WHAT STOCHASTIC INFORMATION MATTERS FOR LEARNING? OBSERVABLE-RELATIVE EQUIVALENCE AND CONTROL

Anonymous authors
Paper under double-blind review

ABSTRACT

Two stochastic learning rules can have the same conditional update mean and full covariance at every common learner state yet lead to different finite-horizon outcomes. We establish that the stochastic information needed to describe learning depends on the future quantity being asked about. We characterize when a summary of today's update is sufficient to predict a future learning outcome. For every finite order k, we construct SGD systems on the same smooth, strongly convex objective whose temporal polynomial statistics agree through order k, while their probabilities of the same finite-horizon event differ by arbitrarily close to one. No fixed finite moment hierarchy is universally sufficient. When summary-equivalent laws are realizable choices, unresolved stochastic structure becomes a control variable that can improve learning. At the same learner state, the law minimizing immediate error differs from the law minimizing error after eight further SGD steps; choosing for future loss reduces terminal risk by 8.35% relative to state-aware immediate-loss control in our exact system. In three neural settings, Gaussianizing updates while preserving their conditional mean and covariance moves boundary-crossing risk toward the Gaussian prediction in 59 of 60 systems, closing 59% to 75% of the discrepancy. In closed-loop on-policy learning, matched local mean and covariance lead to different future evidence distributions. In same-prompt Qwen RLVR, moment-matched gradient laws produce substantially distinguishable applied updates after clipping and AdamW. The future learning question determines both the stochastic resolution needed for prediction and the realizable law worth choosing for control.

1 INTRODUCTION

What does it mean for two stochastic learning rules to be the same? They may have the same expected update and the same covariance at every learner state. That agreement can be enough to predict one quantity but insufficient to predict another. A one-step quadratic loss may assign the two rules exactly the same value, making them indistinguishable, while a future boundary crossing, terminal error, or encounter with new data substantially separates them. The stochastic information needed to describe learning therefore depends on the future quantity in question.

The issue of characterizing a stochastic learning rule matters because much of stochastic optimization is understood through compressed update laws: expected gradients, covariance, diffusions, Gaussian approximations, or finite collections of moments (Mandt et al., 2017; Li et al., 2019; 2021; Wei et al., 2025). These descriptions are effective in important regimes (Cheng et al., 2020; Ankirchner & Perko, 2024; Gheissari & Jagannath, 2026). Existing work also shows that distributional structure beyond variance can affect escape and other optimization behavior (Simsekli et al., 2019; Nguyen et al., 2019; Bajovic et al., 2023; Azizian et al., 2024). What remains unresolved is how to decide, for a particular learning question, whether a chosen stochastic summary retains the information that question requires.

We propose an observable-relative answer. Fix the learner's current state, its future learning process, a horizon, and a quantity to be predicted. Each possible update today then has an expected consequence for that future quantity. Two update laws are equivalent for this question when they assign the

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.

same expectation to those consequences. A retained summary of the update law, such as mean and covariance, is sufficient when the information it preserves can represent that future value. The future quantity to be predicted therefore determines the required stochastic resolution.

This viewpoint yields an exact account of what a summary leaves unresolved. For mean and covariance, the future quantity is determined precisely when the value assigned to today's possible updates is quadratic; the best quadratic approximation determines the largest predictive disagreement otherwise. We extend this result across learning trajectories and identify when smooth future values make low-order descriptions accurate. Yet no fixed finite moment order works for every future question, even in smooth, strongly convex, contractive SGD.

The same unresolved stochastic information can also be used to improve learning. If a learner can choose among update laws with the same retained summary, the part of the law invisible to that summary becomes a control variable. Which law is preferable can depend on the learner's current state and on the learning that follows its next update.

Our evidence follows that progression. Exact examples locate both sides of the sufficiency boundary: the same low-order summary can be sufficient under one future continuation and insufficient under another. In Margin-Routed Minibatches, the learner checks whether it assigns a positive score to a difficult negative example after its first update. That single yes-or-no answer selects the minibatch law for its second update, thus improving terminal error while preserving marginal exposure, pairwise sampling, conditional gradient mean, and full covariance. A separate nonlinear system shows why the horizon matters for control: the law preferred for immediate loss differs from the law preferred after further learning. There, a state-aware immediate-loss controller gains nothing over the best fixed law, while a future-aware controller improves future risk.

The consequences also appear beyond exact constructions. In three nonlinear neural settings, variance-preserving Gaussianization moves finite-horizon path risk toward the covariance-matched Gaussian prediction in 59 of 60 independently generated systems, closing roughly 59% to 75% of the original discrepancy. In a closed-loop policy-gradient learner, locally matched drift and diffusion lead to different future evidence distributions after repeated learning: the probability of missing at least one evidence type differs by 3.074 percentage points. This matters because a stochastic update can change the evidence from which later updates are learned, a mechanism related to decision-dependent data distributions (Perdomo et al., 2020; Drusvyatskiy & Xiao, 2023; Cutler et al., 2024).

Finally, the unresolved freedom is substantial in a modern same-prompt learner. In Qwen2.5-1.5B-Instruct under RLVR (Yang et al., 2025), two gradient laws at the same state and prompt have identical conditional mean and full covariance, yet their probabilities of a nonzero update are 78.66% and 98.68%. The 20.02 percentage-point difference has the same direction across all 32 prompts. Their final applied-update distributions remain distinguishable after clipping and AdamW (Loshchilov & Hutter, 2019). This establishes that substantial stochastic-law freedom beyond mean and covariance exists in a real same-prompt RLVR update and remains visible through realistic optimizer processing in the tested setting. We summarize our main contributions as follows:

1. **Future learning questions define stochastic equivalence.** We characterize when a retained update-law summary suffices for a specified future quantity, exactly measure the predictive ambiguity it hides, and bound how that ambiguity accumulates over complete learning trajectories (Section 2).

