POLICIES ARE PROGRAMS: A FACTORIAL THEORY OF RESOURCE-COUPLED CONTROL

Anonymous authors
Paper under double-blind review

ABSTRACT

A deployed controller changes its environment through both its actions and the resources used to compute them. Reconstructing its value from separate action and execution evaluations assumes that these channels combine additively. We characterize the resulting error by an execution–actuation factorial contrast. An exact Bellman identity separates local interaction from cross-channel propagation: even an additive one-step kernel can have nonadditive long-horizon value. Discount- and mixing-dependent bounds have asymptotically sharp leading constants, attained by two-state constructions. A four-state lower bound further shows that arbitrarily many complete trajectories from the three separate evaluation arms can leave the sign of deployment benefit unidentified, even with additive kernels. An exact stationary queue law identifies when the additive model rejects a beneficial deployment, including families with vanishing execution loss and unbounded rejection regret. Reanalysis of held-out synthetic queue records finds a predicted gain of -176.573 versus an actual gain of $+73.292$ at the same operating point. A randomized-block CPU control independently exhibits interaction with action-release latency removed. Together, these results characterize the accuracy, information requirements, and decision consequences of additive deployment evaluation.

1 INTRODUCTION

Learned schedulers allocate resources while consuming resources themselves (Mao et al., 2019). When a controller shares a processor with its service, two implementations issuing the same command can leave different capacity to execute it. Execution therefore changes the policy's transition law, not just an accounting penalty.

Control–scheduling codesign (Xia and Sun, 2006) and colocated-workload research (Lo et al., 2015) establish that these costs matter. We ask a specific evaluation question: *can deployment value be reconstructed from a separately measured action benefit and execution effect?* The issue is whether the execution effect transfers across actuation regimes.

Fix an executable controller and consider two interventions. The execution factor C chooses reference versus coupled execution. The actuation factor U suppresses versus applies the controller's candidate commands. Let J_{cu} be the resulting closed-loop return. A three-cell additive prediction of deployment value is

$$\widetilde{J}_{11} = J_{01} + J_{10} - J_{00}, \qquad J_{11} - \widetilde{J}_{11} = \Gamma := J_{11} - J_{10} - J_{01} + J_{00}. \tag{1}$$

This is a standard unnormalized factorial interaction (Dasgupta et al., 2015). Its dynamical structure is the substantive problem: subtracting a scalar cost is valid only when the measured execution effect transports to the actuation regime being evaluated.

Our contribution is a decision-relevant theory of this transport assumption. **First**, the Bellman decomposition yields discount- and mixing-dependent error bounds whose leading constants are attained in the small-perturbation limit, even on two states (Theorem 2 and Proposition 3). **Second**, a four-state construction makes all three separate-arm trajectory laws identical in two worlds while deployment is beneficial in one and harmful in the other (Theorem 5). Thus local additivity does not supply the missing information. **Third**, an exact slack law identifies false deployment rejections and supports explicit policy-selection margins (Section 4).

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.

The experiments connect these claims to evaluation decisions. In the existing held-out queue records, the additive predictor can reverse both deployment-benefit signs and a pairwise controller ranking; all controllers at the selected load are reported. Fixed invocation-and-command tapes show that schedule feedback is not necessary for interaction. A fixed-window CPU experiment applies its command before measurement, removing online action-release latency as an explanation. These are synthetic mechanism tests, while the theoretical results apply to the specified effective-kernel classes. The underlying factorial contrast, action-abstraction principles, and resolvent differentiation are established tools; the contribution is their interaction-specific bounds, information obstruction, and decision consequences.

2 Executable policies and effective dynamics

Write an executable policy as $\kappa = (\pi, \chi)$. The decision semantics π map an interaction history h_t to a candidate-action distribution; the execution semantics χ specify observation, invocation, computation, and command release. Two programs are *action-equivalent* when their candidate-action laws agree at every specified common history. Their realized actions can differ after their environments diverge.

Let $Q_{\kappa,t}(\cdot \mid h)$ denote the effective conditional law of the next interaction record, including the action and next observation. If an execution footprint F is a causally sufficient interface, one possible factorization is

$$Q_{\kappa,t}(da, dx' \mid h) = \pi_t(da \mid h) \int P_t(dx' \mid h, a, f)\, \nu_{\chi,t}(df \mid h, a). \tag{2}$$

Sufficiency excludes additional implementation-dependent channels after conditioning. CPU time alone need not suffice; our analysis uses Q directly. Equality of effective kernels is exactly the condition for equality of every bounded next-record payoff at every starting history, and implies equality of all common finite-horizon rewards. Conditional footprint matching in (2) is sufficient; marginal matching is not. Appendix A gives the precise quotient, a tight finite-horizon coupling bound, and a basic coverage example, using standard abstraction and simulation-lemma principles (Ferns et al., 2004; Lobel and Parr, 2024). The main question is what happens when the execution and actuation channels *both* perturb these effective dynamics.

3 A Bellman calculus for execution–actuation interaction

Use a common finite augmented state space containing sufficient environment and controller memory. This is an analysis state and need not be observed by the learner. Let $P_{cu,t}$ and $r_{cu,t}$ be the closed-loop transition matrices and expected one-step rewards under each intervention. Assume a common initial law ρ and common terminal payoff g, with $g = 0$ unless specified. Arm-specific initialization can be included as an initial transition with zero reward. The Bellman equations are

$$V_{cu,t} = r_{cu,t} + P_{cu,t} V_{cu,t+1}, \qquad V_{cu,T} = g, \qquad J_{cu} = \rho V_{cu,0}.$$

Define $\Delta_t^C = V_{10,t} - V_{00,t}$, $\Delta_t^U = V_{01,t} - V_{00,t}$, and $W_t = V_{11,t} - V_{10,t} - V_{01,t} + V_{00,t}$. Also define the signed mixed kernel $K_t = P_{11,t} - P_{10,t} - P_{01,t} + P_{00,t}$ and $\delta r_t = r_{11,t} - r_{10,t} - r_{01,t} + r_{00,t}$.

Theorem 1 (Exact interaction recursion)**.** *The interaction-to-go satisfies $W_T = 0$ and*

$$W_t = P_{11,t} W_{t+1} + \ell_t, \tag{3}$$

$$\ell_t = \delta r_t + K_t V_{00,t+1} + (P_{11,t} - P_{10,t})\Delta_{t+1}^C + (P_{11,t} - P_{01,t})\Delta_{t+1}^U. \tag{4}$$

Consequently,

$$\Gamma = \mathbb{E}_{11,\rho}\left[\sum_{t=0}^{T-1} \ell_t(X_t)\right]. \tag{5}$$

The first two terms of (4) are immediate mixed effects. The remaining terms describe how a change in one channel acts on the future value change generated by the other. They remain even when $K_t = 0$ and $\delta r_t = 0$. The identity is an algebraic decomposition of a total intervention effect; its terms are not automatically separately identified physical mediators.

An additive transition law is insufficient. Take two states, initially state zero, with state one absorbing. Under (c, u), the probability of moving from zero to one is $\varepsilon(c+u)$, where $0 < \varepsilon \leq 1/2$. All intermediate rewards are zero and the common terminal payoff after two transitions is $\mathbf{1}\{X_2 = 1\}$. Then $K_t = 0$ for both transitions, but

$$J_{cu} = 1 - [1 - \varepsilon(c+u)]^2, \qquad \Gamma = -2\varepsilon^2.$$

Nonadditivity here is entirely generated by temporal propagation. No learning instability, delayed command, or mixed local parameter is required.

3.1 Discount, mixing, and the size of the error

For homogeneous kernels, discount $\gamma \in [0, 1)$, and a common current-state reward $r(x) \in [0, 1]$, define

$$\varepsilon_C = \max_u \sup_x \mathrm{TV}(P_{1u}(x, \cdot), P_{0u}(x, \cdot)),$$
$$\varepsilon_U = \max_c \sup_x \mathrm{TV}(P_{c1}(x, \cdot), P_{c0}(x, \cdot)), \qquad k = \tfrac{1}{2} \max_x \sum_y |K(x, y)|.$$

Let $\alpha \in [0, 1]$ upper-bound the Dobrushin coefficients $\sup_{x,z} \mathrm{TV}(P_{cu}(x, \cdot), P_{cu}(z, \cdot))$ of the four kernels. The choice $\alpha = 1$ is always valid.

Theorem 2 (Propagation bound)**.** *With $W = V_{11} - V_{10} - V_{01} + V_{00}$,*

$$\|W\|_\infty \leq \frac{\gamma k}{(1-\gamma)(1-\gamma\alpha)} + \frac{4\gamma^2 \varepsilon_C \varepsilon_U}{(1-\gamma)(1-\gamma\alpha)^2}. \tag{6}$$

In particular, setting $\alpha = 1$ gives $\gamma k/(1-\gamma)^2 + 4\gamma^2 \varepsilon_C \varepsilon_U/(1-\gamma)^3$. If $\alpha < 1$, the average-reward interaction obeys

$$|\overline{\Gamma}| \leq \frac{k}{1-\alpha} + \frac{4\varepsilon_C \varepsilon_U}{(1-\alpha)^2}. \tag{7}$$

When $k = 0$, interaction is second order in the channel perturbations, amplified by the squared memory factor $(1-\gamma\alpha)^{-2}$. Both leading constants below are asymptotically necessary. The bound assumes a common bounded current-state reward; arm-specific rewards remain covered by Theorem 1, not automatically by (6). The unbounded stationary queue in Section 4 is analyzed separately.

Proposition 3 (Asymptotically sharp constants)**.** *For $0 < \gamma < 1$ and $0 < \alpha \leq 1$, the coefficients 1 and 4 in (6) are individually unimprovable in the small-perturbation limit. With common reward $(0, 1)$ and initial state zero, a pure mixed perturbation $k = \eta$ has $\Gamma \sim \gamma\eta/[(1-\gamma)(1-\gamma\alpha)]$. A separate additive-kernel family has $k = 0$, $\varepsilon_C = x$, $\varepsilon_U = y$, and*

$$\Gamma \sim -\frac{4\gamma^2 xy}{(1-\gamma)(1-\gamma\alpha)^2} \quad \text{as } \frac{\gamma(x+y)}{1-\gamma\alpha} \to 0. \tag{8}$$

The same coefficients are asymptotically necessary in (7) for $0 < \alpha < 1$.

For the second claim, set $\beta = 1 - \alpha$ and use

$$P(t) = \begin{pmatrix} 1-t & t \\ \beta + t & 1-\beta-t \end{pmatrix}, \qquad t = cx + uy, \quad 0 < x + y \leq \alpha.$$

Its Dobrushin coefficient is $|\alpha - 2t| \leq \alpha$. Unlike a one-directional perturbation, both rows change, attaining the full coefficient 4. Appendix A.5 derives the exact rational values. This establishes leading-constant sharpness, not pointwise equality for finite perturbations.

3.2 A mixed-resolvent representation

For an affine interpolation $P(c, u) = P_{00} + cA + uB$ that is stochastic on $[0, 1]^2$, let $R(c, u) = (I - \gamma P(c, u))^{-1}$ and $V(c, u) = R(c, u)r$. Differentiating the inverse gives the following exact refinement when the local mixed kernel vanishes.

