\ell)$
Task only	3.5701	1.0582	2.5119	3.5697	1.1220
BODY	3.5678	1.0373	2.5304	3.5898	1.2238
Compression	3.5583	1.0489	2.5094	3.5556	1.1695
Direct	3.5661	1.0554	2.5108	3.5250	1.1108

BODY minus Compression gives overlap -0.011514 $[-0.012212, -0.010710]$ and complement 0.020979 $[0.020343, 0.021577]$. We compute these intervals by resampling the same 4,000 reference pairs while keeping the controllers fixed. Since 2,500 pairs were also used during policy search, this is not a test on entirely new states. The intervals measure uncertainty from the reference sample, not variability across independently trained controllers. The reversal under on-policy evaluation illustrates the ensemble dependence distinguished by Proposition 12.

We assess physical intervention using two diagnostics: squared action $\sum_t \|A_t\|^2$ and the change each command produces in the next state compared with zero action. Let $S'_{t,a}$ and $S'_{t,0}$ denote these two successors from the same current state. For each episode, define

$$D_{\text{trans}} \doteq \tfrac{1}{2} \sum_t (S'_{t,a} - S'_{t,0})^\top \Sigma_{\text{ref}}^{-1} (S'_{t,a} - S'_{t,0}),$$

where Σ_{ref} contains the reference variances of the state coordinates on its diagonal. Thus each coordinate difference is divided by its reference standard deviation before squaring and summing. All controllers use the same normalization, and Table 4 reports averages across episodes. This is a measure of displacement from passive dynamics, not transition KL, since no process noise covariance is specified for the deterministic simulator.

Table 4: **Physical outcomes of the information objectives.** Means over 60 held-out episodes with the same environment seeds for all controllers. Success is the fraction of episodes with a catch; $\sum_t \|A_t\|^2$ measures cumulative squared action. Direct explicitly penalizes squared action and achieves both higher average return and lower average squared action in this evaluation.

Method	Sparse return	Success	$\sum_t \|A_t\|^2$	D_{trans}
Task only	564.6	80.0%	659.0	130.86
BODY	594.3	86.7%	661.2	132.26
Compression	580.8	86.7%	643.4	126.86
Direct	697.9	96.7%	609.1	116.00

In Table 4, paired BODY minus Compression differences are return 13.4 $[-101.1, 130.0]$, squared action 17.73 $[-23.14, 57.66]$, and displacement 5.39 $[-6.91, 17.32]$. We compute these intervals by resampling matched evaluation episodes, keeping each pair of runs with the same environment seed together. The controllers remain unchanged, so the intervals describe variation across episodes rather than across training runs. This comparison establishes no physical advantage of BODY over Compression, consistent with the distinction between controller information and intervention in Theorem 27.

B.4 Separating the effects of controller encoding and state visitation

Stronger feedback changes both what the command reveals about the state and which states the system visits. To separate these effects, we use a scalar linear Gaussian system:

$$S_{t+1} = 0.9S_t + 0.4A_t + N_t, \qquad A_t = KS_t + \epsilon_t,$$

where $N_t \sim \mathcal{N}(0, 0.25^2)$ and $\epsilon_t \sim \mathcal{N}(0, 0.2^2)$ are independent across time and of each other. We vary K from 0 to -2 and calculate each controller's stationary state distribution analytically. The reference state distribution is that induced by $K = 0$.

Figure 4 compares the three quantities in Proposition 12. Behavioral MC uses each controller's state distribution and commands. The occupancy-matched probe keeps that state distribution but replaces the commands by independent Gaussian actions with standard deviation 0.2. Fully standardized MC uses these independent actions and the fixed reference state distribution. The probes change the action used in the information calculation; they do not generate a new stationary state distribution.

As feedback becomes stronger, behavioral MC falls from 0.8702 to 0.0537. At $K = -2$, the occupancy-matched probe instead gives 0.3214, while fully standardized MC remains 0.8702. The gap between behavioral MC and the occupancy-matched probe isolates the effect of action conditioning. The gap between the probe and fully standardized MC isolates the effect of changing the state distribution. Thus a change in behavioral MC need not reflect a change in the body's dynamics.

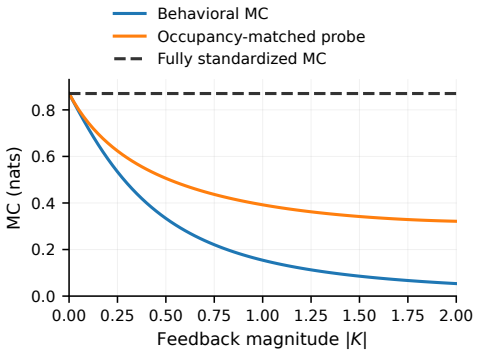

Figure 4: **Separating encoding from visitation in a Gaussian model.** Behavioral MC changes with both the controller's commands and the states it visits. Independent action probes remove the effect of state information encoded in the commands. Fixing the state distribution as well removes the remaining policy dependence. The dynamics are unchanged, and all three curves are calculated analytically.