2. **No fixed finite moment order suffices universally.** For every finite order, SGD systems on the same smooth, strongly convex objective can agree on temporal polynomial statistics yet have almost maximally different finite-horizon event probabilities. Exact examples also locate where mean and covariance suffice, and how changing the continuation or future data budget moves that boundary (Sections 2 and 3).

3. **Unresolved stochastic structure can improve future learning.** Realizable moment-matched laws permit state-dependent and future-aware control that improves terminal loss even when immediate-loss choice does not. Neural experiments establish finite-horizon effects, closed-loop learning changes future evidence, and same-prompt RLVR reveals substantial law freedom that survives clipping and AdamW (Sections 4 and 5).

Together, these results make the stochastic structure of an update a design variable for future learning as the quantity we hope to improve determines which distinctions matter, and realizable distinctions can be used to steer what the learner does next.

2 WHAT STOCHASTIC INFORMATION DOES FUTURE LEARNING REQUIRE?

Two update laws can have the same mean and full covariance yet lead to different futures. Agreement may be enough for one quantity and insufficient for another. Today's update changes the state from which the learner continues to train, encounters data, and is eventually evaluated. Which features of its law must we retain to predict a specified future outcome?

What can today's random update change about the future? Let x be the learner's current state and $z \in \mathcal{Z}$ a possible update. The state includes model parameters, optimizer memory, and any other persistent variables needed to determine subsequent learning. Applying z produces $T_x z$.

Compare two candidate laws μ and ν at this same state. After today's update, both learners follow a common future learning process. Let K be one step of that process and Φ the quantity ultimately measured, such as terminal loss or reward. Then $(K^h\Phi)(y)$ is the expected value of Φ after h further learning steps from state y.

If today's update were z, the learner would start its remaining $H-1$ steps from $T_x z$. The function $v_{x,H}(z) = (K^{H-1}\Phi)(T_x z)$ assigns that update its expected future consequence. We call it the *backward observable*. The difference between the laws on the specified future quantity is exactly

$$\Delta_H^{\Phi}(\mu, \nu; x) = \int v_{x,H}(z)\,(\mu - \nu)(dz). \tag{1}$$

The backward observable converts a future learning question into a function of today's update.

When are two update laws equivalent for that question? Laws μ and ν are *observable-relative equivalent* for (x, H, Φ) when $\mathbb{E}_\mu[v_{x,H}] = \mathbb{E}_\nu[v_{x,H}]$. They may assign very different probabilities to individual updates and still give exactly the same answer to this future learning question.

For a family of future questions, equivalence requires agreement on each corresponding backward observable. The complete update law can contain distinctions that none of those questions uses.

What does a conventional stochastic summary retain? A summary reports expectations of selected functions of the update. Let $\mathcal{S} \subset C(\mathcal{Z})$ be that function class, where $C(\mathcal{Z})$ denotes continuous functions on the update space. Write $\mu \equiv_{\mathcal{S}} \nu$ when $\int s\,d\mu = \int s\,d\nu$ for every $s \in \mathcal{S}$.

Mean and covariance correspond to the quadratic class $\mathcal{S}_2 = \operatorname{span}\{1, z_i, z_i z_j\}$. Once the means agree, matching quadratic expectations is equivalent to matching the full covariance matrix. These statistics determine the expectation of every quadratic function of today's update. Continued learning, however, can assign a nonquadratic future value to those same updates.

How much future disagreement can a summary hide? For a backward observable v, define its *predictive ambiguity* by $\Omega_{\mathcal{S}}(v) = \sup_{\mu \equiv_{\mathcal{S}} \nu} |\mathbb{E}_\mu[v] - \mathbb{E}_\nu[v]|$. This is the largest disagreement between laws sharing the reported summary.

The answer depends on how well the summary can represent v. The summary also determines the expectation of constants and of functions uniformly approximable by those it records. Let $\widetilde{\mathcal{S}} = \overline{\mathcal{S} + \mathbb{R}1}^{\,\|\cdot\|_\infty}$ collect these functions. Then $\operatorname{dist}_\infty(v, \widetilde{\mathcal{S}}) = \inf_{s \in \widetilde{\mathcal{S}}} \sup_{z \in \mathcal{Z}} |v(z) - s(z)|$ is the smallest worst-case error in approximating future value using the summary. For mean and covariance, it is best quadratic approximation.

Theorem 1 (Observable-resolution characterization)**.** *Let $\mathcal{Z}$ be compact Hausdorff, let $\mathcal{S} \subset C(\mathcal{Z})$ be linear, and let $v \in C(\mathcal{Z})$. Over unrestricted probability laws on $\mathcal{Z}$,*

$$\Omega_{\mathcal{S}}(v) = 2\,\operatorname{dist}_\infty\Big(v, \widetilde{\mathcal{S}}\Big). \tag{2}$$

The optimal worst-case error for predicting $\mathbb{E}_\mu[v]$ from the reported summary alone is $\operatorname{dist}_\infty(v, \widetilde{\mathcal{S}})$.

A summary determines the future quantity exactly when it represents its backward observable. Otherwise, the theorem gives the largest disagreement compatible with that summary. For mean and covariance, the boundary is quadratic representability on the update space.

The theorem gives an exact resolution test for a specified future learning quantity. Its proof uses generalized moment duality and uniform approximation (Kemperman, 1968; Tardella, 2001; Bertsimas & Popescu, 2005). Here the function being approximated is generated by the learner's continuation and the future quantity of interest. The theorem ranges over all summary-equivalent probability laws; realizable laws may occupy a smaller part of that class. The exact duality assumes compact update support and a continuous backward observable. A continuous update map and terminal quantity, together with a future learning kernel K that maps continuous functions to continuous functions, ensure that the backward observable is continuous. Appendix A gives the proof and separate unbounded-law bounds under derivative and moment conditions.

