QUERY-POOL GEOMETRY AND BAYES RISK IN NOISY AFFINE PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

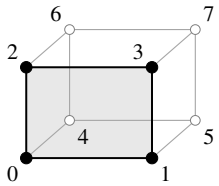
We show that query-pool geometry improves optimal adaptive prediction while preserving the optimal limiting risk of fixed acquisition. We study two regions: one follows a uniformly drawn noisy binary affine rule, and the other has independent fair labels. Let r be the parameter dimension, $0 < \eta < 1/2$ the noise rate, λ the budget ratio, and $c = [1 - h_2(\eta)]^{-1}$ the recovery cost, where h_2 is binary entropy. At budget $\lfloor \lambda r \rfloor$ with $c < \lambda < c + 1$, uniform random pools of size r^2 give the same optimal limiting Bayes Brier risk for fixed, two-batch, and fully adaptive acquisition, including full-pool inspection. A prescribed subcube with the same pool cardinality preserves the fixed-design limit and lets two batches attain the risk when the affine region is known. For the lower bound, we combine uniform early-evidence control with prediction bounds at outcome-dependent counts. The structured policy uses $o(r)$ verification labels before recovery. We study the verification-continuation tradeoff through exact acquisition studies and paired experiments, and establish the separation between acquisition classes under positive noise through finite-risk inequalities. The implementation is available at `https://anonymous.4open.science/r/LATC-74F7`.

1 INTRODUCTION

Suppose a learner must locate a predictive subpopulation and estimate its rule using one label budget. The regional evidence collected early guides where subsequent labels improve prediction. We ask how query-pool geometry affects when this evidence becomes available, holding the target distribution, noise, public regional information, pool cardinality, and budget fixed.

Figure 1 shows two statistical roles of a query set. The face $\{0, 1, 2, 3\}$ gives three independent affine constraints and a parity check, since its clean labels sum to zero modulo two. The set $\{0, 1, 2, 4\}$ gives four independent constraints, and its labels are jointly fair under both regional models. Under independent noise of rate η, a length-k check has expected sign $(1 - 2\eta)^k$ in the affine region and zero in the independent region. The same number of queries can therefore carry different amounts of regional evidence and parameter information.

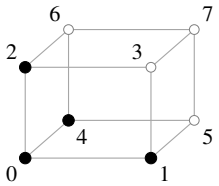
(a) $\{0, 1, 2, 3\}$: an affine relation



$0 \oplus 1 \oplus 2 \oplus 3 = 0$

Augmented rank 3; one parity check

(b) $\{0, 1, 2, 4\}$: independent rows



$\mathrm{rank}_{\mathbb{F}_2}\{(1, x^\top) : x \in \{0, 1, 2, 4\}\} = 4$

Four independent parameter constraints

Figure 1: Four queries with different statistical roles. The left face gives three independent affine constraints and a four-label parity check. The right set gives four independent constraints. Filled points are queried, and integers encode binary coordinates.

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.

Our main result, Theorem 1, compares optimal risks across acquisition classes for recovery cost c and budget ratio λ. Fixed, two-batch, and fully adaptive acquisition have the same limiting risk on random pools throughout $c < \lambda < c + 1$, including full-pool inspection. At the same cardinality and budget, a prescribed subcube preserves the fixed-design limit and lets two batches attain the risk with the affine region known. Table 1 expresses this gain in terms of public regional uncertainty.

A subcube fixes some binary coordinates and varies the others over all configurations. Its location is public and independent of the predictive region; the rest of the pool is sampled uniformly. Using the subcube's parity relations, we can verify the region with a sublinear number of labels before a linear recovery continuation.

For the converse, we compare the label counts needed for verification and prediction. On random pools, regional evidence vanishes below nearly r labels, and useful prediction requires a count near cr. Under the shared budget, the first region reaching the evidence cutoff is the only one that can also exceed the recovery threshold. Lemma 2 covers counts that depend on observed labels, and Proposition 1 gives a risk bound for every adaptive policy.

2 Related Work

Prediction-oriented acquisition evaluates observations through the final predictor, as in variance reduction (Cohn et al., 1996), transductive design (Yu et al., 2006; Hübotter et al., 2024), and expected predictive information gain (Smith et al., 2023). Bayesian and loss-sensitive design similarly assess observations through terminal decisions (Chaloner & Verdinelli, 1995; Krause & Guestrin, 2009; Huang et al., 2026; Sloman et al., 2023). In our setting, choosing where to allocate labels is itself an acquisition decision: early labels guide where estimation continues, and the final predictor uses all acquired labels.

Algebraic self-testing (Blum et al., 1990) and active property testing (Balcan et al., 2012; Alon et al., 2016; Blum & Hu, 2018) study evidence of structural consistency. Pool and stream models, minimax analysis, and batch learning connect query access and acquisition rounds to label requirements (Sabato & Hess, 2018; Hanneke & Yang, 2015; Gentile et al., 2024). Our model couples a regional consistency test to learning parity with noise (Blum et al., 2003; Feldman, 2007). Since the tasks share a budget, prediction depends on how early and how accurately the region is selected.

Our prediction bound uses an information-spectrum likelihood comparison (Polyanskiy et al., 2010). Noisy Bayesian selection and adaptive submodularity also analyze sequential decisions (Golovin et al., 2010; Golovin & Krause, 2011). We apply the comparison to a posterior prediction score weighted by the terminal count event, so the dependence between allocation and outcomes is retained. A uniform bound over the pool covers query sets chosen after full-pool inspection. These two steps give the risks for each acquisition class; further connections are discussed in Appendix J.

3 Model, Access, and Prediction

3.1 Regional Model and Acquisition Classes

Each of the two known regions contains $m = 2^d$ coordinates in $\mathbb{F}_2^d$, where $\mathbb{F}_2$ is the binary field. The affine parameter dimension is $r = d + 1$, and the total population is $N = 2m$. The binary indicator vector $Z = (Z_1, Z_2)$ satisfies $Z_1 + Z_2 = 1$, with $Z_j = 1$ identifying the affine region. Public evidence W determines $p(W) = \mathbb{P}(Z_1 = 1 \mid W)$, the probability that region one is affine. Regional label arrays are conditionally independent given (W, Z), and affine-region labels satisfy

$$Y_{j,x} = \beta_j + u_j^\top x + \xi_{j,x}, \qquad \theta_j = (\beta_j, u_j^\top)^\top \sim \mathrm{Uniform}(\mathbb{F}_2^r), \qquad \xi_{j,x} \sim \mathrm{Bernoulli}(\eta), \quad (1)$$

where the intercept $\beta_j \in \mathbb{F}_2$ and slope $u_j \in \mathbb{F}_2^d$ define the rule, with addition modulo two. The errors $\xi_{j,x}$ have known rate η, are mutually independent, and are independent of the parameter. Labels in the other region are independent fair bits, and each coordinate retains its sampled label throughout acquisition. The model allows $0 \le \eta < 1/2$, with fixed positive noise in the main theorem.

A public pool $\mathcal{P}_j$ specifies the available coordinates in region j. Conditional on W, the pools are independent of the indicators, parameters, and errors. A policy observes $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2)$ and acquires

exactly b distinct labels from $\mathcal{I}_{\mathcal{P}} = \bigcup_j (\{j\} \times \mathcal{P}_j)$, the available region-coordinate pairs. Fixed acquisition selects all queries before observing labels. Two-batch acquisition selects its second set after the first batch. A fully adaptive policy can update each next query after observing earlier labels. Both regions have weight $1/2$ in the population Brier loss (Glenn et al., 1950):

$$R(\widehat{p}, Y) = \frac{1}{2m} \sum_{j=1}^{2} \sum_{x} (\widehat{p}_{j,x} - Y_{j,x})^2. \tag{2}$$

The Bayes predictor assigns each coordinate its conditional probability of label one and predicts acquired labels exactly. We write R^*_{fix}, R^*_2, and R^*_{ad} for the optimal risks of the three acquisition classes, averaged over public information, pools, labels, and policy randomization. Computation is unrestricted in each class. For $b < N$, averaging over unqueried targets multiplies population risk by $N/(N-b)$.

Write $\delta = 1 - 2\eta$ for noise attenuation and $C = 1 - h_2(\eta)$ for binary-channel capacity, where $h_2(a) = -a \log_2 a - (1-a)\log_2(1-a)$ and $0 \log_2 0 = 0$. The reciprocal $c = C^{-1}$ is the recovery cost per parameter dimension. If the affine region and its parameter are known, the population prediction gain is $g = \delta^2/8$ over risk $1/4$. Using public evidence alone, the smallest regional selection error is $e_{\text{pub}} = \mathbb{E} \min\{p(W), 1 - p(W)\}$. Let $\widetilde{W} = (W, \mathcal{P})$ collect the public information, and $\mathcal{H}$ denote the acquired query-label history.

3.2 Pool Relations and Posterior Prediction

To express the evidence carried by a query set, let Q contain n regional coordinates and let $A_Q \in \mathbb{F}_2^{n \times r}$ have augmented rows $(1, x^\top)$. The dependency space $\mathcal{D}(Q) = \ker(A_Q^\top)$ records the parity checks among these rows. Writing $P_{1,Q}$ and $P_{0,Q}$ for the affine and independent label laws, respectively, we obtain the likelihood ratio

$$L_Q(y) = \frac{P_{1,Q}(y)}{P_{0,Q}(y)} = \sum_{v \in \mathcal{D}(Q)} \delta^{|v|} (-1)^{v^\top y}. \tag{3}$$

The outcome vector is $y \in \mathbb{F}_2^n$, and the Hamming weight $|v|$ counts labels in a parity check. Independent rows give $L_Q = 1$. A unique length-k relation gives equal-prior testing error $1/2 - \delta^k/4$. Along an adaptive path, query-selection factors agree under both models and cancel, so the selected coordinates and outcomes have this same likelihood.

Let $Q_j(\mathcal{H})$ be region j's acquired coordinates and $f_\theta(x) = (-1)^{\beta + u^\top x}$ the clean affine sign. The predictor depends on regional confidence $\gamma_j = \mathbb{P}(Z_j = 1 \mid \widetilde{W}, \mathcal{H})$ and conditional sign mean $M_{j,x} = \mathbb{E}[f_{\theta_j}(x) \mid Z_j = 1, \widetilde{W}, \mathcal{H}]$ through

$$\gamma_1 = \frac{p(W) L_{Q_1}}{p(W) L_{Q_1} + (1 - p(W)) L_{Q_2}}, \quad \gamma_2 = 1 - \gamma_1, \quad \widehat{p}_{j,x} = \frac{1}{2} - \frac{\gamma_j \delta}{2} M_{j,x} \quad (x \notin Q_j). \tag{4}$$

Target noise contributes the factor δ, and acquired coordinates have prediction $Y_{j,x}$. At zero affine likelihood, we use the convention $\gamma_j = M_{j,x} = 0$. For augmented target $\widetilde{x} = (1, x^\top)^\top$ and $L_Q(y) > 0$, the conditional sign mean is

$$M_x(y) = \frac{\displaystyle\sum_{v : A_Q^\top v = \widetilde{x}} \delta^{|v|} (-1)^{v^\top y}}{L_Q(y)}. \tag{5}$$

Note that dependencies among queried rows determine regional evidence through the denominator; target representations determine the numerator. We derive these identities and extend them to adaptive transcripts in Appendix C.

4 Query-Pool Geometry and Optimal Risk

For sufficiently large r, a random regional pool contains $q_r = r^2$ coordinates sampled uniformly without replacement from $\mathbb{F}_2^d$. A structured pool contains a prescribed subcube B_j of size $t_r =$

$2^{\lceil \log_2((\log r)^2)\rceil}$, where log is the natural logarithm. Its other $q_r - t_r$ coordinates are sampled uniformly without replacement from the complement of B_j and assigned an independent public random order. Superscripts rand and str distinguish these laws. Since the verification count is $t_r = \Theta((\log r)^2)$, a linear number of labels remains for continuation.

Theorem 1 (Optimal risk under random and structured access)**.** *Fix the law of (W, Z) in the two-region model with weights $1/2$, conditional uniform affine parameters, and common noise $0 < \eta < 1/2$. For $b_r = \lfloor \lambda r \rfloor$ and $\lambda \in [0, c+1) \setminus \{c\}$,*

$$\lim R_{\mathrm{fix}}^{*,\mathrm{rand}} = \lim R_2^{*,\mathrm{rand}} = \lim R_{\mathrm{ad}}^{*,\mathrm{rand}} = \lim R_{\mathrm{fix}}^{*,\mathrm{str}} = \begin{cases} 1/4, & 0 \le \lambda < c, \\ 1/4 - g(1 - e_{\mathrm{pub}}), & c < \lambda < c+1, \end{cases}$$
$$\lim R_2^{*,\mathrm{str}} = \lim R_{\mathrm{ad}}^{*,\mathrm{str}} = \begin{cases} 1/4, & 0 \le \lambda < c, \\ 1/4 - g, & c < \lambda < c+1. \end{cases} \tag{6}$$

All limits are taken as $r \to \infty$.

For $c < \lambda < c+1$, fixed acquisition has the same optimum under both pool laws, and structured two-batch acquisition gains ge_{pub}. Table 1 writes these limits using public selection accuracy $\alpha = 1 - e_{\mathrm{pub}}$.

Pool law	Fixed	Two-batch	Fully adaptive
Uniform random	$1/4 - g\alpha$	$1/4 - g\alpha$	$1/4 - g\alpha$
Prescribed subcube	$1/4 - g\alpha$	$1/4 - g$	$1/4 - g$

Table 1: Exact limiting Bayes Brier risks on $c < \lambda < c+1$, with public selection accuracy $\alpha = 1 - e_{\mathrm{pub}}$. Both pool laws have cardinality r^2 and use the same label budget.

This budget interval comes from two thresholds on label counts. For fixed $\varepsilon \in (0, 1)$, regional evidence on random pools vanishes below $(1-\varepsilon)r$ labels, and prediction gain vanishes below $c(1-\varepsilon)r$. For every $\lambda < c+1$, we can choose margins so that only the region selected at the evidence cutoff can exceed the prediction cutoff. The critical-budget limits and bounds at larger budgets are given in Appendix F.6.

5 Verification, Recovery, and Bayes Risk

5.1 Uniform Evidence over the Random Pool

We first control regional evidence for every early query set in the public pool. We then bound prediction at terminal counts chosen using the observed labels.

Lemma 1 (Uniform early evidence)**.** *Fix $\varepsilon_{\mathrm{early}} \in (0, 1)$ and $T = \lfloor (1 - \varepsilon_{\mathrm{early}})r \rfloor$. For a random pool of size $q_r = r^2$, with probability tending to one, every set of at most T rows has nullity below $k_r^{\mathrm{null}} = O(\log r)$, and every nonzero dependency in the full pool has length greater than $\ell_r^{\mathrm{dep}} = \Omega(r/\log r)$. Nullity is the dimension of the dependency space. At fixed $0 < \eta < 1/2$,*

$$\sup_{\substack{Q \subseteq \mathcal{P}_j,\ |Q| \le T \\ y \in \mathbb{F}_2^{|Q|}}} |L_Q(y) - 1| \le 2^{k_r^{\mathrm{null}}} \delta^{\ell_r^{\mathrm{dep}}} \longrightarrow 0. \tag{7}$$

In a relation, the last coordinate is determined by the others. Several independent relations constrain the points to a low-dimensional affine space. Lemma 3 bounds the probabilities of both events over the full pool. Long checks carry attenuated evidence, and the nullity bound limits their combined contribution. Lemmas 4 and 5 sharpen this argument for the finite bounds in Section 6.3.

We stop the history when a regional count first reaches T and call that region I. If both terminal counts stay below T, we choose I publicly. Under the all-independent fair-label reference law, the stopped history is independent of Z given public information, so selection accuracy is at most $1 - e_{\mathrm{pub}}$. Selection and stopping factors cancel in the likelihood ratio. Equation 7 then gives

$$\mathbb{P}(Z_I = 1) \le 1 - e_{\mathrm{pub}} + o(1). \tag{8}$$

5.2 Prediction at an Outcome-Dependent Count

For a lower bound, we reveal Z to the predictor after acquisition. This improves the information available for prediction. We express the resulting indicator-informed risk through the posterior clean-sign mean and its population squared average:

$$M_{j,x}^{\mathrm{or}} = \mathbb{E}[f_{\theta_j}(x) \mid W, Z, \mathcal{P}, \mathcal{H}], \qquad \mathcal{U}_j^{\mathrm{or}} = \frac{1}{m}\sum_x (M_{j,x}^{\mathrm{or}})^2. \tag{9}$$

The superscript or denotes the revealed indicators, and $\mathcal{U}_j^{\mathrm{or}} \in [0,1]$ measures predictable affine-sign energy. Conditional regional independence gives $M_{j,x}^{\mathrm{or}} = M_{j,x}$ on $Z_j = 1$. For T distinct regional observations, let $B_T \sim \mathrm{Binomial}(T, \eta)$ count label errors. The supplied-parameter log-likelihood against fair labels has distribution S_T, which defines the prediction bound

$$\begin{aligned} S_T &= (T - B_T)\log_2(2(1-\eta)) + B_T \log_2(2\eta), \\ \Psi_\eta(T, r) &= \min\left\{1, \inf_{\zeta \in \mathbb{R}} \left[2^{\zeta - r} + \mathbb{P}(S_T > \zeta)\right]\right\}. \end{aligned} \tag{10}$$

We split the likelihood at a real threshold ζ into a bounded contribution and an upper tail. The resulting $\Psi_\eta(T, r)$ bounds expected prediction energy. At zero noise, $\Psi_0(T, r) = \min(1, 2^{T-r})$.

Lemma 2 (Conditional affine prediction)**.** *For terminal count $n_j = |Q_j(\mathcal{H})|$, every adaptive policy and every integer $0 \le T \le m$ satisfy, on $Z_j = 1$,*

$$\mathbb{E}[\mathbf{1}\{n_j \le T\}\mathcal{U}_j^{\mathrm{or}} \mid W, Z] \le \Psi_\eta(T, r). \tag{11}$$

The lemma applies conditional on public pools and independent policy randomization, with the outcome-dependent count event inside the expectation. For the affine-parameter posterior q_θ, orthogonality gives $\mathcal{U}^{\mathrm{or}} \le \sum_\theta q_\theta^2$. The expected squared posterior mass equals the expected posterior probability assigned to the true parameter.

For an experiment with exactly T distinct regional observations, let A be a transcript-measurable event and q° a transcript-measurable probability vector on the parameter space. We evaluate the score $G = \mathbf{1}_A q_\theta^\circ$ at the true parameter θ. Under the joint fair-label reference law P_0, the parameter is independent of the transcript, so $\mathbb{E}_0 G \le 2^{-r}$. Since $0 \le G \le 1$, comparison with the joint affine law P gives

$$\mathbb{E}_P G \le 2^{\zeta}\mathbb{E}_0 G + \mathbb{P}_P(S_T > \zeta) \le 2^{\zeta - r} + \mathbb{P}(S_T > \zeta). \tag{12}$$

Optimizing over ζ gives Equation 10. To apply the score at an outcome-dependent count, we use an auxiliary experiment with exactly T regional observations.

Follow the policy until termination or immediately before its $(T+1)$-th regional query. A terminated run records its posterior and count event, then adds distinct regional observations to reach T. Averaging over these additional outcomes preserves the original prediction contribution (Appendix B). Since $\mathbb{E}S_T = TC$, the bound vanishes for $T \le c(1-\varepsilon)r$ with fixed $\varepsilon > 0$.

5.3 Converting Verification into Terminal Risk

Choose an evidence cutoff T and a prediction cutoff T_{rec}, both integers, with $1 \le T \le T_{\mathrm{rec}}$, $T < r$, $T_{\mathrm{rec}} < cr$, and $b_r \le T + T_{\mathrm{rec}}$. Let I be the first region reaching T, with a public choice if both terminal counts stay smaller. Every other region has at most T_{rec} labels under the shared budget, so

$$\sum_j Z_j \mathcal{U}_j^{\mathrm{or}} \le Z_I + \sum_j \mathbf{1}\{n_j \le T_{\mathrm{rec}}\} Z_j \mathcal{U}_j^{\mathrm{or}}. \tag{13}$$

Early evidence bounds the probability that I is affine, and Lemma 2 bounds prediction in regions below the recovery cutoff. Revealing Z after acquisition gives the following lower bound for every policy π:

$$R(\pi) \ge \frac{1}{4} - \frac{b}{4N} - g\,\mathbb{E}\sum_j Z_j \mathcal{U}_j^{\mathrm{or}}. \tag{14}$$

The correction $b/(4N)$ accounts for acquired targets. Combining these bounds gives the following finite allocation inequality.

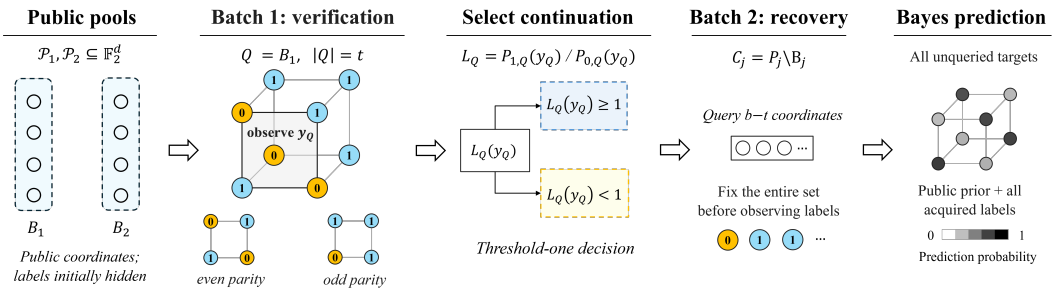


Figure 2: Structured two-batch acquisition. From left to right: public pools, first-batch labels, likelihood-based regional selection, a fixed recovery batch, and Bayes prediction. Colored bits show noisy parities, and grayscale shows prediction probabilities. All second-batch coordinates are selected before observing their labels. Prediction uses the public prior and both batches.

Proposition 1 (Finite Bayes risk bound for adaptive allocation)**.** *In the exclusive two-region model, let* $0 \leq \eta < 1/2$ *and* $1 \leq T \leq T_{\mathrm{rec}} \leq m$*, with* $b \leq T + T_{\mathrm{rec}}$*. Suppose a pool event* $\mathcal{G}_T$*, independent of* (W, Z)*, has probability at least* $1 - \rho_T^{\mathrm{pool}}$*. On this event, assume* $|L_Q(y) - 1| \leq e_T^{\mathrm{ev}}$ *for every regional set of size at most* T *and every outcome. Then*

$$R^*_{\mathrm{ad}}(b) \geq \frac{1}{4} - \frac{b}{4N} - g\left[1 - e_{\mathrm{pub}} + \frac{e_T^{\mathrm{ev}}}{2} + \rho_T^{\mathrm{pool}} + \Psi_\eta(T_{\mathrm{rec}}, r)\right]. \tag{15}$$

For $c < \lambda < c + 1$, we can choose cutoff margins so that the evidence, exceptional-pool, and prediction errors vanish. We attain the lower bound by recovering the rule in the publicly favored region. A fixed structured design chooses its possible high-count region before observing labels, so the same allocation argument gives its matching limit.

5.4 Structured Verification and Recovery

Figure 2 shows the structured policy. Write $\mathcal{P}_j = B_j \,\dot\cup\, \mathcal{C}_j$, with a designated block B_j of size $0 \leq t \leq b$ and a publicly ordered remainder $\mathcal{C}_j$. After observing y_Q on $Q = B_1$, select $\widehat{J} = 1 + \mathbf{1}\{L_Q(y_Q) < 1\}$, assigning ties to region one. The likelihood uses all block labels, including the illustrated noisy parities. Fix the first $b - t$ coordinates of $\mathcal{C}_{\widehat{J}}$ as the second batch before observing their labels. Region-two continuation excludes B_2, and Equation 4 uses both batches and the public prior.

For every public prior, the threshold-one rule minimizes the sum of affine rejection and independent acceptance probabilities. We use it in the attaining policy. A size-t subcube has $2t$ affine restrictions. Comparing its likelihood test with a disagreement test at $\eta < \tau_{\mathrm{test}} < 1/2$ bounds the error sum $\varepsilon_{\mathrm{test}}$ by

$$\varepsilon_{\mathrm{test}} \leq e^{-t\,\mathrm{kl}(\tau_{\mathrm{test}}, \eta)} + 2t\, 2^{-t(1 - h_2(\tau_{\mathrm{test}}))}, \tag{16}$$

where kl is binary relative entropy with natural logarithms. For $t = t_r$, the error sum tends to zero, and the continuation has $n = b_r - t_r = \lambda r - o(r)$ labels. Thus, $n > cr$ for sufficiently large r whenever $\lambda > c$.

For a decoding disagreement threshold $\eta < \tau_{\mathrm{dec}} < 1/2$ and $n + t \leq m$, minimum-distance decoding has error at most $\min(1, \varepsilon_{\mathrm{dec}})$, where

$$\begin{aligned} p_{\mathrm{dist}}(m, n, t) &= \prod_{i=0}^{n-1} \left(1 - \frac{t + i}{m}\right), \\ \varepsilon_{\mathrm{dec}} &= e^{-n\,\mathrm{kl}(\tau_{\mathrm{dec}}, \eta)} + \frac{2^r - 2}{p_{\mathrm{dist}}(m, n, t)} 2^{-n(1 - h_2(\tau_{\mathrm{dec}}))}. \end{aligned} \tag{17}$$

The probability p_{dist} accounts for distinct-coordinate sampling outside the designated block. Conditional on the parameter and block labels, continuation coordinates are uniform on its complement. Since the error bound is uniform in the parameter, it also applies after regional selection. For $\lambda > c$, a decoding threshold sufficiently close to η makes both error terms vanish (Appendix E).

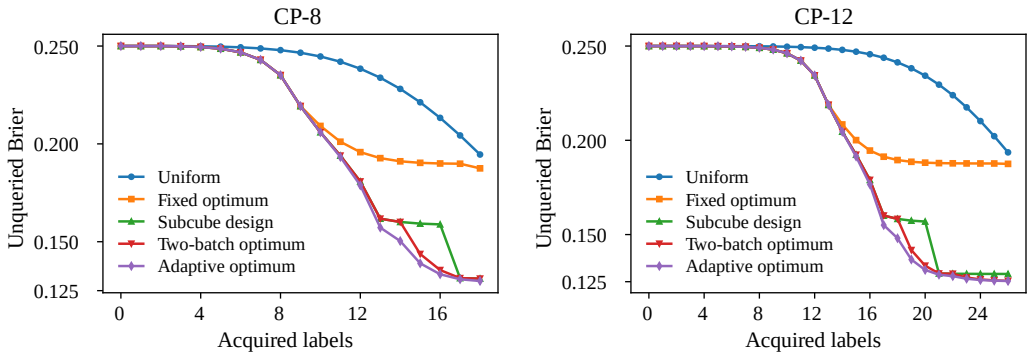


Figure 3: Exact unqueried Brier in noiseless, full-access CP-8 and CP-12. Uniform samples a global b-subset without replacement. Subcube design optimizes its first-set family and the fixed continuation after each outcome. Fixed optimum, Two-batch optimum, and Adaptive optimum give the optimal risks over their respective full acquisition classes.

5.5 Extensions of the Allocation Argument

The same allocation argument applies to bounded regional gains controlled by count-weighted prediction bounds. For a finite sign dictionary f_θ with prior weights ν_θ, define

$$K_{\theta,\theta'} = \frac{1}{m}\sum_x f_\theta(x) f_{\theta'}(x), \qquad \mathfrak{d}_\nu = -\log_2\left(\max_{\theta'}\sum_\theta \nu_\theta (K_{\theta,\theta'})_+\right). \tag{18}$$

The matrix K measures population sign agreement, and $(a)_+ = \max(a, 0)$. The effective dimension $\mathfrak{d}_\nu$ replaces r in Theorem 3 and equals r for uniform affine functions. The general allocation bound is in Appendix A; we illustrate gradual prediction gains in a Gaussian model in Appendix A.5.