Corollary 4 (Ordered propagation)**.**

$$W = \int_0^1 \int_0^1 \gamma^2 R(c,u)\big[AR(c,u)B + BR(c,u)A\big]V(c,u)\,dc\,du. \tag{9}$$

The two orders of composition need not agree. This explains why assigning each channel a context-independent value correction loses information even in an affine transition family. It also distinguishes the claim from a statement about the immediate runtime of a model.

3.3 What three separate evaluation arms cannot identify

A nonzero interaction could in principle be reconstructed from richer data than three scalar values. The next result shows a stronger obstruction: even full separate-arm trajectories can miss the cross-channel transition needed for deployment.

Theorem 5 (Three-arm information limit)**.** *Fix $0 < \gamma < 1$, $0 < \alpha \le 1$, and $0 < x, y \le \alpha$. There exist two four-state worlds, with the same initial state and common reward in $[0,1]$, satisfying $K = 0$, $\varepsilon_C = x$, $\varepsilon_U = y$, and Dobrushin bound α, such that the complete state–reward trajectory law in each arm $00, 01, 10$ is identical between worlds. Nevertheless,*

$$J_{00} = J_{01} = J_{10} = j_0 := \frac{1}{2(1-\gamma)}, \qquad J_{11}^{\pm} = j_0 \pm h,$$
$$h = \frac{\gamma^2 xy}{2(1-\gamma)(z+\gamma x)(z+\gamma y)}, \qquad z = 1 - \gamma\alpha. \tag{10}$$

For any amount of fixed-arm trajectory data from $00, 01, 10$, the minimax expected absolute error of estimating J_{11} over this pair is exactly h. The minimax expected regret of choosing between baseline and deployment is $h/2$, attained by choosing each with probability $1/2$.

Construction. Start at state 0, give states $0, m, +, -$ rewards $(1/2, 1/2, 1, 0)$, and use the baseline kernel $\alpha I + (1-\alpha)\mathbf{1}e_0^{\mathsf{T}}$. Execution moves probability x from $0 \to 0$ to $0 \to m$; actuation moves probability y from $m \to m$ to $m \to +$ or $m \to -$. Only their composition reaches a reward-distinguishing state. All cells retain the same reset probability, proving the mixing bound; solving three Bellman equations gives (10). Appendix A.7 proves the claims.

For small x, y, $h \sim \gamma^2 xy/[2(1-\gamma)(1-\gamma\alpha)^2]$, the same discount/mixing order as the propagation upper bound. More separate-arm episodes cannot close this gap. The obstruction concerns trajectories from the specified initial state, not arbitrary-state transition queries: knowledge of all three kernels would identify $P_{11} = P_{10} + P_{01} - P_{00}$ when $K = 0$.

4 Slack amplification and deployment decisions

An analytically tractable queue makes the propagation issue quantitative. Consider an $M/M/1$ queue with arrival rate $\lambda > 0$ and service rate $\mu + ub - cd$, where $b, d > 0$. Write baseline slack as $s = \mu - \lambda$ and assume $s > d$, so all four cells have stationary distributions. With utility equal to negative stationary mean number in the system,

$$J_{cu} = -\frac{\lambda}{s + ub - cd}. \tag{11}$$

This is a separate continuous-time, unbounded-state example, not a claimed stationary model of our finite-horizon fluid simulation.

Theorem 6 (Exact slack law)**.** *The additive-prediction error is*

$$\Gamma = \frac{\lambda bd(2s + b - d)}{s(s-d)(s+b)(s+b-d)} \tag{12}$$
$$= \int_0^b \int_0^d \frac{2\lambda}{(s+v-w)^3}\,dw\,dv > 0. \tag{13}$$

It obeys $2\lambda bd/(s+b)^3 \le \Gamma \le 2\lambda bd/(s-d)^3$. For fixed positive λ, b, d, as $s \downarrow d$, $\Gamma \sim \lambda/(s-d)$.

Table 1: Exact $M/M/1$ policy-selection witness. Higher returns are better. All four cells are stable for both policies; the three-cell predictor ranks them incorrectly.

Policy	b	d	J_{01}	J_{10}	$\widetilde{J}_{11}$	J_{11}	Γ
L	0.02	0.01	-3.636	-4.211	-3.847	-3.810	0.037
H	0.20	0.15	-2.000	-16.000	-14.000	-3.200	10.800

Here the action restores service capacity, so an execution effect measured at $U = 0$ overstates its harm at $U = 1$. The decision consequence can be characterized exactly, rather than inferred from $\Gamma > 0$ alone.

Corollary 7 (Exact false-rejection region)**.** *Under Theorem 6, the true and additive-predicted gains over baseline are*

$$G := J_{11} - J_{00} = \frac{\lambda(b-d)}{s(s+b-d)}, \qquad \widetilde{G} := \widetilde{J}_{11} - J_{00} = \frac{\lambda[s(b-d) - 2bd]}{s(s+b)(s-d)}. \tag{14}$$

Thus $\widetilde{G} < 0 < G$ holds exactly when

$$b > d, \qquad d < s < \frac{2bd}{b-d}. \tag{15}$$

For any fixed $\lambda > 0$, taking $b = 2d$ and $s = 3d/2$ gives $G = 4\lambda/(15d)$ and $\widetilde{G} = -20\lambda/(21d)$. As $d \downarrow 0$, both capacity perturbations vanish while all cells remain stable and the regret of rejecting deployment diverges.

The divergence uses an unbounded queue reward and shrinking slack; it is not a fixed-discount, bounded-reward claim. In this queue family the error is conservative: $\Gamma > 0$ rules out a predicted positive gain when the true gain is negative. General finite-state interaction can have either sign, as Theorem 5 demonstrates.

Decision margins and policy selection. Write $a_i = \widetilde{J}^i_{11}$ and suppose $|\Gamma_i| \le B_i$. Deployment against a known baseline j_0 is sign-certified if $|a_i - j_0| > B_i$; a pairwise ranking is certified if $|a_i - a_j| > B_i + B_j$. For a finite class, let $\hat{\imath} \in \arg\max_i a_i$ and $i^* \in \arg\max_i J^i_{11}$. Then

$$J^i_{11} - J^j_{11} = a_i - a_j + \Gamma_i - \Gamma_j, \qquad J^{i^*}_{11} - J^{\hat{\imath}}_{11} \le B_{i^*} + B_{\hat{\imath}}. \tag{16}$$

If simultaneous return-estimation errors satisfy $|\hat{a}_i - a_i| \le e_i$, use $B_i + e_i$ instead. These certificates require justified model bounds, not CPU-dose measurements alone. Theorem 5 shows why sampling only the three arms cannot uniformly eliminate the model uncertainty.

Table 1 gives an exact stable example with $\lambda = .8$, $\mu = 1$, and $J_{00} = -4$. Even a *policy-specific* three-cell correction chooses L, although coupled deployment favors H by $64/105 \approx .610$. Comparing J_{11} directly to the additive predictor, rather than merely to an off-core ranking, is essential.

5 Identifying and estimating the deployment comparison

The four cells must denote reproducible interventions, not post-hoc partitions by measured CPU time. In a simulator, $C = 0$ removes the footprint-to-plant channel while retaining the controller's computation and memory updates; $U = 0$ suppresses the candidate action without suppressing invocation. On hardware, $C = 0$ is a specified other-core placement and $C = 1$ a specified same-core placement. Its estimand is a placement contrast, not an ideal free-computation contrast. Residual shared-cache, memory, power, and machine-load effects remain possible.

Feedback is part of the treatment response. The assignments C and U can be independent while realized invocation rates, actions, and CPU doses are dependent. If coupling creates a backlog that activates a feedback controller, its activation contributes to $J_{11} - J_{10}$. This does not invalidate the factorial experiment. Requiring equal action frequencies across C would instead remove part of the total deployment effect. A frozen command-and-invocation tape defines a useful but different open-loop estimand. We report both as distinct estimands; their difference is not an identified natural indirect effect.

Null regimes and exclusion. If changing U leaves the full conditional transition and reward laws unchanged at each C, then $J_{c1} = J_{c0}$ and $\Gamma = 0$. The analogous statement holds for C. A zero average measured footprint is not a substitute for this exclusion assumption. Conversely, positive doses do not force a nonzero interaction: effects can cancel. A null result can establish a small interaction at a tested operating point when its interval is narrow, but it does not certify additivity at other loads or for other implementations.

For diagnostics, we record per-cell invocation counts, applied nonzero-action counts N_A^{cu}, applied boost dose D_U^{cu}, and delivered capacity-loss dose D_C^{cu}. Counts and doses have different units. Quantile calibration uses an independent baseline stream; a τ-quantile is a cutoff, not a promise of a τ firing rate. Atoms at zero, ties, and closed-loop state changes can strongly alter that rate.

Paired estimator. For block i, let $Y_{cu}^{(i)}$ be the complete episode return and $H_i = Y_{11}^{(i)} - Y_{10}^{(i)} - Y_{01}^{(i)} + Y_{00}^{(i)}$. We use

$$\widehat{\Gamma} = \frac{1}{n}\sum_i H_i, \qquad \widehat{\mathrm{SE}} = \frac{s_H}{\sqrt{n}}, \qquad \mathrm{CI}_{.95} = \widehat{\Gamma} \pm t_{.975,n-1}\widehat{\mathrm{SE}}. \tag{17}$$

Shared simulation arrivals give exact common random numbers. Hardware arms share a pre-generated arrival trace and are executed in a seeded random order within each block. Pairing does not cancel arbitrary temporal machine drift; the intervals are approximate block-level sampling intervals. We additionally report block-bootstrap intervals for hardware and multiplicity-adjusted p-values for the 12 primary simulation configurations. The decision reanalysis is post hoc, with a separate 48-comparison Holm family detailed in Appendix B.1.

6 Experiments

We test whether execution interaction changes deployment decisions, whether it survives fixed invocation-and-command schedules, and whether it appears physically when online action-release latency is removed. Tables are generated from saved episode or event-level records. The decision contrasts reanalyze the existing queue runs; historical measurements remain separate in Appendix D.

6.1 Held-out stochastic queue experiments

The discrete-time plant starts at $Q_0 = 0$ and evolves for $T = 1000$ steps:

$$A_t = \max(0, \lambda + Z_t), \quad Z_t \sim \mathrm{Uniform}[-.1, .1],$$
$$Q_{t+1} = \big[Q_t + A_t - \max(0, 1 + ua_t - cf_t)\big]_+, \qquad r_t = -Q_{t+1} - .02ua_t. \tag{18}$$

On invocation, the candidate boost is $a_t = .1\mathbf{1}\{Q_t > 0\}$ and capacity loss is $f_t = .0075$; otherwise both are zero. The latter corresponds to the supplied simulator's CPU coordinate .005 plus cache coordinate .01 weighted by .25. These are model coordinates, not measured hardware seconds. Gates are period five, absolute cutoff one, or a .75 baseline-quantile cutoff. Calibration uses 40 seeds in $[100000, 100039]$; all evaluations use 200 disjoint seeds in $[200000, 200199]$.