2.1 From one update to a learning trajectory

Learning applies stochastic updates repeatedly. Even when two laws match summaries at every common learner state, their learners can visit different states later. To measure the resulting trajectory difference, let P and Q be complete-state learning kernels generated through the same application map by laws μ_x and ν_x, with $\mu_x \equiv_{\mathcal{S}} \nu_x$ at every common state x. For continuation under Q, write $v^Q_{x,r}(z) = (Q^{r-1}\Phi)(T_x z)$.

Replacing one transition at a time makes the trajectory difference a sum of local disagreements. Each local term asks whether the summary represents the value of the learning that remains.

Proposition 1 (Trajectory-resolution accumulation)**.** *For any common initial law λ,*

$$\left|\lambda(P^H - Q^H)\Phi\right| \le 2\sum_{t=0}^{H-1} \mathbb{E}_{X_t\sim\lambda P^t} \operatorname{dist}_\infty\left(v^Q_{X_t,H-t}, \widetilde{\mathcal{S}}\right). \tag{3}$$

The corresponding bound with P and Q exchanged also holds.

The bound evaluates unresolved future value along states the learner actually visits. Local disagreements can accumulate despite statewise summary matching. Each term measures the resolution required by the remaining learning. Time and optimizer memory belong to the complete state.

When do low-order descriptions recover accurate trajectories? Low-order summaries succeed when each relevant backward observable is within δ uniformly of a degree-k polynomial: laws matching moments through degree k then differ on that future quantity by at most 2δ.

A concrete bound follows for additive updates. Suppose $T_x z = x + \eta z$, the statewise laws match moments through degree k, $\|z\| \le R$, and the propagated values $V^Q_r = Q^r\Phi$ satisfy $\|D^{k+1}V^Q_r\|_{\mathrm{op}} \le M_{k+1}$ uniformly. Taylor cancellation and the trajectory proposition give

$$\left|\lambda(P^H - Q^H)\Phi\right| \le \frac{2H\eta^{k+1}R^{k+1}M_{k+1}}{(k+1)!}. \tag{4}$$

At a fixed macroscopic horizon $H = T/\eta$, this is $O(\eta^k)$. For mean and covariance matching, $k = 2$, the bound is $(TR^3M_3/3)\eta^2$.

The bound identifies when low-order descriptions remain accurate over a learning trajectory: the propagated future values must have controlled derivatives over the relevant horizon (Talay & Tubaro, 1990; Li et al., 2019; Ankirchner & Perko, 2024). Smoothness of the terminal quantity Φ alone does not ensure this condition. Discontinuous events can demand finer resolution; the finite-order result below shows that a bounded smooth path observable can do so as well.

2.2 No universal finite moment resolution

Could a sufficiently high moment order settle every future learning question? Even smooth, strongly convex, contractive SGD rules this out.

Consider with-replacement SGD on $\ell_z(\theta) = \frac{1}{2}(\theta - z)^2$, with $\theta_{t+1} = (1-\eta)\theta_t + \eta Z_t$ and $\eta \in (1/2, 1)$. Its population objective is globally 1-smooth and 1-strongly convex, and its deterministic recursion is contractive.

Theorem 2 (Finite-order nonidentification)**.** *For every finite $k \geq 1$ and every $\epsilon > 0$, there exist two distinct finite empirical laws ρ and σ on a common finite support, a shared initialization $\theta_0 = 0$, step size $\eta = 3/4$, and a finite horizon H with the following properties. Their population objectives and minimizers agree. Their statewise raw update moments through degree k agree for every θ. Every finite-horizon temporal polynomial observable of total degree at most k has the same expectation under the two systems. Nevertheless,*

$$\left| \mathbb{P}_\rho \left(\max_{t \leq H} \theta_t \geq \frac{1}{2} \right) - \mathbb{P}_\sigma \left(\max_{t \leq H} \theta_t \geq \frac{1}{2} \right) \right| > 1 - \epsilon. \tag{5}$$

The event indicator can be replaced by a bounded C^∞ function of the finite trajectory that agrees with it on every reachable path.

The construction shifts probability between finite empirical laws while preserving moments through degree k. Because the recursion is affine in IID draws, this also preserves every temporal polynomial of total degree at most k. A top-support draw forces the threshold crossing, while trajectories without it cannot cross; choosing its probability and the horizon separates the event probabilities by more than $1 - \epsilon$. The laws and horizon are chosen separately for each k and ϵ; the theorem specifies no growth rate for H as k increases. Appendix A gives the construction.

Recording whether a crossing has occurred makes the path event terminal, so Proposition 1 forces the accumulated degree-k approximation residual to exceed $(1 - \epsilon)/2$.

2.3 From unresolved information to control

What if the learner can choose among laws that share the same summary? At state x, let $\mathcal{Q}_s(x)$ be the realizable update laws with summary value s. For the future quantity (H, Φ), their *control width* is $\mathfrak{C}^\Phi_{x,H}(s) = \sup_{\mu \in \mathcal{Q}_s(x)} \mathbb{E}_\mu[v_{x,H}] - \inf_{\mu \in \mathcal{Q}_s(x)} \mathbb{E}_\mu[v_{x,H}]$. Write $\Omega_{\mathcal{S}}^{\mathcal{Q}_s(x)}(v)$ for ambiguity restricted to this feasible family.