For J regions with a common noise rate and positive population weights ω_j summing to one, the public and indicator-informed allocation gains are

$$G_{\mathrm{pub}} = \mathbb{E}\max_j g_j p_j(W), \qquad G_{\mathrm{ind}} = \mathbb{E}\max_j g_j Z_j. \tag{19}$$

The regional gain is $g_j = \omega_j \delta^2/4$, and $p_j(W) = \mathbb{P}(Z_j = 1 \mid W)$ is its public affine probability. In Appendix F.2, we establish limits $1/4 - G_{\mathrm{pub}}$ and $1/4 - G_{\mathrm{ind}}$ on the same budget interval, including dependent indicators. Their difference is the value of indicator-informed allocation, equal to ge_{pub} for two exclusive regions. Equality holds if and only if a public optimizer is optimal in almost every compatible indicator state.

6 Finite Acquisition Studies and Risk Certificates

We examine the result through three finite calculations. In Figure 3, we compute exact acquisition-class optima under noiseless full access. Figure 4 shows how verification and continuation rank affect noisy prediction. In Figure 5, finite risk inequalities under Theorem 1 establish separation between structured two-batch acquisition and every adaptive random-pool policy.

6.1 Noiseless Full-Access Acquisition

CubePair (CP-d) consists of two full-access regions with zero noise and a uniform exclusive prior. Since each regional history is summarized by its count, augmented rank, and consistency, we can optimize the acquisition classes in Figure 3 exactly.

The Subcube design curve has risk R_{sub}. Its first-set family contains the empty set, a subcube in region one, and equal-dimensional subcubes in both regions, with subcube dimension $2 \le d_{\mathrm{sub}} \le \min(d, 5)$ subject to the label budget. For each outcome, we use the optimal fixed continuation. Once the

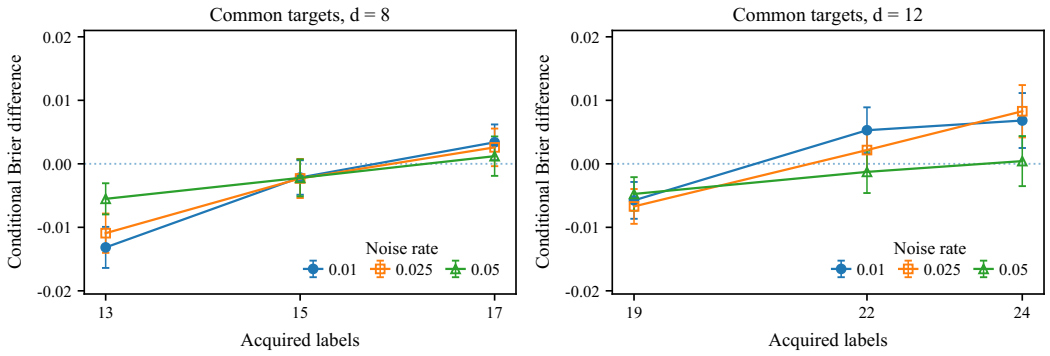


Figure 4: Two-check minus subcube conditional Brier on targets outside both pools. Positive values favor the subcube. Panels correspond to dimensions and curves to noise rates. Pointwise 95% paired Student intervals use 512 independent samples per dimension and noise pair, as reported in Table 14.

budget is large enough for verification and estimation, two batches improve on fixed acquisition, and full adaptation gives further gains. In the full two-batch class, partial subcubes retain useful checks and leave more labels for continuation. Budget curves are given in Appendix K, and we vary public regional evidence in Appendix L.1. In Appendix O, we compare BALD, BatchBALD, EPIG, and one-step Brier on noisy six-point matched-rank pools (Houlsby et al., 2011; Kirsch et al., 2019).

6.2 Checks and Continuation Rank under Noise

The noisy paired study uses $d \in \{8, 12\}$, $\eta \in \{.01, .025, .05\}$, pool size $q = (d+1)^2$, a uniform regional prior, and budgets $\lfloor \lambda(d+1) \rfloor$ for $\lambda \in \{1.5, 1.7, 1.9\}$. Each dimension-noise pair has 512 independent samples shared across designs and budgets. We compare a random prefix, two disjoint four-point checks selected from the full random pool, and a prescribed subcube. Each first batch has eight labels and uses threshold-one selection, continues in public order, and uses the same posterior predictor. The inequality $\binom{q-4}{2} > 2^d - 1$ guarantees two disjoint checks in every pool at these dimensions.

For each sample, define common targets $\mathcal{V}_j^{\mathrm{eval}} = \mathbb{F}_2^d \setminus (\mathcal{P}_j^{\mathrm{rand}} \cup \mathcal{P}_j^{\mathrm{str}})$, with total count $M_{\mathrm{eval}} = |\mathcal{V}_1^{\mathrm{eval}}| + |\mathcal{V}_2^{\mathrm{eval}}|$. This population is fixed across policies and budgets. Let R^{common} denote the realized Brier score on these targets and include both pool pairs in $\mathcal{P}$. Integrating target labels yields

$$\mathbb{E}[R^{\mathrm{common}} \mid \mathcal{H}, \mathcal{P}] = \frac{1}{4} - \frac{\delta^2}{4M_{\mathrm{eval}}} \sum_{j=1}^{2} \sum_{x \in \mathcal{V}_j^{\mathrm{eval}}} \gamma_j^2 M_{j,x}^2. \tag{20}$$

The conditional risk has the same expectation as the realized Brier and reduces variation from target noise. Let $J_\star$ identify the affine region and $k_\star$ denote its acquired augmented rank. For a nonempty acquired set, Equation 5 implies

$$\frac{1}{m} \sum_{x} M_{J_\star, x}^2 \le 2^{k_\star - r}. \tag{21}$$

The acquired span contains $2^{k_\star - 1}$ target rows; posterior sign means are zero elsewhere. For continuation count $n = b - 8$, correct selection of affine region two gives $k_\star \le n$, so continuation rank matters when $n < r$. If region one is affine, the first batch also contributes parameter constraints. Two disjoint checks can give up to six independent constraints, compared with four for the subcube.

Figure 4 shows that two checks have lower mean risk at short continuations. As the continuation grows, the comparison shifts toward the subcube. The intervals are pointwise at each budget. The positive Bonferroni-adjusted endpoint interactions support this increase in the relative value of subcube verification (Appendix L.5).

As an independent control, we search each random pool for an eight-point affine flat, a translate of a three-dimensional subspace, and otherwise use two checks. Every sampled region-one pool at $d = 8$

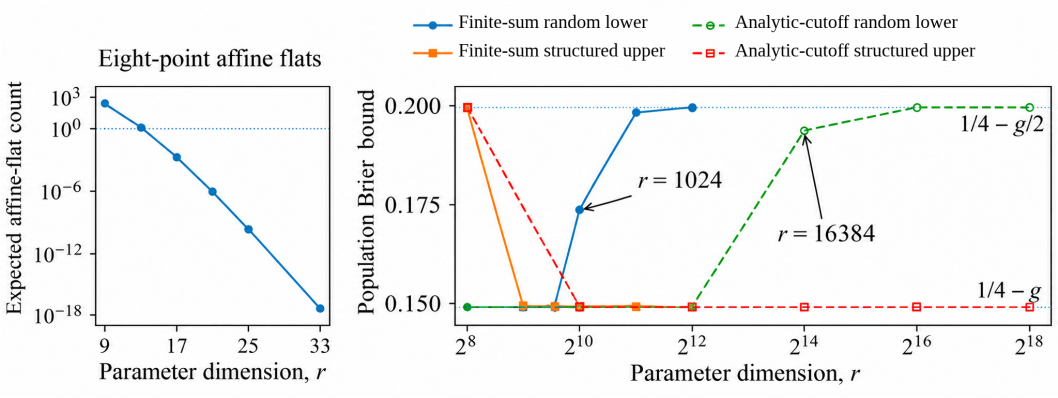


Figure 5: Finite risk bounds at $\eta = .05$, $p = 1/2$, and $b_r = \lfloor (c + 1/2)r \rfloor$. **Left:** expected eight-point affine-flat counts. **Right:** random-pool adaptive lower bounds and structured two-batch upper bounds. Tables 33 and 29 report the finite-sum and analytic-cutoff bounds, respectively. For either calculation, separation holds wherever a lower curve lies above its corresponding upper curve.

contains a flat with the subcube's rank and testing law. Risk contrasts have both signs, and all jointly Holm-adjusted p-values exceed the significance threshold (Appendix P.1).

6.3 Finite Bounds under the Theorem's Assumptions

For population size $m = 2^{r-1}$, we evaluate the theorem's finite risk inequalities using combinatorial probabilities and likelihood tails. At an early-count cutoff $T < r$ and integer support parameter $\kappa \geq 1$, the bounds based on dependency support give

$$e_T^{\mathrm{ev}} = (1 + \delta^{\kappa})^{\lfloor T/\kappa \rfloor} - 1, \qquad \rho_T^{\mathrm{pool}} \leq 2\Gamma_{r,q,T}(\kappa). \tag{22}$$

The parameter κ specifies support per independent dependency, and $\Gamma_{r,q,T}(\kappa)$ is the one-pool failure bound in Equation 75. The resulting event controls every available query set of size at most T. Proposition 1, with $T_{\mathrm{rec}} = b_r - T$, gives a random-pool lower bound L_r. The testing and decoding bounds give a structured two-batch upper bound U_r, so

$$L_r \leq R_{\mathrm{ad}}^{*,\mathrm{rand}}(b_r), \qquad R_2^{*,\mathrm{str}}(b_r) \leq U_r. \tag{23}$$

Both bounds use the same population model, prior, noise, pool cardinality, and budget.

Whenever $L_r > U_r$, structured two-batch acquisition has lower risk than every adaptive random-pool policy. Figure 5 shows that expected flat counts decline and that finite-sum evaluation establishes separation earlier than the analytic cutoff $T = \lfloor 3r/4 \rfloor$. The cutoffs and arithmetic with outward rounding are given in Appendix P.4.

In Appendix Q, we evaluate verification with positive noise at the prescribed subcube sizes. We integrate out the ambient parameter using affine symmetry, quantify sampling uncertainty with exact-binomial intervals, and compare testing with a bound for random access to the whole pool.

7 Discussion and Conclusion

We have shown how query-pool geometry affects when regional evidence can guide a shared label budget. With a prescribed subcube, the learner can verify the region with a sublinear number of labels while preserving both the recovery threshold and the limiting fixed-design risk. The adaptive gain comes from spending the remaining labels on recovery in the affine region.

In Appendix F.7, we relate this limit to $a_\star(\eta)$, defined as the supremum of query ratios at which the optimal testing advantage vanishes. There we establish the bounds $1 \leq a_\star(\eta) \leq c$ and obtain the prediction interval $c < \lambda < c + a_\star(\eta)$. We give the computational costs of exact inference and acquisition in Appendix I.1.

AI USE STATEMENT

We used generative AI tools for language editing and to assist with proof development, including suggesting alternative derivations and checking intermediate mathematical arguments. The authors developed the theorem statements and core proof strategies. All AI-assisted mathematical arguments and proof steps were independently reviewed and verified by the authors before inclusion in the paper. The authors take full responsibility for the contents of the submission.

ETHICS STATEMENT

This work develops mathematical theory and experiments on synthetic populations. The experiments use generated labels and publicly specified query geometries.

REPRODUCIBILITY STATEMENT

The appendices provide the proofs, exact acquisition recursions, sampling distributions, and numerical checks. Appendices H and N derive posterior risks and acquisition-class optima, while Appendix I describes their independent validation. Appendix L.2 specifies the random streams. The paired prediction study and budget interactions appear in Appendices L.4 and L.5, and Appendix P.1 covers the independent affine-flat control. Appendix P.4 gives the integer cutoffs and outward-rounded arithmetic for the finite certificates, and Appendix Q describes the subcube verification experiment and exact-binomial intervals.

REFERENCES

Noga Alon, Rani Hod, and Amit Weinstein. On active and passive testing. *Combinatorics, Probability and Computing*, 25(1):1–20, 2016.

Erdal Arikan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on information Theory*, 55(7): 3051–3073, 2009.

Maria-Florina Balcan, Eric Blais, Avrim Blum, and Liu Yang. Active property testing. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pp. 21–30. IEEE, 2012.

Rémi Bardenet and Odalric-Ambrym Maillard. Concentration inequalities for sampling without replacement. *Bernoulli*, 2015.

David Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pp. 265–272, 1953.

Avrim Blum and Lunjia Hu. Active tolerant testing. In *Conference On Learning Theory*, pp. 474–497. PMLR, 2018.

Avrim Blum, Adam Kalai, and Hal Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. *Journal of the ACM (JACM)*, 50(4):506–519, 2003.

Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proceedings of the twenty-second annual ACM symposium on Theory of computing*, pp. 73–83, 1990.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical science*, pp. 273–304, 1995.

David A Cohn, Zoubin Ghahramani, and Michael I Jordan. Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145, 1996.

Jack Edmonds. Lehman's switching game and a theorem of tutte and nash-williams. *J. Res. Nat. Bur. Standards Sect. B*, 69:73–77, 1965.

Vitaly Feldman. Attribute-efficient and non-adaptive learning of parities and dnf expressions. *Journal of Machine Learning Research*, 8(7), 2007.

Claudio Gentile, Zhilei Wang, and Tong Zhang. Fast rates in pool-based batch active learning. *Journal of Machine Learning Research*, 25(262):1–42, 2024.

W Brier Glenn et al. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011.

Daniel Golovin, Andreas Krause, and Debajyoti Ray. Near-optimal bayesian active learning with noisy observations. *Advances in Neural Information Processing Systems*, 23, 2010.

Venkatesan Guruswami, Andrii Riazanov, and Min Ye. Arikan meets shannon: Polar codes with near-optimal convergence to channel capacity. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 552–564, 2020.

Steve Hanneke and Liu Yang. Minimax analysis of active learning. *The Journal of Machine Learning Research*, 16(1):3487–3602, 2015.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*, 2011.

Zhuoyue Huang, Freddie Bickford Smith, and Tom Rainforth. Loss-driven bayesian active learning. *arXiv preprint arXiv:2604.11995*, 2026.

Jonas Hübotter, Bhavya Sukhija, Lenart Treven, Yarden As, and Andreas Krause. Transductive active learning: Theory and applications. *Advances in Neural Information Processing Systems*, 37: 124686–124755, 2024.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

Andreas Kirsch, Joost Van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. *Advances in neural information processing systems*, 32, 2019.

Adam Klivans, Konstantinos Stavropoulos, and Arsen Vasilyan. Testable learning with distribution shift. In *The Thirty Seventh Annual Conference on Learning Theory*, pp. 2887–2943. PMLR, 2024.

Jannik Kossen, Sebastian Farquhar, Yarin Gal, and Tom Rainforth. Active testing: Sample-efficient model evaluation. In *International Conference on Machine Learning*, pp. 5753–5763. PMLR, 2021.

Jannik Kossen, Sebastian Farquhar, Yarin Gal, and Thomas Rainforth. Active surrogate estimators: An active learning approach to label-efficient model evaluation. *Advances in Neural Information Processing Systems*, 35:24557–24570, 2022.

Andreas Krause and Carlos Guestrin. Optimal value of information in graphical models. *Journal of Artificial Intelligence Research*, 35:557–591, 2009.

Mohammad Naghshvar and Tara Javidi. Active sequential hypothesis testing. *Ann. Statist.*, 2013.

Yury Polyanskiy, H Vincent Poor, and Sergio Verdú. Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory*, 56(5):2307–2359, 2010.

Sivan Sabato and Tom Hess. Interactive algorithms: Pool, stream and precognitive stream. *Journal of Machine Learning Research*, 18(229):1–39, 2018.

Sabina J Sloman, Ayush Bharti, Julien Martinelli, and Samuel Kaski. Bayesian active learning in the presence of nuisance parameters. *arXiv preprint arXiv:2310.14968*, 2023.

Freddie Bickford Smith, Andreas Kirsch, Sebastian Farquhar, Yarin Gal, Adam Foster, and Tom Rainforth. Prediction-oriented bayesian active learning. In *International conference on artificial intelligence and statistics*, pp. 7331–7348. PMLR, 2023.

Ido Tal and Alexander Vardy. How to construct polar codes. *IEEE Transactions on Information Theory*, 59(10):6562–6582, 2013.

Wei Tan, Lan Du, and Wray Buntine. Diversity enhanced active learning with strictly proper scoring rules. *Advances in Neural Information Processing Systems*, 34:10906–10918, 2021.

Ian Waudby-Smith and Aaditya Ramdas. Confidence sequences for sampling without replacement. *Advances in neural information processing systems*, 33:20204–20214, 2020.

Victor K Wei. Generalized hamming weights for linear codes. *IEEE Transactions on information theory*, 37(5):1412–1418, 1991.

Dayou Yu, Weishi Shi, and Qi Yu. Actively testing your model while it learns: realizing label-efficient learning in practice. *Advances in Neural Information Processing Systems*, 36:31404–31427, 2023.

Kai Yu, Jinbo Bi, and Volker Tresp. Active learning via transductive experimental design. In *Proceedings of the 23rd international conference on Machine learning*, pp. 1081–1088, 2006.

Appendix Table of Contents

A **General Verification and Recovery Bounds** **14**

B **Conditional Prediction for a Finite Dictionary** **18**

C **Relation Likelihoods and Adaptive Transcripts** **20**

D **Uniform Random-Pool Converse** **21**

E **Verification and Parameter Recovery** **27**

F **Proofs of the Optimal-Risk Results** **28**

G **Affine-Basis Prediction** **36**

H **Exact Noiseless Acquisition** **37**

I **Numerical Validation** **41**

J **Related Work** **43**

K **Budget Curves and Paired Inference** **44**

L **Public Evidence and Finite-Budget Experiments** **46**

M **Finite Testing and Prediction** **52**

N **Exact Noisy Acquisition** **54**

O **Acquisition-Objective Comparisons** **62**

P **Affine-Flat Controls and Dimension-Dependent Bounds** **63**

Q **Subcube Verification under Positive Noise** **68**

R **Potential limitations** **71**

TECHNICAL APPENDICES

A GENERAL VERIFICATION AND RECOVERY BOUNDS

We extend the allocation argument to J regions with positive weights ω_j summing to one and noise rates η_j. The weighted risk is $R_\omega = \sum_j (\omega_j/m) \sum_x (\widehat{p}_{j,x} - Y_{j,x})^2$, and the population contains $N = Jm$ coordinates. In region j, write $p_j(W) = \mathbb{P}(Z_j = 1 \mid W)$ for the public affine probability, $\delta_j = 1 - 2\eta_j$ for noise attenuation, and $c_j = [1 - h_2(\eta_j)]^{-1}$ for the recovery cost per parameter dimension. The maximal prediction gain in that region is $g_j = \omega_j \delta_j^2/4$. Indicators may be dependent, with regional arrays conditionally independent given (W, Z). The allocation values using public information and the indicators are

$$G_{\mathrm{pub}} = \mathbb{E} \max_j g_j p_j(W), \qquad G_{\mathrm{ind}} = \mathbb{E} \max_j g_j Z_j. \tag{24}$$

The allocation proof uses regional gain bounds independently of their affine specialization. Let $\widetilde{W}$ collect public information, including the pools, and let S be the latent state. Suppose each policy's risk $R(\pi)$ admits regional gains D_j satisfying

$$R(\pi) \geq B - \rho_b - \mathbb{E} \sum_j D_j, \quad 0 \leq D_j \leq v_j(\widetilde{W}, S) \leq v_{\max}, \tag{25}$$

where B is the baseline risk, ρ_b corrects for observed targets, and v_j caps the regional gain D_j. The corresponding allocation values are

$$\mathcal{V}_{\mathrm{pub}} = \mathbb{E} \max_j \mathbb{E}[v_j \mid \widetilde{W}], \quad \mathcal{V}_{\mathrm{state}} = \mathbb{E} \max_j v_j, \tag{26}$$

where $\mathcal{V}_{\mathrm{pub}}$ uses public information and $\mathcal{V}_{\mathrm{state}}$ uses the latent state as well.

Theorem 2 (Verification before recovery)**.** *Let $1 \leq u \leq v$ be integer cutoffs with $b \leq u + v$. Stop when a regional count first reaches u, or at termination. Suppose every adaptive policy has a conditional stopped-transcript law, given $(\widetilde{W}, S)$, within total variation ε_u of a reference law depending on $\widetilde{W}$ and the policy alone. If every such policy's terminal counts n_j satisfy*

$$\mathbb{E} \sum_j \mathbf{1}\{n_j \leq v\} D_j \leq \varepsilon_v, \tag{27}$$

then

$$R^*_{\mathrm{ad}}(b) \geq B - \mathcal{V}_{\mathrm{pub}} - \rho_b - \varepsilon_v - v_{\max} \varepsilon_u. \tag{28}$$

For the matching upper bound, let $\widehat{J}$ be the region selected by the first batch. Its expected selection regret is $e_{\mathrm{ver}} = \mathbb{E}[\max_j v_j - v_{\widehat{J}}]$. If recovery in that region has expected excess loss e_{rec}, the two-batch risk is at most $B - \mathcal{V}_{\mathrm{state}} + e_{\mathrm{ver}} + e_{\mathrm{rec}}$.

A.1 PREDICTION EXPERIMENTS AND GAIN BOUNDS

The transcript records b distinct queries, their outcomes, and independent policy randomization. The latent-state observation model admits the conditional distributions used below, with query access included in $\widetilde{W}$. Revealing S after acquisition gives the gains and allocation values in Equations 25 and 26.

For the early-evidence comparison, define the stopping time

$$\sigma_u = \min \left(\{t \leq b : \max_j n_j(t) = u\} \cup \{b\} \right), \tag{29}$$

where $n_j(t)$ counts observations from region j among the first t queries. Let $P^{\pi,u}_{w,s}$ be the law of $\mathcal{H}_{\sigma_u}$ conditional on $(\widetilde{W}, S) = (w, s)$. The early-evidence hypothesis requires that, for every policy π, a conditional reference law $Q^{\pi,u}_w$ satisfies

$$\mathrm{TV}(P^{\pi,u}_{w,s}, Q^{\pi,u}_w) \leq \varepsilon_u \quad \text{for almost every } (w, s), \tag{30}$$

where $\mathrm{TV}(P,Q) = \sup_A |P(A) - Q(A)|$ is total variation. The reference law depends on the policy and public information, including the entire pool, and is independent of the latent state. The argument also applies after averaging the left-hand side of Equation 30 over $(\widetilde{W}, S)$.

For attainment, let $\mathcal{H}_1$ be the first-batch history and $\widehat{J}$ its selected region. It is enough that a fixed continuation of the remaining length and a terminal predictor have conditional expected loss at most

$$B - v_{\widehat{J}}(\widetilde{W}, S) + e(\widetilde{W}, S, \mathcal{H}_1), \quad e \geq 0, \quad \mathbb{E}e \leq e_{\mathrm{rec}}. \tag{31}$$

The function e bounds excess loss averaged over second-batch observations and evaluation targets at the given history and state. An unconditional expected bound gives the same result. Fixing continuation queries before observing their labels and averaging yields

$$R_2^*(b) \leq B - \mathcal{V}_{\mathrm{state}} + e_{\mathrm{ver}} + e_{\mathrm{rec}}. \tag{32}$$

A.2 Proof and Fixed-Design Consequences

Proof of Theorem 2. Let I be the first region reaching u, with a public choice if every terminal count is below u. This index is measurable from the stopped transcript. Once I reaches u, any region $j \neq I$ finishing above v would bring the two counts to at least $u + v + 1$, exceeding the budget. If every count stays below u, every region already satisfies the cutoff v. In either case,

$$\sum_j D_j \leq v_I(\widetilde{W}, S) + \sum_j \mathbf{1}\{n_j \leq v\} D_j. \tag{33}$$

For fixed public information w and state s, the function $v_I(w,s)$ takes values in $[0, v_{\max}]$ on the stopped transcript. The total-variation comparison then yields

$$\mathbb{E}_{P^{\pi,u}_{w,s}} v_I(w,s) \leq \mathbb{E}_{Q^{\pi,u}_w} v_I(w,s) + v_{\max}\varepsilon_u. \tag{34}$$

Under $Q^{\pi,u}_w$, the index I has the same distribution for every compatible state. Write $q_j(w) = Q^{\pi,u}_w(I = j)$ for its selection probabilities. Integrating over the actual conditional state prior gives

$$\begin{aligned} \mathbb{E}\, v_I(\widetilde{W}, S) &\leq \mathbb{E} \sum_j q_j(\widetilde{W}) \mathbb{E}[v_j(\widetilde{W}, S) \mid \widetilde{W}] + v_{\max}\varepsilon_u \\ &\leq \mathcal{V}_{\mathrm{pub}} + v_{\max}\varepsilon_u. \end{aligned} \tag{35}$$

Combining this inequality with Equations 25, 27, and 33 proves Equation 28, uniformly over policies.

For the second access geometry, the first-batch allocation regret satisfies

$$\mathbb{E}\, v_{\widehat{J}}(\widetilde{W}, S) = \mathcal{V}_{\mathrm{state}} - \mathbb{E}[\max_j v_j(\widetilde{W}, S) - v_{\widehat{J}}(\widetilde{W}, S)] \geq \mathcal{V}_{\mathrm{state}} - e_{\mathrm{ver}}. \tag{36}$$

Integrating Equation 31 establishes Equation 32. Optimizing the two-batch predictor can only improve this upper bound. □

The same gain and low-count inequalities also bound fixed and adaptive acquisition directly. If $b \leq 2v$, at most one region has more than v labels, and we obtain

$$\begin{aligned} R^*_{\mathrm{fix}}(b) &\geq B - \mathcal{V}_{\mathrm{pub}} - \rho_b - \varepsilon_v, \\ R^*_{\mathrm{ad}}(b) &\geq B - \mathcal{V}_{\mathrm{state}} - \rho_b - \varepsilon_v. \end{aligned} \tag{37}$$

A fixed design chooses its possible high-count region from $\widetilde{W}$ and independent randomization, limiting its expected gain to $\mathcal{V}_{\mathrm{pub}}$. For an adaptive design, the realized high-count gain is at most $\max_j v_j(\widetilde{W}, S)$. Equation 27 handles the other regions in both cases, after S is revealed for prediction.

A.3 Finite Separation and Decision-Relevant Verification

Consider access geometries A and B with the same latent-state experiment, target distribution, gain caps, budget, and public allocation value. Their public values agree when the pools are independent

of the state conditional on shared public evidence and the caps depend on that evidence and state. Applying the converse to A and the two-batch construction to B yields

$$R_{\mathrm{ad}}^{*,\mathrm{A}}(b) - R_{2}^{*,\mathrm{B}}(b) \geq \mathcal{V}_{\mathrm{state}} - \mathcal{V}_{\mathrm{pub}} - \rho_b - \varepsilon_v - v_{\max}\varepsilon_u - e_{\mathrm{ver}} - e_{\mathrm{rec}}. \tag{38}$$

The two acquisition classes are strictly separated when the state-informed allocation gain exceeds the observation, verification, and recovery errors.

Now consider a sequence of experiments in which these errors tend to zero and fixed acquisition achieves $B - \mathcal{V}_{\mathrm{pub}} + o(1)$ under both geometries. Equation 37 also applies to geometry B. If B, $\mathcal{V}_{\mathrm{pub}}$, and $\mathcal{V}_{\mathrm{state}}$ are fixed, inclusion of the acquisition classes yields

$$\begin{aligned} \lim R_{\mathrm{fix}}^{*,\mathrm{A}} = \lim R_{2}^{*,\mathrm{A}} = \lim R_{\mathrm{ad}}^{*,\mathrm{A}} = \lim R_{\mathrm{fix}}^{*,\mathrm{B}} = B - \mathcal{V}_{\mathrm{pub}}, \\ \lim R_{2}^{*,\mathrm{B}} = \lim R_{\mathrm{ad}}^{*,\mathrm{B}} = B - \mathcal{V}_{\mathrm{state}}. \end{aligned} \tag{39}$$

The same conclusion holds with convergent baselines and allocation values.

Verification may identify the best regional decision directly. For example, suppose a first batch estimates a state $\widehat{S}$ and $\widehat{J}$ maximizes $v_j(\widetilde{W}, \widehat{S})$. Then

$$e_{\mathrm{ver}} \leq v_{\max}\mathbb{P}(\widehat{S} \neq S) \tag{40}$$

whenever equal states imply the same gain-maximizing decision. More generally, a test that consistently selects a gain-maximizing region has vanishing regret even when several states lead to that choice.