Deployment decisions, not just large contrasts. At $\lambda = .99$, the calibrated controller has $\widehat{\Gamma} = 249.865$ with interval $[216.106, 283.624]$ (Table 2). Its baseline, additive-predicted, and actual values are -133.174, -309.747, and -59.882. Thus the predicted deployment gain is -176.573, whereas the actual gain is $+73.292$; both paired intervals exclude zero in opposite directions (Table 3). This is an empirical false rejection in the finite-horizon fluid model, not a fit to the stationary $M/M/1$ law.

For Quantile .75 minus Cutoff 1, the predicted difference is -175.858 with interval $[-205.978, -145.738]$, but the actual difference is $+71.718$ with interval $[67.460, 75.975]$. Both this pairwise reversal and the calibrated deployment-sign reversal survive Holm adjustment over the 48 reported decision contrasts. Period 5 remains best among all three under both evaluations: the result is a wrong *pairwise* ranking, not failure to select the best of the full grid. Appendix B.1 reports all loads and pairs.

The zero-dose and low-load cases remain informative. At $\lambda = .95$, Cutoff 1 never invokes on the held-out trajectories, whereas the calibrated gate has $\widehat{\Gamma} = 2.084$ with interval $[2.014, 2.154]$. At

Table 2: Held-out queue results, 200 paired seeds and 1000 steps per cell. $\widehat{\Gamma}$ is in total-return units; intervals are pointwise. N_A counts applied nonzero commands; D_C sums delivered capacity loss. A zero-dose cell is reported, not discarded.

λ	Gate	$\widehat{\Gamma}$	95% CI	N_A^{01}/N_A^{11}	D_C^{10}/D_C^{11}
0.70	Period 5	0.000	[0.000, 0.000]	0.0/0.0	1.50/1.50
0.70	Cutoff 1	0.000	[0.000, 0.000]	0.0/0.0	0.00/0.00
0.70	Quantile .75	0.000	[0.000, 0.000]	0.0/0.0	7.50/7.50
0.90	Period 5	0.000	[0.000, 0.000]	0.0/0.0	1.50/1.50
0.90	Cutoff 1	0.000	[0.000, 0.000]	0.0/0.0	0.00/0.00
0.90	Quantile .75	-0.059	[-0.061, -0.057]	0.0/35.8	7.50/7.50
0.95	Period 5	0.391	[0.376, 0.405]	59.3/59.4	1.50/1.50
0.95	Cutoff 1	0.000	[0.000, 0.000]	0.0/0.0	0.00/0.00
0.95	Quantile .75	2.084	[2.014, 2.154]	167.7/167.7	2.03/1.26
0.99	Period 5	23.724	[21.925, 25.523]	132.2/133.2	1.50/1.50
0.99	Cutoff 1	2.289	[0.065, 4.513]	0.2/0.2	0.03/0.00
0.99	Quantile .75	249.865	[216.106, 283.624]	40.3/42.3	3.85/0.32

Table 3: Deployment gains at $\lambda = .99$, relative to the same baseline. These are post-hoc contrasts of the same 200 seeds as Table 2; intervals are pointwise paired 95% intervals. All three tested gates are shown.

Gate	$\widetilde{G}$	95% CI	G	95% CI
Period 5	63.883	[60.611, 67.155]	87.607	[82.915, 92.299]
Cutoff 1	-0.715	[-1.903, 0.472]	1.574	[0.455, 2.693]
Quantile .75	-176.573	[-206.925, -146.221]	73.292	[68.486, 78.097]

$\lambda = .70$, even coupled service exceeds every possible arrival, making the queue exactly empty. At $\lambda = .90$, coupling activates the calibrated actuator (0 versus 35.775 mean applied commands), producing the small negative interaction -0.05917 once the action penalty is included. These are treatment responses, not reasons to discard cells.

Frozen tapes and footprint sensitivity. An additional grid uses seven loads and four per-invocation losses, with feedback and a fixed J_{00} invocation-and-command tape: 56 configurations beyond the 12 primary ones, totaling 54,400 raw cell episodes. The tape is generated from the baseline counterfactual in the simulator, not from an implementable oracle available to a deployed controller. Figure 1 shows that interaction survives even with matched command and footprint schedules, so endogenous invocation is not necessary for nonadditivity. All zero-loss controls have exactly zero interaction; the full grid, doses, and cutoffs are provided. Fixed-tape interaction does not isolate the $K = 0$ propagation term: the queue truncation can itself induce local mixing. The two-state examples establish the pure-propagation claim.

6.2 A FIXED-WINDOW PHYSICAL EXECUTION CONTROL

Two Python processes run on Linux in a shared CPU container. The service is pinned to logical CPU 0; the controller runs on CPU 0 or 1. The visible machine reports an AMD EPYC 9V74, distinct core identifiers, and a four-CPU cgroup quota. The synthetic controller emits a constant batch-size-eight command before measurement and then executes redundant CPU work, nominally 3 ms every 10 ms. $U = 0$ leaves batch size one; $U = 1$ applies the command before the common start time. There is therefore no online command-release latency to explain the measured interaction. The static command keeps action semantics fixed across placement.

Each service batch consumes .5 ms of CPU overhead plus .15 ms per actual item. Arrival times are Poisson and fixed in wall-clock time, rather than generated per completed service round. Measurement lasts .30 s; incomplete batches receive no completion credit. Utility is negative time-average number in the system, reconstructed exactly from the saved arrival and completion events. The three nonzero-work scenarios use 24 randomized blocks each; an idle-work control uses 12, for 336 trials in total.

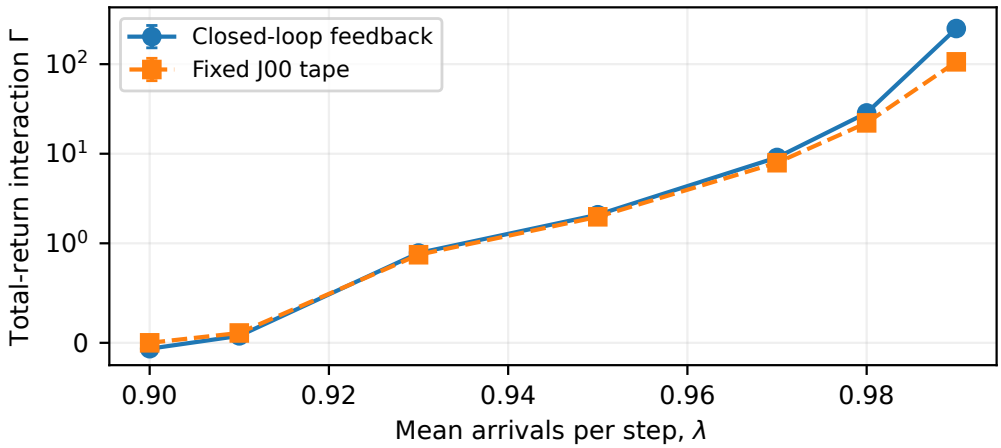

Figure 1: Calibrated queue gate with loss .0075 per invocation. Freezing both invocation and candidate commands removes schedule feedback but does not generally remove value interaction. Error bars are paired 95% intervals; the vertical axis is symmetric-logarithmic with a linear region near zero. The two curves are different intervention protocols, not an identified mediation decomposition.

Table 4: Physical execution control. Higher returns are better. Load is normalized by the specified batch-one CPU demand, not measured utilization. "Idle" removes synthetic busy work but retains the controller process and timer. Intervals are paired, pointwise 95% intervals.

Load	n	J_{00}	J_{01}	J_{10}	J_{11}	$\widehat{\Gamma}$	95% CI
0.60	24	-1.25	-0.89	-4.01	-1.38	2.26	[1.57, 2.94]
0.95	24	-8.96	-1.63	-64.77	-2.84	54.59	[50.64, 58.55]
1.40	24	-97.74	-2.78	-166.98	-5.45	66.57	[63.74, 69.40]
0.95 (idle)	12	-12.05	-1.76	-13.24	-1.65	1.29	[-1.11, 3.70]

At nominal load .95, the same-core placement effect is approximately -55.807 when actions are suppressed but only -1.212 when batching is applied, giving $\widehat{\Gamma} = 54.595$ (Table 4). The controller's mean CPU dose is approximately .091–.093 s across the nonzero-work arms, about 30% of the measurement window. Thus this experiment tests nonadditivity without command-release latency; the small-footprint claim is supported separately by theory and simulation. The positive interaction is consistent with batching reducing sensitivity to lost processor capacity; it does not establish a cache, bandwidth, or power mediator. The idle-work interval $[-1.113, 3.698]$ includes zero, rather than proving an exact null.

An event-trace audit finds no invalid or post-deadline credited completion; independently reconstructed queue values agree within 5.7×10^{-14}. The fixed deadline controls admission, integration, and completion credit despite scheduler lateness. One interrupted and resumed block is retained; excluding it leaves the low-load interaction positive. Clock diagnostics, bootstrap intervals, and the interruption record are in Appendix C.

7 Related Work

Control and shared resources. Control–scheduling codesign integrates controller performance with computing-resource allocation (Xia and Sun, 2006), while Heracles studies colocated workload interference and coordinated isolation (Lo et al., 2015). Learned schedulers such as Decima motivate deployment-sensitive evaluation (Mao et al., 2019). Our question concerns transporting an execution effect across actuation regimes: the same implementation can have different conditional value effects even when its candidate commands are fixed.

Performative learning. Performative prediction studies decision-induced distribution shift (Perdomo et al., 2020). Performative reinforcement learning allows policies to influence rewards and transition dynamics (Mandal et al., 2023); policy-gradient approaches address performative optimality (Basu et al., 2026). Execution coupling is an implementation-dependent source of such changes, including between action-equivalent programs. We study evaluation error and identifiability rather than policy-gradient convergence.

Perturbation, abstraction, and factorial inference. Markov perturbation analysis provides stationary-distribution sensitivity bounds and series expansions (Abbas et al., 2016). Bisimulation metrics and simulation lemmas relate transition/reward discrepancies to value (Ferns et al., 2004; Lobel and Parr, 2024); factorial causal inference supplies the four-cell contrast (Dasgupta et al., 2015). Building on these tools, we characterize the *mixed* long-horizon error with sharp leading constants and a separate-arm information obstruction of matching parameter order. The queue law then turns the error into an exact deployment-decision boundary. The elementary quotient and one-step support result in Appendix A provide foundations rather than the main novelty.

8 Limitations and conclusion

The finite-state certificates require kernel discrepancies and mixing bounds that ordinary logs may not identify. The three-arm lower bound assumes fixed-arm trajectories from a specified initial state; targeted transition queries can defeat that obstruction. The stationary queue assumes exponential interarrival and service times, all-cell stability, and an unbounded queue reward. Experiments use synthetic controllers and one shared, short-window CPU environment, not trained neural policies or production workloads. Physical mediators are not separately identified, and pairing does not remove arbitrary machine drift.

The resulting characterization is concrete: local additivity does not imply value additivity; the mixed error has asymptotically sharp propagation constants; and even unlimited separate-arm trajectories can leave its decision-relevant sign unidentified. Near a queue's slack boundary, the error can reject a beneficial deployment despite vanishing resource perturbations. Evaluation should therefore justify cross-regime coverage or measure the coupled cell, compare error bounds to decision margins, and keep closed-loop and fixed-tape estimands distinct.