Proposition 2 (Resolution-to-control bridge)**.** *For every backward observable $v_{x,H}$ and every realizable summary fiber $\mathcal{Q}_s(x)$,*

$$\mathfrak{C}^\Phi_{x,H}(s) = \Omega_{\mathcal{S}}^{\mathcal{Q}_s(x)}(v_{x,H}) \leq \Omega_{\mathcal{S}}(v_{x,H}) = 2 \operatorname{dist}_\infty \left(v_{x,H}, \widetilde{\mathcal{S}} \right). \tag{6}$$

A zero width means that choosing among these laws cannot change this future quantity at this state and horizon. The same unresolved stochastic structure therefore has two roles. If the law is unknown, it creates predictive ambiguity. If summary-equivalent laws are realizable choices, the reachable part of that ambiguity becomes control authority. Section 4 computes the widths in an exact nonlinear learner and shows how the preferred law depends on the learner's state and on the learning that follows its selection.

3 Exact sufficiency boundaries

Can the same stochastic summary be sufficient for one learning future and insufficient for another? We locate two exact boundaries by changing the future training rule or the number of future observations. In each case, sufficiency on one side is an algebraic equality.

When the future learning rule changes. Consider a learner estimating two Bernoulli means. A base gradient draws an outcome from one of the populations. Native applies one gradient, $G_{\mathrm{N}} = G_0$; Shape combines four independent base gradients, $G_{\mathrm{S}} = (G_1 + G_2 + G_3 - G_4)/2$. Their update laws differ, but their conditional mean and full covariance agree exactly.

What happens next determines whether that agreement is enough. If later steps choose either population with probability $1/2$ regardless of learner state, the dynamics are affine and backward propagation keeps the quadratic terminal loss quadratic. Native and Shape then have *exactly the same expected terminal loss at every finite horizon*.

Now let the next four sampling choices depend on the learner's parameters. Importance weighting preserves the objective and conditional expected gradient; today's laws and the terminal loss remain

unchanged. Exact finite-tree calculation predicted Shape's normalized terminal loss to be 0.06899 lower than Native's. Fresh paired validation with $N = 32{,}768$ found it 0.06007 lower.

Two update laws with the same mean and full covariance have identical expected terminal loss at every horizon under fixed future sampling, yet different expected terminal losses when later sampling responds to the learner's state. Full details are in Appendix B.

When the future data budget changes. Two empirical training datasets assign different frequencies to records, yet induce the same gradient mean and full covariance at every common state. The learner's parameter p specifies probabilities for three future labeled-example types; one ordinary SGD step makes this policy random. The resulting policy laws differ while their first and second moments agree. Each data type maps to the same labeled example in both systems.

Future examples share one realized policy p. The law of one example depends on average type probabilities, and that of a pair depends on pairwise probabilities. Both agree, so the *complete labeled-data distributions are identical* for one or two examples. Every common downstream learner therefore has the same output distribution at either budget.

A third example can expose a difference in the policies' third moments. Dataset A has an exactly $9/160 = 0.05625$ higher probability than Dataset B of supplying all three data types. Yet a fixed ridge-regression learner has 0.01153 higher normalized prediction risk under A by exact calculation. Fresh validation with $N = 65{,}536$ found a risk gap of 0.01219 and a coverage gap of 0.05620. A also produces more single-type triples and fewer two-plus-one triples; the high ridge error on single-type triples outweighs its extra complete coverage.

If a new policy is generated before each example, the complete three-example labeled-data distributions become identical again. **With at most two future examples, every downstream learner has the same output law under both systems. A third exposes dependence created by reusing the learned policy and changes the fixed ridge learner's prediction risk.** The boundary lies in higher-order dependence in future evidence, not marginal exposure or pairwise data statistics. Full details are in Appendix C.

4 From resolution to stochastic-law control

Section 2 established that unresolved predictive ambiguity can become control authority when summary-equivalent laws are realizable choices. We compute that authority in two exact learning systems. The first shows why the best law depends on the learner's state. The second shows that the law best for immediate loss can be the wrong law after learning continues.

State-relative control. Can a learner improve final prediction by choosing among update laws with the same local mean and full covariance? Margin-Routed Minibatches tests this in a two-parameter nonlinear classifier trained for two SGD steps. Each step draws four records from three example types: an easy negative, a hard negative, and a nearby positive. We choose a law over the $\binom{4+3-1}{4} = 15$ possible minibatch compositions. Every admissible law preserves ordinary-IID single-example exposure, the full ordered pairwise within-batch sampling law, and the conditional gradient mean and covariance. Only higher-order composition changes.

The best second-step law depends on the state reached after the first update. We optimize over all 121 extreme admissible laws and exhaustively evaluate their $121^2 = 14{,}641$ two-step pairs to find the globally best state-blind schedule. The adaptive optimum uses the *same first-step law*, isolating the value of observing the intermediate state.

A single bit captures most of that value. The deployed rule reads the sign of the hard-negative preactivation, $\gamma = w/2 + b$, to choose the second-step law. It captures 85.99% of the full Bellman advantage and, in independent confirmation, lowers terminal MSE by 4.64% relative to the best state-blind schedule. A state-independent control uses essentially the same two laws at the same overall frequencies but assigns them to the wrong states, producing 0.07343 higher terminal MSE. **Hidden minibatch structure becomes useful when its choice is aligned with the learner's current state.** Replacing the nonlinear activation with a linear one gives exactly zero control width at every finite horizon under the same moment matching.

Nonmyopic control. Should the learner choose the law that gives the best immediate update, or the one that gives the best result after further learning? We answer this in a two-coordinate ReLU regression learner. It predicts $\max(w_j, 0)$ on each of two one-hot inputs. Each input appears with four target values $Y \in \{1/4, 1/2, 3/4, 1\}$, giving eight labeled records. A minibatch pairs one record for each input, so there are 16 possible two-record minibatches. Native samples them uniformly. For one controlled update, we change their probabilities while preserving the conditional gradient mean and full covariance. The learner then takes eight further SGD steps using Native sampling.