Choose a measurable public optimizer $j^*(\widetilde{W})$ of the conditional gain in Equation 26. Its state-informed advantage is

$$\mathcal{V}_{\mathrm{state}} - \mathcal{V}_{\mathrm{pub}} = \mathbb{E}[\max_j v_j(\widetilde{W}, S) - v_{j^*(\widetilde{W})}(\widetilde{W}, S)] \geq 0. \tag{41}$$

Equality holds if and only if the public optimizer is state-optimal almost surely. Learning the state has allocation value through changes in the best regional decision.

A.4 Specialization to the Affine Prediction Model

For the affine model, take $\widetilde{W} = (W, \mathcal{P})$, $S = Z$, $B = 1/4$, $v_j(\widetilde{W}, S) = g_j Z_j$, and

$$D_j = g_j Z_j \mathcal{U}_j^{\mathrm{or}}(\mathcal{H}), \quad \rho_b = \frac{b\omega_{\max}}{4m}, \quad \omega_{\max} = \max_j \omega_j, \tag{42}$$

where $\mathcal{U}_j^{\mathrm{or}}$ averages squared posterior affine signs after revealing Z. We derive this risk decomposition in Appendix B.4. Because the pools and Z are conditionally independent given W, the allocation values are $\mathcal{V}_{\mathrm{pub}} = G_{\mathrm{pub}}$ and $\mathcal{V}_{\mathrm{state}} = G_{\mathrm{ind}}$.

Set $u = T$ and $v = T_{\mathrm{rec}}$ using the cutoffs in Equation 96. Theorem 3 gives

$$\varepsilon_v = \Psi_\eta(T_{\mathrm{rec}}, r)\, \mathbb{E} \sum_j g_j Z_j. \tag{43}$$

The random-pool likelihood bound, with its exceptional-pool probability, establishes the averaged form of Equation 30. If every regional likelihood differs from one by at most $e_T^{\mathrm{ev}} \leq 1$ on an event of probability at least $1 - \rho_T^{\mathrm{pool}}$, the stopped product likelihood implies

$$\varepsilon_u \leq \frac{(1 + e_T^{\mathrm{ev}})^J - 1}{2} + \rho_T^{\mathrm{pool}}. \tag{44}$$

Structured verification and recovery make e_{ver} and e_{rec} vanish, completing the weighted affine separation.

A.5 Independent Gaussian Coordinates and Gradual Prediction

A Gaussian model illustrates how the allocation theorem handles gradual prediction gains. Let $\vartheta \sim \mathcal{N}(0, I_D)$ be a D-dimensional coefficient vector. Querying coordinate $i \in \{1, \ldots, D\}$ once

observes $Y_i = \vartheta_i + \epsilon_i$, with independent $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and $\sigma^2 > 0$. Evaluation uses $Y'_X = \vartheta_X + \epsilon'_X$, where X is uniform on the coordinates and $\epsilon'_X \sim \mathcal{N}(0, \sigma^2)$ is independent of acquisition. An observed feature and its evaluation response share a coefficient and have independent observation errors.

Let $Q(\mathcal{H})$ be the queried coordinates and $n(\mathcal{H}) = |Q(\mathcal{H})|$ their terminal count. At a realized history, query-selection factors agree across candidate coefficient vectors and cancel in Bayes' formula. The posterior is therefore still a product of Gaussian distributions:

$$\begin{aligned} \mathbb{E}[\vartheta_i \mid \mathcal{H}] &= \begin{cases} Y_i/(1+\sigma^2), & i \in Q(\mathcal{H}), \\ 0, & i \notin Q(\mathcal{H}), \end{cases} \\ \mathrm{Var}(\vartheta_i \mid \mathcal{H}) &= \begin{cases} \sigma^2/(1+\sigma^2), & i \in Q(\mathcal{H}), \\ 1, & i \notin Q(\mathcal{H}). \end{cases} \end{aligned} \tag{45}$$

The variance of each coefficient depends only on whether that coordinate has been observed. Averaging the predictive variance over the uniform target X yields

$$\mathbb{E}[(Y'_X - \mathbb{E}[Y'_X \mid \mathcal{H}, X])^2 \mid \mathcal{H}] = 1 + \sigma^2 - \frac{n(\mathcal{H})}{D(1+\sigma^2)}. \tag{46}$$

Risk decreases linearly with the observed fraction. At $n = \lfloor aD \rfloor$ for $a \in (0,1)$, the population gain tends to $a/(1+\sigma^2)$.

This expression also determines the low-count term when the count depends on outcomes. Define $D_{\mathrm{lin}}(\mathcal{H}) = n(\mathcal{H})/[D(1+\sigma^2)]$, the reduction from prior predictive risk. For each integer $0 \le T \le D$,

$$\mathbb{E}[\mathbf{1}\{n(\mathcal{H}) \le T\} D_{\mathrm{lin}}(\mathcal{H})] = \frac{\mathbb{E}[\mathbf{1}\{n(\mathcal{H}) \le T\} n(\mathcal{H})]}{D(1+\sigma^2)} \le \frac{T}{D(1+\sigma^2)}. \tag{47}$$

The deterministic count $n = T$ attains this bound. A count proportional to D gives a constant-order low-count contribution, as the following two-region example illustrates.

Two regions with identical query laws. Let $Z_1 + Z_2 = 1$, with $\mathbb{P}(Z_1 = 1) = p$ and $p_1 = p$, $p_2 = 1 - p$. Each region has D query responses $Y_{j,i}$ and D held-out target responses $Y'_{j,i}$, indexed by $i \in \{1, \dots, D\}$. Put $v_\sigma = 1 + \sigma^2$, their common marginal variance. In the predictive region, generate independent coefficients $\vartheta_{j,i} \sim \mathcal{N}(0,1)$ and set

$$Y_{j,i} = \vartheta_{j,i} + \epsilon_{j,i}, \qquad Y'_{j,i} = \vartheta_{j,i} + \epsilon'_{j,i}, \tag{48}$$

where the errors are mutually independent $\mathcal{N}(0, \sigma^2)$ variables. In the other region, every query and target response is an independent $\mathcal{N}(0, v_\sigma)$ variable. The regional arrays are conditionally independent given Z. A policy acquires b distinct query responses, and risk averages squared error over the fixed $2D$ held-out targets.

Under either indicator state, query responses are independent $\mathcal{N}(0, v_\sigma)$ variables. Every adaptive transcript has the same law under both states, so the regional posterior stays at p_j. Let Q_j be the acquired indices and $n_j = |Q_j|$ their terminal count. An index with an unobserved query response has target mean zero and variance v_σ. At an acquired index, the corresponding quantities are

$$\begin{aligned} \mathbb{E}[Y'_{j,i} \mid \mathcal{H}] &= \frac{p_j Y_{j,i}}{v_\sigma}, \\ \mathrm{Var}(Y'_{j,i} \mid \mathcal{H}) &= v_\sigma - \frac{p_j}{v_\sigma} + \frac{p_j(1-p_j) Y_{j,i}^2}{v_\sigma^2}. \end{aligned} \tag{49}$$

We select each query before observing its response. That response has second moment v_σ conditional on the preceding history, which implies

$$\mathbb{E} \sum_{i \in Q_j} Y_{j,i}^2 = v_\sigma \mathbb{E} n_j, \qquad \mathcal{R}_{\mathrm{pair}}(\pi) = v_\sigma - \frac{p^2 \mathbb{E} n_1 + (1-p)^2 \mathbb{E} n_2}{2D v_\sigma}, \tag{50}$$

where $\mathcal{R}_{\mathrm{pair}}(\pi)$ denotes Bayes squared risk after acquisition by policy π. For $0 \le b \le D$, assigning all queries to a region maximizing p_j achieves

$$\mathcal{R}^*_{\mathrm{pair,fix}}(b) = \mathcal{R}^*_{\mathrm{pair},2}(b) = \mathcal{R}^*_{\mathrm{pair,ad}}(b) = v_\sigma - \frac{b \max\{p^2, (1-p)^2\}}{2D v_\sigma}. \tag{51}$$

Thus, every label improves optimal risk, and regional confidence stays at the public prior. With Z revealed after acquisition, the gains $D_j = Z_j n_j/(2Dv_\sigma)$ are bounded by $Z_j/(2v_\sigma)$ and satisfy Theorem 2 with a gradual low-count contribution. In the uniform affine model, Lemma 2 makes that contribution vanish below the channel threshold.

B Conditional Prediction for a Finite Dictionary

We now develop the prediction bound for a finite sign dictionary $f_\theta(x) \in \{-1, 1\}$ with prior ν. Define

$$K_{\theta,\theta'} = \frac{1}{m}\sum_x f_\theta(x) f_{\theta'}(x), \quad \Lambda_\nu = \max_{\theta'} \sum_\theta \nu_\theta (K_{\theta,\theta'})_+, \quad \mathfrak{d}_\nu = -\log_2 \Lambda_\nu, \tag{52}$$

where K records population agreement, $(a)_+ = \max(a, 0)$, and Λ_ν is the largest prior-weighted positive column sum. The effective dimension $\mathfrak{d}_\nu$ equals r for uniform affine parameters. For any nonnegative real dimension argument $\mathfrak{d}$, define $\Psi_\eta(T, \mathfrak{d})$ by replacing r with $\mathfrak{d}$ in Equation 10. At zero noise this becomes $\Psi_0(T, \mathfrak{d}) = \min(1, 2^{T-\mathfrak{d}})$.

Theorem 3 (Conditional prediction bound)**.** *Fix $(W, Z) = (w, z)$ with $z_j = 1$. Under conditional parameter prior ν_j, conditionally independent Bernoulli(η_j) flips, and conditionally independent regions, every adaptive policy satisfies*

$$\mathbb{E}[\mathbf{1}\{n_j \le T\}\mathcal{U}_j^{\mathrm{or}}(\mathcal{H}) \mid W = w, Z = z] \le \Psi_{\eta_j}(T, \mathfrak{d}_{\nu_j}), \quad 0 \le T \le m. \tag{53}$$

The integer $n_j = |Q_j(\mathcal{H})|$ is the terminal regional count. Public pools are arbitrary and conditionally independent of the regional parameter and noise. The inequality also holds conditional on their realized values and independent policy randomization.

B.1 Reference Score

Fix the public information, including the realized pools, an indicator vector with an affine region, and the positive support of that region's parameter prior ν. Combine identical dictionary entries because their Gram columns agree. If K_+ is the entrywise positive part of the population Gram matrix, every posterior vector q satisfies

$$\frac{1}{m}\sum_x \left(\sum_\theta q_\theta f_\theta(x)\right)^2 = q^\top K q \le q^\top K_+ q. \tag{54}$$

The entries of K_+ lie in $[0, 1]$, with prior-weighted column sums bounded by Λ_ν. We use these properties to construct a bounded prediction score and bound its expectation under the fair-label reference law.

Consider a procedure making exactly T distinct regional observations, with auxiliary information X_{aux} independent of the parameter and label errors. Let P be the joint law of the parameter, auxiliary information, and transcript. The reference law P_0 keeps the same prior and acquisition rule and returns independent fair labels at distinct identities. Under P_0, the parameter is independent of both X_{aux} and the transcript.

For a transcript-measurable indicator A and a transcript-measurable probability vector q° supported on the same dictionary, define

$$G(\theta, \mathcal{H}, X_{\mathrm{aux}}) = \mathbf{1}_A \sum_{\theta'} q^\circ_{\theta'} (K_{\theta,\theta'})_+. \tag{55}$$

Since $G \in [0, 1]$ and the reference transcript is independent of the parameter,

$$\mathbb{E}_0 G = \mathbb{E}_0\left[\mathbf{1}_A \sum_{\theta'} q^\circ_{\theta'} \sum_\theta \nu_\theta (K_{\theta,\theta'})_+\right] \le \Lambda_\nu. \tag{56}$$

The acquisition factors cancel from the likelihood ratio of P to P_0. A distinct observation contributes $2(1-\eta)$ when it agrees with the supplied parameter and 2η otherwise. Since queried errors remain independent along an adaptive path, the true-model log-likelihood has distribution S_T from Equation 10. Splitting at $\zeta \in \mathbb{R}$ yields $\mathbb{E}_P G \le 2^\zeta \mathbb{E}_0 G + \mathbb{P}_P(S_T > \zeta) \le 2^\zeta \Lambda_\nu + \mathbb{P}(S_T > \zeta)$. Optimizing over ζ and using $G \le 1$ gives $\Psi_\eta(T, \mathfrak{d}_\nu)$. At zero noise, every true-model history has likelihood ratio 2^T, so the bound is $\min(1, 2^T \Lambda_\nu)$.

B.2 Outcome-Dependent Regional Counts

The count-event extension uses the same score. Fix $(W, Z) = (w, z)$ with $z_j = 1$, and let $A = \{n_j \le T\}$ be the low-count event. Include the other regions' complete target arrays and the policy's independent randomization in X_{aux}. Under this conditioning, they are independent of region j's parameter and label errors. Write q for the posterior with the regional indicator known and q' for the posterior after also revealing X_{aux}. Positive semi-definiteness of K implies $\mathbb{E}[\mathbf{1}_A q^\top K q \mid w, z] \le \mathbb{E}[\mathbf{1}_A (q')^\top K q' \mid w, z]$. Because the transcript determines A, conditional Jensen's inequality applies with the event indicator.

Simulate the policy using X_{aux} for queries outside region j, and stop at termination or immediately before its $(T+1)$st regional observation. A terminated run records its posterior, then observes additional distinct coordinates from the full regional population until the count reaches T. The condition $T \le m$ ensures their availability, including when the public pool has fewer than T coordinates. The additional observations are part of the auxiliary experiment and leave the original policy's recorded transcript unchanged.

In the reference score, let A indicate a completed run and take $q^\circ = q'$ to be its posterior before the additional observations. Use any probability vector on the prior support for interrupted or zero-probability histories. Integrating out the additional observations then yields $\mathbb{E}_P G = \mathbb{E}[\mathbf{1}_A (q')^\top K_+ q' \mid w, z] \ge \mathbb{E}[\mathbf{1}_A (q')^\top K q' \mid w, z]$. Applying $\mathbb{E}_P G \le \Psi_\eta(T, \mathfrak{d}_\nu)$ to this completed experiment proves Theorem 3.

B.3 Finite Information Threshold

For $0 < \eta < 1/2$, set

$$C_\eta = 1 - h_2(\eta), \quad V_\eta = \eta(1-\eta)\left(\log_2 \frac{1-\eta}{\eta}\right)^2, \tag{57}$$

where C_η and V_η are the mean and variance of a binary-channel information increment (Polyanskiy et al., 2010). Independence of the increments gives mean TC_η and variance TV_η for S_T. If $TC_\eta < \mathfrak{d}$, the threshold $\zeta = (\mathfrak{d} + TC_\eta)/2$ lies halfway between these information levels. Applying the variance inequality yields

$$\Psi_\eta(T, \mathfrak{d}) \le 2^{-(\mathfrak{d} - TC_\eta)/2} + \frac{4TV_\eta}{(\mathfrak{d} - TC_\eta)^2}. \tag{58}$$

At zero noise, the inequality holds with $V_0 = 0$. Information variance is continuous at zero and bounded on $[0, \overline{\eta}]$ for each $\overline{\eta} < 1/2$. Thus, the expected energy on $\{n_j \le T\}$ vanishes whenever $TC_\eta \le (1-\varepsilon)\mathfrak{d}$, $T = O(\mathfrak{d})$, and $\mathfrak{d}\varepsilon^2 \to \infty$.

For M uniformly weighted parameters, let $\Gamma = \max_{\theta'} \sum_\theta (K_{\theta,\theta'})_+$ be the largest positive Gram column sum. Converting prediction to parameter decoding through $q^\top K q \le \Gamma \max_\theta q_\theta$ yields $\min(1, \Gamma\Psi_\eta(T, \log_2 M))$. Bounding the positive-overlap score directly instead keeps the coefficient on the likelihood tail equal to one, leading to

$$\Psi_\eta(T, \log_2(M/\Gamma)) \le \min(1, \Gamma\Psi_\eta(T, \log_2 M)). \tag{59}$$

At a common threshold, the likelihood terms agree, and the direct score assigns unit weight to the upper tail. Optimizing over the threshold proves the comparison.

B.4 Shared-Budget Allocation Bounds

For integer cutoffs $T = (T_1, \dots, T_J) \in \{0, \dots, m\}^J$ and a vector $v \in [0,1]^J$, define

$$\mathcal{F}_{T,b}(v) = \max_{S \subseteq [J],\, \sum_{j \in S}(T_j+1) \le b} \sum_{j \in S} v_j g_j. \tag{60}$$

A feasible set S contains regions whose counts can exceed their cutoffs within budget b. To average their gains, let $\pi_z(w) = \mathbb{P}(Z = z \mid W = w)$ be the conditional regional prior and $p_j(w) = \mathbb{P}(Z_j = 1 \mid W = w)$ its marginals. Define

$$\mathcal{E}_T(w) = \sum_z \pi_z(w) \sum_{j : z_j = 1} g_j \Psi_{\eta_j}(T_j, \mathfrak{d}_j(w, z)), \tag{61}$$

where $\mathcal{E}_T(w)$ bounds the combined low-count gain and $\mathfrak{d}_j(w,z)$ uses the corresponding conditional parameter prior. Writing $\omega_{\max} = \max_j \omega_j$ for the largest regional weight, we obtain

$$\begin{aligned} R^*_{\mathrm{ad}}(b) &\geq \frac{1}{4} - \frac{b\omega_{\max}}{4m} - \mathbb{E}\mathcal{F}_{T,b}(Z) - \mathbb{E}\mathcal{E}_T(W), \\ R^*_{\mathrm{fix}}(b) &\geq \frac{1}{4} - \frac{b\omega_{\max}}{4m} - \mathbb{E}\mathcal{F}_{T,b}(p(W)) - \mathbb{E}\mathcal{E}_T(W). \end{aligned} \tag{62}$$

To prove these bounds, reveal the true indicator vector after acquisition. Acquired targets are known exactly, and unqueried independent targets remain fair. An unqueried affine target has posterior variance $1/4 - \delta_j^2 (M^{\mathrm{or}}_{j,x})^2/4$, since its label error is independent of the history conditional on the parameter.

Acquired targets contribute at most $b\omega_{\max}/(4m)$. Regions above their cutoffs form a feasible set in $\mathcal{F}_{T,b}$, each contributing at most $Z_j g_j$, and Theorem 3 bounds the remaining energy. A fixed design chooses the high-count set before observing labels, so conditional averaging replaces Z by $p(W)$.

C Relation Likelihoods and Adaptive Transcripts

C.1 Affine Restriction Likelihood

For distinct coordinates $Q = (x_1, \ldots, x_n)$, let A_Q have rows $(1, x_i^\top)$. A uniform affine parameter induces a uniform clean restriction on the code $\mathrm{im}(A_Q) \subseteq \mathbb{F}_2^n$. Adding independent Bernoulli label errors produces $P_{1,Q}$. Under the independent regional model, every outcome has probability $P_{0,Q}(y) = 2^{-n}$.

For $v \in \mathbb{F}_2^n$, its sign moment is

$$\sum_y P_{1,Q}(y)(-1)^{v^\top y} = \mathbb{E}_\theta (-1)^{v^\top A_Q \theta}\, \mathbb{E}_\xi (-1)^{v^\top \xi} = \mathbf{1}\{A_Q^\top v = 0\}\delta^{|v|}. \tag{63}$$

Averaging over the parameter leaves only moments in the dependency space, each attenuated by its label-noise weight. Fourier inversion yields Equation 3, and orthogonality under uniform labels gives

$$\chi^2(P_{1,Q} \| P_{0,Q}) = \sum_{v \in \mathcal{D}(Q)\setminus\{0\}} \delta^{2|v|}. \tag{64}$$

For the conditional affine-sign mean $M_x(y)$, the Fourier coefficient is

$$\begin{aligned} \sum_y P_{1,Q}(y) M_x(y) (-1)^{v^\top y} &= \mathbb{E}[f_\theta(x)(-1)^{v^\top Y_Q}] \\ &= \mathbf{1}\{A_Q^\top v = \widetilde{x}\}\delta^{|v|}. \end{aligned} \tag{65}$$

The vector $\widetilde{x}$ is the augmented target, and Y_Q contains the observed labels. Fourier inversion and division by $P_{1,Q}(y) = 2^{-n} L_Q(y)$ establish Equation 5. An empty representation set gives a sign mean of zero, and zero affine likelihood gives the model zero posterior mass. The following pathwise cancellation extends these formulas to adaptive selection.

For fixed Q, total variation satisfies

$$\mathrm{TV}(P_{1,Q}, P_{0,Q}) = \frac{1}{2}\mathbb{E}_0 |L_Q(Y) - 1| \leq \frac{1}{2}\sqrt{\chi^2(P_{1,Q} \| P_{0,Q})}. \tag{66}$$

For adaptive acquisition, the observation also records the selected identities. Canceling the selection factors yields the same likelihood ratio for the joint transcript.

C.2 Adaptive Transcript Likelihood

On a realized transcript (Q, y), the public pool, the policy's independent randomization, and preceding outcomes determine the next query. The selection factors agree under the affine and independent laws, so the transcript likelihood ratio is $L_Q(y)$.

The same cancellation holds at a bounded stopping time determined by the acquired history. Under the independent reference law, the likelihood is a nonnegative mean-one martingale. Using conditional regional independence, we obtain the indicator-specific ratio

$$\frac{dP_z}{dP_0}(\mathcal{H}) = \prod_{j:z_j=1} L_{Q_j}(y_j). \tag{67}$$

If each queried augmented row is independent of its regional predecessors, every dependency space contains only zero. The product in Equation 67 then equals one for every indicator vector, so the regional posterior stays at its prior.

C.3 Matched-Rank Testing Comparison

In dimension seven, consider the eight-coordinate sets $Q_4 = (0, 1, 2, 3, 4, 8, 16, 32)$, $Q_8 = (0, 1, 2, 4, 8, 16, 32, 63)$, where integers encode binary coordinates. Both augmented matrices have rank seven and a unique nonzero dependency, of weight four in Q_4 and eight in Q_8. For a unique dependency of weight k, the likelihood reduces to

$$L_Q(y) = 1 + \delta^k(-1)^{v^\top y}, \quad \chi^2(P_{1,Q}\|P_{0,Q}) = \delta^{2k}, \quad \mathrm{TV}(P_{1,Q}, P_{0,Q}) = \frac{\delta^k}{2}. \tag{68}$$

The equal-prior testing error is $(1 - \mathrm{TV})/2 = 1/2 - \delta^k/4$. Independent rows have likelihood one. An eight-point subcube has fourteen nonzero dependencies of weight four and one of weight eight. Table 2 compares the resulting tests at matched query count and augmented rank.

Eight-label first set	Augmented rank	Dependency length	Testing error
Q_4	7	4	0.39760
Q_8	7	8	0.45806
Independent rows	8	None	0.50000

Table 2: Exact equal-prior testing errors at $\eta = .10$. The first two query sets match label count and augmented rank, isolating dependency length.

C.4 Minimal Informative Parity

Any set of at most three distinct coordinates in an affine region has linearly independent augmented vectors $(1, x)$. Indeed, an odd number of rows sums to a vector with first coordinate one, and two distinct rows have a nonzero sum. A uniform affine parameter therefore induces independent fair labels on such a set. Independent label errors leave this distribution unchanged, so it agrees with the independent regional model.

For linearly independent $a, b \in \mathbb{F}_2^d$, the four points $x, x + a, x + b, x + a + b$ have even clean affine parity. Their target-sign product has expectation δ_j^4 in an affine region and zero in an independent region. Therefore, four distinct labels are necessary for regional evidence from the algebraic label distribution on known coordinates, and this parity check achieves that count.

D Uniform Random-Pool Converse

D.1 Uniform Dependency Geometry

Because the learner can inspect the entire public pool, the evidence bound must hold for every available early query set. Fix $\varepsilon_{\mathrm{early}} \in (0, 1)$ and let $\mathcal{P}$ contain $q = r^2$ distinct uniform coordinates from $\mathbb{F}_2^{r-1}$, with regional population size $m = 2^{r-1}$. Set

$$T = \lfloor (1 - \varepsilon_{\mathrm{early}})r \rfloor, \quad k_r^{\mathrm{null}} = \left\lceil \frac{4 \log_2 q}{\varepsilon_{\mathrm{early}}} \right\rceil, \quad \ell_r^{\mathrm{dep}} = \left\lfloor \frac{r}{4 \log_2 q} \right\rfloor, \tag{69}$$

where T is the early-query cutoff, k_r^{null} bounds nullity, and ℓ_r^{dep} bounds dependency length. For sufficiently large r, we have $q < m/2$, $1 \le k_r^{\mathrm{null}} < T$, and $\ell_r^{\mathrm{dep}} \ge 1$.

Lemma 3 (Uniform dependency geometry of a random pool)**.** *With probability at least* $1 - \rho_r^{\mathrm{pool}}$*, every subset* $Q \subseteq \mathcal{P}$ *of size at most* T *has* $|Q| - \operatorname{rank}(A_Q) < k_r^{\mathrm{null}}$*, and every nonzero dependency among the entire pool's augmented rows has weight greater than* ℓ_r^{dep}*. A valid failure bound is*

$$\rho_r^{\mathrm{pool}} \le \frac{\ell_r^{\mathrm{dep}} q^{\ell_r^{\mathrm{dep}}}}{m-q} + T(2q)^T 2^{-k_r^{\mathrm{null}}(r-T-1)}, \quad \rho_r^{\mathrm{pool}} \longrightarrow 0. \tag{70}$$

Proof. Fix a subset of w sampled rows. Its augmented sum can be zero only if w is even. Once the first $w-1$ coordinates are fixed, their binary sum determines the final coordinate, whose conditional probability is at most $(m-q)^{-1}$. Summing over subsets of size at most ℓ_r^{dep} produces the first term in Equation 70. It vanishes because $q^{\ell_r^{\mathrm{dep}}} \le 2^{r/4}$ and $m = 2^{r-1}$.

Fix $n \le T$ and put $k = n - k_r^{\mathrm{null}}$. If k is negative, the rank event is empty. For nonnegative k, a row set of rank at most k has a spanning subset of k rows, obtained by extending a basis within the set when necessary. There are at most 2^n such subsets. Conditional on one of them, its k-dimensional span contains at most 2^{k-1} augmented rows with first coordinate one for $k \ge 1$; for $k = 0$, it contains none. Each remaining distinct coordinate, of which there are k_r^{null}, lies in this span with conditional probability at most

$$\frac{2^{k-1}}{m-q} \le 2^{k-r+1} \tag{71}$$

The inequality uses $q \le m/2$. Thus, a fixed n-row subset has rank-deficiency probability at most $2^n 2^{-k_r^{\mathrm{null}}(r-k-1)}$. The pool contains at most q^n candidate subsets. Summing over $n \le T$ and using $r-k-1 \ge r-T-1$ produces the second term.

For all sufficiently large r, $r - T - 1 \ge \varepsilon_{\mathrm{early}} r/2$. Since $k_r^{\mathrm{null}} \ge 4 \log_2 q / \varepsilon_{\mathrm{early}}$, the logarithm to base two of the second term is at most $\log_2 T + T \log_2(2q) - 2r \log_2 q \le \log_2 r + r - r \log_2 q$, which tends to negative infinity. Both geometric properties hold outside the two exceptional events. □

On the event of Lemma 3, every available Q with $|Q| \le T$ satisfies

$$\sup_y |L_Q(y) - 1| \le \sum_{v \in \mathcal{D}(Q) \setminus \{0\}} \delta^{|v|} \le 2^{k_r^{\mathrm{null}}} \delta^{\ell_r^{\mathrm{dep}}} =: e_r. \tag{72}$$

At fixed positive noise, $\delta^{\ell_r^{\mathrm{dep}}}$ decays faster than $2^{k_r^{\mathrm{null}}}$ grows. Thus, $e_r \to 0$ uniformly over all available sets and outcomes.

D.2 Dependency-Support Bounds

Checks with overlapping supports share observations. Disjoint information sets control the total likelihood contribution of these dependencies.

For a binary linear code $\mathcal{C}$, the kth generalized Hamming weight $d_k(\mathcal{C})$ is the smallest support of a k-dimensional subcode (Wei, 1991). The condition $\kappa \dim \mathcal{C}_S \le |S|$ for every S, where $\mathcal{C}_S$ contains codewords supported on S, is equivalent to $d_k(\mathcal{C}) \ge \kappa k$ for $1 \le k \le \dim \mathcal{C}$. In words, the support must grow proportionally to the number of independent checks.