AI use statement

A large language model assisted with developing and revising theoretical statements and proofs, designing and implementing numerical checks and experiments, analyzing recorded outputs, and drafting and editing the manuscript. Synthetic observations were generated by the executable simulators and measurement programs described here, not invented as language-model text. Historical source data were preserved separately from new measurements. Automated checks complement rather than replace author review; responsibility for the correctness, originality, and final submission rests with the authors.

Reproducibility statement

The supplement contains all new raw episode and hardware event records, seed ranges, configurations, scripts for regenerating every table and figure, numerical falsification checks, and complete proofs. Appendix E gives the entry points. Historical source aggregates are explicitly labeled and are not presented as independently rerun hardware trials. Separate data directories distinguish historical observations from the new experiments.

Ethics statement

The experiments use synthetic arrivals and CPU workloads, with no human subjects or personal training data. Resource-aware control can affect the availability and fairness of shared infrastructure; the microbenchmarks are not safety or fairness certifications. Measurements should be repeated on authorized resources before drawing deployment decisions in a different environment.

REFERENCES

Karim Abbas, Joost Berkhout, and Bernd Heidergott. A critical account of perturbation analysis of Markov chains. *Markov Processes and Related Fields*, 22:227–265, 2016. URL https://arxiv.org/abs/1609.04138.

Debabrota Basu, Udvas Das, Brahim Driss, and Uddalak Mukherjee. Performative policy gradient: Optimality in performative reinforcement learning. *arXiv preprint arXiv:2512.20576v2*, 2026. URL https://arxiv.org/abs/2512.20576v2.

Tirthankar Dasgupta, Natesh S. Pillai, and Donald B. Rubin. Causal inference from 2^k factorial designs by using potential outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(4):727–753, 2015. URL https://arxiv.org/abs/1211.2481.

Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite Markov decision processes. In *Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence*, pages 162–169, 2004. URL https://arxiv.org/abs/1207.4114.

David Lo, Liqun Cheng, Rama Govindaraju, Parthasarathy Ranganathan, and Christos Kozyrakis. Heracles: Improving resource efficiency at scale. In *Proceedings of the 42nd Annual International Symposium on Computer Architecture*, 2015. doi: 10.1145/2749469.2749475. URL https://research.google.com/pubs/archive/43792.pdf.

Sam Lobel and Ronald Parr. An optimal tightness bound for the simulation lemma. *arXiv preprint arXiv:2406.16249v2*, 2024. URL https://arxiv.org/abs/2406.16249v2.

Debmalya Mandal, Stelios Triantafyllou, and Goran Radanovic. Performative reinforcement learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 23642–23680, 2023. URL https://proceedings.mlr.press/v202/mandal23a.html.

Hongzi Mao, Malte Schwarzkopf, Shaileshh Bojja Venkatakrishnan, Zili Meng, and Mohammad Alizadeh. Learning scheduling algorithms for data processing clusters. In *Proceedings of the ACM Special Interest Group on Data Communication*, pages 270–288, 2019. doi: 10.1145/3341302.3342080. URL https://arxiv.org/abs/1810.01963.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7599–7609, 2020. URL https://proceedings.mlr.press/v119/perdomo20a.html.

Feng Xia and Youxian Sun. Control-scheduling codesign: A perspective on integrating control and computing. *Dynamics of Continuous, Discrete and Impulsive Systems, Series B*, 13(S1):1352–1358, 2006. URL https://arxiv.org/abs/0806.1385.

A COMPLETE PROOFS AND COUNTEREXAMPLES

The kernel arguments below apply on finite or standard Borel record spaces so that the conditional distributions and couplings used in the proofs exist. Matrix statements use the finite augmented state space specified in Section 3. The augmentation is an assumption about the analysis model, not a claim that ordinary action logs observe all relevant resource state.

A.1 EXACT AND APPROXIMATE EQUIVALENCE

Proposition 8 (Exact action quotient)**.** *For action-equivalent programs, equality of their effective conditional kernels at every specified starting history is equivalent to equality of the expected value of every bounded one-step payoff on the next interaction record at every such history. Equal kernels also imply equal finite-horizon values for every common bounded history-dependent reward. Unequal kernels imply that* some *payoff distinguishes the programs, not that an arbitrary fixed reward does.*

Proof of Proposition 8. Fix a specified history h and let Q and Q' be the two laws of the next record. If $Q = Q'$, then their integrals against every bounded measurable payoff agree. Conversely, indicator payoffs $g_B = \mathbf{1}_B$ recover the probabilities of all measurable sets B. Agreement for all such payoffs implies $Q(B) = Q'(B)$ for all B, which is equality of the probability measures.

If the conditional kernels agree at each history, induction gives equality of the complete finite-horizon record laws from a common initial history. Integrating any common bounded trajectory reward then gives equal values. Under the sufficient-interface assumption in (2), equality of π and of $\nu_\chi(\cdot \mid h, a)$ at each relevant pair implies equality of the integrated kernels. This is a sufficient condition, not a necessary one: different footprints can have identical plant effects.

The quantifier over payoffs cannot be removed. For instance, the two next-state distributions $(1, 0)$ and $(.8, .2)$ are distinct, and the common reward $(0, 1)$ distinguishes them. The common reward $(0, 0)$ does not. Nor can the quantifier over starting histories be replaced by a fixed initial law without restricting attention to its reachable histories: differences confined to unreachable histories do not change that deployment's value. □

Why a marginal match is insufficient. Let S be uniform on $\{0, 1\}$, let the command always be a no-op, and set the next physical state to $X' = \mathbf{1}\{F = S\}$. The primitive reward is the same function $r(X') = X'$ in both worlds. Under the first implementation, $F = S$ and $J = 1$. Under the second, $F = 1 - S$ and $J = 0$. The distributions of F are both uniform on $\{0, 1\}$, so even matching the full marginal footprint law fails; matching its mean also fails. The missing information is the coupling between the current state and the footprint. If unmeasured implementation-dependent channels remain after conditioning on F, even conditional matching of measured coordinates is insufficient.

Proposition 9 (Quantitative action abstraction)**.** *Suppose the common per-step payoffs after each transition lie in* $[0, 1]$*, the initial histories agree, and the conditional total variation distance between two effective kernels is at most* ε *at every common history. Over* T *transitions,*

$$|J_\kappa - J_{\kappa'}| \le B_T(\varepsilon) := \sum_{t=1}^{T} \big[1 - (1 - \varepsilon)^t\big] \le \frac{T(T+1)}{2}\varepsilon. \tag{19}$$

The first bound is attained by a two-state absorbing construction.

Proof of Proposition 9. Couple the two processes from the same initial history. Whenever the histories still agree, use a maximal coupling of their next-record kernels. By the assumed total variation bound, their next records disagree with probability at most ε, conditional on agreement so far. Once histories disagree, couple arbitrarily. The probability that histories remain identical through transition t is at least $(1 - \varepsilon)^t$. Therefore the absolute difference between expected common $[0, 1]$ payoffs evaluated after transition t is at most $1 - (1 - \varepsilon)^t$. Summing these bounds gives the first inequality. Bernoulli's inequality gives $1 - (1 - \varepsilon)^t \le t\varepsilon$, proving the second.

For tightness, take states 0 and 1, start at zero, and use the same no-op action in both programs. The reference stays at zero forever. The other program moves from zero to absorbing state one with probability ε at each transition. Its reward after transition t is the common function $r(X_t) = X_t$.

The row discrepancy at state zero is exactly ε; at state one the kernels agree. The expected reward difference at transition t is exactly $1-(1-\varepsilon)^t$, so their return difference equals $B_T(\varepsilon)$. This also verifies the endpoint cases $\varepsilon = 0$ and $\varepsilon = 1$ without an undefined closed-form division. □

Scaling and timing. For common payoffs in an interval of length R, multiply (19) by R. The first possible reward difference in Proposition 9 is after the first transition. A reward received before the first transition from a shared initial state would contribute zero initial difference. The distinction prevents an off-by-one horizon claim. The absorbing construction is a specialization of the overlap-based simulation-lemma argument, not an independent claim to its general novelty.

A.2 The exact coverage obstruction

Even logs containing actions, states, rewards, and footprints need not identify a new execution regime. Consider a one-step family indexed by $\theta \in \{0,1\}$ with a common no-op action, logged footprint $F \sim \mathrm{Bernoulli}(p)$, and reward $Y = \theta F$. The target always executes with $F = 1$, so $J_\theta = \theta$. From n independent logs,

$$\inf_{\widehat{J}} \max_{\theta\in\{0,1\}} \mathbb{E}_\theta|\widehat{J} - J_\theta| = \tfrac{1}{2}(1-p)^n. \tag{20}$$

This elementary two-world result is exact for the stated family; it is not a general sample-complexity characterization. At $p = 0$, unlimited action-only data cannot resolve the deployment comparison. Recording a resource coordinate does not, by itself, create support at its target value.

Proof of (20). Let E be the event that none of the n logged footprints equals one. Under both $\theta = 0$ and $\theta = 1$, $\Pr(E) = (1-p)^n$. The conditional law of all observations on E is identical in the two worlds. For any estimator, including one with private randomness, let Z denote its output on this shared conditional law. The triangle inequality gives

$$\mathbb{E}|Z| + \mathbb{E}|Z-1| \geq 1.$$

Consequently the sum of the two unconditional risks is at least $(1-p)^n$, and their maximum is at least half of that number.

This lower bound is attained by returning $1/2$ on E and otherwise returning the reward of any observed record with $F = 1$. In the latter case the reward identifies θ exactly. Both risks are thus $(1-p)^n/2$. The equality uses logs that include the footprint; removing that coordinate cannot make estimation easier. At $p = 0$, the complete logging laws agree for every sample size while the target values differ by one. More generally, two indistinguishable logging worlds with target-value separation Δ force worst-case expected absolute error at least $\Delta/2$. □

The result is a deliberately minimal information obstruction. It makes no claim that observing each footprint coordinate separately gives joint support or that coverage alone resolves hidden confounding. An identifiable deployment-value model needs both adequate support and valid assumptions relating execution, state, and subsequent outcomes.

A.3 Exact interaction recursion and occupancy expansion

Proof of Theorem 1. Suppress the time index on the kernels and attach a prime to next-time values. From the definitions,

$$V'_{10} = V'_{00} + \Delta^{C\prime}, \quad V'_{01} = V'_{00} + \Delta^{U\prime}, \quad V'_{11} = V'_{00} + \Delta^{C\prime} + \Delta^{U\prime} + W'.$$

Subtracting the four Bellman equations and substituting these identities gives

$$\begin{aligned} W &= \delta r + P_{11}(V'_{00} + \Delta^{C\prime} + \Delta^{U\prime} + W') - P_{10}(V'_{00} + \Delta^{C\prime}) \\ &\quad - P_{01}(V'_{00} + \Delta^{U\prime}) + P_{00}V'_{00} \\ &= \delta r + KV'_{00} + P_{11}W' + (P_{11} - P_{10})\Delta^{C\prime} + (P_{11} - P_{01})\Delta^{U\prime}, \end{aligned}$$

This is exactly (4). All four terminal values equal g, hence $W_T = 0$. Iterating $W_t = \ell_t + P_{11,t}W_{t+1}$ gives

$$W_0 = \sum_{t=0}^{T-1} P_{11,0}P_{11,1}\cdots P_{11,t-1}\ell_t,$$

with the empty product equal to the identity. Left multiplication by the common initial law ρ is the expectation in (5). □

Zero one-step interaction with nonzero terminal interaction. In the two-state example from the main text,

$$P_{cu} = \begin{pmatrix} 1-\varepsilon(c+u) & \varepsilon(c+u) \\ 0 & 1 \end{pmatrix}.$$

All entries are valid probabilities for $\varepsilon \le 1/2$, and direct subtraction gives $K = 0$. The chance of not yet reaching state one after two transitions is $(1-\varepsilon(c+u))^2$. Thus $J_{00} = 0$, $J_{10} = J_{01} = 2\varepsilon - \varepsilon^2$, and $J_{11} = 4\varepsilon - 4\varepsilon^2$. The interaction is $-2\varepsilon^2$. Since all intermediate rewards are zero and the terminal payoff is common, the example's interaction must come from the two propagation terms of Theorem 1.