At the same learner state, immediate squared error favors one endpoint of the feasible moment-matched minibatch laws; error after eight further Native steps favors the other endpoint. Changing the horizon reverses which minibatches should receive more probability, while both laws retain the same conditional gradient mean and full covariance. This reversal changes final performance. State-aware immediate-loss control ties the globally best fixed law at final risk 0.1855057. Choosing for future learning lowers risk to 0.1700154, an exact gain of 0.01549, or 8.35%. Paired validation over $524{,}288$ learners finds a gain of 0.01564 over the best fixed law. Assigning the future-aware recommendations to unrelated states removes at least 80% of their Native-relative benefit under the frozen comparison. **The law that is best now is different from the law that is best after learning continues, even at the same state and with the same local mean and covariance.**

Why the preferred law reverses. At the common pre-update coordinate $u = w_j = 9/8$, the controlled update can leave that coordinate at four values: $w_j^{\text{next}} = -\frac{3}{4}u + \frac{7}{4}Y \in \{-13/32, 1/32, 15/32, 29/32\}$. The negative outcome deactivates its ReLU; the other three can continue learning. Choosing a minibatch law changes the probabilities of these outcomes. Native assigns each probability $1/4$. We note that mean and covariance leave one probability shift among the four outcomes unresolved: $h = (-1, 3, -3, 1)$. A feasible shift changes their Native probabilities by δh. If v_H lists the expected future loss after each outcome, the shift changes expected loss by δc_H, where $c_H = h^\top v_H$. Because h cancels every quadratic, both the best quadratic approximation error and unrestricted predictive ambiguity are exact: $\text{dist}_\infty(v_H, \Pi_2) = \frac{|c_H|}{8}, \Omega_{\Pi_2}(v_H) = \frac{|c_H|}{4}$. For immediate squared error, $c_0 = 689/1024 > 0$; after eight further updates, $c_8 \approx -0.36892$. Its sign reversal selects the opposite endpoint law. At eight steps, unrestricted ambiguity is approximately 0.09223. Fixing Native's mean and covariance reduces the width to 0.06149; requiring every original minibatch to retain at least half its Native probability leaves 0.03074 of realizable control width (Appendix E). The three widths separate what the quadratic summary can miss, what remains possible at Native's summary value, and what the learner can actually choose.

The choice is recoverable from finite continuation data. With $8{,}192$ independent eight-step rollouts per outcome, the estimated quadratic approximation error is 0.04605, against the exact 0.04611. At 99% simultaneous confidence, the estimates still certify which endpoint law has lower future loss. More generally, if the estimated future values have uniform error at most δ and the selected law is within ξ of the estimated optimum, its true future loss is within $2\delta + \xi$ of the best realizable law with those moments. **The future value identifies both the stochastic information hidden by the summary and the law that uses it to improve learning.**

5 Consequences and modern-system realization

The exact systems establish when unresolved stochastic structure can matter and when a learner can control it. Does that structure change neural learning trajectories? Can it alter the evidence an adaptive learner later receives? Is substantial law freedom present in a modern RLVR update? We test these three consequences separately. Together, they show that distinctions hidden by mean and covariance are measurable beyond the exact constructions.

Neural path consequence. To isolate the effect of gradient-law shape, we change the noise distribution while preserving its conditional mean and full covariance. For a finite-data gradient $g(\theta; Z)$, write $\bar{g}(\theta) = \mathbb{E}[g(\theta; Z)]$ and $\xi = g - \bar{g}$, and define the variance-preserving (VP) gradient

$$\widetilde{g}_b(\theta) = \bar{g}(\theta) + \frac{1}{\sqrt{b}} \sum_{j=1}^{b} \xi_j(\theta), \tag{7}$$

where the ξ_j are conditionally IID. VP-1 is ordinary batch-one SGD. Increasing b makes the residual law progressively more Gaussian, while the factor $1/\sqrt{b}$ keeps its covariance fixed: $\mathbb{E}[\widetilde{g}_b \mid \theta] = \bar{g}(\theta),\ \mathrm{Cov}(\widetilde{g}_b \mid \theta) = \mathrm{Cov}(g \mid \theta)$. Centered cumulants of order $r \geq 3$ scale as $b^{1-r/2}$. Thus the drift and full diffusion agree at every common learner state, while higher-order law structure changes. Related stochastic-gradient constructions also manipulate noise (Li et al., 2021; Xie et al., 2021), and moment-preserving transform freedom appears in ensemble methods (Sakov & Oke, 2008).

The outcome is the probability that evaluation loss crosses a specified boundary during 200 training updates. In a 793-parameter tanh classifier, path risk rises from 2.40% under ordinary SGD to 5.46% under VP-4 and 8.05% under VP-16, toward the fixed covariance-matched Gaussian prediction of 9.95%. All 20 independently generated systems move toward that prediction; VP-16 closes 74.1% of the mean per-system absolute gap.

The pattern holds across architectures and data. In a ten-class ReLU CNN, VP-16 closes 59.3% of the Gaussian gap, with all 20 systems moving toward the prediction. On Fashion-MNIST, whose boundary is defined independently of the Gaussian comparator, it closes 74.7% of the gap, a mean absolute reduction of 13.37 percentage points; 19 of 20 systems move toward the prediction. **Across three nonlinear neural settings, changing the update law while preserving its conditional mean and full covariance moves finite-horizon path risk toward the fixed Gaussian prediction in 59 of 60 systems.**

Closed-loop evidence consequence. A second question arises when the learner's policy determines what data it will encounter next. Can matched local drift and covariance lead to different future evidence? We use a three-action Bernoulli policy-gradient learner and two laws constructed from IID complete gradient estimates. More generally, if G_i have common conditional mean $\bar{G}$ and covariance C, a coefficient law $G_c = \sum_i c_i G_i$ preserves both when $\sum_i c_i = 1,\ \sum_i c_i^2 = 1$. Here Native applies one complete estimate, $G_N = G_0$, while Shape applies $G_S = (G_1 + G_2 + G_3 - G_4)/2$. Shape combines four independent complete estimates into one applied update. These laws have exactly the same conditional mean and full covariance at every common policy state, while their higher-order structure differs. VP and coefficient-law matching are different interventions with the same purpose: they move within a fixed local summary class.