Lemma 4 (Disjoint information sets and the dependency spectrum)**.** *Let* $\mathcal{C} \subseteq \mathbb{F}_2^n$ *be a binary linear code of dimension* $k_{\mathcal{C}}$*. For* $S \subseteq [n] = \{1, \dots, n\}$*, define its shortened subcode by* $\mathcal{C}_S = \{v \in \mathcal{C} : \operatorname{supp}(v) \subseteq S\}$*, where* $\operatorname{supp}(v)$ *is the set of nonzero coordinates. If an integer* $\kappa \ge 1$ *satisfies* $\kappa \dim \mathcal{C}_S \le |S|$ *for every* S*, then*

$$\sum_{v \in \mathcal{C}} \delta^{|v|} \le (1 + \delta^\kappa)^{k_{\mathcal{C}}}, \quad 0 \le \delta \le 1. \tag{73}$$

Proof. Let G be a full-row-rank $k_{\mathcal{C}} \times n$ generator matrix of $\mathcal{C}$. For a column set I, write G_I for its restriction to those columns. Rank and nullity give $\dim \mathcal{C}_S = k_{\mathcal{C}} - \operatorname{rank}(G_{S^c}), \quad n - |I| + \kappa \operatorname{rank}(G_I) \ge \kappa k_{\mathcal{C}}$. Complements are within $[n]$. The second inequality is the column-packing condition for κ disjoint bases (Edmonds, 1965), so the packing theorem yields disjoint full-rank column sets $I_1, \dots, I_\kappa$, each of cardinality $k_{\mathcal{C}}$. A codeword is uniquely determined by its restriction to any I_i. Thus, these sets are disjoint information sets.

For V uniform on $\mathcal{C}$, each restriction V_{I_i} is uniform on $\mathbb{F}_2^{k_\mathcal{C}}$. Since the sets are disjoint, Hölder's inequality gives

$$\begin{aligned}\sum_{v\in\mathcal{C}} \delta^{|v|} = 2^{k_\mathcal{C}}\mathbb{E}\delta^{|V|} &\le 2^{k_\mathcal{C}}\mathbb{E}\prod_{i=1}^{\kappa}\delta^{|V_{I_i}|} \\ &\le 2^{k_\mathcal{C}}\prod_{i=1}^{\kappa}\left(\mathbb{E}\delta^{\kappa|V_{I_i}|}\right)^{1/\kappa} = (1+\delta^\kappa)^{k_\mathcal{C}}.\end{aligned} \tag{74}$$

For $\delta = 0$, the same expression follows by continuity. Equality holds directly in dimension zero. □

Lemma 5 (Finite support density in a random pool)**.** *Let $m = 2^{r-1}$, $1 \le q \le m$, $1 \le T \le \min(q, r-1)$, and let $\kappa \ge 1$ be an integer. For each $1 \le n \le T$, set $\ell_n = \lfloor n/\kappa \rfloor + 1$ and $k_n = n - \ell_n$. These are the smallest violating nullity and the corresponding rank bound. Define*

$$\begin{aligned}B_{r,q,n}(\kappa) &= \begin{cases}\binom{q}{n}\binom{n}{k_n}\dfrac{\binom{2^{k_n-1}-k_n}{\ell_n}}{\binom{m-k_n}{\ell_n}}, & k_n \ge 1,\\ 0, & k_n \le 0,\end{cases} \\ \Gamma_{r,q,T}(\kappa) &= \min\left\{1, \sum_{n=1}^{T} B_{r,q,n}(\kappa)\right\}.\end{aligned} \tag{75}$$

The size-specific bounds $B_{r,q,n}(\kappa)$ sum to $\Gamma_{r,q,T}(\kappa)$. We take a binomial coefficient to be zero when its lower argument exceeds its nonnegative upper argument. For a uniform pool of q distinct coordinates,

$$|S| - \operatorname{rank}(A_S) \le |S|/\kappa \quad \textit{for every } S \subseteq \mathcal{P} \textit{ with } |S| \le T \tag{76}$$

with probability at least $1 - \Gamma_{r,q,T}(\kappa)$. In particular, $\Gamma_{r,q,T}(1) = 0$ when $\kappa = 1$.

Proof. A violating n-row set has rank at most k_n. The event is empty for $k_n \le 0$, since each augmented row has first coordinate one. For $k_n \ge 1$, choose k_n rows spanning the set, extending a basis with selected rows as needed. Their positions can be chosen in $\binom{n}{k_n}$ ways.

Conditional on these spanning coordinates, their span contains at most 2^{k_n-1} augmented rows with first coordinate one, including the k_n already selected. The remaining ℓ_n coordinates are uniform without replacement among the other $m - k_n$ population coordinates. The probability that all lie in the span is at most

$$\frac{\binom{2^{k_n-1}-k_n}{\ell_n}}{\binom{m-k_n}{\ell_n}}. \tag{77}$$

Union bounds over the spanning positions, the $\binom{q}{n}$ pool subsets, and all $n \le T$ establish Equation 75. At $\kappa = 1$, every $k_n = -1$, in agreement with the deterministic bound that nullity is at most the number of rows. □

For the asymptotic specialization, take $q = r^2$, $T = \lfloor (1 - \varepsilon_{\text{early}})r \rfloor$ with fixed $0 < \varepsilon_{\text{early}} < 1$, and set

$$\kappa_r = \left\lfloor \frac{r - T - 1}{2\log_2(2q)} \right\rfloor. \tag{78}$$

The conditions $q \le m/2$ and $r - T - 1 \ge 2\log_2(2q)$ ensure $\kappa_r \ge 1$. Since $\ell_n > n/\kappa_r$, the spanning-set argument in Lemma 3 bounds the contribution at size n by $(2q)^n 2^{-\ell_n(r-n+\ell_n-1)} \le (2q)^{-n}$. By the definition of κ_r and the inequality $r - n + \ell_n - 1 \ge r - T - 1$, the magnitude of the negative exponent is at least $2n\log_2(2q)$. Summation bounds the failure probability by $(2q-1)^{-1}$. At a specified dimension, we can also select κ directly using the finite sum in Equation 75.

Suppose $\kappa_r \geq 1$. Shortening the dependency code of an available Q to coordinates S gives the nullspace of the corresponding rows. The uniform condition in Equation 76 therefore meets the hypothesis of Lemma 4. With $k_Q^{\mathrm{dep}} = \dim \mathcal{D}(Q)$ denoting query nullity, we obtain

$$\begin{aligned} \sup_y |L_Q(y) - 1| &\leq (1 + \delta^{\kappa_r})^{k_Q^{\mathrm{dep}}} - 1 \\ &\leq \exp\left(\frac{T}{\kappa_r}\delta^{\kappa_r}\right) - 1 =: \widetilde{e}_r. \end{aligned} \tag{79}$$

D.3 Pool Construction and Cardinality

The limits also hold for $q_r = \lceil r^{1+a_{\mathrm{pool}}} \rceil$ with any fixed $a_{\mathrm{pool}} > 0$, keeping the specified subcube size and budget. The positive exponent makes the pool larger than the linear label budget, and

$$\log q_r = \Theta(\log r), \quad q_r - t_r \gg b_r, \quad \frac{t_r}{q_r} \longrightarrow 0, \quad \frac{q_r}{2^{r-1}} \longrightarrow 0. \tag{80}$$

The first relation yields $k_r^{\mathrm{null}} = O(\log r)$ and $\ell_r^{\mathrm{dep}} = \Omega(r/\log r)$ in Lemma 3, so both the failure probability and evidence bound vanish. The support argument also yields $\kappa_r = \Theta(r/\log r)$. The remaining relations ensure enough coordinates for verification and recovery, and the distinct-coordinate correction tends to one.

At $q_r = r^2$, the subcube has $t_r = \Theta((\log r)^2)$ coordinates and $1 + \log_2 t_r$ independent augmented rows. As the subcube grows, the testing error in Equation 16 tends to zero. Since $t_r = o(r)$ and $t_r = o(\sqrt{r})$, the above-capacity and critical recovery limits are preserved.

D.4 Finite-Pool Evidence Bounds

For finite pools, define

$$E_T(\mathcal{P}, \eta) = \max_{\substack{Q \subseteq \mathcal{P} \\ |Q| \leq T}} \sum_{v \in \mathcal{D}(Q) \setminus \{0\}} (1 - 2\eta)^{|v|}, \tag{81}$$

where E_T is the largest total attenuated dependency weight available in the realized pool within T labels.

For $0 \leq \eta < 1/2$, every coefficient in Equation 3 is nonnegative before multiplication by its observed sign. At the all-zero outcome, these signs align, so we have the exact identity

$$\max_{\substack{Q \subseteq \mathcal{P},\ |Q| \leq T \\ y \in \mathbb{F}_2^{|Q|}}} |L_Q(y) - 1| = E_T(\mathcal{P}, \eta). \tag{82}$$

The triangle inequality gives the upper bound, attained at the all-zero outcome.

For a realized pool pair, put $e_T^{\mathrm{ev}} = \max_{j=1,2} E_T(\mathcal{P}_j, \eta)$, which controls both regional histories before either count reaches T. Fix $W = w$ and write $p = p(w)$ for the affine-region prior. For a feasible budget and cutoffs $1 \leq T \leq T_{\mathrm{rec}} \leq m$ with $b \leq T + T_{\mathrm{rec}}$, the conditional optimal risk satisfies

$$R_{\mathrm{ad}}^*(b \mid w, \mathcal{P}_1, \mathcal{P}_2) \geq \frac{1}{4} - \frac{b}{4N} - g \min\left\{1, \max(p, 1-p) + \frac{e_T^{\mathrm{ev}}}{2} + \Psi_\eta(T_{\mathrm{rec}}, r)\right\}. \tag{83}$$

The stopped-transcript comparison bounds affine selection by $\max(p, 1-p) + e_T^{\mathrm{ev}}/2$. The budget allows the selected region to finish above T_{rec}, and Theorem 3 bounds the other region's contribution by $\Psi_\eta(T_{\mathrm{rec}}, r)$. Since total affine-region energy is at most one, including the observed-target correction proves Equation 83. Averaging gives the unconditional risk bound.

Equation 83 applies to both the exact E_T and every valid upper bound on it. For $T < r$, Lemma 5 and Lemma 4 give, outside an event of probability at most $\Gamma_{r,q,T}(\kappa)$ for each regional pool,

$$E_T(\mathcal{P}, \eta) \leq (1 + \delta^{\kappa})^{\lfloor T/\kappa \rfloor} - 1 \leq \exp\left(\frac{T}{\kappa}\delta^{\kappa}\right) - 1. \tag{84}$$

The support condition bounds every available dependency code's dimension by $\lfloor T/\kappa \rfloor$. Disjoint information sets give the first inequality, and the failure probability determines the admissible choice of κ.

At lower dimensions, a policy inspecting the full pool can select short relations. For $r = 9$ and $r = 13$, the inequality $\binom{q-4}{2} > m - 1$ guarantees two disjoint four-point checks in every pool of size $q = r^2$. Two distinct pairs sharing a nonzero binary sum have four distinct endpoints. Delete them and apply the same argument to obtain a second check. The two dependency vectors and their sum contribute $E_T(\mathcal{P}, \eta) \geq 2\delta^4 + \delta^8 \quad (T \geq 8)$. A single observed check yields $E_T \geq \delta^4$ for $T \geq 4$. Since relations may occur anywhere in the pool, we compare a control that selects two checks from the full pool with a prefix in the public order (Appendix L.3).

D.5 Finite Verification and Decoding Bounds

To complement Equation 83, consider the structured two-batch policy with subcube size t and continuation count $n = b - t$. Assume that each regional ordered remainder contains at least n coordinates. For thresholds $\eta < \tau_{\mathrm{test}}, \tau_{\mathrm{dec}} < 1/2$, define

$$\begin{aligned} e_{\mathrm{p}} &= \min\left\{1, \frac{1}{2}\left[e^{-t\,\mathrm{kl}(\tau_{\mathrm{test}},\eta)} + 2t\,2^{-t[1-h_2(\tau_{\mathrm{test}})]}\right]\right\}, \\ e_{\mathrm{d}} &= \min\left\{1, e^{-n\,\mathrm{kl}(\tau_{\mathrm{dec}},\eta)} + \frac{2^r - 2}{p_{\mathrm{dist}}(m,n,t)}\,2^{-n[1-h_2(\tau_{\mathrm{dec}})]}\right\}, \end{aligned} \tag{85}$$

where e_{p} bounds equal-prior region-selection error and e_{d} bounds parameter-decoding error. The probability p_{dist} is defined in Equation 17. Uniformity in the parameter makes the decoding bound valid after selection, averaging over the ordered remainder. With equal regional weights and public prior $1/2$, we obtain

$$R_2^{*,\mathrm{str}}(b) \leq \min\left\{\frac{1}{4}, \frac{1}{4} - g + e_{\mathrm{p}} + e_{\mathrm{d}}\right\}. \tag{86}$$

Use the decoded noisy affine rule in the selected region, predict $1/2$ elsewhere, and assign acquired targets their observed labels. Correct selection and decoding achieve the risk when the region and parameters are known. The complementary event contributes at most $e_{\mathrm{p}} + e_{\mathrm{d}}$ to expected squared error, proving the bound.

D.6 Noise Scales near the Zero-Noise Regime

For dimension-dependent noise η_r, write $\delta_r = 1 - 2\eta_r$ for attenuation. A length-w check has expected sign δ_r^w under the affine model and zero under independence. The dependency-support argument controls the aggregate evidence from these checks. We also construct a verification procedure using disjoint checks.

A sufficient condition for vanishing early evidence. Take $T_r/r \to a$ with $0 < a < 1$, and let

$$\kappa_r = \left\lfloor \frac{r - T_r - 1}{2\log_2(2r^2)} \right\rfloor. \tag{87}$$

The support parameter satisfies $\kappa_r = \Theta(r/\log r)$. Outside an event of probability at most $(2r^2 - 1)^{-1}$ for each regional pool, the support bound gives

$$\sup_{\substack{Q \subseteq \mathcal{P}_j,\ |Q| \leq T_r \\ y \in \mathbb{F}_2^{|Q|}}} |L_Q(y) - 1| \leq \exp\left(\frac{T_r}{\kappa_r}\delta_r^{\kappa_r}\right) - 1. \tag{88}$$

Thus, $(T_r/\kappa_r)\delta_r^{\kappa_r} \to 0$ is sufficient for the stopped-transcript comparison. Since $\delta_r^{\kappa_r} \leq e^{-2\eta_r\kappa_r}$ and $T_r/\kappa_r = O(\log r)$, a convenient sufficient rate is

$$\frac{\eta_r r}{\log r\,\log\log r} \longrightarrow \infty. \tag{89}$$

When $\eta_r \to 0$ at this rate and $1 < \lambda < 2$, set $T_r = \lfloor b_r/2 \rfloor$ and $T_{\mathrm{rec},r} = \lceil b_r/2 \rceil$. Both counts remain a fixed fraction below r. The supplied-parameter likelihood is at most $2^{T_{\mathrm{rec},r}}$, so

$$\Psi_{\eta_r}(T_{\mathrm{rec},r}, r) \leq 2^{T_{\mathrm{rec},r} - r} \longrightarrow 0. \tag{90}$$

Proposition 1 and the recovery argument in Appendix E then yield

$$\lim R_{\mathrm{fix}}^{*,\mathrm{rand}} = \lim R_{2}^{*,\mathrm{rand}} = \lim R_{\mathrm{ad}}^{*,\mathrm{rand}} = \lim R_{\mathrm{fix}}^{*,\mathrm{str}} = \frac{1}{4} - \frac{1-e_{\mathrm{pub}}}{8}, \qquad \lim R_{2}^{*,\mathrm{str}} = \lim R_{\mathrm{ad}}^{*,\mathrm{str}} = \frac{1}{8}. \tag{91}$$

Structured verification uses the prescribed $t_r = \Theta((\log r)^2)$ subcube. Fixed positive disagreement thresholds give vanishing testing and recovery errors as $\eta_r \to 0$.

Disjoint relations at low noise. A counting argument produces many disjoint relations of sublinear length in any pool of quadratic cardinality. Define

$$h_r = \left\lceil \frac{2r}{\log_2 r} \right\rceil, \qquad s_r = \lfloor \sqrt{\log_2 r} \rfloor, \qquad v_r = \lfloor r^2/2 \rfloor, \tag{92}$$

where h_r is the subset size used to find relations, s_r is the number of disjoint checks, and v_r is the reserved pool size. For all sufficiently large r, the subset counts satisfy

$$\binom{\lfloor r^2/4 \rfloor}{h_r} > 2^{r-1}. \tag{93}$$

Indeed, $\binom{n}{h} \geq (n/h)^h$ for $1 \leq h \leq n$, and the base-two logarithm of the left-hand side is at least $(2 + o(1))r$.

Consider the augmented sums of all h_r-subsets in a set of at least $\lfloor r^2/4 \rfloor$ coordinates. The first component is always the parity of h_r, so at most 2^{r-1} sums are possible. Thus, two distinct subsets share a sum, and their nonempty symmetric difference is a dependency of even length at most $2h_r$. Removing its support and repeating produces s_r pairwise disjoint dependencies from v_r initial coordinates, since $2h_r s_r = o(r)$. Each support is selected entirely from public coordinates.

Proposition 2 (Random-pool recovery with vanishing noise)**.** *Fix the law of (W, Z) in the exclusive, equal-weight model with conditional uniform affine parameters and the two pool constructions of Theorem 1. If $\eta_r = O(\log r/r)$, then for every $1 < \lambda < 2$ and $b_r = \lfloor \lambda r \rfloor$,*

$$\lim R_{2}^{*,\mathrm{rand}} = \lim R_{\mathrm{ad}}^{*,\mathrm{rand}} = \lim R_{2}^{*,\mathrm{str}} = \lim R_{\mathrm{ad}}^{*,\mathrm{str}} = \frac{1}{8}, \qquad \lim R_{\mathrm{fix}}^{*,\mathrm{rand}} = \lim R_{\mathrm{fix}}^{*,\mathrm{str}} = \frac{1}{4} - \frac{1-e_{\mathrm{pub}}}{8}. \tag{94}$$

The conclusion includes $\eta_r = 0$ for every r.

Proof. Assign each random pool an independent public uniform order and find the disjoint dependencies among its first v_r coordinates. Extend their union with distinct coordinates from this block to the deterministic first-set size $t_r^{\mathrm{low}} = 2h_r s_r = o(r)$. Order the complementary coordinates with the unused second block first. With this decomposition, we can apply the threshold-one test and fixed continuation in Algorithm 1.

For the ith selected dependency support D_i, let $V_i = (-1)^{\sum_{x \in D_i} Y_x}$ be its observed parity sign. The signs are independent because their supports are disjoint. Under the affine model, they depend only on the independent label errors and have means $\delta_r^{|D_i|} \geq a_r$, where $a_r = \delta_r^{2h_r}$. Under the independent model, their means are zero. The assumption $\eta_r = O(\log r/r)$ gives $\eta_r h_r = O(1)$ and $\inf_{r \geq r_0} a_r > 0$ for some r_0. The test $s_r^{-1} \sum_i V_i \geq a_r/2$ has error sum at most

$$2\exp(-s_r a_r^2/8) \longrightarrow 0 \tag{95}$$

by the bounded-variable concentration inequality (Hoeffding, 1963). The full threshold-one likelihood test minimizes the error sum and is at least as accurate. At zero noise, accepting exactly when all checks are positive has affine rejection probability zero and independent acceptance probability 2^{-s_r}.

Conditional on the first public block and verification history, the second block is uniformly ordered outside its v_r exclusions. It provides the $b_r - t_r^{\mathrm{low}} = \lambda r - o(r)$ continuation coordinates in either selected region. The parameter-uniform decoder in Appendix E applies with exclusion size

$v_r = O(r^2)$ and noise $\eta_r \to 0$. Its distinct-coordinate correction tends to one, and $\lambda > 1$ makes the decoding error vanish. Correct verification and recovery achieve population risk tending to $1/8$, with a matching lower bound obtained by revealing the affine region and its parameter. The prescribed structured subcube policy achieves the same limit.

For fixed acquisition, take $T_r = \lfloor b_r/2 \rfloor$. At most one regional count exceeds T_r, and we choose that region before outcomes, with affine probability at most $1 - e_{\mathrm{pub}}$. The prediction remainder is at most $2^{T_r - r} = o(1)$. The fixed-design allocation bound gives the stated lower limit, and uniform acquisition in the publicly favored region attains it by the same recovery argument. □

The support-density bound and the disjoint-check construction describe the noise scale at which early regional evidence becomes useful from two directions.

E Verification and Parameter Recovery

E.1 Subcube and Disjoint-Check Tests

A subcube of size $t = 2^{d_{\mathrm{sub}}}$ and binary dimension d_{sub} has $2^{d_{\mathrm{sub}}+1} = 2t$ affine restrictions. For $\eta < \tau_{\mathrm{test}} < 1/2$, accept the affine model if a restriction has at most $t\tau_{\mathrm{test}}$ disagreements. Under the affine model, comparison with the true restriction bounds rejection by $e^{-t\,\mathrm{kl}(\tau_{\mathrm{test}}, \eta)}$. Under independence, a union bound limits acceptance to $2t\,2^{-t(1-h_2(\tau_{\mathrm{test}}))}$. The threshold-one likelihood rule minimizes the sum of the two errors, so Equation 16 follows. At zero noise, the true restriction is always accepted.

For a positive integer s with $3s \le d$, an alternative uses s disjoint faces $F_i = \{e_{3i-2}, e_{3i-2} + e_{3i-1}, e_{3i-2} + e_{3i}, e_{3i-2} + e_{3i-1} + e_{3i}\}$, $1 \le i \le s$, where e_k denotes a coordinate basis vector. Each face contributes one independent anchor and two independent directions, so their union has $4s$ coordinates and augmented rank $3s$. The dependency space is generated by the disjoint face parities. Thus, the observed signs $V_i = (-1)^{\sum_{x \in F_i} Y_x}$ are independent under either distribution, with mean δ^4 under the affine model and zero under independence. The test $s^{-1} \sum_i V_i \ge \delta^4/2$ has error sum at most $2\exp(-s\delta^8/8)$. At fixed positive signal, $s = \lceil (\log r)^2 \rceil$ ensures vanishing error and $4s = o(r)$. The coordinate condition holds for sufficiently large r.

E.2 Distinct-Coordinate Recovery after Selection

We next prove Equation 17 uniformly in the parameter. Fix a public excluded set of t coordinates and a threshold $\eta < \tau_{\mathrm{dec}} < 1/2$. Conditional on the parameter and verification history, the continuation samples n coordinates uniformly without replacement from the complement, with $n + t \le m$.

For each parameter, averaging over the ordered remainder and its error bounds reduces the decoding error by $\min(1, \varepsilon_{\mathrm{dec}})$. Under independent labels, acceptance of an affine parameter within Hamming radius $n\tau_{\mathrm{dec}}$ has probability at most $2^r 2^{-n(1-h_2(\tau_{\mathrm{dec}}))}$.

Begin with independent uniform coordinates and label errors, then condition on the coordinates being distinct. The error-count tail for the true parameter is the first term of Equation 17. Each of the $2^r - 2$ competitors with a different slope has fair-binomial distance, entering the radius-$n\tau_{\mathrm{dec}}$ ball with probability at most $2^{-n(1-h_2(\tau_{\mathrm{dec}}))}$. The complementary parameter is outside the ball whenever the true distance is at most $n\tau_{\mathrm{dec}} < n/2$.

Conditioning on distinct permitted coordinates divides the competitor bound by $p_{\mathrm{dist}}(m, n, t)$ and preserves the true error-count tail. A unique candidate within the ball minimizes disagreement. The bound holds for each parameter, so it also holds after averaging over the selected region's verification posterior.

For fixed η and $n/r \to \lambda > [1 - h_2(\eta)]^{-1}$, choose $\tau_{\mathrm{dec}} > \eta$ with $\lambda[1 - h_2(\tau_{\mathrm{dec}})] > 1$. If t is polynomial in r, then $n(t+n)/m \to 0$, $p_{\mathrm{dist}}(m, n, t) \to 1$, and both decoding and independent-model false acceptance errors vanish. For $\eta_r \to 0$ and $n/r \to \lambda > 1$, choose a fixed positive τ_{dec} with $\lambda[1 - h_2(\tau_{\mathrm{dec}})] > 1$. Eventually $\eta_r < \tau_{\mathrm{dec}}/2$, and the true label error tail vanishes uniformly, for example by the bounded-variable concentration bound (Hoeffding, 1963). These conclusions include $\eta_r = 0$.

F Proofs of the Optimal-Risk Results

F.1 Finite Allocation Bound

For $c < \lambda < c + 1$, choose positive $\varepsilon_{\mathrm{early}}, \varepsilon_{\mathrm{rec}}$ satisfying

$$\varepsilon_{\mathrm{early}} + c\varepsilon_{\mathrm{rec}} < c + 1 - \lambda, \quad 1 - \varepsilon_{\mathrm{early}} \le c(1 - \varepsilon_{\mathrm{rec}}). \tag{96}$$

These margins give $T = \lfloor (1 - \varepsilon_{\mathrm{early}})r \rfloor$ and $T_{\mathrm{rec}} = \lfloor c(1 - \varepsilon_{\mathrm{rec}})r \rfloor$, with $T < r$, $T \le T_{\mathrm{rec}} < cr$, and $b_r \le T + T_{\mathrm{rec}}$ for sufficiently large r. Both cutoffs stay a fixed proportion below their respective evidence and recovery thresholds.

Proof of Proposition 1. Stop when a regional count first reaches T, or at termination if both counts stay smaller. Let I be the first region reaching the cutoff, with a fixed public choice on early termination. Both stopped counts are at most T. Conditional on $W = w$ and a pool in $\mathcal{G}_T$, the likelihood under either exclusive indicator vector differs from the all-independent reference likelihood by at most e_T^{ev}. The corresponding total variation is at most $e_T^{\mathrm{ev}}/2$. Under the reference law, I is independent of the indicator vector given W, so its affine probability is at most $\max(p(w), 1 - p(w))$. Averaging over public information and accounting for exceptional pools gives $\mathbb{P}(Z_I = 1) \le 1 - e_{\mathrm{pub}} + e_T^{\mathrm{ev}}/2 + \rho_T^{\mathrm{pool}}$. If the policy terminates before either count reaches T, both counts are at most T_{rec}. Otherwise, $n_I \ge T$, and any other region with $n_j > T_{\mathrm{rec}}$ would require at least $T + T_{\mathrm{rec}} + 1$ labels. Thus, I is the only possible high-count region, including at $b = T + T_{\mathrm{rec}}$. Reveal the true indicator vector for prediction. The high-count contribution to posterior energy is at most $\mathbb{P}(Z_I = 1)$, and Theorem 3 bounds the remaining contribution by $\Psi_\eta(T_{\mathrm{rec}}, r)$ under either exclusive indicator state. Averaging over the states leaves a single copy of this remainder because exactly one region is affine. Consequently,

$$\mathbb{E} \sum_j Z_j \mathcal{U}_j^{\mathrm{or}} \le 1 - e_{\mathrm{pub}} + e_T^{\mathrm{ev}}/2 + \rho_T^{\mathrm{pool}} + \Psi_\eta(T_{\mathrm{rec}}, r). \tag{97}$$

Each unit of affine-region energy contributes Brier gain g, and acquired targets contribute at most $b/(4N)$. Summing proves Equation 15. Since revealing the indicator improves the information available to the terminal predictor, the lower bound also holds for the original experiment. □

F.2 Weighted Regional Separation

Theorem 4 (Optimal-risk separation for weighted regions)**.** *Fix $J \ge 2$, positive weights summing to one, common positive noise, and the law of (W, Z), allowing arbitrary indicator dependence. Let*

$$G_{\mathrm{pub}} = \mathbb{E} \max_j g_j p_j(W), \quad G_{\mathrm{ind}} = \mathbb{E} \max_j g_j Z_j \tag{98}$$

be the allocation gains using public information and the revealed indicators, respectively. Under the conditional uniform affine prior, for $b_r = \lfloor \lambda r \rfloor$ and $c < \lambda < c + 1$,

$$\begin{aligned} \lim R_{\mathrm{fix}}^{*,\mathrm{rand}} = \lim R_2^{*,\mathrm{rand}} = \lim R_{\mathrm{ad}}^{*,\mathrm{rand}} = \lim R_{\mathrm{fix}}^{*,\mathrm{str}} = \frac{1}{4} - G_{\mathrm{pub}}, \\ \lim R_2^{*,\mathrm{str}} = \lim R_{\mathrm{ad}}^{*,\mathrm{str}} = \frac{1}{4} - G_{\mathrm{ind}}. \end{aligned} \tag{99}$$

The limits use the expected risks defined in Section 3 as $r \to \infty$.