B.5 Pendulum: nonlinear geometry and observation horizon

A freely moving pendulum separates immediate persistence from motion over a longer physical interval. With angle ϑ_n, angular velocity ω_n, and no applied torque, we simulate

$$\omega_{n+1} = \omega_n + \Delta t \tfrac{3g_{\text{grav}}}{2} \sin \vartheta_n + 0.02\xi_n, \qquad \xi_n \sim \mathcal{N}(0,1),$$
$$\vartheta_{n+1} = \text{wrap}(\vartheta_n + \Delta t \omega_{n+1}), \qquad \Delta t = 0.05\,\text{s}.$$

Here g_{grav} is gravitational acceleration and angles are wrapped to their periodic interval. Both canonical modes nearly saturate at one step, but the weaker squared correlation falls to 0.292 at 0.50 s (Figure 2c). Horizon summaries use interquartile means over 20 simulation seeds with 95% bootstrap intervals. This motivates looking beyond fine timescales, as in Theorem 38.

Figure 5 uses a separate pendulum variant to isolate approximation error. We replace $\sin\vartheta$ by $(1-\kappa)\vartheta + \kappa\sin\vartheta$, add velocity damping -0.05ω inside the acceleration term and independent angle noise of standard deviation 0.002, and leave angles unwrapped. Varying κ preserves the derivative at the origin. Over eight steps, narrow angle ensembles remain accurately described locally; wider ensembles and greater nonlinearity increase error, as anticipated by Proposition 37. A separate nearest-neighbor information check for the horizon study retains greater overlap in the stronger mode at 0.50 s (Kraskov et al., 2004).

B.6 MountainCar: limits of informational and geometric proxies

A car must inject energy to climb out of a valley. This makes MountainCar a useful contrast to coasting: simpler feedback need not mean less intervention (Theorem 27).

The controller observes the car's position and velocity and chooses how strongly to push. We search over simple controllers that use these two observations. Small Gaussian noise is added to the motion, making the next state slightly random. The task cost penalizes failure to reach $x \geq 0.45$ and rewards reaching it sooner. It contains no action penalty, so task performance and intervention are assessed separately. Controller information is estimated from sampled states and actions using the nearest-neighbor estimator of Kraskov et al. (2004).

For intervention, we report $\mathcal{T}_{\text{G}} \doteq 12.5\,\mathbb{E}[\sum_t A_t^2]$. This equals cumulative transition KL for the Gaussian dynamics without clipping; we use it as an intervention proxy for the clipped simulator.

Table 5 compares a baseline controller with candidates selected for lower information or lower intervention. The controller selected for lower information intervenes more than Baseline, whereas Direct intervenes less at a similar task cost. The task costs differ, so this comparison does not hold task performance fixed.

Table 5: **Less information can accompany more intervention in MountainCar.** All policies reach the goal in every evaluation episode. Task cost and intervention are averaged over 20 independent batches of 192 episodes; controller information is estimated from five separate datasets. Brackets give 95% bootstrap intervals across batches or datasets, respectively.

Policy	Task cost	Intervention proxy $\mathcal{T}_{\mathrm{G}}$	Controller information $I(S;A)$
Baseline	0.01899	827.2 $[826.4, 828.1]$	1.310 $[1.299, 1.320]$
Low information	0.02060	930.1 $[929.4, 930.9]$	1.193 $[1.183, 1.203]$
Direct	0.01902	815.7 $[813.3, 817.9]$	1.353 $[1.342, 1.363]$

Passive rollouts also expose the limits of canonical correlation (Figure 6). The weaker squared canonical correlation falls, then rises beyond approximately 20 steps. This does not contradict Theorem 39: nonlinear motion can make dependence more nearly linear without increasing mutual information about the starting state.

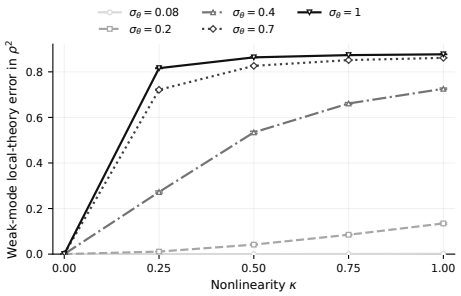

Figure 5: **Local Gaussian geometry in the nonlinear pendulum.** The absolute error in the predicted weaker squared canonical correlation grows with nonlinearity κ and initial angle spread σ_ϑ. Empirical means use 20 independently simulated datasets.

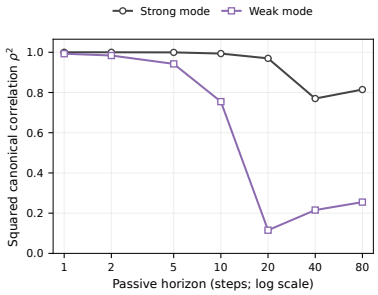

Figure 6: **Correlation can rise without information recovery.** Passive MountainCar squared canonical correlations on a fixed state ensemble, averaged over ten passive rollout repetitions with approximate 95% normal intervals. The later rise does not imply increased mutual information.