After 32 successive updates under each law, we freeze the learned policies and ask whether nine future queries supply all three action-specific evidence types. We selected one operating regime from 75 development settings and froze it before independent confirmation. Across 2^{21} fresh trajectories per law, Shape increases the chance of missing at least one evidence type from 25.88% under Native to 28.96%: a 3.074 percentage-point difference (11.88% relative; 99% confidence interval $[3.028, 3.120]$ percentage points). The confirmed effect is within 0.009 percentage points of the 3.066-point development estimate, and fractional evidence coverage falls by 1.738 percentage points.

Repeated updates produce different distributions over learned policies, and those policies govern the evidence available after training. **Matching the conditional gradient mean and full covariance at every common policy state does not fix the evidence the learner can obtain next.**

Real RLVR stochastic resolution. Substantial unresolved law freedom also exists inside a modern language-model training update. In Qwen2.5-1.5B-Instruct under RLVR (Yang et al., 2025), we compare Native and Shape at exactly the same model state and prompt. Native uses one complete four-completion RLOO gradient. Shape combines four independent same-prompt RLOO gradients using Shape, G_S, as defined in paragraph 5. Their conditional gradient mean and full covariance agree exactly by construction. In a prospectively frozen 32-prompt test, the pre-optimizer gradient laws are distinguishable by a paired randomization test ($p = .001$). The difference has a direct physical interpretation: Native produces a nonzero gradient with probability 78.66%, while Shape does so with probability 98.68%. This 20.02 percentage-point shift has the same direction for all 32 prompts. Even conditional on a nonzero gradient, the projected update distributions differ. **Equal conditional mean and covariance conceal both a large difference in how often the learner moves and a difference in the moves it makes.**

The distinction survives optimizer processing. Starting from a nonzero AdamW state produced by 512 ordinary RLVR updates (Loshchilov & Hutter, 2019), we pass both laws through the same global clipping rule and frozen AdamW transition. Across 4,096 fresh tuples from 64 prompts, the final applied-update distributions remain distinguishable ($p = .001$). The normalized discrepancy is

0.26637 at the raw-gradient stage and 0.29402 at the final applied-update stage. **Real same-prompt RLVR therefore contains substantial stochastic-law freedom beyond mean and covariance, and realistic history-dependent optimizer processing preserves a measurable distinction between the resulting applied updates.**

6 Related work

Value-aware model learning and Value Equivalence ask what an environment model must preserve for planning (Farahmand et al., 2017; Farahmand, 2018; Grimm et al., 2020; 2021; 2022; Grimm, 2022; Ayoub et al., 2020; Arumugam & Van Roy, 2022; Kastner et al., 2023). We ask which distinctions in the learner's own stochastic update law remain visible to a specified future learning outcome. We characterize the resulting equivalence, the exact ambiguity left by a retained summary, and the control available when unresolved laws are realizable choices.

Weak approximation, stochastic modified equations, and goal-oriented stochastic dynamics study errors for selected observables; generalized moment theory studies what partial distributional information can determine (Talay & Tubaro, 1990; Chatterjee, 2006; Li et al., 2019; Ankirchner & Perko, 2024; Li et al., 2021; Zou et al., 2026; Kemperman, 1968; Winkler, 1988; Tardella, 2001; Bertsimas & Popescu, 2005; Han et al., 2018). Low-order descriptions can be accurate in particular SGD regimes, while richer stochastic flows retain additional statistics (Bhattacharya & Jiang, 2026; Gess et al., 2024). Our learning-specific criterion is the future value that continued learning assigns to today's possible updates. A summary resolves the future question exactly when it can represent that value; no fixed finite moment order does so universally.

Non-Gaussian stochastic structure is known to affect SGD stability, escape, fluctuations, and long-horizon behavior (Simsekli et al., 2019; Nguyen et al., 2019; Bajovic et al., 2023; Azizian et al., 2024; Ma & Ying, 2021; Battash et al., 2024; Liu et al., 2021; Block et al., 2024). We identify when laws with the same conditional mean and full covariance remain distinguishable by future learning, quantify that distinction, and demonstrate its neural consequences.

Adaptive sampling and gradient-noise design already offer ways to influence optimization (Gopal, 2016; Gower et al., 2019; Wu et al., 2020; Xie et al., 2021; Chung et al., 2021; Mignacco & Mori, 2026). Decision-dependent stochastic approximation and performative prediction study data distributions altered by decisions (Drusvyatskiy & Xiao, 2023; Cutler et al., 2024; Mendler-Dünner et al., 2020). Here, the controlled choice lies within a realizable family of update laws sharing the same local mean and full covariance. Its best member can depend on the learner's state and on what happens after the next update; in closed-loop learning, the chosen law also changes future evidence. Our exact systems establish and audit this control over complete feasible families, while the neural and Qwen results show consequential unresolved structure in larger learners. Scalable future-value estimation and law selection remain open.