Proof of Theorem 4. Choose the cutoffs in Equation 96. They satisfy $b_r \le T + T_{\mathrm{rec}}$, $T < r$, $T \le T_{\mathrm{rec}} < cr$, and $\Psi_\eta(T_{\mathrm{rec}}, r) \to 0$. Write $g_{\max} = \max_j g_j$ and $\omega_{\max} = \max_j \omega_j$ for the largest regional gain and weight. We first establish finite bounds that allow dependent indicators.

Suppose a pool event $\mathcal{G}_T$, independent of (W, Z), has probability at least $1 - \rho_T^{\mathrm{pool}}$. On this event, every regional history with at most T labels satisfies $|L_Q(y) - 1| \le e_T^{\mathrm{ev}} \le 1$. Stop at the first regional count of T and denote that region by I, using a fixed public choice if the policy terminates earlier. For a fixed indicator vector z, the stopped likelihood against all-independent targets is

$$\prod_{j : z_j = 1} L_{Q_j}(y_j). \tag{100}$$

Every factor lies in $[1-e_T^{\mathrm{ev}}, 1+e_T^{\mathrm{ev}}]$, so the product differs from one by at most $e_T^{\mathrm{joint}} = (1+e_T^{\mathrm{ev}})^J - 1$, because $1-(1-e_T^{\mathrm{ev}})^k \le (1+e_T^{\mathrm{ev}})^k - 1$ for $0 \le k \le J$. Therefore, conditional on the realized good pools and $W = w$, each indicator-specific stopped distribution is within total variation $e_T^{\mathrm{joint}}/2$ of the common reference distribution.

Under the reference law, the selected index is independent of Z conditional on W. Since $0 \le g_I z_I \le g_{\max}$, averaging the total-variation comparison over the conditional indicator distribution yields

$$\mathbb{E}[g_I Z_I] \le G_{\mathrm{pub}} + g_{\max}(e_T^{\mathrm{joint}}/2 + \rho_T^{\mathrm{pool}}). \tag{101}$$

On the exceptional pool event, use $g_I Z_I \le g_{\max}$.

The budget now connects selection to prediction. A region other than I finishing above T_{rec} would require at least $T+T_{\mathrm{rec}}+1$ labels, and early termination leaves all regions below the cutoff. Reveal the true indicators after acquisition and use $\mathcal{U}_j^{\mathrm{or}}$ from Equation 9. The weighted high-count contribution is at most $g_I Z_I$. Conditional on each indicator state and public observation, Theorem 3 bounds the remaining contribution by

$$\Psi_\eta(T_{\mathrm{rec}}, r)\,\mu, \quad \mu := \mathbb{E}\sum_j g_j Z_j \le \sum_j g_j. \tag{102}$$

Each conditional uniform affine prior has dimension argument r. Directly observed targets contribute at most $b\omega_{\max}/(4m)$, so

$$R_{\mathrm{ad}}^{*,\mathrm{rand}}(b) \ge \frac{1}{4} - \frac{b\omega_{\max}}{4m} - G_{\mathrm{pub}} - g_{\max}(e_T^{\mathrm{joint}}/2 + \rho_T^{\mathrm{pool}}) - \mu\Psi_\eta(T_{\mathrm{rec}}, r). \tag{103}$$

For fixed J, Lemma 3 makes the evidence and exceptional-pool terms vanish across all regional pools.

Because $b \le T + T_{\mathrm{rec}} \le 2T_{\mathrm{rec}}$, at most one region finishes above T_{rec}. A fixed design chooses this region using public information, the pools, and independent policy randomization. Its conditional expected gain is therefore at most $\max_j g_j p_j(W)$. Under either pool geometry,

$$R_{\mathrm{fix}}^{*}(b) \ge \frac{1}{4} - \frac{b\omega_{\max}}{4m} - G_{\mathrm{pub}} - \mu\Psi_\eta(T_{\mathrm{rec}}, r). \tag{104}$$

For an adaptive design, revealing the indicator vector bounds the high-count contribution by $\max_j g_j Z_j$. Therefore,

$$R_{\mathrm{ad}}^{*}(b) \ge \frac{1}{4} - \frac{b\omega_{\max}}{4m} - G_{\mathrm{ind}} - \mu\Psi_\eta(T_{\mathrm{rec}}, r). \tag{105}$$

Matching acquisition policies. For fixed acquisition, choose a public maximizer of $g_j p_j(W)$ and query b_r remaining pool coordinates there. The parameter-uniform decoding and independent-model rejection bounds in Appendix E apply because $\lambda > c$. On acceptance, predict with the decoded noisy rule, using $1/2$ in every other unqueried region. The gain tends to G_{pub} under either pool law.

For structured acquisition, use $Jt_r = o(r)$ labels to verify all regions. Select an accepted region with maximal g_j, or a public fallback if all tests reject, and allocate the remaining labels to its remaining pool. A union bound makes the testing error vanish under dependent indicator states, and the parameter-uniform recovery bound applies after selection. Predict with the decoded noisy rule in the selected accepted region and with $1/2$ elsewhere. If all tests reject, use $1/2$ at every unqueried target. Correct testing and decoding produce gain $\max_j g_j Z_j + o(1)$, including for the all-zero indicator vector, and achieve G_{ind}. Conditional Jensen's inequality gives $G_{\mathrm{ind}} \ge G_{\mathrm{pub}}$. □

These allocation values also characterize exactly when the gap is zero. Fix the joint law of (W, Z), choose a public maximizer $j^*(W) \in \arg\max_j g_j p_j(W)$, and write

$$G_{\mathrm{ind}} - G_{\mathrm{pub}} = \mathbb{E}\left[\max_j g_j Z_j - g_{j^*(W)} Z_{j^*(W)}\right]. \tag{106}$$

Equality holds exactly when $g_{j^*(W)} Z_{j^*(W)} = \max_j g_j Z_j$ almost surely. A public choice may remain optimal across compatible indicator states even with positive model-indicator entropy.

For example, take equal weights and uninformative public evidence. The uniform laws on $\{(1,0),(0,1)\}$ and $\{(0,0),(1,1)\}$ share marginals $1/2$, one bit of entropy, and public gain $g/2$. Their indicator-informed gains are g and $g/2$, respectively. The difference comes from whether learning the indicators changes the best regional choice.

With nonuniform weights, the remaining-target mass differs from one by at most $b\omega_{\max}/m = o(1)$, so the unqueried risks have the same limits. We extend the argument to general noncritical budgets in Appendix F.4.

F.3 Two-Region Separation

Proof of Theorem 1. For $0 \le \lambda < c$, reveal the affine-region indicator after acquisition. Its regional count is at most b_r, so Lemma 2 gives

$$\frac{1}{4} - \frac{b_r}{4N} - g\Psi_\eta(b_r, r) \le R^*_{\mathrm{ad}}(b_r) \le \frac{1}{4} - \frac{b_r}{4N}. \tag{107}$$

The upper bound uses prediction $1/2$ at unqueried coordinates. Both correction terms vanish, proving the common limit under either pool law. For $c < \lambda < c+1$, apply Theorem 4 with $J = 2$, $\omega_1 = \omega_2 = 1/2$, and $Z_1 + Z_2 = 1$. The two gains become

$$G_{\mathrm{pub}} = g\mathbb{E}\max\{p(W), 1-p(W)\} = g(1-e_{\mathrm{pub}}), \quad G_{\mathrm{ind}} = g. \tag{108}$$

Substitution establishes the stated limits, and $N/(N-b_r) \to 1$ establishes the unqueried versions. □

F.4 Budget-Dependent Prediction Risks

For a nonnegative binary decision loss ℓ, let H_ℓ be its continuous, symmetric, concave Bayes risk, with attained zero-loss endpoint actions and finite interior Bayes actions. Put $B_\ell = H_\ell(1/2)$, $g^\ell_j = \omega_j[B_\ell - H_\ell(\eta_j)]$, and

$$F^\ell_\lambda(v) = \max_{S\subseteq[J]:\,\sum_{j\in S} c_j < \lambda} \sum_{j\in S} g^\ell_j v_j, \tag{109}$$

where $[J] = \{1,\dots,J\}$ indexes regions and $v \in [0,1]^J$ contains regional probabilities. The empty set is feasible. The Bayes entropies for Brier, log loss, and classification are $p(1-p)$, $h_2(p)$, and $\min(p, 1-p)$, respectively.

Theorem 5 (Budget-dependent prediction risks)**.** *In the conditional affine model, fix J, positive weights, known $0 < \eta_j < 1/2$, and the law of (W, Z). For $b_r = \lfloor \lambda r \rfloor$, with $\lambda > 0$ distinct from every subset sum of the c_j,*

$$\begin{aligned} \lim R^{*,\mathrm{rand}}_{\ell,\mathrm{fix}} &= \lim R^{*,\mathrm{str}}_{\ell,\mathrm{fix}} = B_\ell - \mathbb{E}F^\ell_\lambda(p(W)), \\ \lim R^{*,\mathrm{str}}_{\ell,2} &= \lim R^{*,\mathrm{str}}_{\ell,\mathrm{ad}} = B_\ell - \mathbb{E}F^\ell_\lambda(Z). \end{aligned} \tag{110}$$

The vector $p(W) = (p_1(W),\dots,p_J(W))$ contains the public model probabilities, and $R^*_{\ell,\mathcal{A}}$ is optimal weighted loss in class $\mathcal{A}$. For common noise η and $c < \lambda < c+1$, the random-pool adaptive classes also satisfy

$$\lim R^{*,\mathrm{rand}}_{\ell,\mathrm{ad}} = \lim R^{*,\mathrm{rand}}_{\ell,2} = \lim R^{*,\mathrm{rand}}_{\ell,\mathrm{fix}} = B_\ell - \mathbb{E}\max_j g^\ell_j p_j(W). \tag{111}$$

A feasible subset contains regions that fit within the recovery budget, with a strict margin for verification and integer rounding.

Proof of Theorem 5. For action a and binary outcome y, the nonnegative loss $\ell(a,y)$ has Bayes entropy

$$H_\ell(p) = \inf_a\{(1-p)\ell(a,0) + p\ell(a,1)\}. \tag{112}$$

As an infimum of affine functions, H_ℓ is concave. We use the entropy assumptions and allocation quantities above. The indicators may be arbitrarily dependent, and regional arrays are conditionally independent. Proper scores and loss-sensitive design provide the decision framework (Gneiting & Raftery, 2007; Huang et al., 2026).

Prediction at low label counts. For $0 < s \le 1/2$, set

$$\omega_\ell(s) = \sup_{|p-1/2|\le s} [B_\ell - H_\ell(p)]. \tag{113}$$

The modulus $\omega_\ell(s)$ is nonnegative and tends to zero as $s \to 0$. After revealing the indicator vector, an unqueried affine target has label-one probability $(1 - \delta_j M^{\mathrm{or}}_{j,x})/2$. Splitting according to $|M^{\mathrm{or}}_{j,x}| \le a$ for $a \in (0, 1]$ yields

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}\{n_j \le T_j\}\frac{1}{m}\sum_{x\notin Q_j}\left\{B_\ell - H_\ell\left(\frac{1-\delta_j M^{\mathrm{or}}_{j,x}}{2}\right)\right\}\,\middle|\, W = w, Z = z\right] \\ \le \omega_\ell(a/2) + \frac{B_\ell}{a^2}\Psi_{\eta_j}(T_j, r), \quad z_j = 1.\end{aligned} \tag{114}$$

On the complementary event, the indicator is at most $(M^{\mathrm{or}}_{j,x})^2/a^2$, whose average weighted by the count event is bounded by Theorem 3. When $\Psi_{\eta_j}(T_j, r) \to 0$, choose $a = \Psi_{\eta_j}(T_j, r)^{1/4}$, or let $a \downarrow 0$ if $\Psi = 0$. Continuity of H_ℓ makes the resulting bound vanish, including for classification with its kink at $1/2$.

In a high-count affine region, every unqueried conditional target probability lies in $[\eta_j, 1 - \eta_j]$. Concavity and symmetry imply $H_\ell(p) \ge H_\ell(\eta_j)$ throughout this interval, bounding the full weighted gain by g^ℓ_j. Unqueried independent targets remain fair, and direct observations contribute at most $B_\ell b\omega_{\max}/m$. Since conditional Jensen's inequality makes Bayes risk smaller when the indicators are revealed, these bounds apply under the stated information model.

Let $\lambda > 0$ differ from every subset sum $c(S) = \sum_{j\in S} c_j$. Because there are finitely many subsets, choose a fixed $\varepsilon \in (0, 1)$ small enough that $c(S) > \lambda \implies (1 - \varepsilon)c(S) > \lambda$. Take $T_j = \lfloor(1 - \varepsilon)c_j r\rfloor$. Every high-count set $S_H = \{j : n_j > T_j\}$ obeys

$$(1 - \varepsilon)r \sum_{j\in S_H} c_j \le \sum_{j\in S_H}(T_j + 1) \le b_r \le \lambda r. \tag{115}$$

It follows that $c(S_H) < \lambda$, so S_H is feasible in Equation 109. Each stopped remainder vanishes because $T_j[1 - h_2(\eta_j)] \le (1 - \varepsilon)r$. For sufficiently large r, these cutoffs are valid regional counts, with $T_j = O(r) \ll m$.

For an adaptive design, the high-count gain with known indicators is at most $F^\ell_\lambda(Z)$. In a fixed design, S_H depends only on W, the independent pools, and policy randomization. Averaging conditionally, we replace Z_j by $p_j(W)$ and bound the gain by $F^\ell_\lambda(p(W))$. Combining these facts with Equation 114 proves both lower bounds in Theorem 5.

Attainment under the shared budget. Choose $\alpha > 0$ so that every feasible subset satisfies $\sum_{j\in S}(c_j + \alpha) < \lambda$. Such a common margin exists because the feasible family is finite and each inequality is strict. For a fixed acquisition, select a public optimizer of $F^\ell_\lambda(p(W))$ and query $\lceil(c_j + \alpha)r\rceil$ independently ordered remaining pool coordinates in each selected region. Complete the exact budget with additional distinct available identities, using their observed labels for zero-loss predictions. Each remaining pool contains $r^2 - O((\log r)^2)$ coordinates, so completing the budget this way is feasible for sufficiently large r.

For each selected region, choose $\tau_j > \eta_j$ with $(c_j + \alpha)[1 - h_2(\tau_j)] > 1$. The decoding and independent-model acceptance errors vanish by the bounds in Appendix E. Predict with the noisy-rule Bayes action on acceptance and the fair-bit action elsewhere. This finite set of actions has finite loss at both outcomes, so failure events contribute vanishing excess risk. We use additional labels to complete the budget for prediction at their acquired identities. Equation 114 and Cauchy-Schwarz applied to the effective transcript give the expected unselected-region loss $B_\ell + o(1)$, including after selection, which depends on the first batch. Completing the budget changes only $O(b_r)$ evaluated identities, so the gain is $\sum_{j\in S} g^\ell_j Z_j + o(1)$ under either pool.

For structured acquisition, verify all regions with $Jt_r = o(r)$ first-batch labels and choose an optimizer of $F^\ell_\lambda(\widehat{Z})$, removing zero-gain regions in ties when possible. A union bound makes the

model-selection error vanish. The positive budget margins cover verification and rounding, so the second batch has the same recovery counts. Conditional on the parameters and selection, every remaining pool is uniformly ordered outside its exclusion. The parameter-uniform recovery bound then establishes gain $F_\lambda^\ell(Z) + o(1)$ and proves Theorem 5.

□

For random pools, revealing the indicators gives a lower bound, and fixed designs belong to the adaptive class. These observations imply

$$B_\ell - \mathbb{E}F_\lambda^\ell(Z) \le \liminf_r R_{\ell,\mathrm{ad}}^{*,\mathrm{rand}} \le \limsup_r R_{\ell,\mathrm{ad}}^{*,\mathrm{rand}} \le B_\ell - \mathbb{E}F_\lambda^\ell(p(W)). \tag{116}$$

For common η and $c < \lambda < c+1$, choose T and T_{rec} as in Equation 96. Let I be the first region reaching T, with a public choice on early termination. Under this budget, only I can have more than T_{rec} terminal labels. Define D_j^ℓ as the weighted reduction from B_ℓ on region j's unqueried targets after revealing Z. Concavity gives $0 \le D_j^\ell \le g_j^\ell Z_j$, and Equation 114 yields

$$\mathbb{E}\sum_j \mathbf{1}\{n_j \le T_{\mathrm{rec}}\} D_j^\ell = o(1). \tag{117}$$

The stopped likelihood comparison in Appendix F.2 holds uniformly over the full pool and every policy. Under the reference law, I is independent of Z given W. With bounded regional gain caps and vanishing total variation, we obtain

$$\mathbb{E}[g_I^\ell Z_I] \le \mathbb{E}\max_j g_j^\ell p_j(W) + o(1). \tag{118}$$

Consequently, $R_{\ell,\mathrm{ad}}^{*,\mathrm{rand}} \ge B_\ell - \mathbb{E}\max_j g_j^\ell p_j(W) - o(1)$ after adding the observed-target and low-count corrections. The fixed acquisition construction attains this value. Inclusion of fixed and two-batch policies in the adaptive class proves Equation 111.

A budget-dependent adaptation premium. For a fixed public distribution choose $S^*(W) \in \arg\max_{c(S)<\lambda} \sum_{j\in S} g_j^\ell p_j(W)$. The structured adaptation premium is

$$\Delta_\ell(\lambda) = \mathbb{E}\left[F_\lambda^\ell(Z) - \sum_{j\in S^*(W)} g_j^\ell Z_j\right] \ge 0. \tag{119}$$

The premium is zero exactly when the public optimizer is optimal for almost every compatible indicator vector. This occurs, in particular, for $\lambda < \min_j c_j$ and $\lambda > \sum_j c_j$. At intermediate budgets, the feasible subsets of regions determine the premium.

For example, take J equally weighted regions, common noise, no public evidence, and independent model probability p. If $kc < \lambda < (k+1)c$ with $0 \le k < J$, at most k regions can be completed. Set $K = \sum_j Z_j \sim \mathrm{Binomial}(J,p)$ and $g^\ell = [B_\ell - H_\ell(\eta)]/J$. Then

$$\Delta_\ell(\lambda) = g^\ell\{\mathbb{E}\min(K,k) - kp\}. \tag{120}$$

The increment when the allowed count increases from k to $k+1$ is $g^\ell[\mathbb{P}(K \ge k+1) - p]$, which decreases with k. For $J = 4$, $p = 1/2$ and Brier noise $\eta = .10$, the premiums for counts $k = 0,1,2,3,4$ are exactly $0, \quad 0.0175, \quad 0.0250, \quad 0.0175, \quad 0$. For heterogeneous noise or dependent indicators, we obtain the corresponding values by enumerating feasible regional subsets. Rational arithmetic verifies the displayed example at every noncritical allocation count.

F.5 Critical Recovery Budgets

For the two-region Brier problem, we can also obtain exact limits at the recovery thresholds. Assume fixed positive noise, equal weights, and one affine region. Set

$$\begin{gathered} C = 1 - h_2(\eta), \quad c = C^{-1}, \quad V = \eta(1-\eta)\left(\log_2\frac{1-\eta}{\eta}\right)^2, \\ g = \frac{(1-2\eta)^2}{8}, \quad \alpha = \mathbb{E}\max\{p(W), 1-p(W)\}, \end{gathered} \tag{121}$$

where V is the information variance, positive for $0 < \eta < 1/2$, g is the gain from affine recovery, and $\alpha = 1 - e_{\mathrm{pub}}$ is public localization accuracy. Let Φ denote the standard normal cumulative distribution function.

Proposition 3 (Critical Brier budgets)**.** *Under the preceding two-region assumptions and notation, let $s \in \mathbb{R}$ be fixed and $b_r = \lfloor cr + s\sqrt{r} \rfloor$. Then*

$$\begin{aligned} \lim R_{\mathrm{fix}}^{*,\mathrm{rand}} = \lim R_{\mathrm{fix}}^{*,\mathrm{str}} &= \frac{1}{4} - g\alpha\Phi\left(\frac{sC}{\sqrt{cV}}\right), \\ \lim R_{2}^{*,\mathrm{str}} = \lim R_{\mathrm{ad}}^{*,\mathrm{str}} &= \frac{1}{4} - g\Phi\left(\frac{sC}{\sqrt{cV}}\right). \end{aligned} \tag{122}$$

Write $\psi_s = \Phi(sC/\sqrt{cV})$ for the limiting recovery probability. Then the random-pool adaptive risk obeys

$$\frac{1}{4} - g\psi_s \le \liminf_r R_{\mathrm{ad}}^{*,\mathrm{rand}}(b_r) \le \limsup_r R_{\mathrm{ad}}^{*,\mathrm{rand}}(b_r) \le \frac{1}{4} - g\alpha\psi_s. \tag{123}$$

At $b_r = \lfloor 2cr \rfloor$, the fixed limits are $1/4 - g\alpha$ and the structured adaptive limits are $1/4 - g$.

Proof of Proposition 3. The information bound and decoder acceptance match near cr. We use the former for the converse and the latter for attainment.

Information near the recovery threshold. For $n_r = cr + s\sqrt{r} + o(\sqrt{r})$, use $\zeta_r = r - \log_2 r$ in Equation 10. The information increments are independent, with mean C and variance V. The central limit theorem implies

$$\limsup_r \Psi_\eta(n_r, r) \le \lim_r \mathbb{P}(S_{n_r} > r - \log_2 r) = \Phi\left(\frac{sC}{\sqrt{cV}}\right). \tag{124}$$

The likelihood term left outside the tail is $1/r$, and integer rounding changes the normalized threshold by $o(1)$.

At $b_r = cr + O(\sqrt{r})$, at least one regional count in a fixed design is at most $b_r/2$, so its mean squared posterior sign vanishes by Theorem 3. The other region is chosen before observing the labels. Conditional on public information, pools, and policy randomization, its affine probability is at most $\max\{p(W), 1 - p(W)\}$. Applying the same theorem with cutoff b_r bounds its energy by $\Psi_\eta(b_r, r)$. Since observed targets contribute $o(1)$ to population gain, averaging over public information proves the fixed-design lower bound.

For any adaptive policy, the affine region has at most b_r labels. Apply Theorem 3 with that cutoff under each exclusive indicator state and average over their probabilities. Total affine-region energy is at most $\Psi_\eta(b_r, r)$, proving the lower bound for structured adaptation and the corresponding lower bound for random adaptation.

Attainment by minimum-distance decoding. Let $n_r = cr + s\sqrt{r} + o(\sqrt{r})$ be the continuation count. The fixed policy chooses the publicly favored region. The structured policy first verifies with the specified subcube, then uses $n_r = b_r - t_r$ labels in the selected region. Since $t_r = O((\log r)^2) = o(\sqrt{r})$, both designs have the same normalized recovery budget, and the structured verification error tends to zero.

Write $D = h_2'(\eta) = \log_2((1-\eta)/\eta) > 0$, and choose $a_r = (\log r)^3$. Accept a unique affine parameter within Hamming radius

$$k_r = \left\lfloor n_r\eta + \frac{n_r C - r}{D} - a_r \right\rfloor. \tag{125}$$

For sufficiently large r, the radius k_r lies strictly between zero and $n_r/2$. Using the fixed continuation in Appendix E, the decoder accepts a unique parameter within this radius and predicts with its noisy rule. On rejection, it assigns $1/2$ to unqueried targets.

The true error count is $B_{n_r} \sim \mathrm{Binomial}(n_r, \eta)$. Since $a_r = o(\sqrt{r})$ and $V = \eta(1-\eta)D^2$,

$$\mathbb{P}(B_{n_r} \le k_r) \longrightarrow \Phi\left(\frac{sC}{\sqrt{cV}}\right). \tag{126}$$

Every different-slope competitor has a fair binomial distance under independent uniform coordinates. Conditioning on distinct coordinates outside the public exclusion gives the same correction $p_{\text{dist}}^{-1} \to 1$ as in Appendix E. The probability of any such competitor in the acceptance ball is at most

$$\frac{2^r - 2}{p_{\text{dist}}(m, n_r, t)} 2^{-n_r[1-h_2(k_r/n_r)]}, \tag{127}$$

where the excluded set size t is polynomial in r. Expanding at η yields

$$r - n_r[1 - h_2(k_r/n_r)] = -Da_r + O(1) \longrightarrow -\infty. \tag{128}$$

The remainder is bounded because $n_r C - r = O(\sqrt{r}) + o(\sqrt{r})$ and $a_r = o(\sqrt{r})$. Whenever the true parameter lies inside the ball, its complement is outside. Erroneous acceptance of the complement with the true parameter outside requires $B_{n_r} \geq n_r - k_r$, whose probability tends to zero since $\eta < 1/2$. Under independent labels, a union bound over all parameters has the same vanishing exponent.

Conditional on the parameter, continuation coordinates and label errors are independent of verification outcomes. Therefore, the parameter-uniform bounds apply after selection. Correct acceptance contributes population gains g, and erroneous acceptance has vanishing probability. The fixed policy selects the affine region with probability α; structured verification makes this probability tend to one. These selection probabilities establish the two limits, and the Bayes predictor using the same observations has risk at most that of the decoder.

The boundary at two recovery costs. For nonnegative integer counts with $n_1 + n_2 = \lfloor 2cr \rfloor$,

$$\sup_{n_1+n_2=\lfloor 2cr \rfloor} [\Psi_\eta(n_1, r) + \Psi_\eta(n_2, r)] \leq 1 + o(1). \tag{129}$$

To establish uniformity, consider any sequence of maximizing count pairs and a subsequence on which n_1/r converges. If its limit differs from c, one count stays a fixed fraction below recovery and has a vanishing information bound. The other bound is at most one. If both ratios converge to c, take a further subsequence on which $(n_1 - cr)/\sqrt{r}$ converges in the extended real line. A finite limit gives two normal tails with opposite arguments, whose sum is one. At an infinite limit, concentration makes one bound vanish and the other remains at most one. The claimed uniform bound follows in every case.

For a fixed design, Theorem 3 bounds its conditional weighted energy by

$$p(W)\Psi_\eta(n_1, r) + (1 - p(W))\Psi_\eta(n_2, r) \leq \max\{p(W), 1 - p(W)\}(1 + o(1)). \tag{130}$$

Averaging bounds the gain by $g\alpha$. Concentrating queries in the publicly favored region achieves this gain because its count exceeds cr by a linear margin. Structured verification followed by recovery achieves gain g at budget $2cr$, as in Appendix E. □

F.6 Two-Region Budget Regimes

Budget ratio	Fixed both pools	Structured two-batch/adaptive	Random two-batch/adaptive
$0 \leq \lambda < c$	$1/4$	$1/4$	$1/4$
$\lambda = c$	$1/4 - g\alpha/2$	$1/4 - g/2$	$[1/4 - g/2,\ 1/4 - g\alpha/2]$
$c < \lambda < c + 1$	$1/4 - g\alpha$	$1/4 - g$	$1/4 - g\alpha$
$c + 1 \leq \lambda \leq 2c$	$1/4 - g\alpha$	$1/4 - g$	$[1/4 - g,\ 1/4 - g\alpha]$
$\lambda > 2c$	$1/4 - g$	$1/4 - g$	$1/4 - g$

Table 3: Exact limits and benchmark bounds for $b_r = \lfloor \lambda r \rfloor$, with public selection accuracy $\alpha = 1 - e_{\text{pub}}$. Scalar entries give limits. Each interval combines the indicator-informed lower bound and fixed-design upper bound, with $L \leq \liminf R^* \leq \limsup R^* \leq U$.

Table 3 combines the critical-budget results of Proposition 3 with the noncritical limits of Theorem 5, using $g = \delta^2/8$ and $\alpha = 1 - e_{\text{pub}}$.