A.4 Discounted and mixing-dependent bounds

For a vector v, write $\operatorname{span}(v) = \max_x v(x) - \min_x v(x)$. Two elementary inequalities will be used. If a signed row vector q has zero sum, then

$$|qv| \le \tfrac{1}{2} \|q\|_1 \operatorname{span}(v). \tag{21}$$

Indeed, subtract the midpoint of v's range before applying the ℓ_1 inequality. For a stochastic matrix P with Dobrushin coefficient at most α,

$$\operatorname{span}(Pv) \le \alpha \operatorname{span}(v). \tag{22}$$

To see this, apply (21) to $P(x,\cdot) - P(z,\cdot)$ and take the supremum over x, z.

Proof of Theorem 2. The homogeneous Bellman equations are $V_{cu} = r + \gamma P_{cu} V_{cu}$. Taking spans in the baseline equation and using $\operatorname{span}(r) \le 1$ gives

$$\operatorname{span}(V_{00}) \le \frac{1}{1-\gamma\alpha}. \tag{23}$$

Subtracting the 00 equation from the 10 equation gives

$$\Delta^C = \gamma P_{10} \Delta^C + \gamma (P_{10} - P_{00}) V_{00}.$$

Each component of $(P_{10} - P_{00})V_{00}$ is bounded in absolute value by $\varepsilon_C \operatorname{span}(V_{00})$. Its span is therefore at most twice that number. It follows that

$$\operatorname{span}(\Delta^C) \le \frac{2\gamma\varepsilon_C}{(1-\gamma\alpha)^2}, \qquad \operatorname{span}(\Delta^U) \le \frac{2\gamma\varepsilon_U}{(1-\gamma\alpha)^2}. \tag{24}$$

The stationary discounted version of the exact recursion is

$$(I - \gamma P_{11}) W = \gamma K V_{00} + \gamma (P_{11} - P_{10}) \Delta^C + \gamma (P_{11} - P_{01}) \Delta^U. \tag{25}$$

Since K has zero row sums, (21) bounds its term by $k \operatorname{span}(V_{00})$. The two other terms are bounded by $\varepsilon_U \operatorname{span}(\Delta^C)$ and $\varepsilon_C \operatorname{span}(\Delta^U)$, respectively. The nonnegative resolvent satisfies $\left\|(I - \gamma P_{11})^{-1}\right\|_{\infty\to\infty} = 1/(1-\gamma)$ by its Neumann series. Combining these inequalities proves (6). Setting $\alpha = 1$ gives the stated generic bound. Independently, r cancels at time zero and $0 \le V_{cu} \le 1/(1-\gamma)$, so the bound may also be capped by $2\gamma/(1-\gamma)$.

For the average-reward statement, $\alpha < 1$ implies a unique invariant distribution for each finite-state kernel and geometric convergence to it in total variation. Let $\overline{J}_{cu}$ be its expectation of r. The bounded geometric transient implies

$$(1-\gamma) V_{cu}(x) \longrightarrow \overline{J}_{cu} \quad \text{as } \gamma \uparrow 1$$

uniformly in x. Multiply (6) by $1-\gamma$ and take this limit to obtain (7). No unbounded queue reward is used in this step. □

What the bound does and does not certify. A useful sufficient regime is small k and small channel perturbations relative to the memory factor $1/(1-\gamma\alpha)$. The bound is not a data-driven certificate unless its kernel and mixing quantities are themselves justified. It can be loose, including for large perturbations. Numerical checks of randomly sampled systems below are falsification checks, not substitutes for the proof. Sharpness of the leading constants follows from the explicit constructions in Appendix A.5.

A.5 TWO-STATE WITNESSES ATTAINING BOTH LEADING CONSTANTS

Proof of Proposition 3. Fix $\gamma \in (0,1)$, $0 \le \beta < 1$, and $\alpha = 1-\beta$. Write $\delta = 1-\gamma$ and $z = \delta+\gamma\beta = 1-\gamma\alpha$. All values start at state zero and use the same current-state reward $r = (0,1)^{\mathsf{T}}$.

For the mixed-kernel term, let

$$P_{\mathrm{m}}(p) = \begin{pmatrix} 1-p & p \\ \beta & 1-\beta \end{pmatrix}, \qquad p_{cu} = cu\eta, \quad 0 < \eta \le \alpha.$$

The Dobrushin coefficient is $\alpha - p \le \alpha$. Its two Bellman equations give $V_p(0) = \gamma p/[\delta(z+\gamma p)]$. The first three values vanish, $k = \eta$, and

$$\Gamma_{\mathrm{m}} = \frac{\gamma\eta}{\delta(z+\gamma\eta)}.$$

As $\gamma\eta/z \to 0$, its ratio to $\gamma k/(\delta z)$ tends to one; the second term of the bound is lower order since $\varepsilon_C = \varepsilon_U = \eta$. Thus the coefficient 1 cannot be uniformly decreased while keeping a finite second-term coefficient.

For the propagation term, perturb both rows:

$$P(t) = \begin{pmatrix} 1-t & t \\ \beta+t & 1-\beta-t \end{pmatrix}, \qquad t_{cu} = cx+uy, \qquad x,y>0, \quad x+y \le \alpha.$$

All entries are nonnegative, all rows sum to one, and the Dobrushin coefficient is $|\alpha - 2t| \le \alpha$. Direct subtraction gives $K = 0$, $\varepsilon_C = x$, and $\varepsilon_U = y$. Subtracting the two Bellman equations and then using the state-zero equation yields

$$V_t(1) - V_t(0) = \frac{1}{z+2\gamma t}, \qquad V_t(0) = \frac{\gamma t}{\delta(z+2\gamma t)}.$$

Taking the four-cell difference gives

$$\Gamma = -\frac{4\gamma^2 xy[z+\gamma(x+y)]}{\delta(z+2\gamma x)(z+2\gamma y)(z+2\gamma(x+y))}. \tag{26}$$

Its ratio to $-4\gamma^2 xy/(\delta z^2)$ is

$$\frac{z^2[z+\gamma(x+y)]}{(z+2\gamma x)(z+2\gamma y)(z+2\gamma(x+y))} \longrightarrow 1 \quad \text{as } \gamma(x+y)/z \to 0.$$

Since $|\Gamma| \le \|W\|_\infty$, this initial-state witness forces the coefficient 4 in the vector bound. The admissible parameter range permits this limit for every fixed $\gamma \in (0,1)$ and $\alpha \in (0,1]$. For $\alpha = 1$, $z = 1-\gamma$, giving the necessary cubic discount dependence.

For $\beta > 0$, the invariant rewards are $p/(\beta+p)$ and $t/(\beta+2t)$ in the two families. Hence

$$\overline{\Gamma}_{\mathrm{m}} = \frac{\eta}{\beta+\eta}, \qquad \overline{\Gamma}_{\mathrm{p}} = -\frac{4xy(\beta+x+y)}{(\beta+2x)(\beta+2y)(\beta+2(x+y))}.$$

As perturbations vanish relative to $\beta = 1-\alpha$, these approach η/β and $-4xy/\beta^2$ to first nonzero order, establishing both average-reward coefficients. Replacing r by $1-r$ reverses either interaction without changing any kernels, so sharpness does not depend on a prescribed interaction sign. These are sharp leading constants, not finite-parameter equality claims. □

A.6 ORDERED PROPAGATION BY DIFFERENTIATING THE RESOLVENT

Proof of Corollary 4. The kernels in the affine family are stochastic throughout the square and $\gamma < 1$, so the inverse exists with a uniformly convergent Neumann series. Differentiating $(I-\gamma P)R = I$ gives

$$\partial_c R = \gamma RAR, \qquad \partial_u R = \gamma RBR.$$

For $V = Rr$ with fixed r, $\partial_c V = \gamma RAV$. A second derivative yields

$$\partial_u\partial_c V = \gamma^2 RBRAV + \gamma^2 RARBV = \gamma^2 R(BRA + ARB)V.$$

Integrating the mixed derivative over $[0,1]^2$ and applying the fundamental theorem of calculus twice yields $V(1,1) - V(0,1) - V(1,0) + V(0,0)$, exactly W. No commutativity of A, B, or the resolvent is assumed. When $K = 0$ at the four corners, $A = P_{10} - P_{00}$ and $B = P_{01} - P_{00}$ give such an interpolation: it is also the convex bilinear interpolation of the four stochastic corner matrices, hence is stochastic throughout the square. □

A.7 A MIXED-KERNEL-FREE THREE-ARM INFORMATION OBSTRUCTION

Proof of Theorem 5. Let the common state space be $\{0, m, +, -\}$, initial law e_0^T, and reward $r = (1/2, 1/2, 1, 0)^\mathsf{T}$. Put $\beta = 1 - \alpha$ and

$$P_0 = \alpha I + \beta \mathbf{1} e_0^\mathsf{T}, \qquad A = x e_0 (e_m - e_0)^\mathsf{T}, \qquad B_\pm = y e_m (e_\pm - e_m)^\mathsf{T}.$$

World $\theta \in \{+, -\}$ uses $P_{cu}^\theta = P_0 + cA + uB_\theta$. The perturbations change different rows, and $x, y \le \alpha$, so every kernel is stochastic. Each has a common component $\beta \mathbf{1} e_0^\mathsf{T}$ of mass β; its remaining component is α times a stochastic matrix. Thus every Dobrushin coefficient is at most α. Affinity gives $K = 0$, and the half-row-ℓ_1 norms of A and B_θ are x and y.

Under 00 and 01, the chain remains at zero forever, so the altered actuation row at m is never observed. Under 10, it remains in $\{0, m\}$ and its full kernel is the same in both worlds. Equality of the kernels on these recursively reachable sets proves equality of all three fixed-arm trajectory laws, for every finite horizon and for the infinite trajectories. Rewards on these sets are always $1/2$, so each of the three values is $j_0 = 1/[2(1-\gamma)]$. This equality holds jointly for any collection of independent episodes, and for adaptive choices of which separate arm to sample next when each episode restarts at zero. No switch to a different treatment within an episode is included in this data model.