7 Conclusion: the future question defines the stochastic resolution

Two stochastic learning rules are equivalent only relative to a future learning question. We characterize exactly when a summary of today's update resolves that question, how much predictive ambiguity it leaves, and how that ambiguity propagates through learning. No fixed finite moment order suffices for every future question, even in smooth, strongly convex, contractive SGD. The unresolved structure matters in practice. Summary-matched but distinct laws change neural learning trajectories and the evidence an adaptive learner later encounters; in same-prompt RLVR, substantial law differences survive clipping and AdamW. When these laws are realizable choices, the hidden structure also becomes control: choosing a law for future loss improves learning in our exact systems, even where choosing for immediate loss does not. The resulting design principle is: specify the future question, identify the distinctions in today's update law that it can see, and choose among realizable laws accordingly. Different future learning questions induce different stochastic effective theories.

REPRODUCIBILITY STATEMENT

The supplementary appendix gives complete proofs, exact learning-system constructions, and the experimental protocols underlying the reported results. It specifies sampling laws, evaluation quantities, inferential units, sample sizes, and independent validation procedures. The supplementary materials also provide figure-ready data, reproduction scripts, and manifests linking reported results to their supporting artifacts.

AI USE STATEMENT

Generative AI tools assisted with drafting and polishing selected passages, discovering and checking literature beyond the immediately adjacent works. The research ideas, theory, proofs, experimental designs, analyses, and reported findings are the work of the human authors. The authors verified the AI-assisted text and references and take responsibility for the entire submission.

REFERENCES

Stefan Ankirchner and Stefan Perko. A comparison of continuous-time approximations to stochastic gradient descent. *Journal of Machine Learning Research*, 25(13):1–55, 2024. URL https://www.jmlr.org/papers/v25/23-0237.html.

Dilip Arumugam and Benjamin Van Roy. Deciding what to model: Value-equivalent sampling for reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-0656. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/3b18d368150474ac6fc9bb665d3eb3da-Abstract-Conference.html.

Alex Ayoub, Zeyu Jia, Csaba Szepesvari, Mengdi Wang, and Lin Yang. Model-based reinforcement learning with value-targeted regression. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 463–474. PMLR, 2020. URL https://proceedings.mlr.press/v119/ayoub20a.html.

Waïss Azizian, Franck Iutzeler, Jerome Malick, and Panayotis Mertikopoulos. What is the long-run distribution of stochastic gradient descent? a large deviations analysis. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 2168–2229. PMLR, 2024. URL https://proceedings.mlr.press/v235/azizian24a.html.

Dragana Bajovic, Dusan Jakovetic, and Soummya Kar. Large deviations rates for stochastic gradient descent with strongly convex functions. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 10095–10111. PMLR, 2023. URL https://proceedings.mlr.press/v206/bajovic23a.html.

Barak Battash, Lior Wolf, and Ofir Lindenbaum. Revisiting the noise model of stochastic gradient descent. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 4780–4788. PMLR, 2024. URL https://proceedings.mlr.press/v238/battash24a.html.

Dimitris Bertsimas and Ioana Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804, 2005. doi: 10.1137/S1052623401399903.

Riddhiman Bhattacharya and Tiefeng Jiang. Non-asymptotic analysis of online noisy stochastic gradient descent. *Advances in Applied Probability*, 58:1226–1275, 2026. doi: 10.1017/apr.2026.10056.

Adam Block, Dylan Foster, Akshay Krishnamurthy, Max Simchowitz, and Cyril Zhang. Butterfly effects of sgd noise: Error amplification in behavior cloning and autoregression. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/f8b06ac8601473c3232aec9facbe79ad-Abstract-Conference.html.

Sourav Chatterjee. A generalization of the lindeberg principle. *The Annals of Probability*, 34(6):2061–2076, 2006. doi: 10.1214/009117906000000575.

Xiang Cheng, Dong Yin, Peter Bartlett, and Michael Jordan. Stochastic gradient and langevin processes. In *ICML*, 2020.

Wesley Chung, Valentin Thomas, Marlos C. Machado, and Nicolas Le Roux. Beyond variance reduction: Understanding the true impact of baselines on policy optimization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1999–2009. PMLR, 2021. URL https://proceedings.mlr.press/v139/chung21a.html.

Joshua Cutler, Mateo Díaz, and Dmitriy Drusvyatskiy. Stochastic approximation with decision-dependent distributions: Asymptotic normality and optimality. *Journal of Machine Learning Research*, 25(90):1–49, 2024. URL https://www.jmlr.org/papers/v25/22-0832.html.

Dmitriy Drusvyatskiy and Lin Xiao. Stochastic optimization with decision-dependent distributions. *Mathematics of Operations Research*, 48(2):954–998, 2023. doi: 10.1287/moor.2022.1287.

Amir-Massoud Farahmand. Iterative value-aware model learning. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/7a2347d96752880e3d58d72e9813cc14-Abstract.html.

Amir-Massoud Farahmand, Andre Barreto, and Daniel Nikovski. Value-aware loss function for model-based reinforcement learning. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1486–1494. PMLR, 2017. URL https://proceedings.mlr.press/v54/farahmand17a.html.

Benjamin Gess, Sebastian Kassing, and Vitalii Konarovskyi. Stochastic modified flows, mean-field limits and dynamics of stochastic gradient descent. *Journal of Machine Learning Research*, 25(30):1–27, 2024. URL https://jmlr.org/papers/v25/23-0220.html.

Reza Gheissari and Aukosh Jagannath. Universality of high-dimensional scaling limits of stochastic gradient descent (extended abstract). In *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 2638–2638. PMLR, 2026. URL https://proceedings.mlr.press/v336/gheissari26a.html. Full version: arXiv:2512.13634.

Siddharth Gopal. Adaptive sampling for sgd by exploiting side information. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 364–372. PMLR, 2016. URL https://proceedings.mlr.press/v48/gopal16.html.