For random-pool two-batch and adaptive acquisition, Theorem 4 gives the common fixed-acquisition limit on $c < \lambda < c + 1$. At $\lambda = c$, the critical information bound gives the lower endpoint of the displayed interval. For $c + 1 \leq \lambda \leq 2c$, the lower endpoint comes from indicator-informed prediction. Fixed acquisition gives both upper endpoints.

F.7 Operational Verification and the Transition

To describe the transition beyond the main theorem's interval, consider one regional random pool of size r^2. Let $P_{r,1}^{\pi,a}$ and $P_{r,0}^{\pi,a}$ be the joint laws of the public pool and transcript under affine and independent labels. A tester π may inspect the full pool and acquire at most $\lfloor ar \rfloor$ distinct labels. Define

$$\varepsilon_r(a) = \sup_{\pi} \mathrm{TV}(P_{r,1}^{\pi,a}, P_{r,0}^{\pi,a}), \qquad a_\star(\eta) = \sup\{a \geq 0 : \varepsilon_r(a) \longrightarrow 0\}, \tag{131}$$

where $\varepsilon_r(a)$ is the optimal testing advantage in total-variation units and $a_\star$ is the supremum of asymptotically uninformative query ratios. The tester may use independent auxiliary randomization, and $\varepsilon_r(a)$ is nondecreasing in the allowed query ratio a.

The uniform early-evidence lemma implies $\varepsilon_r(a) \to 0$ for each $a < 1$. For $a > c$, the recovery and independent-model rejection bounds in Appendix E, uniform in the parameter, yield a consistent test along a uniform regional continuation. Consequently,

$$1 \leq a_\star(\eta) \leq c. \tag{132}$$

From verification to a prediction interval. Fix $a < c$ with $\varepsilon_r(a) \to 0$ and $c < \lambda < c + a$. Choose $a' < a$ and $v < c$ such that $a' \leq v$ and $\lambda < a' + v$, and take $T = \lfloor a'r \rfloor$ and $T_{\mathrm{rec}} = \lfloor vr \rfloor$. Under either exclusive indicator state, the stopped two-region experiment is a one-region test using at most T labels, with independent auxiliary observations from the other region. Its total variation from the all-independent reference is at most $\varepsilon_r(a)$. The selection step in Proposition 1 then yields

$$\mathbb{P}(Z_I = 1) \leq 1 - e_{\mathrm{pub}} + \varepsilon_r(a). \tag{133}$$

The same count inequality and conditional prediction lemma then give

$$R_{\mathrm{ad}}^{*,\mathrm{rand}}(b_r) \geq \frac{1}{4} - \frac{b_r}{4N} - g\{1 - e_{\mathrm{pub}} + \varepsilon_r(a) + \Psi_\eta(T_{\mathrm{rec}}, r)\}. \tag{134}$$

Both remainder terms vanish, and fixed recovery in the publicly favored region matches the bound. Thus, the fixed-acquisition limit holds throughout $c < \lambda < c + a_\star$.

For the equal prior, define departure from the single-region recovery value by

$$\lambda_{\mathrm{ad}}(\eta) = \inf\left\{\lambda > c : \liminf_{r\to\infty} R_{\mathrm{ad}}^{*,\mathrm{rand}}(\lfloor \lambda r \rfloor) < \frac{1}{4} - \frac{g}{2}\right\}. \tag{135}$$

The preceding argument and recovery in both regions above $2c$ establish

$$c + a_\star(\eta) \leq \lambda_{\mathrm{ad}}(\eta) \leq 2c. \tag{136}$$

The allocation converse gives the lower endpoint, and recovery in both regions gives the upper endpoint. The bound $a_\star \geq 1$ recovers the interval in Theorem 1.

The second-moment comparison. For a second-moment calculation, sample independent uniform coordinates $x_1, \ldots, x_n \in \mathbb{F}_2^{r-1}$ with independent channel errors at each row; coordinates may repeat. For a nonzero vector v of even weight, the coordinate sum in $A_Q^\top v$ is uniform on $\mathbb{F}_2^{r-1}$. Odd-weight vectors have nonzero augmented first component, so Fourier orthogonality gives

$$\begin{aligned} \mathbb{E}\chi^2(P_{1,Q} \| P_{0,Q}) &= 2^{-(r-1)} \sum_{\substack{2 \leq w \leq n \\ w \text{ even}}} \binom{n}{w} \delta^{2w} \\ &= 2^{-r}\left[(1+\delta^2)^n + (1-\delta^2)^n - 2\right], \end{aligned} \tag{137}$$

where w indexes the dependency weight. Define $a_\chi = [\log_2(1+\delta^2)]^{-1}$ as the crossover ratio of the expected divergence. Comparing order-two and Shannon entropy yields $1 < a_\chi < c$. Below a_χ with a fixed ratio margin, the second moment vanishes. Above that ratio, rare outcomes can contribute to its growth. The operational threshold $a_\star$ instead concerns optimal adaptive testing over the full pool and determines the prediction interval established above.

G AFFINE-BASIS PREDICTION

In an affine-basis pool, every feasible query gives an independent constraint. Regional confidence stays at its prior, yet odd combinations of acquired constraints predict the targets they represent. At a fixed public observation, write $p_j = p_j(W)$ and let $n = (n_j)_j$ collect the regional counts. Define

$$H_\delta(n) = \sum_{\substack{3 \le k \le n \\ k \text{ odd}}} \binom{n}{k} \delta^{2k+2}, \quad B(n) = \frac{1}{4} - \sum_j \frac{\omega_j}{4m} \left[n_j + p_j^2 H_{\delta_j}(n_j) \right], \tag{138}$$

where H_δ sums the contributions of unqueried odd combinations of acquired basis rows and $B(n)$ is the population risk at that public observation.

Theorem 6 (Finite basis-pool prediction formula)**.** *If each regional pool is an affine basis of r coordinates, then for every indicator prior, noise vector, positive weight vector, and integer $0 \le b \le Jr$,*

$$R_{\mathrm{ad}}^{*,\mathrm{basis}}(b) = R_2^{*,\mathrm{basis}}(b) = R_{\mathrm{fix}}^{*,\mathrm{basis}}(b) = \min_{\substack{0 \le n_j \le r \\ \sum_j n_j = b}} B(n). \tag{139}$$

With public evidence, optimize at each W and average. Only the prior marginals p_j enter B.

Terminal risk depends on regional counts, so every adaptive policy averages values attained by fixed designs.

Proof of Theorem 6. Fix a public observation and let π denote its regional prior. Every subset of an affine basis has independent augmented rows. By Equation 67, the adaptive transcript law is the same for every indicator vector, so the posterior stays at π.

Consider a positive-probability transcript with regional count n_j. Under $Z_j = 1$, the clean signs on the independent acquired rows have a uniform product prior and independent binary observation channels. Each observed clean sign has posterior mean $\delta_j(-1)^{Y_{j,x}}$, and these signs remain conditionally independent. Query-selection factors cancel pathwise, so the same posterior applies when earlier outcomes determine the acquired identities.

An augmented evaluation row in the acquired span has a unique representation by acquired rows. Its first coordinate is one, so the number k of rows in the representation is odd. Conversely, every odd subset represents a distinct augmented evaluation row. The n_j singletons correspond exactly to acquired identities, and the unqueried represented coordinates correspond to odd $k \ge 3$, with $\binom{n_j}{k}$ possibilities.

For an unqueried coordinate represented by k rows, the conditional affine-sign mean is δ_j^k times the observed parity sign. The target's independent label error adds a factor δ_j, and the affine model has posterior weight p_j. The independent component has sign mean zero, so the squared marginal mean is $p_j^2 \delta_j^{2k+2}$. Outside the acquired span, both models have zero sign mean. At observed targets, the labels are known exactly.

A binary target with conditional sign mean s has variance $(1 - s^2)/4$. Summing over the regional cubes yields $B(n)$, whose squared parity terms depend only on the counts. Every adaptive policy therefore averages feasible values of $B(n)$, and a fixed design can realize a minimizing count vector. This proves equality across the acquisition classes. □

The binomial theorem gives the equivalent closed form

$$H_\delta(n) = \frac{\delta^2}{2} \left[(1+\delta^2)^n - (1-\delta^2)^n - 2n\delta^2 \right]. \tag{140}$$

The final term removes singleton contributions from the odd binomial sum. For numerical stability, we evaluate the nonnegative combinatorial expression. At $\eta = 0$, we have $H_1(r) = 2^{r-1} - r = m - r$, so querying the full basis yields

$$R_{\mathrm{basis}}(Jr) = \frac{1}{4} \left(1 - \sum_j \omega_j p_j^2 \right) - \frac{r}{4m} \sum_j \omega_j (1 - p_j^2). \tag{141}$$

At zero noise, revealing the indicator vector and clean parameter after complete-basis acquisition gives

$$R^{\mathrm{full}}_{\mathrm{oracle}} = \Big(1 - \frac{r}{m}\Big)\,\frac{1 - \sum_j \omega_j p_j}{4},$$
$$\Delta^{\mathrm{full}}_{\mathrm{ind}} = \Big(1 - \frac{r}{m}\Big)\,\frac{1}{4}\sum_j \omega_j p_j (1 - p_j). \tag{142}$$

The risk $R^{\mathrm{full}}_{\mathrm{oracle}}$ comes from uncertainty on unqueried independent targets. Its difference from $R_{\mathrm{basis}}(Jr)$ is the model-uncertainty gap $\Delta^{\mathrm{full}}_{\mathrm{ind}}$. For $m > r$, dividing by $1 - r/m$ gives the unqueried versions.

When every regional rate is fixed and strictly positive,

$$\frac{H_{\delta_j}(r)}{m} \le \delta_j^2 \left(\frac{1 + \delta_j^2}{2}\right)^r \longrightarrow 0. \tag{143}$$

Since $H_\delta(n)$ increases with n, the prediction gain also vanishes for every partial basis acquisition at fixed positive noise.

G.1 Likelihood and Terminal-Risk Ties

Equal regional likelihoods can lead to different continuation risks because the first batch also estimates the rule. For a first-batch history H, let $V_j(H)$ be the expected terminal Brier when the prescribed continuation is assigned to region j. Consider two conventions that agree off $\{L_Q = 1\}$ and select region one with probabilities $a(H)$ and $a_0(H)$ on that event. Conditioning on H gives $R_a - R_{a_0} = \mathbb{E}\,[\mathbf{1}\{L_Q = 1\}(a(H) - a_0(H))(V_1(H) - V_2(H))]$. The expectation averages first-batch outcomes and public randomization, with $a_0 = 1$ for the prescribed policy.

The affine-basis model gives a concrete example of this distinction. Take equal regional weights, public prior $1/2$, $1 \le t < b \le r$, and t independent first-batch rows in region one. Every history has $L_Q = 1$. Continuing there produces counts $(b, 0)$; selecting region two produces $(t, b - t)$. Both choices have localization probability $1/2$, and their prediction risks satisfy

$$R^{\mathrm{full}}_{\mathrm{switch}} - R^{\mathrm{full}}_{\mathrm{stay}} = \frac{H_\delta(b) - H_\delta(t) - H_\delta(b - t)}{32m} \ge 0, \tag{144}$$

where the superscripts indicate full-population loss and H_δ is the representation sum in Equation 138. Partition the b rows into the two batches. Every odd subset within one part appears in the sum for all b rows, and subsets meeting both parts contribute the remaining nonnegative terms. The inequality is strict for $b \ge 3$ and $0 \le \eta < 1/2$. Thus, continuing in the region already queried can improve prediction at a tie in model evidence. Multiplication by $2m/(2m - b)$ gives the unqueried-risk difference.

A choice among continuations with equal terminal risk preserves the optimized value. This follows by averaging for two batches and by backward recursion for full adaptation.

H Exact Noiseless Acquisition

Definition 1 (Restricted two-batch design)**.** *For each public observation $(W, \mathcal{P})$, fix a finite first-set family $\mathcal{B}_b \subseteq 2^{\mathcal{I}_\mathcal{P}}$ containing $\varnothing$, with $|S| \le b$ for every member. A design a chooses $S \in \mathcal{B}_b$ as a measurable function of $(W, \mathcal{P})$ and assigns every outcome y_S a fixed continuation $C(y_S) \subseteq \mathcal{I}_\mathcal{P} \setminus S$ of size $b - |S|$. Its value and the family optimum are*

$$V(\mathsf{a}) = \mathbb{E}\sum_j \frac{\omega_j}{m} \sum_{x \notin Q_j(\mathcal{H}_\mathsf{a})} \widehat{p}_{j,x}(1 - \widehat{p}_{j,x}), \quad R_{\mathcal{B}_b}(b) = \min_{\mathsf{a} \in \mathcal{A}_b(\mathcal{B}_b)} V(\mathsf{a}). \tag{145}$$

The terminal history is $\mathcal{H}_\mathsf{a}$, and $\mathcal{A}_b(\mathcal{B}_b)$ is the induced design class, abbreviated $\mathcal{A}_b$. Optimization is conditional on public information before averaging.

For the subcube family, Definition 1 specializes to R_{sub}. Allowing every first set yields R^*_2, and including the empty set establishes $R^*_{\mathrm{ad}} \le R^*_2 \le R_{\mathrm{sub}} \le R^*_{\mathrm{fix}}$. The threshold-one policy in Section 5.4 uses a single-region continuation in the public order.

Counts, augmented ranks, and consistency determine terminal risk and feasible actions under noiseless full access. We use these sufficient statistics in the recursions below and validate them by enumerating the full target array. We develop the noisy restricted-pool calculation in Appendix N.

H.1 Sufficient State

A zero-noise regional history is summarized by its count n_j, augmented rank k_j, and consistency flag o_j. Consistent histories have $o_j = 1$. For an inconsistent history, set $o_j = 0$ and record rank zero, since the affine model has zero posterior mass. The joint state is $s = ((n_j, k_j, o_j))_{j=1}^{J}$.

A rank-k augmented span contains $h_{\mathrm{span}}(k)$ regional coordinate rows, with $h_{\mathrm{span}}(0) = 0$ and $h_{\mathrm{span}}(k) = 2^{k-1}$ for $1 \le k \le r$. To see the latter count, the rows with first coordinate one form a coset of the first-coordinate-zero subspace, which has dimension $k - 1$.

A consistent history has affine likelihood ratio $L_j^{(0)} = 2^{n_j - k_j}$ against independent fair labels. An inconsistent history has ratio zero. These values follow by counting 2^{r-k_j} compatible parameters among 2^r and comparing with independent-label probability 2^{-n_j}. Along an adaptive transcript, query-selection factors agree across candidate indicator states and cancel. Thus, the regional posterior is

$$\alpha_z(s) = \frac{\pi_z \prod_{j:z_j=1} L_j^{(0)}}{\sum_{z'} \pi_{z'} \prod_{j:z'_j=1} L_j^{(0)}}, \quad \gamma_j(s) = \sum_z z_j \alpha_z(s), \tag{146}$$

where $\pi_z = \mathbb{P}(Z = z)$ is the prior at the fixed public observation, $\alpha_z(s)$ is its posterior at state s, and $\gamma_j(s)$ is the marginal affine probability. Every positive-probability history has a positive denominator.

Within a consistent closure, an unqueried coordinate has an identified affine value L_v conditional on the affine model. Its marginal mean is $1/2 + \gamma_j(L_v - 1/2)$, with variance $(1 - \gamma_j^2)/4$. Coordinates outside the closure remain fair. Summing their variances gives the terminal weighted Bayes loss

$$\mathcal{L}(s) = \sum_j \frac{\omega_j}{4m} \left[m - n_j - \mathbf{1}\{o_j = 1\} \gamma_j(s)^2 (h_{\mathrm{span}}(k_j) - n_j) \right]. \tag{147}$$

H.2 Adaptive Dynamic Programming

We can expand a consistent region whenever $k_j < r$. Querying an unused coordinate outside the span increases count and rank by one and leaves the history consistent. Both label values are fair under every surviving indicator state and lead to the same reduced successor.

If $h_{\mathrm{span}}(k_j) > n_j$, we can also make a relation check at an unused coordinate in its closure. Agreement occurs with probability $(1 + \gamma_j)/2$, increasing the count while preserving rank and consistency. A contradiction makes the region inconsistent, after which further queries increase only its recorded count.

Every unqueried identity belongs to one of these action classes. Since the counts also determine the remaining budget, the optimal value $V(s)$ obeys the recursion below, with terminal loss from Equation 147:

$$V(s) = \min_{\mathsf{u}} \sum_{s'} P(s' \mid s, \mathsf{u}) V(s'), \tag{148}$$

where u selects a region and an action class, and $P(s' \mid s, \mathsf{u})$ specifies the successor probabilities. Each feasible action corresponds to an available coordinate, so the recursion gives the full adaptive optimum.

H.3 Fixed Designs and Continuations

A fixed regional query set with count n and rank k is feasible exactly when

$$0 \le k \le \min(n, r), \quad n \le h_{\mathrm{span}}(k). \tag{149}$$

We can construct such a set from k independent augmented rows by adding distinct coordinates from their closure until the count reaches n. An independent region is consistent with probability

2^{k-n}, and an affine region is always consistent. Integrating Equation 147 over indicator states and consistency flags gives the fixed-design risk. Enumerating feasible regional counts and ranks with total count b then gives the full fixed optimum.

After first-batch state s, a fixed continuation adds q_j queries at unobserved identities and Δk_j independent directions in a consistent region. These additions are feasible exactly when

$$0 \le \Delta k_j \le \min(q_j, r - k_j), \quad n_j + q_j \le h_{\mathrm{span}}(k_j + \Delta k_j). \tag{150}$$

To realize a continuation, choose its independent extensions and then distinct identities from the expanded closure. All identities are fixed before we observe second-batch outcomes. An independent region that is currently consistent stays consistent with probability $2^{\Delta k_j - q_j}$; an affine region stays consistent with probability one. For an already inconsistent region, every $q_j \le m - n_j$ is feasible, and the region remains inconsistent.

Average over continuation outcomes using $\alpha_z(s)$, then minimize over feasible $(q_j, \Delta k_j)$ with $\sum_j q_j = b - \sum_j n_j$. Affine offsets determine which outcomes agree with the current constraints, and the integrated risk depends only on the reduced state.

For the subcube family in Section 6.1, integrating the best fixed continuation for each first-batch outcome gives $R_{\mathrm{sub}}(b)$ from Definition 1. At the operating budgets $b = 17$ on CP-8 and $b = 21$ on CP-12, the selected first set is an eight-point subcube in region one.

H.4 Uniform Acquisition

Uniform acquisition selects a global b-subset of distinct identities, so its regional counts are hypergeometric. Equivalently, each next identity is uniform among those remaining. The balanced fixed design specifies its regional counts in advance.

To average over the random query set and its labels, let $V_b^{\mathrm{unif}}(s)$ be the expected terminal population Brier risk of continuing uniformly from the reduced state s until total count b, and put $n = n_1 + n_2$. At $n = b$, set $V_b^{\mathrm{unif}}(s) = \mathcal{L}(s)$ using Equation 147. For $n < b$, write $s_{j,+}$ for the successor that adds an independent augmented row, $s_{j,=}$ for agreement with a dependent row, and $s_{j,\times}$ for a contradiction. The successor $s_{j,0}$ adds a label to an already inconsistent region. Then

$$\begin{aligned} V_b^{\mathrm{unif}}(s) = \frac{1}{2m-n} \sum_{j=1}^{2} \Bigg[& \mathbf{1}\{o_j = 1\} \Bigg\{ (m - h_{\mathrm{span}}(k_j)) V_b^{\mathrm{unif}}(s_{j,+}) \\ & + (h_{\mathrm{span}}(k_j) - n_j) \left[\frac{1+\gamma_j(s)}{2} V_b^{\mathrm{unif}}(s_{j,=}) + \frac{1-\gamma_j(s)}{2} V_b^{\mathrm{unif}}(s_{j,\times}) \right] \Bigg\} \\ & + \mathbf{1}\{o_j = 0\} (m - n_j) V_b^{\mathrm{unif}}(s_{j,0}) \Bigg]. \end{aligned} \tag{151}$$

The coefficients count available identities in each action class, omitting zero-coefficient terms. Starting from $s_0 = ((0,0,1),(0,0,1))$, unqueried Brier is $2mV_b^{\mathrm{unif}}(s_0)/(2m-b)$. The successor laws match those of adaptive optimization, with actions weighted by uniform coordinate selection.

We check this result independently by integrating query-set ranks exactly. Let $c_{n,k}$ count coordinate subsets of size n and augmented rank k. Initialize $c_{0,0} = 1$ and assign zero to infeasible entries. Counting possible final additions yields

$$n c_{n,k} = c_{n-1,k} \big[h_{\mathrm{span}}(k) - n + 1 \big] + c_{n-1,k-1} \big[m - h_{\mathrm{span}}(k-1) \big]. \tag{152}$$

The first term adds an unused row within the current span, and the second adds a row outside a rank-$(k-1)$ span. Every final subset is counted n times, with $\sum_k c_{n,k} = \binom{m}{n}$.

For a fixed count-rank tuple (n_1, k_1, n_2, k_2), put $a_j = 2^{k_j - n_j}$ and $D = pa_2 + (1-p)a_1$, where $p = \mathbb{P}(Z_1 = 1)$ is the exclusive regional prior. Under independent labels, consistency has probability a_j in region j, and D is the probability that both histories are consistent. Averaging squared posterior model probability over the labels yields

$$G_1 = p(1-a_2) + \frac{(pa_2)^2}{D}, \quad G_2 = (1-p)(1-a_1) + \frac{((1-p)a_1)^2}{D}. \tag{153}$$

With $p = 1/2$ in Figure 3, the exact unqueried risk is

$$R^{\mathrm{unq}}_{\mathrm{unif}}(b) = \frac{1}{4} - \frac{\sum_{n_1+n_2=b} \sum_{k_1,k_2} c_{n_1,k_1} c_{n_2,k_2} \sum_{j=1}^{2} \big[h_{\mathrm{span}}(k_j) - n_j\big] G_j}{4(2m-b)\binom{2m}{b}}. \tag{154}$$

The sums cover feasible counts and ranks. Rational evaluation produces the Uniform columns in Tables 5 and 6. Appendix I summarizes the independent complete-array validation.

H.5 Two-Batch Optimization

Let $F_b(s)$ be the best fixed-continuation risk from state s at total budget b. For a feasible first-batch tuple $a = (n_1, k_1, n_2, k_2)$, integrate its consistency outcomes over the prior indicator states. Optimizing the first batch yields

$$R^*_2(b) = \min_{a:\, n_1+n_2 \le b} \sum_s \mathbb{P}(s \mid a) F_b(s), \tag{155}$$

where the tuples satisfy Equation 149, $\mathbb{P}(s \mid a)$ averages the affine and independent consistency outcomes, and F_b uses Equation 150. Each feasible tuple corresponds to a coordinate set, so this minimization covers the full two-batch class.

At CP-8 budget $b = 16$, seven coordinates from an eight-point subcube have augmented rank four. A consistent first-batch outcome leads to nine additional queries in that region, reaching rank nine. After a contradiction, the second batch instead queries an affine basis of nine coordinates in the other region. Each continuation uses its full budget, and direct integration yields

$$R_{\mathrm{unq}} = \frac{184981}{1364992}. \tag{156}$$

The partial subcube preserves useful relations and leaves additional labels for continuation, so two-batch risk is lower. Tables 5 and 6 give the complete budget curves.

H.6 Batch-Size Optimization

The recursion also optimizes the batch split. Consider the subcube family in Section 6.1, for which we allow every fixed continuation. For total budget b, let $\mathcal{A}_{b,t}$ contain designs in $\mathcal{A}_b$ with exactly t first-batch labels. Define

$$R_{\mathrm{sub}}(b,t) = \min_{\mathsf{a} \in \mathcal{A}_{b,t}} V(\mathsf{a}), \quad R^*_2(b,t) = \min_{\substack{a=(n_1,k_1,n_2,k_2) \text{ feasible} \\ n_1+n_2=t}} \sum_s \mathbb{P}(s \mid a) F_b(s). \tag{157}$$

At first-batch size t, the first minimization uses the subcube family and the second covers all feasible sets. Both use F_b from Appendix H.5. Minimizing over t recovers $R_{\mathrm{sub}}(b)$ and $R^*_2(b)$, assigning $+\infty$ to an empty family.

We evaluate every first-batch count around the change in selected subcube size on CP-8 and CP-12, optimizing each fixed continuation. Table 4 reports the selected subcube count, every optimal full-class count, and risk at the equal split $\lfloor b/2 \rfloor$.

Partial subcubes retain verification relations with fewer first-batch labels, explaining the improvement available to the full two-batch class.

H.7 Complete Target-Array Validation

For small populations, we validate the recursions by enumerating every complete target array y with its probability. A history H records queried identities and observed labels. Let $P_H = \mathbb{P}(H)$ and $A_{H,v} = \mathbb{P}(H, Y_v = 1)$. With coordinate weight $\mu_v = \omega_j/m$ in region j, the history-weighted terminal loss is

$$\mathcal{B}(H) = \sum_v \mu_v \frac{A_{H,v}(P_H - A_{H,v})}{P_H}. \tag{158}$$

d	b	t_{sub}	t_2^*	R_{sub}	$R_2^*(b, \lfloor b/2 \rfloor)$	R_2^*
8	14	4	$4, 5$	0.160099	0.164561	0.160099
8	15	4	6	0.159257	0.144387	0.143727
8	16	4	7	0.158861	0.135732	0.135518
8	17	8	$8, 9$	0.131388	0.131388	0.131388
8	18	8	$8, \ldots, 11$	0.131148	0.131148	0.131148
12	18	4	$4, 5$	0.158308	0.176721	0.158308
12	19	4	6	0.157381	0.147696	0.141770
12	20	4	7	0.156909	0.136629	0.133489
12	21	8	8	0.129345	0.130032	0.129345
12	22	8	$8, 9$	0.129225	0.129556	0.129225

Table 4: Exact unqueried Brier for first set and batch allocation in the noiseless experiment. The count t_{sub} minimizes risk within the optimized subcube design family. The set t_2^* contains every optimal first-batch count in the full two-batch class. The equal-split column still optimizes all first sets and continuations compatible with that split.

We evaluate this expression directly from the complete target probability table, assigning zero contribution to zero-probability histories.

The fully adaptive calculation considers every unqueried identity and sums the values of its two successors. For a fixed remaining set, it instead sums terminal values over all label assignments on that set. Two-batch optimization enumerates first-batch subsets and outcomes, then minimizes fixed-continuation risk at each resulting history. Either batch may be empty.

I Numerical Validation

Likelihoods and posterior predictions in Equations 3, 4, and 5 agree across direct parameter summation, dependency expansion, Walsh-Hadamard transforms, and syndrome inference. The checks cover normalization, target-noise attenuation, zero-likelihood histories, and acquired-target removal, with syndrome calculations in dimensions two through seven. Complete target-array enumeration with rational arithmetic validates the reduced-state acquisition recursions, including outcome-dependent continuations and inconsistent histories. For Uniform acquisition, it also agrees with rational rank-count integration at every feasible unqueried budget in dimensions one and two.

We evaluate the conditional prediction bound on correlated finite dictionaries with unequal priors, four acquisition rules, and the specified count cutoffs. Integer subset sums verify dependency and target-representation identities, and rational integration verifies the unqueried and common-target Brier decompositions. We report exhaustive binary-code checks of Lemma 4 in Appendix N.6 and validate the higher-dimensional confidence formulas in Appendix N.8.

All acquired sets satisfy distinctness, exact budget, and pool membership, and are disjoint from the common evaluation targets. The risk calculations satisfy acquisition-class ordering and expected-risk monotonicity on fixed target populations. Our paired uncertainty estimates account for covariance across policies and budgets, and we evaluate numerical integration error separately from sampling variation.

I.1 Algorithm and Computational Cost

Algorithm 1 applies to any designated first set used in Section 5.4.

Exact transforms for likelihoods and prediction. For a parameter θ, let $e_\theta(Q)$ count disagreements with the n acquired labels and set $w_\theta(Q) = (1-\eta)^{n-e_\theta(Q)}\eta^{e_\theta(Q)}$, its observation likelihood. Under parameter prior ν_θ, the likelihood ratio and posterior are

$$L_Q = 2^n \sum_\theta \nu_\theta w_\theta(Q), \quad q_\theta(Q) = \frac{\nu_\theta w_\theta(Q)}{\sum_{\theta'} \nu_{\theta'} w_{\theta'}(Q)}. \tag{159}$$

At zero noise, the weight is the consistency indicator.