Center the reward by $1/2$. In world $+$, only the reachable terminal state $+$ has nonzero centered reward $1/2$. Write the centered deployment values on $0, m, +$ as v_0, v_m, v_+, and set $\delta = 1 - \gamma$, $z = 1 - \gamma\alpha$. Their Bellman equations are

$$\begin{aligned} (\delta + \gamma x) v_0 &= \gamma x v_m, \\ (z + \gamma y) v_m &= \gamma\beta v_0 + \gamma y v_+, \\ z v_+ &= 1/2 + \gamma\beta v_0. \end{aligned}$$

Substituting the last two equations into the first gives

$$\left[(\delta + \gamma x)(z + \gamma y) - \frac{\gamma^2 \beta x (z + \gamma y)}{z} \right] v_0 = \frac{\gamma^2 x y}{2z}.$$

The bracket equals $\delta(z + \gamma x)(z + \gamma y)/z$, because $z - \gamma\beta = \delta$. Therefore $v_0 = h$ as in (10). In world $-$ the nonzero centered reward is $-1/2$, giving $v_0 = -h$ by linearity.

Let D be any data from the three fixed arms. Its law is identical in the two worlds, so any possibly randomized estimator has a common output law Z. The triangle inequality gives

$$\mathbb{E}|Z - (j_0 + h)| + \mathbb{E}|Z - (j_0 - h)| \ge 2h.$$

The maximum risk is at least h and is attained by the constant estimate j_0. For a deployment decision, let p be its probability of deploying under the common data law. Its regret is $(1-p)h$ in world $+$ and ph in world $-$. Their maximum is minimized at $p = 1/2$ with value $h/2$. A deterministic function of the exact three-arm laws makes the same fixed choice in both worlds and incurs worst-case regret h. A deterministic procedure using finite random data can use that data as a randomizer, so the h claim is deliberately restricted to exact-law deterministic decisions. □

Order and scope. For $\gamma(x + y)/z \to 0$, the exact information radius is $h \sim \gamma^2 xy/(2\delta z^2)$. For $\alpha < 1$, the same reset construction yields average-reward radius

$$\overline{h} = \lim_{\gamma \uparrow 1} (1 - \gamma) h = \frac{xy}{2(1 - \alpha + x)(1 - \alpha + y)}.$$

It has the same small-perturbation mixing order as the average-reward upper bound. The equality of separate-arm laws is compatible with strict mixing: mixing within an arm does not create coverage of an actuation row at a state unreachable under that arm. The construction has unreachable states in some arms and does not assert nonidentifiability under full-support arbitrary-state queries. In particular, when three complete kernels are known and $K = 0$, the fourth kernel is algebraically fixed. The exact minimax constants here concern the stated two-world family, whereas Proposition 3 concerns the uniform interaction bound; the two sharpness statements are distinct.

A.8 QUEUEING LAW AND THE RANKING WITNESS

Proof of Theorem 6. For cell (c, u), the birth–death chain has arrival rate λ and service rate $\nu = \mu + ub - cd > \lambda$. Put $z = \lambda/\nu < 1$. The probabilities $p_q = (1 - z)z^q$ satisfy detailed balance $p_q\lambda = p_{q+1}\nu$, and sum to one. Their mean is $z/(1 - z) = \lambda/(\nu - \lambda)$, which proves (11).

Let $F(v, w) = -\lambda/(s + v - w)$ on $[0, b] \times [0, d]$. Every denominator is positive because $s > d$, and

$$\partial_v\partial_w F(v, w) = \frac{2\lambda}{(s + v - w)^3}.$$

The fundamental theorem of calculus gives (13). Combining the four rational terms in $F(b, d) - F(0, d) - F(b, 0) + F(0, 0)$ over a common denominator gives (12). On the rectangle, $s - d \le s + v - w \le s + b$, so the integrand bounds imply the stated upper and lower bounds. Finally, only $-J_{10} = \lambda/(s - d)$ diverges as $s \downarrow d$, while the other three terms remain finite. This proves the asymptotic formula. □

Proof of Corollary 7. Subtract $J_{00} = -\lambda/s$ from the coupled value and from the three-cell prediction. Combining denominators gives (14). Every denominator is positive under $s > d$, so $G > 0$ exactly when $b > d$. Under this condition $\widetilde{G} < 0$ exactly when $s(b - d) < 2bd$, proving (15); at the upper endpoint the additive predictor is indifferent. For $b \le d$, $G \le 0$ and $\widetilde{G} < 0$, so the family has no false positive deployment decision. Substituting $b = 2d$, $s = 3d/2$ gives the two displayed gains. The four slacks are $3d/2, 7d/2, d/2, 5d/2$, all positive. Consequently, for any prescribed capacity tolerance $\varepsilon > 0$ and regret level $M > 0$, choose

$$0 < d < \min\{\varepsilon/2,\ 4\lambda/(15M)\}, \qquad b = 2d, \qquad \mu = \lambda + 3d/2.$$

Both perturbations are smaller than ε, every cell is stable, and rejecting deployment incurs regret $G > M$. The reward is unbounded and the smallest slack tends to zero, explaining why this does not contradict bounded-reward guarantees. □

Proof of the decision-margin certificates. The exact relation is $J^i_{11} = a_i + \Gamma_i$. Thus the deployment gain against known j_0 has the sign of $a_i - j_0$ whenever its absolute magnitude exceeds B_i. Likewise, $|\Gamma_i - \Gamma_j| \le B_i + B_j$ gives the pairwise certificate. For $\hat{\imath} \in \arg\max_i a_i$, subtract values to obtain

$$J^{i^*}_{11} - J^{\hat{\imath}}_{11} = (a_{i^*} - a_i) + (\Gamma_{i^*} - \Gamma_{\hat{\imath}}) \le B_{i^*} + B_{\hat{\imath}}.$$

On an event with simultaneous $|\hat{a}_i - a_i| \le e_i$, the same proof uses $J^i_{11} = \hat{a}_i + [(a_i - \hat{a}_i) + \Gamma_i]$ and total error $e_i + B_i$. No independence between policy estimates is needed. An uncertain external baseline with error e_0 adds e_0 to the deployment threshold; it cannot be treated as known for free. These are model-conditional sufficient certificates, not estimators for $k, \varepsilon_C, \varepsilon_U$, or α from ordinary CPU counters.

Exact model-selection calculation. The example uses $\lambda = 4/5$, $\mu = 1$, $(b_L, d_L) = (1/50, 1/100)$, and $(b_H, d_H) = (1/5, 3/20)$. The smallest slack among either policy's four cells is $1/20 > 0$. The coupled values are $J^L_{11} = -80/21$ and $J^H_{11} = -16/5$, so H is better by $64/105$. The additive predictions are $\widetilde{J}^L_{11} = -804/209$ and $\widetilde{J}^H_{11} = -14$, which rank L higher. Equation (16) follows by subtracting two instances of (1). The sufficient margin condition follows from $|\Gamma_H - \Gamma_L| \le B_H + B_L$.

An off-core/on-core ranking reversal is a different property. Within this stationary family, $b_H > b_L$ makes H better under reference execution, whereas $b_H - d_H < b_L - d_L$ makes it worse under coupled execution, provided all cells remain stable. This can occur even when a policy-specific additive correction ranks the coupled deployments correctly. A claimed failure of the three-cell model must therefore compare to $\widetilde{J}_{11}$, not merely display a placement reversal.

B SIMULATION PROTOCOL AND COMPLETE AUXILIARY COMPARISONS

Randomness and calibration. Each seed initializes a Python `random.Random` instance and supplies exactly one uniform draw per time step. The arrival sequence does not depend on invocation, applied action, execution mode, or the number of calls in another arm. The vectorized implementation

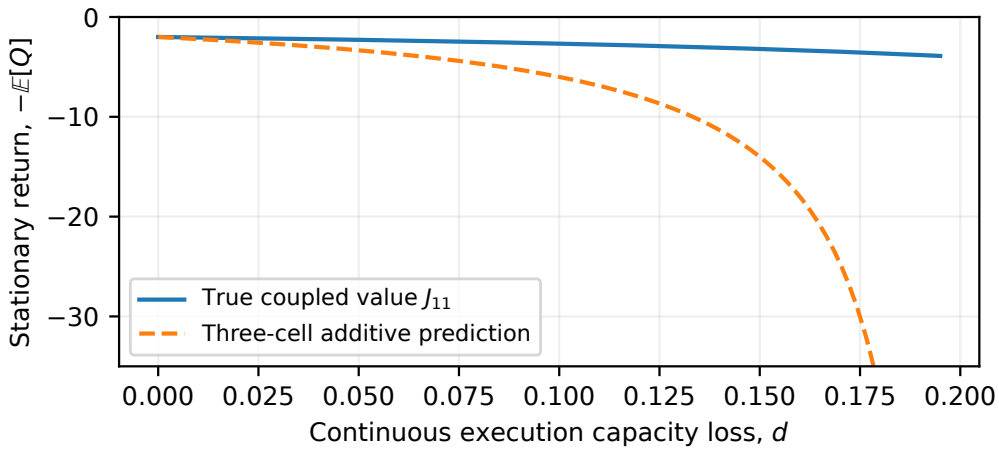

Figure 2: Analytical stationary values, not simulation measurements, for $\lambda = .8$, $s = .2$, and $b = .2$. The three-cell prediction diverges as the execution-only cell approaches its stability boundary, while the action restores slack in the coupled deployment. The plotted vertical range clips increasingly negative predictions near that boundary.

uses the same draws and arithmetic order as the supplied scalar implementation and is tested against it across 12 load–gate combinations and four seed values. The scalar reference remains in the release.

Calibration uses 40 baseline episodes per load, with no applied action and no capacity loss, to collect post-transition queue values. The cutoff is the sorted sample at index $\lfloor .75m \rfloor$, where $m = 40{,}000$. This is an empirical cutoff, not a proof of the largest reachable queue. In particular, observing a maximum below an absolute threshold in a finite calibration sample does not prove that the threshold will never be reached on a new trajectory. The gate checks $Q_t \geq q$, including equality; a zero cutoff therefore invokes at every step. The candidate boost is still zero at $Q_t = 0$.

Action and footprint bookkeeping. The invocation count records all invocations, including those producing zero candidate boost. N_A^{cu} counts applied strictly positive boosts, so it is identically zero when $u = 0$. The total applied boost is $D_U^{cu} = .1N_A^{cu}$ in these experiments. D_C^{cu} sums the actual delivered per-step service losses and is zero under reference execution. CPU and cache coordinates in the reference implementation are synthetic scalar inputs with a specified linear mapping into capacity loss. No claim equates them to physical CPU seconds, cache misses, or bandwidth measurements.

Why low-load zeros occur. At $\lambda = .70$, the maximum arrival is .80, below the smallest possible service .9925 in the primary grid, so $Q_t = 0$ follows inductively from $Q_0 = 0$. At $\lambda = .90$ without coupling, arrivals never exceed one, again giving $Q_t = 0$. Under continuous coupled invocation, however, service falls to .9925 and some arrivals create a positive queue; the action can then activate. Merely assuming $\lambda < \mu$ would not prove an always-empty queue because random bursts can still create backlog. These statements are bounded-arrival arguments, not generic stability arguments.

Frozen-tape construction. For each evaluation seed and load, we run the same gate on the 00 baseline queue and save its invocation indicator g_t^0 and candidate command a_t^0. In each replay arm, delivered loss is cdg_t^0 and applied action is ua_t^0, regardless of that replay arm's current queue. Thus both coupled arms have the same loss schedule, and both actuated arms have the same applied command schedule. The tape is seed specific and is correlated with the shared exogenous arrival stream. It is a simulator-defined counterfactual intervention, not a feasible prediction method for unknown future arrivals. Because the action schedule is the same across C, its additive penalty cancels from the replay contrast. Table 5 reports the paired estimates for the primary loss level.