Robert Mansel Gower, Nicolas Loizou, Xun Qian, Alibek Sailanbayev, Egor Shulgin, and Peter Richtárik. SGD: General analysis and improved rates. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 5200–5209. PMLR, 2019. URL https://proceedings.mlr.press/v97/qian19b.html.

Christopher Grimm, André Barreto, Satinder P. Singh, and David Silver. The value equivalence principle for model-based reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/3bb585ea00014b0e3ebe4c6dd165a358-Abstract.html.

Christopher Grimm, André Barreto, Gregory Farquhar, David Silver, and Satinder P. Singh. Proper value equivalence. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/400e5e6a7ce0c754f281525fae75a873-Abstract.html.

Christopher Grimm, André Barreto, and Satinder P. Singh. Approximate value equivalence. In *Advances in Neural Information Processing Systems*, volume 35, pp. 33029–33040, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/d53538ba21c05fa361d2b21704172753-Abstract-Conference.html.

Christopher Ryan Grimm. *The Value Equivalence Principle for Model-Based Reinforcement Learning*. PhD thesis, University of Michigan, 2022. URL https://deepblue.lib.umich.edu/bitstream/handle/2027.42/174289/crgrimm_1.pdf?sequence=1.

Yanjun Han, Jiantao Jiao, and Tsachy Weissman. Local moment matching: A unified methodology for symmetric functional estimation and distribution estimation under wasserstein distance. In *Proceedings of the 31st Conference on Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 3189–3221. PMLR, 2018. URL https://proceedings.mlr.press/v75/han18b.html.

Tyler Kastner, Murat A. Erdogdu, and Amir-Massoud Farahmand. Distributional model equivalence for risk-sensitive reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2467. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/b0cd0e8027309ea050951e758b70d60e-Abstract-Conference.html.

J. H. B. Kemperman. The general moment problem, a geometric approach. *The Annals of Mathematical Statistics*, 39(1):93–122, 1968. doi: 10.1214/aoms/1177698508.

Qianxiao Li, Cheng Tai, and Weinan E. Stochastic modified equations and dynamics of stochastic gradient algorithms i: Mathematical foundations. *JMLR*, 20(40):1–47, 2019.

Zhiyuan Li, Sadhika Malladi, and Sanjeev Arora. On the validity of modeling sgd with stochastic differential equations (sdes). In *NeurIPS*, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/69f62956429865909921fa916d61c1f8-Abstract.html.

Kangqiao Liu, Liu Ziyin, and Masahito Ueda. Noise and fluctuation of finite learning rate stochastic gradient descent. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7045–7056. PMLR, 2021. URL https://proceedings.mlr.press/v139/liu21ad.html.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Chao Ma and Lexing Ying. On linear stability of sgd and input-smoothness of neural networks. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/8c26d2fad09dc76f3ff36b6ea752b0e1-Abstract.html.

Stephan Mandt, Matthew D. Hoffman, and David M. Blei. Stochastic gradient descent as approximate bayesian inference. *Journal of Machine Learning Research*, 18(134):1–35, 2017. URL https://www.jmlr.org/papers/v18/17-214.html.

Celestine Mendler-Dünner, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/33e75ff09dd601bbe69f351039152189-Abstract.html.

Francesca Mignacco and Francesco Mori. A statistical physics framework for optimal learning. *PNAS Nexus*, 5 (6):pgag182, 2026. doi: 10.1093/pnasnexus/pgag182.

Thanh Huy Nguyen, Umut Simsekli, Mert Gurbuzbalaban, and Gael Richard. First exit time analysis of stochastic gradient descent under heavy-tailed gradient noise. In *NeurIPS*, 2019.

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*, pp. 7599–7609. PMLR, 2020. URL https://proceedings.mlr.press/v119/perdomo20a.html.

Pavel Sakov and Peter R. Oke. Implications of the form of the ensemble transformation in the ensemble square root filters. *Monthly Weather Review*, 136(3):1042–1053, 2008. doi: 10.1175/2007MWR2021.1.

Umut Simsekli, Levent Sagun, and Mert Gurbuzbalaban. A tail-index analysis of stochastic gradient noise in deep neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 5827–5837. PMLR, 2019. URL https://proceedings.mlr.press/v97/simsekli19a.html.

Denis Talay and Luciano Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8(4):483–509, 1990. doi: 10.1080/07362999008809220.

Luca Tardella. A note on estimating the diameter of a truncated moment class. *Statistics & Probability Letters*, 54(2):115–124, 2001. doi: 10.1016/S0167-7152(00)00205-4.

Ziyang Wei, Jiaqi Li, Zhipeng Lou, and Wei Biao Wu. Gaussian approximation and concentration of constant learning-rate stochastic gradient descent. In *NeurIPS*, 2025.

Gerhard Winkler. Extreme points of moment sets. *Mathematics of Operations Research*, 13(4):581–587, 1988. doi: 10.1287/moor.13.4.581.

Jingfeng Wu, Wenqing Hu, Haoyi Xiong, Jun Huan, Vladimir Braverman, and Zhanxing Zhu. On the noisy gradient descent that generalizes as SGD. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 10367–10376. PMLR, 2020. URL https://proceedings.mlr.press/v119/wu20c.html.

Zeke Xie, Li Yuan, Zhanxing Zhu, and Masashi Sugiyama. Positive-negative momentum: Manipulating stochastic gradient noise to improve generalization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11448–11458. PMLR, 2021. URL https://proceedings.mlr.press/v139/xie21h.html.

An Yang et al. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2025. URL https://arxiv.org/abs/2412.15115.

Joanna Zou, Han Cheng Lie, and Youssef Marzouk. Goal-oriented learning of stochastic differential equations using error bounds on path-space observables. *arXiv preprint arXiv:2603.20467*, 2026. URL https://arxiv.org/abs/2603.20467.