Algorithm 1 Prescribed two-batch acquisition and exact prediction

Input: Public decompositions $\mathcal{P}_j = B_j \dot{\cup} \mathcal{C}_j$ for $j = 1, 2$, with designated blocks $|B_j| = t$, ordered recovery remainders $\mathcal{C}_j$, integers $0 \le t \le b$, noise rates, and public prior $p(W)$. Each remainder contains at least $b - t$ coordinates. The first batch is B_1.

1. Set $Q = B_1$. Query its coordinates and observe y_Q.
2. Evaluate the affine likelihood ratio $L_Q(y_Q)$ by Equation 3 or the equivalent parameter likelihood sum.
3. Set $\widehat{J} = 1$ when $L_Q(y_Q) \ge 1$, and set $\widehat{J} = 2$ otherwise.
4. Select the first $b - t$ identities in the public order on $\mathcal{C}_{\widehat{J}}$. Query this entire set before using any of its outcomes.
5. Form each regional parameter posterior from all its observed labels. Combine regional likelihoods with $p(W)$ to compute γ_1, γ_2.
6. Compute all $M_{j,x}$ and return Equation 4, using the observed label at every acquired identity.

For one region, let $a(x) = (-1)^{Y_x}\mathbf{1}\{x \in Q\}$ and $n = |Q|$. The unnormalized Walsh-Hadamard transform is

$$C_u = \sum_{x \in \mathbb{F}_2^d} a(x)(-1)^{u^\top x}, \quad e_{\beta,u} = \frac{n - (-1)^\beta C_u}{2}, \quad u \in \mathbb{F}_2^d,\ \beta \in \mathbb{F}_2, \tag{160}$$

where C_u is the observed correlation with slope u and $e_{\beta,u}$ is its candidate disagreement count. These counts determine the likelihood weights $w_{\beta,u}$ and posterior probabilities $q_{\beta,u}$. A second transform yields

$$M_x = \sum_{u \in \mathbb{F}_2^d} (q_{0,u} - q_{1,u})(-1)^{u^\top x}. \tag{161}$$

At positive noise, we stabilize normalization by subtracting the largest log-likelihood. At zero noise, we normalize the consistency indicators.

Two transforms and normalization require $O(m \log m + n)$ arithmetic operations and $O(m)$ memory per region. Computing all predictions for J regions requires $O(Jd2^d + b)$ operations and $O(J2^d)$ memory. On a subcube of size $t = 2^{d_{\mathrm{sub}}}$, the transform evaluates its $2t$ affine restrictions in $O(t \log t)$ operations and $O(t)$ memory. The same correlations implement minimum-distance decoding by maximizing $|C_u|$ over slopes and then selecting the intercept.

Optimizing acquisition classes. In the noiseless full-access model, a regional state records count, augmented rank, and consistency. Up to budget b, there are at most $[2(b+1)(r+1)]^J$ joint states and at most three action types per region. The recursions in Appendix H optimize adaptive acquisition and enumerate feasible fixed continuations, with a state count polynomial in b and r for fixed J.

A noisy restricted pool with $K = |\mathcal{I}_\mathcal{P}|$ available identities has 3^K partially labeled sets. Given their posterior terminal costs, adaptive backward recursion uses at most $O(K3^K)$ transitions for a specified budget and $O(3^K)$ stored values. Fixed and two-batch optimization also enumerate query subsets and outcome assignments (Appendix N).

I.2 Dependency-Syndrome Inference

The dependency space offers another exact route to the posterior. For one region, let $n = |Q|$ be the query count, $k_Q = \mathrm{rank}(A_Q)$ its augmented rank, and $\ell_Q = n - k_Q$ its nullity. A basis matrix $H_Q \in \mathbb{F}_2^{\ell_Q \times n}$ of $\mathcal{D}(Q)$ maps labels to the syndrome $H_Q y$. The clean restriction is uniform on $\mathrm{im}(A_Q) = \ker H_Q$, so $L_{\eta,Q}(y) = 2^{\ell_Q}\mathbb{P}_\eta(H_Q\xi = H_Q y)$. We compute this syndrome probability with nonnegative dynamic programming on 2^{ℓ_Q} states. Systematic dependency coordinates initialize independent Bernoulli components. Each remaining coordinate mixes the current distribution with its shift by the corresponding column of H_Q. The resulting cost is $O(n2^{\ell_Q})$ time and $O(2^{\ell_Q})$ memory.

Set $\alpha_\eta = (1-\eta)^2 + \eta^2$ and $\eta' = \eta^2/\alpha_\eta$, the normalization and effective error rate obtained by squaring a channel likelihood. The population mean squared posterior sign can then be written as

$$\mathcal{U}_Q := \frac{1}{m}\sum_x M_x^2 = \frac{2^{n-r}\alpha_\eta^n L_{\eta',Q}(y) - 2^{2n-r}[\eta(1-\eta)]^n}{L_{\eta,Q}(y)^2}. \tag{162}$$

To derive the identity, let q_θ be the conditional affine-parameter posterior and $\bar{\theta}$ the parameter with the same slope and complementary intercept. Affine orthogonality gives $\mathcal{U}_Q = \sum_\theta q_\theta^2 - \sum_\theta q_\theta q_{\bar{\theta}}$. The squared likelihood, normalized by α_η^n, produces the first numerator term. The second follows from the product $[\eta(1-\eta)]^n$ of complementary parameter likelihoods.

For an observed coordinate x_i, the affine-sign mean is

$$M_{x_i} = (-1)^{y_i}\left[\frac{\delta+\delta^{-1}}{2} - \frac{\delta^{-1}-\delta}{2}\frac{L_{\eta,Q}(y+e_i)}{L_{\eta,Q}(y)}\right], \tag{163}$$

where e_i flips the ith observed bit under addition in $\mathbb{F}_2^n$ and $\delta = 1-2\eta$. Grouping dependency terms by whether they contain coordinate i proves the formula. Removing these queried-coordinate energies, with zero loss at acquired targets, yields the exact population risk:

$$R(\mathcal{H}) = \frac{1}{4} - \frac{b}{8m} - \frac{\delta^2}{8m}\sum_{j=1}^{2}\gamma_j^2\left(m\mathcal{U}_{Q_j} - \sum_{x\in Q_j} M_{j,x}^2\right). \tag{164}$$

The conditional risk $R(\mathcal{H})$ assigns zero loss to acquired targets and applies Equation 4 to the remaining coordinates.

J Related Work

Decision states and prediction objectives. Proper scoring rules connect uncertainty with predictive performance (Gneiting & Raftery, 2007; Tan et al., 2021). Our exact comparisons use the same posterior for acquisition directed at latent information and at target prediction. The confidence-order result distinguishes the associated decision states: a Blackwell comparison for the regional indicator and a comparison for an unqueried target label (Blackwell, 1953).

Information bounds and computation. Theorem 3 applies binary-channel information increments from finite-blocklength converses (Polyanskiy et al., 2010) to prediction on outcome-dependent count events. Adaptive identification also links observations and decisions through likelihood comparisons (Kaufmann et al., 2016). For random pools, generalized Hamming weights (Wei, 1991) and matroid packing (Edmonds, 1965) control overlapping dependencies. Polar constructions establish capacity and convergence guarantees for designed codes (Arikan, 2009; Tal & Vardy, 2013; Guruswami et al., 2020). We analyze verification and recovery on the coordinates available in a public pool. Exact computational costs are given in Appendix I.1.

Testing predictors and finite-population evaluation. Active testing evaluates a specified predictor (Kossen et al., 2021), with surrogate estimators and evaluation during training extending the connection between acquisition and assessment (Kossen et al., 2022; Yu et al., 2023). Sequential hypothesis testing selects observations for a declaration (Naghshvar & Javidi, 2013). Our regional declaration determines where the remaining labels estimate the predictor. Sampling-without-replacement analysis supports finite-population uncertainty quantification (Hoeffding, 1963; Bardenet & Maillard, 2015; Waudby-Smith & Ramdas, 2020), and testable learning under distribution shift studies prediction certification across distributions (Klivans et al., 2024). In this comparison, we change query access while keeping the prediction distribution fixed.

J.1 Learning Parity with Noise and Query Access

In an affine region, querying x observes the parity specified by $(1, x^\top)$ with an independent Bernoulli error. This is learning parity with noise, including an intercept. Label-independent distinct uniform queries follow the random-example law conditional on distinct inputs. With access to the full public pool, we can also use coordinates and earlier outcomes to guide subsequent queries.

Combining k parity observations attenuates sign correlation by $(1-2\eta)^k$ under independent errors (Blum et al., 2003). Our likelihood formulas distinguish zero-sum relations, which provide regional evidence, from target representations, which support prediction. Uniform relation bounds and prediction control at outcome-dependent counts give the allocation converse in Theorem 4. Appendix I.1 details the attaining decoder's exponential dependence on dimension.

Parity-learning analysis connects random examples with random linear-code decoding and includes non-adaptive membership-query algorithms under persistent classification noise (Feldman, 2007). Our labels are persistent because each coordinate retains its corrupted label. Membership queries allow designed coordinates throughout the domain; our learner acquires labels from an announced pool. The structured construction includes a prescribed subcube at the same cardinality as the random pool. Our uniform prior includes every slope, whereas attribute-efficiency measures complexity through the number of relevant coordinates. The additional regional verification task determines how the shared budget is allocated to parity recovery.

K Budget Curves and Paired Inference

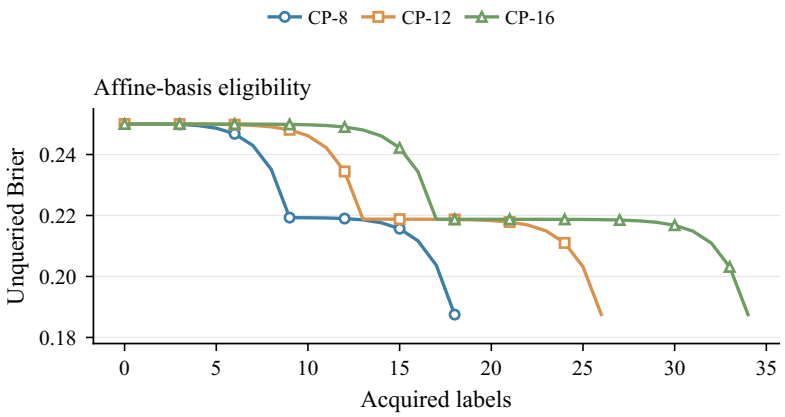


Figure 6: Exact basis-pool control across every feasible budget. Each curve is simultaneously the fixed, two-batch, and adaptive optimum for the indicated population.

K.1 Exact Budget Curves

Tables 5 and 6 give the 46 budget configurations for $d \in \{8, 12\}$ and $0 \le b \le 2(d+1)$, evaluated by the recursion in Appendix H.5. Figure 6 gives the basis-pool optima on CP-8, CP-12, and CP-16 over their feasible budgets.

K.2 Risk Normalization

For population size N and b acquired labels, write R_b^{full} for full-population Brier and R_b^{unq} for Brier averaged over the remaining coordinates. Acquired targets are predicted exactly, so

$$R_b^{\mathrm{unq}} = \frac{N}{N-b} R_b^{\mathrm{full}}. \tag{165}$$

With general regional weights, renormalizing the regional weights on unqueried instances gives

$$R_\omega^{\mathrm{unq}}(\widehat{p}, Y) = \frac{R_\omega(\widehat{p}, Y)}{1 - \sum_j \omega_j n_j / m}. \tag{166}$$

The denominator is the remaining target mass, determined by counts n_j. We normalize within each history with positive remaining mass, then average over histories.

b	Uniform	Fixed optimum	Subcube design	Two-batch optimum	Adaptive optimum
0	0.250000	0.250000	0.250000	0.250000	0.250000
1	0.250000	0.250000	0.250000	0.250000	0.250000
2	0.250000	0.250000	0.250000	0.250000	0.250000
3	0.249969	0.249877	0.249877	0.249877	0.249877
4	0.249878	0.249508	0.249508	0.249508	0.249508
5	0.249688	0.248644	0.248644	0.248644	0.248644
6	0.249348	0.246789	0.246789	0.246789	0.246789
7	0.248787	0.242946	0.242946	0.242946	0.242946
8	0.247907	0.235119	0.235119	0.235119	0.235119
9	0.246583	0.219309	0.219309	0.219309	0.219309
10	0.244665	0.209163	0.205926	0.205926	0.205926
11	0.241992	0.201098	0.193987	0.193987	0.193301
12	0.238415	0.195778	0.180778	0.180778	0.178809
13	0.233814	0.192709	0.161772	0.161772	0.157216
14	0.228109	0.191098	0.160099	0.160099	0.150341
15	0.221266	0.190319	0.159257	0.143727	0.138956
16	0.213301	0.189985	0.158861	0.135518	0.133372
17	0.204319	0.189864	0.131388	0.131388	0.130814
18	0.194564	0.187500	0.131148	0.131148	0.129923

Table 5: Exact CP-8 unqueried Brier. Uniform averages global sampling without replacement, and the subcube column optimizes its specified first-set family. The other columns optimize the full fixed, two-batch, and adaptive classes.

b	Uniform	Fixed optimum	Subcube design	Two-batch optimum	Adaptive optimum
0	0.250000	0.250000	0.250000	0.250000	0.250000
1	0.250000	0.250000	0.250000	0.250000	0.250000
2	0.250000	0.250000	0.250000	0.250000	0.250000
3	0.249998	0.249992	0.249992	0.249992	0.249992
4	0.249992	0.249969	0.249969	0.249969	0.249969
5	0.249980	0.249916	0.249916	0.249916	0.249916
6	0.249959	0.249801	0.249801	0.249801	0.249801
7	0.249923	0.249565	0.249565	0.249565	0.249565
8	0.249866	0.249084	0.249084	0.249084	0.249084
9	0.249776	0.248113	0.248113	0.248113	0.248113
10	0.249638	0.246165	0.246165	0.246165	0.246165
11	0.249430	0.242261	0.242261	0.242261	0.242261
12	0.249117	0.234444	0.234444	0.234444	0.234444
13	0.248654	0.218800	0.218800	0.218800	0.218800
14	0.247977	0.208405	0.204568	0.204568	0.204568
15	0.247007	0.200092	0.192349	0.192349	0.191405
16	0.245649	0.194553	0.178989	0.178989	0.176599
17	0.243798	0.191299	0.160083	0.160083	0.155015
18	0.241347	0.189527	0.158308	0.158308	0.148051
19	0.238201	0.188605	0.157381	0.141770	0.136762
20	0.234281	0.188136	0.156909	0.133489	0.131229
21	0.229535	0.187903	0.129345	0.129345	0.128704
22	0.223934	0.187790	0.129225	0.129225	0.127931
23	0.217479	0.187737	0.129165	0.127212	0.126565
24	0.210200	0.187714	0.129135	0.126206	0.125882
25	0.202180	0.187706	0.129120	0.125702	0.125541
26	0.193600	0.187500	0.129105	0.125443	0.125363

Table 6: Exact CP-12 budget curve, with acquisition classes and normalization as in Table 5.

Condition on the public pools and fix a prediction population with weights $w_v \geq 0$ summing to one. Define

$$V_t = \sum_v w_v \operatorname{Var}(Y_v \mid \mathcal{H}_t), \quad \widehat{p}_{t,v} = \mathbb{E}[Y_v \mid \mathcal{H}_t], \tag{167}$$

where $\mathcal{H}_t$ records the first t acquired labels and V_t is the conditional risk on the fixed evaluation population. The conditional variance decomposition gives

$$V_t - \mathbb{E}[V_{t+1} \mid \mathcal{H}_t] = \sum_v w_v \mathbb{E}\big[(\widehat{p}_{t+1,v} - \widehat{p}_{t,v})^2 \mid \mathcal{H}_t\big] \geq 0. \tag{168}$$

Thus, each additional label reduces or preserves conditional expected risk on a fixed target population, including within a fixed batch. Unqueried risk also changes its evaluation population: removing a predictable target can raise average uncertainty among the remaining coordinates. The common-target construction in Appendix L.2 keeps the evaluation population fixed.

K.3 Paired Uncertainty

For paired sample i, compare policy a and control c on their shared target array. Define

$$\Delta_i^{\mathsf{a},\mathsf{c}} = R_i(\mathsf{c}) - R_i(\mathsf{a}), \quad \overline{\Delta}^{\mathsf{a},\mathsf{c}} = \frac{1}{n}\sum_{i=1}^{n} \Delta_i^{\mathsf{a},\mathsf{c}}, \quad \widehat{\sigma}_\Delta^2 = \frac{1}{n-1}\sum_{i=1}^{n} (\Delta_i^{\mathsf{a},\mathsf{c}} - \overline{\Delta}^{\mathsf{a},\mathsf{c}})^2, \tag{169}$$

where R_i uses the specified target population on array i and n is the number of paired arrays. Positive $\overline{\Delta}^{\mathsf{a},\mathsf{c}}$ favors a. Our studies use $n = 128$ or $n = 512$ independent arrays, with pointwise interval $\overline{\Delta}^{\mathsf{a},\mathsf{c}} \pm t_{n-1,.975}\widehat{\sigma}_\Delta/\sqrt{n}$. The nominal 95% Student intervals have $n-1$ degrees of freedom and use within-array differences to account for covariance. We apply Bonferroni adjustment to budget interactions (Appendix L.5) and Holm adjustment to affine-flat contrasts (Appendix P.1).

L Public Evidence and Finite-Budget Experiments

Design	Available regional coordinates	Quantity evaluated
Noiseless CubePair	Full cube, $d = 8, 12, 16$	Exact acquisition-class and batch-allocation risks
Paired pool comparison	$q = r^2$, $d = 8, 12$	Two-check, affine-flat, and subcube verification with fixed continuations
Matched-rank pools	Six coordinates, $d = 4$	Exact common-target risks and acquisition objectives
Dimension-dependent bounds	$q = r^2$, $r = 256$ to 262144	Numerical evaluation of finite risk inequalities

Table 7: Finite experimental designs. Exact calculations integrate the label outcomes. Paired studies share target arrays and evaluation coordinates across policies.

L.1 Public-Prior Sensitivity

We vary $p(W) \in \{.50, .60, .70, .80, .90, .95, .99\}$ on CP-8, CP-12, and CP-16, with respective budgets 17, 21, and 25. The model uses full access, zero label noise, and uniformly distributed affine parameters. Table 8 compares the fixed, optimized subcube, full two-batch, and adaptive risks.

Stronger public evidence improves fixed acquisition. The remaining differences show how first-batch geometry and additional acquisition rounds improve prediction after public evidence locates the affine region.

L.2 Random-Prefix and Subcube Comparison

The paired pool study uses $d \in \{8, 12\}$, $r = d + 1$, noise rates $\eta \in \{.01, .025, .05\}$, and budgets $b = \lfloor \lambda r \rfloor$ for $\lambda \in \{1.5, 1.7, 1.9\}$. Each pool contains $q = r^2$ coordinates and a designated first set of $t_r = 8$ labels.

With first-batch count $t = t_r = 8$, the continuation length is $n = b - t$, so

d	r	t	$n = b - t$
8	9	8	$(5, 7, 9)$
12	13	8	$(11, 14, 16)$

(170)

Population	b	$p(W)$	Fixed optimum	Optimized subcube design	Two-batch optimum	Adaptive optimum
CP-8	17	0.50	0.18986	0.13139	0.13139	0.13081
		0.60	0.17765	0.13087	0.13087	0.13041
		0.70	0.16536	0.13028	0.13028	0.12995
		0.80	0.15288	0.12959	0.12959	0.12939
		0.90	0.14012	0.12861	0.12861	0.12815
		0.95	0.13345	0.12794	0.12794	0.12721
		0.99	0.12747	0.12692	0.12681	0.12613
CP-12	21	0.50	0.18790	0.12935	0.12935	0.12870
		0.60	0.17539	0.12854	0.12854	0.12803
		0.70	0.16287	0.12774	0.12774	0.12735
		0.80	0.15035	0.12693	0.12693	0.12667
		0.90	0.13784	0.12613	0.12613	0.12600
		0.95	0.13157	0.12572	0.12572	0.12566
		0.99	0.12652	0.12534	0.12534	0.12533
CP-16	25	0.50	0.18776	0.12917	0.12917	0.12852
		0.60	0.17521	0.12834	0.12834	0.12782
		0.70	0.16266	0.12751	0.12751	0.12712
		0.80	0.15012	0.12668	0.12668	0.12642
		0.90	0.13757	0.12585	0.12585	0.12572
		0.95	0.13130	0.12544	0.12544	0.12537
		0.99	0.12628	0.12511	0.12511	0.12509

Table 8: Public-Prior Sensitivity. The optimized subcube design uses the specified subcube query sets and optimal fixed continuations for each outcome. The fixed, two-batch, and adaptive benchmarks optimize their full classes. All reported risks use unqueried normalization.

The budgets are $(13, 15, 17)$ at $d = 8$ and $(19, 22, 24)$ at $d = 12$. After correct selection, prediction depends on the continuation's rank and on any constraints already acquired in the affine region. The subcube has augmented rank $k_{\text{test}} = 1 + \log_2 t_r = 4$ and nullity $t_r - k_{\text{test}} = 4$. An independent eight-row set has rank eight. Writing $k_\star$ for the acquired augmented rank in the affine region, correct localization implies

$$k_\star \le \begin{cases} \min\{r, k_{\text{test}} + b - t_r\}, & J_\star = 1, \\ \min\{r, b - t_r\}, & J_\star = 2, \end{cases} \tag{171}$$

where $J_\star$ identifies the affine region. For a nonempty acquired set, Equation 5 also implies

$$\frac{1}{m} \sum_{x \in \mathbb{F}_2^d} M_{J_\star, x}^2 \le 2^{k_\star - r}. \tag{172}$$

The acquired span contains $2^{k_\star - 1}$ target rows with at most one squared sign mean. All other means are zero, so the bound follows. An empty acquired set has zero energy.

In each region, a common uniform permutation couples the pools. The random pool contains the first q coordinates, and its initial t_r form the designated set. The structured pool contains $\{0, \dots, t_r - 1\}$ and the first $q - t_r$ permutation coordinates outside that subcube. Each remainder retains the permutation order. The two-batch policy queries the region-one first set and assigns the continuation using its threshold-one decision. Concentrated Fixed instead queries the first b coordinates in region one's ordered remainder. Balanced Fixed assigns $\lfloor b/2 \rfloor$ labels to region one and the rest to region two. Every acquired label is used in Equation 4.

The common evaluation population and realized Brier are

$$\mathcal{V}_j^{\text{eval}} = \mathbb{F}_2^d \setminus \left(\mathcal{P}_j^{\text{rand}} \cup \mathcal{P}_j^{\text{str}} \right), \quad R^{\text{common}} = \frac{\displaystyle\sum_{j=1}^{2} \sum_{x \in \mathcal{V}_j^{\text{eval}}} (\widehat{p}_{j,x} - Y_{j,x})^2}{|\mathcal{V}_1^{\text{eval}}| + |\mathcal{V}_2^{\text{eval}}|}, \tag{173}$$

where $\mathcal{V}_j^{\text{eval}}$ is nonempty and fixed across policies and budgets within each sample, and R^{common} is the realized Brier score on these targets. We also record full-population and unqueried risks.

Each dimension-noise pair uses 128 independent target arrays, shared with their public pools across the three budgets. Arrays are the sampling units for means and standard errors, and random-minus-structured contrasts use within-array differences with pointwise Student intervals on 127 degrees of freedom.

In Appendix L.3, we compare the prefix with two disjoint checks selected from the full pool; their existence follows from Appendix D.4.

The prefix has lower mean risk at several short-continuation budgets, and longer continuations shift the comparison toward the subcube. Tables 10 and 12 relate this pattern to regional selection and parameter precision.

For the selection and prediction analysis, define

$$A_{\mathrm{loc}} = \mathbb{P}(\widehat{J} = J_\star), \quad \rho_{\mathrm{param}} = \mathbb{E}\left[\frac{1}{m}\sum_{x\in\mathbb{F}_2^d} M^2_{J_\star,x}\right], \tag{174}$$

where A_{loc} is the probability of correct regional selection and ρ_{param} measures posterior precision in the affine rule. We also evaluate the empirical selection fraction and equal-prior testing accuracy:

$$\widehat{A}_{\mathrm{sel}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\widehat{J}_i = J_{\star,i}\}, \qquad A_{\mathrm{bal}}(Q) = \frac{1}{2} + \frac{1}{4}\mathbb{E}_{0,Q}|L_Q(Y) - 1|, \tag{175}$$

where $\widehat{J}_i$ and $J_{\star,i}$ are the selected and affine regions in sample i, and n is the sample count. For a fixed first set, $A_{\mathrm{bal}}(Q)$ integrates both label laws. An independent prefix has $L_Q = 1$ and $A_{\mathrm{bal}}(Q) = 1/2$ exactly. Its empirical selection fraction depends on the sampled regional composition because ties select region one. Equation 20 connects testing and prediction by weighting target precision with squared regional confidence. Tables 9 and 11 give realized Brier on the same common targets.

		Concentrated Fixed		Balanced Fixed		Prescribed two-batch policy	
η	b	Random	Structured	Random	Structured	Random	Structured
		0.19111	0.19190	0.24113	0.24255	0.19205	0.20796
0.010	13	(0.00555)	(0.00561)	(0.00077)	(0.00077)	(0.00565)	(0.00439)
		0.18895	0.18708	0.23148	0.23341	0.18499	0.18527
0.010	15	(0.00550)	(0.00532)	(0.00143)	(0.00159)	(0.00533)	(0.00473)
		0.18473	0.18367	0.21355	0.21937	0.17734	0.16219
0.010	17	(0.00531)	(0.00537)	(0.00282)	(0.00269)	(0.00528)	(0.00437)
		0.21085	0.20716	0.24459	0.24397	0.20744	0.22341
0.025	13	(0.00537)	(0.00507)	(0.00076)	(0.00066)	(0.00507)	(0.00369)
		0.20579	0.20418	0.23879	0.23925	0.20472	0.20416
0.025	15	(0.00540)	(0.00520)	(0.00152)	(0.00140)	(0.00521)	(0.00417)
		0.20097	0.20001	0.22769	0.23032	0.19925	0.18860
0.025	17	(0.00520)	(0.00532)	(0.00278)	(0.00247)	(0.00541)	(0.00487)
		0.22679	0.22403	0.24642	0.24715	0.22502	0.23769
0.050	13	(0.00427)	(0.00440)	(0.00064)	(0.00064)	(0.00438)	(0.00286)
		0.21780	0.21620	0.24379	0.24391	0.21725	0.23028
0.050	15	(0.00448)	(0.00464)	(0.00116)	(0.00124)	(0.00458)	(0.00380)
		0.21247	0.21305	0.23831	0.23844	0.21132	0.21776
0.050	17	(0.00480)	(0.00502)	(0.00212)	(0.00217)	(0.00478)	(0.00445)

Table 9: Common-population Brier for the $d = 8$ short-continuation control. Cells give the mean and standard error over 128 paired arrays. All policies and budgets share each array's evaluation population outside both query pools.

The identical entries arise from the shared permutation (Table 11). Let $\pi_1, \pi_2, \ldots$ denote the region-one order and B the prescribed subcube. A prefix with $L_Q = 1$ leads the two-batch policy to acquire $\{\pi_1, \ldots, \pi_b\}$ in region one. Structured Concentrated Fixed takes the first b permutation coordinates outside B. On the event

$$L_Q = 1, \qquad \{\pi_1, \ldots, \pi_b\} \cap B = \varnothing. \tag{176}$$

The acquired sets coincide, as do their posterior predictions and target-wise losses. Random Concentrated Fixed starts after the designated prefix and acquires $\{\pi_9, \ldots, \pi_{b+8}\}$, so its risk can differ.