Multiplicity and precision. The 12 primary load–gate comparisons form a single reported family. Table 6 gives Holm-adjusted two-sided p-values. Displayed confidence intervals elsewhere are pointwise, not simultaneous. The ratio in this table is the sum of the four marginal-mean variances

Table 5: Feedback and fixed-tape comparisons at loss .0075. Both use 200 held-out seeds, the same calibrated cutoffs, and 1000 steps. All values are generated from the saved episode records.

λ	Protocol	Cutoff	$\widehat{\Gamma}$	95% CI	N_A^{01}/N_A^{11}
0.90	Feedback	0.00000	-0.059	[-0.061, -0.057]	0.0/35.8
0.90	Replay	0.00000	0.000	[0.000, 0.000]	0.0/0.0
0.91	Feedback	0.00000	0.070	[0.063, 0.077]	47.3/80.9
0.91	Replay	0.00000	0.099	[0.093, 0.105]	51.0/51.0
0.93	Feedback	0.00000	0.903	[0.875, 0.930]	129.8/157.1
0.93	Replay	0.00000	0.885	[0.859, 0.911]	163.7/163.7
0.95	Feedback	0.00963	2.084	[2.014, 2.154]	167.7/167.7
0.95	Replay	0.00963	1.976	[1.911, 2.041]	250.3/250.3
0.97	Feedback	0.04824	9.088	[8.627, 9.548]	123.9/125.2
0.97	Replay	0.04824	7.936	[7.566, 8.307]	251.5/251.5
0.98	Feedback	0.08009	28.565	[26.497, 30.633]	83.0/85.4
0.98	Replay	0.08009	22.106	[20.576, 23.636]	253.2/253.2
0.99	Feedback	0.19084	249.865	[216.106, 283.624]	40.3/42.3
0.99	Replay	0.19084	106.161	[93.682, 118.640]	252.5/252.5

Table 6: Additional diagnostics for the primary grid. "—" denotes undefined variance ratios in deterministic-zero cases. Rates are mean invocations divided by 1000.

λ	Gate	Holm p	Variance ratio	Invocation rate $10/11$
0.70	Period 5	1	—	0.200/0.200
0.70	Cutoff 1	1	—	0.000/0.000
0.70	Quantile .75	1	—	1.000/1.000
0.90	Period 5	1	—	0.200/0.200
0.90	Cutoff 1	1	—	0.000/0.000
0.90	Quantile .75	2.5e-129	9.23	1.000/1.000
0.95	Period 5	4.8e-119	328.18	0.200/0.200
0.95	Cutoff 1	1	—	0.000/0.000
0.95	Quantile .75	1.4e-126	14.19	0.271/0.168
0.99	Period 5	5.8e-65	23.41	0.200/0.200
0.99	Cutoff 1	0.31	23.45	0.004/0.000
0.99	Quantile .75	2.2e-32	1.28	0.513/0.042

divided by the variance of the paired four-cell estimator. It is a *variance* ratio, not a standard-error ratio, and is undefined when the paired variance is zero. Common random numbers can help or hurt in general; the actual ratio is reported rather than presumed.

The sensitivity grid crosses loads $.90, .91, .93, .95, .97, .98, .99$, per-invocation losses $0, .0025, .0075, .015$, and the two feedback/replay protocols. All 56 configurations are included in the release, including deterministic-zero and negative effects. Figure 3 shows the nonzero-loss feedback cases. The grid reuses evaluation seeds to facilitate comparisons; its configurations are not statistically independent replications of the primary table.

B.1 Post-hoc deployment and ranking comparisons

These comparisons are reanalyses of the same 200 held-out evaluation seeds used in the 12 primary queue configurations. They are neither preregistered hypotheses nor additional independent experiments. For policy j and seed i, define

$$\widetilde{G}_j^{(i)} = Y_{01,j}^{(i)} + Y_{10,j}^{(i)} - 2Y_{00,j}^{(i)}, \qquad G_j^{(i)} = Y_{11,j}^{(i)} - Y_{00,j}^{(i)}.$$

The paired difference $G_j^{(i)} - \widetilde{G}_j^{(i)}$ is exactly the interaction H_i of (17). For a same-load policy pair, compute differences of the per-seed additive predictions and of the per-seed coupled returns *before* calculating a standard error. Shared arrivals and exactly equal baseline returns preserve the intended pairing. Subtracting independently estimated marginal confidence intervals would not reproduce these intervals.

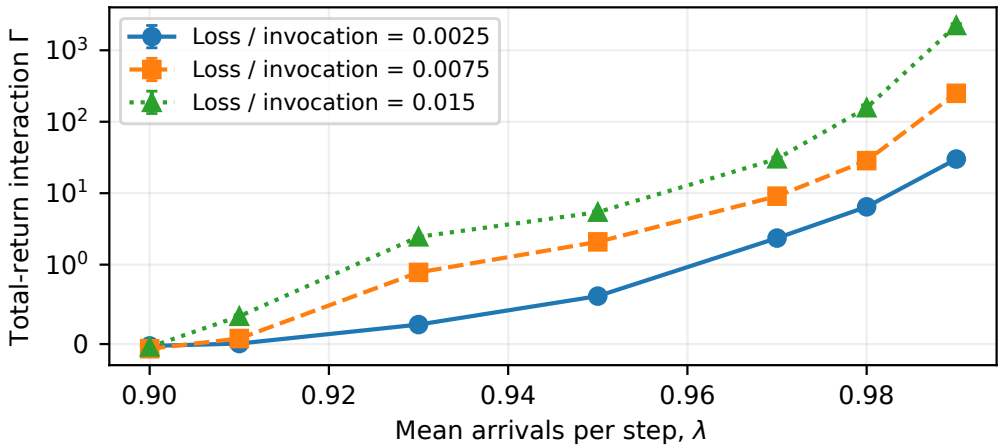

Figure 3: Sensitivity of the finite-horizon feedback interaction to offered load and per-invocation capacity loss. The return scale is symmetric-logarithmic. These simulated footprints are not calibrated hardware costs. The full zero-loss grid is exactly null and is omitted from the plot only for legibility, not from the dataset.

Table 7: All primary deployment-gain reanalyses. $\widetilde{G}$ is the additive prediction and G the measured coupled gain. Parenthetic intervals are paired pointwise 95% intervals in total-return units.

λ	Gate	$\widetilde{G}$ [95% CI]	G [95% CI]
0.70	Period 5	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
0.70	Cutoff 1	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
0.70	Quantile .75	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
0.90	Period 5	-0.027 [-0.029, -0.026]	-0.027 [-0.029, -0.026]
0.90	Cutoff 1	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
0.90	Quantile .75	-0.146 [-0.149, -0.142]	-0.205 [-0.209, -0.200]
0.95	Period 5	0.918 [0.845, 0.992]	1.309 [1.225, 1.394]
0.95	Cutoff 1	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
0.95	Quantile .75	2.370 [2.298, 2.441]	4.454 [4.330, 4.577]
0.99	Period 5	63.883 [60.611, 67.155]	87.607 [82.915, 92.299]
0.99	Cutoff 1	-0.715 [-1.903, 0.472]	1.574 [0.455, 2.693]
0.99	Quantile .75	-176.573 [-206.925, -146.221]	73.292 [68.486, 78.097]

Table 7 gives both gains for all 12 primary configurations. Table 8 gives every within-load policy pair in the stated direction. The 24 gains and 24 pairwise differences form a single, post-hoc 48-comparison family for Holm adjustment. The displayed intervals are pointwise 95% Student intervals; adjusted p-values are in `results/decision_analysis.json`. This declared family is distinct from the 12 primary interaction tests and is not claimed to cover every comparison considered during research.

At $\lambda = .99$, the calibrated controller's additive-predicted gain is -176.573 with interval $[-206.925, -146.221]$, while its coupled gain is 73.292 with interval $[68.486, 78.097]$. The corresponding Holm-adjusted p-values are 1.96×10^{-22} and 1.44×10^{-74}. For calibrated minus cutoff-one, the predicted difference is -175.858 with interval $[-205.978, -145.738]$ and the actual difference is 71.718 with interval $[67.460, 75.975]$; their adjusted p-values are 1.54×10^{-22} and 1.12×10^{-81}. These are differences in expected return under the synthetic protocol, not a claim that every individual trajectory has the same signs.

Period 5 has a larger predicted and actual gain than either alternative at this load. Accordingly, these data demonstrate a deployment false rejection and a pairwise ranking error, not a wrong top-ranked selection over all three gates. The cutoff-one actual gain of 1.574 has a pointwise interval $[.455, 2.693]$, but its adjusted p-value is $.109$; we do not treat it as a familywise-significant positive

Table 8: All within-load ranking contrasts. Direction is the second gate minus the first. P = Period 5, A = Cutoff 1, Q = Quantile .75. The paired intervals are pointwise 95% intervals.

λ	Direction	Predicted [95% CI]	Actual [95% CI]
0.70	A − P	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
0.70	Q − P	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
0.70	Q − A	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
0.90	A − P	0.027 [0.026, 0.029]	0.027 [0.026, 0.029]
0.90	Q − P	-0.118 [-0.122, -0.115]	-0.178 [-0.182, -0.173]
0.90	Q − A	-0.146 [-0.149, -0.142]	-0.205 [-0.209, -0.200]
0.95	A − P	-0.918 [-0.992, -0.845]	-1.309 [-1.394, -1.225]
0.95	Q − P	1.451 [1.391, 1.511]	3.144 [3.063, 3.226]
0.95	Q − A	2.370 [2.298, 2.441]	4.454 [4.330, 4.577]
0.99	A − P	-64.598 [-68.518, -60.678]	-86.033 [-90.186, -81.880]
0.99	Q − P	-240.456 [-272.833, -208.079]	-14.315 [-14.786, -13.845]
0.99	Q − A	-175.858 [-205.978, -145.738]	71.718 [67.460, 75.975]

benefit. All underlying episode rows are unchanged, and a SHA-256 source digest plus seed-level contrasts accompany the reconstruction script.

C PHYSICAL MICROBENCHMARK PROTOCOL, CHECKS, AND UNCERTAINTY

Interventions. The service and controller are separate newly forked processes. Each is pinned with `sched_setaffinity`; the driver uses another allowed logical CPU. Both workers signal readiness before a common future start time is released. Service state is reset between trials. The two controller placements are same-core and other-core; neither is called an ideal zero-cost execution. The initial batch command is configured before the measurement window and never changes within it. This removes the online command-release path, but does not remove delays experienced by the service while competing for CPU.

Work and offered load. Synthetic work consists of a deterministic integer arithmetic loop, with progress measured against `process_time`. A service batch of n actual pending items consumes a target $.0005 + .00015n$ CPU seconds, and n never exceeds the available backlog or selected batch size. The controller targets .003 CPU seconds at each .010-second release; missed releases are skipped, not retrospectively charged to earlier windows. The achieved CPU usage and invocation counts are recorded. Work is deliberately redundant so that placement alters execution without changing action semantics. No trained model, validation accuracy, or network-forward-pass measurement is involved.

Nominal batch-one capacity is $1/(.0005 + .00015)$ items per second. Arrival rates multiply this number by $.60, .95, 1.40$, yielding approximately 923.08, 1461.54, and 2153.85 items per second. These rates need not equal actual utilizations: clock checks, Python overhead, scheduler behavior, and the shared environment affect achieved capacity. The idle control retains the process, timers, and synchronization but uses zero busy-loop budget.

A genuinely fixed observation window. For each block, arrival times are sampled before any arm is run and reused exactly across its four cells. The service checks a common end time rather than extending a round after finishing work. A batch interrupted by the deadline receives no completion credit. All arrivals before the deadline count, including requests not yet processed. Let the common window length be $H = .30$, arrival times be $a_i \in (0, H)$, and completed batches have completion time $t_j \leq H$ and actual size n_j. The time-average number in the system is

$$\overline{Q} = \frac{1}{H}\left[\sum_i (H - a_i) - \sum_j n_j (H - t_j)\right]. \tag{27}$$

This counts in-service as well as waiting requests. The identity follows by integrating the step-function counts of cumulative arrivals and cumulative completions. It provides a second, event-level reconstruction of each saved return $-\overline{Q}$, independent of the worker's online queue bookkeeping.

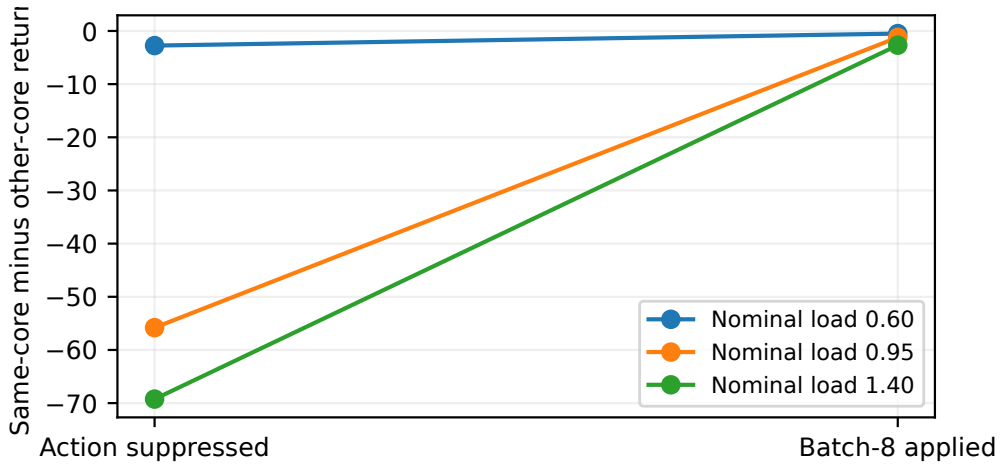

Figure 4: Conditional placement effects in the physical experiment, in return units. Subtracting each line's left value from its right value gives the corresponding factorial interaction. These are mean contrasts; paired uncertainty for the differences is reported in Table 4. The effect of the same execution program depends strongly on whether batching is applied.

Clock lateness and measured CPU dose. The largest start lateness in the release is $.002464811$ s and the largest final deadline-check lateness is $.017103697$ s. Late wakeups and preemptions are real components of the treatment response. Crucially, no arrival or completion after the predetermined end time is credited, and the integration endpoint in (27) is unchanged. Process CPU totals include ordinary clock and accounting overhead and can include a small final-check overhead; they are diagnostics rather than an exactly matched resource-dose intervention.

Randomization, interruption, and statistical units. There are 24 blocks at each nonzero-work load and 12 idle-work blocks. Seeds are $300000 + 1000s + i$ for scenario index s and block i. A separate seeded shuffle determines the four-arm order. Within a block the arrivals are identical, while the operating system's scheduling noise is not shared or controlled. Trials restart the processes, but resource conditions such as cache and frequency are not guaranteed independent across sequential trials.

A tool execution timeout interrupted scenario 0, block 22, after two arms. Those observations were retained, and the other two arms were collected when execution resumed. The metadata explicitly records the 90-trial checkpoint; machine load was recorded at resumption rather than retrospectively invented for the initial start. Primary estimates include the interrupted block. Excluding it changes the low-load estimate from 2.2565 to 2.3015, with a paired interval $[1.5916, 3.0114]$. This sensitivity is disclosed rather than selecting an exclusion rule after seeing a preferred result.

The paired Student intervals treat blocks as the sampling units. A percentile bootstrap with 20,000 block resamples gives intervals $[1.6550, 2.9307]$, $[50.8961, 58.2406]$, and $[63.8773, 69.1184]$ for the three nonzero-work scenarios, and $[-.4979, 3.5363]$ for the idle-work control. These support the same qualitative conclusions as the paired Student intervals, but neither method guarantees coverage under arbitrary time-dependent host interference.

Integrity checks. The audit verifies 336 unique scenario–block–arm keys, complete four-cell blocks, identical arrival lists within each block, strictly increasing arrivals inside the observation window, valid batch sizes, and cumulative completed items no greater than cumulative arrivals at every completion. It independently reconstructs all mean queue values and checks that reward is their negative. There are no invalid or post-deadline credited completions. The largest numerical reconstruction residual is 5.6843×10^{-14}. These checks establish bookkeeping integrity; they do not identify a unique physical mechanism or certify the environment as noise free.

D HISTORICAL EVIDENCE AND PROVENANCE BOUNDARIES

The release preserves the supplied historical JSON files separately from newly generated observations. We distinguish three evidence levels: deterministic simulation that can be rerun; historical hardware with per-trial records that can be reanalyzed; and historical hardware with summary statistics only. The primary experimental claims in Section 6 use new records, not reconstructed pseudo-observations.

Historical queue reproduction. The original queue grid used evaluation seeds 0–199 and baseline calibration seeds 0–39, so the two ranges overlapped. Running the reference-equivalent simulator with exactly those original seeds reproduces all 12 reported interaction means, with a maximum absolute discrepancy of 5.56×10^{-17}. That confirms the stored means, not the original inferential interpretation. The main text replaces the overlapping calibration with the disjoint ranges described above. Some archived JSON fields contain nonstandard `NaN` or `Infinity` for zero-variance ratios; those source files are retained unchanged, whereas all newly generated JSON uses finite numbers or explicit null values.

Historical service ranking. A supplied 20-seed summary compares a cheap synthetic controller to a heavier one. The archived heavy-minus-cheap reference-placement difference is 1.2575 with a reported normal-approximation interval $[-3.3081, 5.8231]$. Under coupling it is -16.2758 with interval $[-20.2287, -12.3230]$. The first interval includes zero, so these records do not establish a statistically supported reversal of both rankings. Only aggregate statistics were supplied: we cannot reconstruct raw paired data, verify the original standard error from episodes, or generate a defensible new paired bootstrap for that study. Those reported intervals are not relabeled as the new experiment's Student intervals.

The historical delay-matching study matches an average inference-duration difference and, in its strongest reported contrast, suppresses actuation. It is not a trajectory-level action-release match and cannot establish that deployed command latency is irrelevant. Its synthetic integer workload is also not evidence about the accuracy of a learned controller. The new fixed-command experiment addresses a narrower mechanism question by removing online command release by construction.

Historical service timing limitations. Inspection of the supplied code identifies per-round deadlines reset from the current wall time, possible batch completion beyond a nominal round, and per-round rather than fixed-wall-clock arrival generation. These features can change offered load and completion credit as treatments alter runtime. Sequential fixed arm order also leaves a potential drift confound. We therefore retain those outputs only as exploratory source material, not as an independently validated fixed-window factorial experiment. The newly written measurement program changes the observation window and records complete events rather than trying to infer corrected historical episodes from summary values.

Historical Linux placement records. The Linux pilot contains 21 raw trials: seven repetitions each of a no-controller baseline, other-core controller, and same-core controller. Reanalysis gives a mean paired same-core minus other-core wall-time difference of $.58435$ s, with a Student interval $[.50807, .66062]$. The mean of the seven paired wall-time ratios is 1.37611, with interval $[1.30594, 1.44628]$. It is not a ratio of aggregate means. The other-core reference is itself slower than the no-controller baseline, with mean paired difference $.44074$ s, so it should not be described as free execution.

These historical contrasts have fixed run order and characterize an association under that protocol; they do not isolate all sources of slowdown. There is no independently crossed actuation factor in this pilot, so no four-cell Γ is inferred from it. A separate supplied priority-actuation pilot changes achieved controller resource shares. Such post-treatment dose changes do not by themselves invalidate a randomized factorial design, but the available pilot is not substituted for the fixed-window four-cell measurements in the main paper.

E NUMERICAL PROOF CHECKS AND REPRODUCIBILITY ENTRY POINTS

The proof-check script samples 2,000 finite-horizon systems with 2–7 states, 2–11 transitions, and arbitrary arm-specific rewards. The largest residual in Theorem 1's identity is 2.6646×10^{-15}. An-

other 2,000 discounted systems with common bounded rewards yield no violations of the generic or mixing-dependent bounds. One hundred affine-kernel systems check the mixed-resolvent derivative against a centered finite difference, with largest discrepancy 1.62×10^{-10}. Symbolic simplification verifies the queue rational formula, the two-state $-2\varepsilon^2$ witness, and the original one-directional propagation witness. Additional symbolic checks verify the constant-4 witness, the three-arm information radius, its average-reward limit, the exact rejection boundary, and the vanishing-perturbation family. A 64-case parameter grid independently solves the new Bellman systems, checks stochasticity and mixing, and verifies equality of the separate-arm kernels on their full reachable sets. It gives maximum value residuals below 10^{-11}. Another 1,000 random comparisons check the implementation of the decision-margin inequality. These finite checks are not formal machine-checked proofs and cannot establish a theorem by sampling.

The test suite covers reference-equivalent simulation, a transition-mediated action-equivalence witness, the constant-reward counterexample, conditional versus marginal matching, the tight total-variation witness, replay-dose equality, zero-compute null controls, multiple-comparison adjustment, the saved data-integrity receipt, sharp leading constants, the information construction, and the raw-record decision reanalysis. The tests and analytic checks are separate from the hardware experiment; no GPU, privileged scheduler changes, or remote service is required.

Contents and commands. From the supplement root, install the listed Python dependencies, then use the following entry points:

```
make verify      # numerical theory checks, data audit, tests
make assets      # regenerate tables and figures
make paper       # compile the anonymous manuscript
make simulation  # rerun held-out queue experiments
make hardware    # new measurements in a fresh directory
```

The scalar historical simulator is retained for reference-equivalence tests. The new code resides under `code/rccontrol`; new observations under `data/new`; and unchanged source data under `data/historical`. Hardware execution supports checkpoint resumption, so a fresh measurement run must use an empty output directory. The Makefile creates one for the hardware target. The notebook provides an inspectable audit of the saved results. Exact hardware numbers are environment dependent and are not promised to repeat bit for bit.

All main-text values are either algebraic examples or generated from the supplied raw observation files. Complete trial records, configuration IDs, and seed ranges are provided. The source-generation and author-preparation records are separated from the anonymous review material; neither author identity nor private filesystem paths are required to reproduce the reported analyses.