		$\widehat{A}_{\mathrm{sel}}$		ρ_{param}		Paired risk gain
η	b	Random	Structured	Random	Structured	Random minus structured
0.010	13	0.6562	0.9531	0.5383	0.3673	-0.01592 ± 0.00843
0.010	15	0.6562	0.9531	0.5813	0.5593	-0.00028 ± 0.00798
0.010	17	0.6562	0.9531	0.6183	0.7462	0.01515 ± 0.01049
0.025	13	0.5781	0.9141	0.4306	0.2743	-0.01598 ± 0.00806
0.025	15	0.5781	0.9141	0.4822	0.4284	0.00055 ± 0.00751
0.025	17	0.5781	0.9141	0.5189	0.5986	0.01064 ± 0.01029
0.050	13	0.5703	0.7656	0.3639	0.1661	-0.01267 ± 0.00770
0.050	15	0.5703	0.7656	0.4099	0.2744	-0.01304 ± 0.00869
0.050	17	0.5703	0.7656	0.4611	0.4064	-0.00644 ± 0.01122

Table 10: Prescribed two-batch policy selection and prediction summaries for the $d = 8$ short-continuation control. The realized affine-selection fraction and mean squared posterior sign use 128 paired arrays. Independent prefixes follow the region-one tie convention, with equal-prior testing accuracy $1/2$. The final column reports the common-population risk contrast and its pointwise 95% interval half-width. Positive contrast favors structured pools.

		Concentrated Fixed		Balanced Fixed		Prescribed two-batch policy	
η	b	Random	Structured	Random	Structured	Random	Structured
		0.19410	0.19447	0.24562	0.24547	0.19447	0.19365
0.010	19	(0.00533)	(0.00596)	(0.00040)	(0.00034)	(0.00596)	(0.00407)
		0.18862	0.18736	0.23790	0.23742	0.18736	0.15969
0.010	22	(0.00560)	(0.00560)	(0.00082)	(0.00069)	(0.00560)	(0.00436)
		0.18570	0.18507	0.22774	0.22528	0.18507	0.14395
0.010	24	(0.00527)	(0.00543)	(0.00175)	(0.00170)	(0.00543)	(0.00363)
		0.20417	0.20517	0.24707	0.24743	0.20542	0.21726
0.025	19	(0.00493)	(0.00501)	(0.00035)	(0.00036)	(0.00500)	(0.00409)
		0.19324	0.19478	0.24230	0.24352	0.19478	0.19119
0.025	22	(0.00512)	(0.00506)	(0.00082)	(0.00086)	(0.00506)	(0.00509)
		0.19100	0.19273	0.23634	0.23817	0.19273	0.17597
0.025	24	(0.00504)	(0.00505)	(0.00165)	(0.00167)	(0.00505)	(0.00514)
		0.22760	0.22511	0.24829	0.24838	0.22536	0.23004
0.050	19	(0.00439)	(0.00412)	(0.00028)	(0.00027)	(0.00414)	(0.00335)
		0.22290	0.21848	0.24635	0.24583	0.21837	0.21236
0.050	22	(0.00474)	(0.00462)	(0.00060)	(0.00063)	(0.00462)	(0.00435)
		0.21601	0.21643	0.24357	0.24268	0.21614	0.19905
0.050	24	(0.00452)	(0.00465)	(0.00107)	(0.00118)	(0.00466)	(0.00451)

Table 11: Common-population Brier at $d = 12$, with means and standard errors over 128 paired arrays. Policies and budgets share the evaluation population outside both pools. Under the event in Equation 176, the random two-batch policy and structured Concentrated Fixed acquire the same set.

Random streams. All stochastic studies use seed 2026. Independent PCG64 streams generate targets and pools, with indices for the study, dimension, noise rate, sample, and experimental component. Each sample is shared across policies and budgets to preserve the pairing used for uncertainty calculations.

L.3 Two-Check Selection from the Full Pool

The two-check control selects eight first-batch labels from the full region-one pool. It uses the same threshold-one decision, fixed continuation, and posterior predictor as the prefix and subcube designs.

Scan coordinate pairs lexicographically by position in the public permutation, recording the first pair for each binary sum. The first repeated sum identifies a four-point check. Remove its coordinates and repeat to obtain a disjoint second check. The remaining coordinates retain their public order. The pair-count argument guarantees both checks, and the exact likelihood includes any additional dependencies. Their eight-point union has augmented rank between four and six. Rank four occurs when the union is an eight-point affine flat.

		$\widehat{A}_{\mathrm{sel}}$		ρ_{param}		Paired risk gain
η	b	Random	Structured	Random	Structured	Random minus structured
0.010	19	0.5547	0.9375	0.5433	0.4555	0.00082 ± 0.00923
0.010	22	0.5547	0.9375	0.5493	0.7593	0.02767 ± 0.01207
0.010	24	0.5547	0.9375	0.5543	0.8898	0.04112 ± 0.01282
0.025	19	0.5469	0.8672	0.4674	0.3186	-0.01185 ± 0.00891
0.025	22	0.5469	0.8672	0.5209	0.5517	0.00359 ± 0.01215
0.025	24	0.5469	0.8672	0.5392	0.6857	0.01676 ± 0.01333
0.050	19	0.4609	0.8359	0.3137	0.2035	-0.00468 ± 0.00710
0.050	22	0.4609	0.8359	0.3914	0.3992	0.00601 ± 0.00966
0.050	24	0.4609	0.8359	0.4165	0.5157	0.01709 ± 0.01101

Table 12: Prescribed two-batch policy selection and prediction summaries at $d = 12$. The realized affine-selection fraction and mean squared posterior sign use 128 paired arrays. Independent prefixes follow the region-one tie convention, with equal-prior testing accuracy $1/2$. The final column reports the common-population risk contrast and its pointwise 95% interval half-width. Positive contrast favors structured pools.

We evaluate the controls on the shared arrays and common targets of the 128-sample study. Table 13 gives all eighteen dimension, noise, and budget configurations, with pointwise paired intervals.

To connect the risk contrasts to their sources of uncertainty, evaluate the variance decomposition on the same common targets. For a terminal history, define

$$\begin{aligned} D_0 &= \frac{1}{4} - \frac{\delta^2}{4M_{\mathrm{eval}}} \sum_j \gamma_j |\mathcal{V}_j^{\mathrm{eval}}|, \\ D_{\mathrm{param}} &= \frac{\delta^2}{4M_{\mathrm{eval}}} \sum_j \gamma_j \sum_{x \in \mathcal{V}_j^{\mathrm{eval}}} (1 - M_{j,x}^2), \\ D_{\mathrm{ind}} &= \frac{\delta^2}{4M_{\mathrm{eval}}} \sum_j \gamma_j (1 - \gamma_j) \sum_{x \in \mathcal{V}_j^{\mathrm{eval}}} M_{j,x}^2, \end{aligned} \tag{177}$$

where D_0, D_{param}, and D_{ind} are the average residual variance, parameter uncertainty, and indicator uncertainty over the common targets. Equation 180 gives $R_{\mathrm{cond}} := \mathbb{E}[R^{\mathrm{common}} \mid \mathcal{H}, \mathcal{P}] = D_0 + D_{\mathrm{param}} + D_{\mathrm{ind}}$. The indicator component is the value of revealing the region after acquisition. By comparison, A_{loc} concerns the first-batch selection decision. Conditional and realized Brier have the same expectation.

L.4 Paired Prediction with 512 Independent Samples

The 512-sample study uses the sampling design in Appendix L.2, with target arrays and random streams independent of the 128-sample study. We draw the affine-region indicator uniformly and its parameter uniformly on $\mathbb{F}_2^r$. Each coordinate is initially assigned a fair binary label. In the affine region, the rule value plus an independent Bernoulli flip replaces this label. An independent pool stream generates permutations in regional order. The affine-flat study uses the same generation procedure with its own independent samples.

All three designs use the pool coupling in Appendix L.2 and Algorithm 1. Their posteriors include every acquired label, and their evaluation targets are the common coordinates outside both pools.

Tables 14 and 15 report conditional and realized Brier. The conditional-risk contrasts also determine the paired budget interactions in Appendix L.5.

L.5 Regional Contributions and Budget Interactions

Sharing target arrays across budgets allows us to measure the change in risk contrast within each sample. For paired array i, define

$$\Delta_i(b) = R_{\mathrm{cond},i}^{\mathrm{check}}(b) - R_{\mathrm{cond},i}^{\mathrm{subcube}}(b), \quad I_i = \Delta_i(b_+) - \Delta_i(b_-), \tag{178}$$

d	η	b	Two-check Brier	$\widehat{A}_{\mathrm{sel}}$	Prefix minus two-check	Two-check minus subcube
8	0.010	13	0.19664 (0.00527)	0.8594	-0.00459 ± 0.00463	-0.01133 ± 0.00837
8	0.010	15	0.18402 (0.00514)	0.8594	0.00097 ± 0.00406	-0.00125 ± 0.00788
8	0.010	17	0.16596 (0.00496)	0.8594	0.01138 ± 0.00854	0.00377 ± 0.01052
8	0.025	13	0.21208 (0.00471)	0.7656	-0.00464 ± 0.00584	-0.01134 ± 0.00749
8	0.025	15	0.20343 (0.00475)	0.7656	0.00129 ± 0.00621	-0.00073 ± 0.00740
8	0.025	17	0.19610 (0.00535)	0.7656	0.00315 ± 0.00971	0.00750 ± 0.01150
8	0.050	13	0.22883 (0.00405)	0.7422	-0.00381 ± 0.00566	-0.00885 ± 0.00734
8	0.050	15	0.22114 (0.00408)	0.7422	-0.00389 ± 0.00670	-0.00914 ± 0.00787
8	0.050	17	0.20771 (0.00417)	0.7422	0.00361 ± 0.00815	-0.01004 ± 0.00989
12	0.010	19	0.18966 (0.00510)	0.7891	0.00481 ± 0.00600	-0.00399 ± 0.00828
12	0.010	22	0.17348 (0.00567)	0.7891	0.01388 ± 0.00987	0.01379 ± 0.01205
12	0.010	24	0.16603 (0.00555)	0.7891	0.01904 ± 0.01105	0.02208 ± 0.01264
12	0.025	19	0.20304 (0.00471)	0.8125	0.00238 ± 0.00546	-0.01423 ± 0.00776
12	0.025	22	0.18510 (0.00492)	0.8125	0.00967 ± 0.00990	-0.00608 ± 0.01107
12	0.025	24	0.17767 (0.00512)	0.8125	0.01506 ± 0.01160	0.00170 ± 0.01253
12	0.050	19	0.22762 (0.00383)	0.7031	-0.00226 ± 0.00547	-0.00242 ± 0.00751
12	0.050	22	0.21848 (0.00447)	0.7031	-0.00011 ± 0.00822	0.00612 ± 0.00996
12	0.050	24	0.20800 (0.00455)	0.7031	0.00814 ± 0.00938	0.00895 ± 0.01081

Table 13: Two-check eight-label control on the random pools. Brier entries give means with standard errors. Contrasts use the same targets outside both pool interventions and give mean differences with pointwise 95% interval half-widths. Positive values favor the subtracted policy. The realized selection fraction $\widehat{A}_{\mathrm{sel}}$ is shared across the three budgets at each dimension and noise pair.

where b_- and b_+ are the smallest and largest budgets. The contrast $\Delta_i(b)$ is two-check minus subcube risk, and I_i is its within-array change across the budget endpoints. A positive interaction indicates increasing relative value for the subcube. We average over the 512 samples in Table 14.

Table 16 reports Student intervals with nominal 95% family-wise coverage after Bonferroni adjustment across the six dimensions and noise pairs. Each half-width is $t_{n-1,1-.05/(2\cdot 6)}$ times the standard error of I_i, accounting for covariance across budgets. Since every interval lies above zero, the results support an increase in the subcube's relative value between the budget endpoints.

Since the first batch is in region one, we also examine risk by affine region identity. Let $\Delta_i^{\mathrm{real}}(b) = R_i^{\mathrm{check}}(b) - R_i^{\mathrm{subcube}}(b)$ be the realized-Brier contrast, where $R_i^{\mathsf{a}}(b)$ is common-target squared error for design a on array i. Write $\mathcal{I}_j = \{i : J_{\star,i} = j\}$ for the arrays whose affine region is j. Then

$$\overline{\Delta}_j^{\mathrm{real}}(b) = \frac{1}{|\mathcal{I}_j|} \sum_{i \in \mathcal{I}_j} \Delta_i^{\mathrm{real}}(b), \quad \overline{\Delta}^{\mathrm{real}}(b) = \sum_{j=1}^{2} \frac{|\mathcal{I}_j|}{n} \overline{\Delta}_j^{\mathrm{real}}(b). \tag{179}$$

The stratum mean $\overline{\Delta}_j^{\mathrm{real}}(b)$ includes every sample with affine region j, covering both correct and incorrect selection. Weighting by the observed stratum proportions recovers the overall contrast.

Table 17 connects this interaction to the affine region's location. At shorter budgets, two checks improve prediction primarily when region one is affine, where the first batch can give additional

d	η	b	ΔD_0	ΔD_{param}	ΔD_{ind}	ΔR_{cond}
8	0.010	13	0.000019	−0.015315	0.002133	-0.013163 ± 0.003233
8	0.010	15	0.000023	−0.003038	0.000861	-0.002153 ± 0.002714
8	0.010	17	0.000018	0.003349	0.000021	0.003387 ± 0.002814
8	0.025	13	−0.000006	−0.013532	0.002616	-0.010922 ± 0.003099
8	0.025	15	−0.000006	−0.004416	0.002117	-0.002305 ± 0.003077
8	0.025	17	0.000001	0.001298	0.001292	0.002591 ± 0.002962
8	0.050	13	−0.000002	−0.008184	0.002672	-0.005514 ± 0.002454
8	0.050	15	−0.000004	−0.004401	0.002194	-0.002210 ± 0.002911
8	0.050	17	−0.000012	−0.000330	0.001547	0.001205 ± 0.003103
12	0.010	19	0.000000	−0.007069	0.001312	-0.005757 ± 0.002900
12	0.010	22	0.000000	0.004968	0.000321	0.005289 ± 0.003606
12	0.010	24	0.000000	0.006799	0.000020	0.006819 ± 0.004327
12	0.025	19	0.000000	−0.009118	0.002410	-0.006708 ± 0.002742
12	0.025	22	0.000000	0.001463	0.000703	0.002166 ± 0.003410
12	0.025	24	0.000000	0.007807	0.000472	0.008279 ± 0.004122
12	0.050	19	0.000000	−0.006802	0.002058	-0.004743 ± 0.002642
12	0.050	22	0.000000	−0.002794	0.001522	-0.001271 ± 0.003330
12	0.050	24	0.000000	−0.000424	0.000852	0.000428 ± 0.003942

Table 14: Conditional-risk decomposition in the 512-sample study. Differences are two checks minus subcube. Component means sum to the final contrast before rounding, and the last column gives pointwise 95% paired Student half-widths.

d	η	b	$\widehat{A}_{\text{prefix}}$	$\widehat{A}_{\text{check}}$	$\widehat{A}_{\text{subcube}}$	$\Delta R_{\text{realized}}$
8	0.010	13	0.6289	0.8633	0.9395	-0.015685 ± 0.004183
8	0.010	15	0.6289	0.8633	0.9395	-0.001739 ± 0.003446
8	0.010	17	0.6289	0.8633	0.9395	0.006706 ± 0.005370
8	0.025	13	0.5840	0.7988	0.8945	-0.010869 ± 0.003931
8	0.025	15	0.5840	0.7988	0.8945	-0.002304 ± 0.003923
8	0.025	17	0.5840	0.7988	0.8945	0.003317 ± 0.005742
8	0.050	13	0.5723	0.7109	0.8184	-0.007756 ± 0.003440
8	0.050	15	0.5723	0.7109	0.8184	-0.001864 ± 0.003987
8	0.050	17	0.5723	0.7109	0.8184	0.000984 ± 0.005223
12	0.010	19	0.5000	0.8320	0.9082	-0.003938 ± 0.003960
12	0.010	22	0.5000	0.8320	0.9082	0.003984 ± 0.005938
12	0.010	24	0.5000	0.8320	0.9082	0.006834 ± 0.005567
12	0.025	19	0.5020	0.7773	0.8848	-0.004177 ± 0.003743
12	0.025	22	0.5020	0.7773	0.8848	0.005979 ± 0.005812
12	0.025	24	0.5020	0.7773	0.8848	0.008020 ± 0.006134
12	0.050	19	0.5000	0.7246	0.7910	-0.004521 ± 0.003736
12	0.050	22	0.5000	0.7246	0.7910	-0.003657 ± 0.005124
12	0.050	24	0.5000	0.7246	0.7910	-0.000934 ± 0.005733

Table 15: Realized affine-selection fractions and Brier in the same 512-array study as Table 14. The last column is two checks minus subcube with a pointwise 95% paired Student half-width. The selection fractions share their first batch across budgets. Independent prefixes use the region-one tie convention.

independent parameter constraints. In region two, parameter information comes from the continuation. As that continuation grows, the subcube's more frequent correct selection improves prediction in this stratum, especially at lower noise.

M Finite Testing and Prediction

At an unqueried target, regional evidence and parameter information affect different components of posterior variance. The decomposition is

$$\operatorname{Var}(Y_{j,x} \mid \mathcal{H}) = \frac{1}{4} - \frac{\gamma_j \delta^2}{4} + \frac{\gamma_j \delta^2}{4}(1 - M_{j,x}^2) + \frac{\gamma_j(1-\gamma_j)\delta^2}{4} M_{j,x}^2. \tag{180}$$

d	η	b_-	b_+	Paired budget interaction
8	0.010	13	17	0.016550 ± 0.005147
8	0.025	13	17	0.013513 ± 0.004324
8	0.050	13	17	0.006719 ± 0.003325
12	0.010	19	24	0.012577 ± 0.005898
12	0.025	19	24	0.014987 ± 0.005341
12	0.050	19	24	0.005171 ± 0.004264

Table 16: Changes in two-check minus subcube conditional Brier between the smallest and largest tested budgets. Each estimate uses paired differences across both designs and both budgets on the same 512 arrays. Intervals use a common Bonferroni correction across all six comparisons. Positive values indicate an increasing relative advantage for the subcube.

d	η	b	$\lvert\mathcal{I}_1\rvert$	$\overline{\Delta}_1^{\mathrm{real}}(b)$	$\lvert\mathcal{I}_2\rvert$	$\overline{\Delta}_2^{\mathrm{real}}(b)$
8	0.010	13	281	-0.030179 ± 0.006511	231	0.001945 ± 0.003769
8	0.010	15	281	-0.006330 ± 0.005361	231	0.003845 ± 0.003902
8	0.010	17	281	0.003917 ± 0.006445	231	0.010098 ± 0.008992
8	0.025	13	264	-0.026105 ± 0.006452	248	0.005351 ± 0.003333
8	0.025	15	264	-0.009817 ± 0.006800	248	0.005694 ± 0.003416
8	0.025	17	264	-0.001565 ± 0.008258	248	0.008515 ± 0.007962
8	0.050	13	253	-0.019571 ± 0.006257	259	0.003784 ± 0.002271
8	0.050	15	253	-0.008307 ± 0.007172	259	0.004429 ± 0.003499
8	0.050	17	253	-0.005208 ± 0.008483	259	0.007032 ± 0.006128
12	0.010	19	254	-0.014140 ± 0.006523	258	0.006106 ± 0.004228
12	0.010	22	254	-0.007356 ± 0.007749	258	0.015149 ± 0.008819
12	0.010	24	254	-0.005070 ± 0.007089	258	0.018554 ± 0.008365
12	0.025	19	254	-0.015791 ± 0.006460	258	0.007256 ± 0.003335
12	0.025	22	254	-0.006374 ± 0.008159	258	0.018141 ± 0.008056
12	0.025	24	254	-0.005633 ± 0.008767	258	0.021462 ± 0.008317
12	0.050	19	256	-0.011840 ± 0.006969	256	0.002798 ± 0.002452
12	0.050	22	256	-0.010188 ± 0.008644	256	0.002875 ± 0.005451
12	0.050	24	256	-0.008228 ± 0.009300	256	0.006360 ± 0.006651

Table 17: Two-check minus subcube realized Brier stratified by the true affine region. The first batch is in region one. Each cell reports a stratum mean and pointwise 95% paired Student interval half-width. Stratum counts are shared across budgets. Positive contrasts favor the subcube.

Public information is implicit in the conditioning. The first two terms give residual variance with the region and parameter known, and the last two give parameter and indicator uncertainty. An affine target has conditional mean $1/2-\delta_j M_{j,x}/2$ and variance $1/4-\delta_j^2 M_{j,x}^2/4$. An independent target has mean $1/2$ and variance $1/4$. The variance between these conditional means is $\gamma_j(1-\gamma_j)\delta_j^2 M_{j,x}^2/4$, so Equation 180 follows. Conditional independence of unqueried target errors also applies along adaptive histories.

For an exact comparison, take $d = 4$ and use either $\mathcal{P}^{(4)} = (0, 1, 2, 3, 4, 8)$ or $\mathcal{P}^{(6)} = (0, 1, 2, 4, 8, 15)$ in both regions. Each pool has six coordinates, augmented rank five, and one dependency of the indicated length. Both are evaluated on the same regional targets $\mathcal{V}_\cap = \mathbb{F}_2^4 \setminus (\mathcal{P}^{(4)} \cup \mathcal{P}^{(6)}) = \{5, 6, 7, 9, 10, 11, 12, 13, 14\}$. The length-four check is more informative about the regional model in Blackwell's sense (Blackwell, 1953), since multiplying its sign by an independent sign of mean δ^2 produces the length-six law. For prediction, the length-six pool represents common targets with fewer noisy labels. The resulting comparison distinguishes information about the region from information useful for predicting its targets.

Proposition 4 (Confidence ordering for the constructed pair)**.** *Fix the regional prior $p \in [0, 1]$, noise $0 < \eta < 1/2$, and acquire all twelve labels. For a uniform common target $X \in \{1, 2\} \times \mathcal{V}_\cap$, let $C_k = |2\mathbb{P}(Y_X = 1 \mid \mathcal{H}, X) - 1|$ under $\mathcal{P}^{(k)}$. Then*

$$C_4 \preceq_{\mathrm{cx}} C_6, \qquad \mathbb{E}C_4 = \mathbb{E}C_6 = \delta^4/2. \tag{181}$$

The convex order $\preceq_{\mathrm{cx}}$ compares expectations of every convex function. Every continuous symmetric concave Bayes entropy has weakly lower risk under the length-six pool, strictly lower for strictly concave entropy. Both classification errors equal $1/2 - \delta^4/4$.

With equal regional prior, write $R_{\cap}^{(k)}(12, \eta)$ for common-target Brier and set $z = \delta^2$. The exact difference is

$$R_{\cap}^{(4)}(12,\eta) - R_{\cap}^{(6)}(12,\eta) = \frac{z^4(1-z)\Pi_{\cap}(z)}{96(1+z^2)(1+z^3)} > 0, \qquad \Pi_{\cap}(z) = z^6 + z^5 + z^4 + 4z^3 + 6z^2 + 2z + 2. \tag{182}$$

The coefficients of $\Pi_{\cap}$ are positive, so the sign follows. We derive the comparison in Appendix N and embed the construction in larger affine dimensions in Appendix N.8.

N Exact Noisy Acquisition

N.1 Matched-Size and Matched-Rank Pools

We optimize all acquisition classes for two regions with $d = 4$, $m = 16$, $r = 5$, and a uniform exclusive prior. One region is affine, and the other has independent fair labels. The regional pools are $\mathcal{P}^{(4)} = (0, 1, 2, 3, 4, 8), \quad \mathcal{P}^{(6)} = (0, 1, 2, 4, 8, 15)$. Integers encode coordinates in $\mathbb{F}_2^4$, and each construction uses the displayed pool in both regions. Both pools have six coordinates, augmented rank five, and a one-dimensional dependency space. In $\mathcal{P}^{(4)}$, the length-four dependency uses the first four coordinates. In $\mathcal{P}^{(6)}$, the dependency uses all six. The target distribution and public regional information are identical in both constructions.

We integrate posterior risk for $0 \le b \le 12$ at noise rates 0, .05, .10, and .20, covering 104 pool, noise, and budget configurations. Exhaustive optimization includes the full fixed, two-batch, and adaptive classes.

Two batches achieve the adaptive optimum for the short-check pool throughout the grid, and all three classes agree for the length-six pool. At complete acquisition, target representations determine the remaining parameter precision.

Let $R_{\mathrm{unq}}^{(k)}(12, \eta)$ denote expected Brier over the unqueried population after complete acquisition, with k indicating dependency length. For $z = (1-2\eta)^2 \in (0, 1)$, we obtain

$$R_{\mathrm{unq}}^{(4)}(12,\eta) - R_{\mathrm{unq}}^{(6)}(12,\eta) = \frac{z^4(1-z)\Pi(z)}{80(1+z^2)(1+z^3)} > 0, \qquad \Pi(z) = z^6 + z^5 + z^4 + 4z^3 + 5z^2 + 2z + 2. \tag{183}$$

The positive coefficients of Π prove the ordering. Proposition 5 separates its parameter and indicator contributions. Tables 18 to 21 give the risks; we derive the comparison on common targets in Appendix N.2.

Proposition 5 (Prediction ordering for the constructed pool pair)**.** *For the same pools, public prior $1/2$, and $0 < \eta < 1/2$, complete acquisition also gives*

$$R_{\mathrm{unq}}^{(6)}(12,\eta) < R_{\mathrm{unq}}^{(4)}(12,\eta). \tag{184}$$

The score $R_{\mathrm{unq}}^{(k)}$ averages Brier over targets outside pool $\mathcal{P}^{(k)}$. The length-six pool has a smaller expected parameter component and a larger prediction-weighted indicator component in Equation 180, with a net risk reduction. Complete acquisition gives the same value in every acquisition class.

Proof of Proposition 5. Suppress the region index and write $\delta = 1 - 2\eta$ and $z = \delta^2$. The dependency space of a six-point pool is generated by its unique nonzero vector v of weight k. For labels $y \in \mathbb{F}_2^6$, let $s = (-1)^{v^\top y}$ be the observed parity sign and $u_k = \delta^k$ the check signal. The affine likelihood is then $1 + u_k s$.

Every unqueried coordinate x has two representations by queried augmented rows. If w indicates one representation, the other is $w + v$, with addition in the binary field. We can therefore write Equation

5 as

$$M_x(y) = \frac{\delta^{|w|}(-1)^{w^\top y} + \delta^{|w+v|}(-1)^{(w+v)^\top y}}{1 + u_k s}. \tag{185}$$

Each representation length determines its noise-attenuation power. Among the ten unqueried coordinates, the length-four pool has six representation pairs of weights $(3, 3)$ and four of weights $(3, 5)$. The length-six pool has ten $(3, 3)$ weight pairs. Squaring the sign means and summing yields

$$\begin{aligned} \mathcal{U}_k(s) &:= \sum_{x \notin \mathcal{P}^{(k)}} M_x(y)^2 = \frac{A_k + B_k s}{(1 + u_k s)^2}, \\ A_4 &= 16\delta^6 + 4\delta^{10}, \quad B_4 = 12\delta^6 + 8\delta^8, \\ A_6 &= B_6 = 20\delta^6, \end{aligned} \tag{186}$$

where $\mathcal{U}_k(s)$ sums squared conditional affine-sign means over the unqueried coordinates. The coefficients A_k and B_k collect the squared terms and parity-dependent cross terms, respectively.

Let s_1, s_2 denote the regional check signs. Under the all-independent reference law, they are independent fair signs. The exclusive regional mixture has density $1 + u_k(s_1 + s_2)/2$, and its posterior probabilities are $\gamma_j = (1 + u_k s_j)/(2 + u_k(s_1 + s_2))$. Averaging the four sign pairs gives

$$\begin{aligned} F_k &:= \mathbb{E} \sum_{j=1}^{2} \gamma_j \mathcal{U}_k(s_j) = \frac{A_k - B_k u_k}{1 - u_k^2}, \\ G_k &:= \mathbb{E} \sum_{j=1}^{2} \gamma_j^2 \mathcal{U}_k(s_j) = \frac{A_k + F_k}{4}, \end{aligned} \tag{187}$$

where F_k is the expected energy with the indicator known and G_k weights energy by squared posterior model probability. Applying Equation 180 to the twenty unqueried targets gives

$$\begin{aligned} R^{(k)}_{\mathrm{unq}}(12, \eta) &= R_{\mathrm{irr}} + D^{(k)}_{\mathrm{param}} + D^{(k)}_{\mathrm{ind}} = \frac{1}{4} - \frac{\delta^2 G_k}{80}, \\ R_{\mathrm{irr}} &= \frac{1}{4} - \frac{\delta^2}{8}, \quad D^{(k)}_{\mathrm{param}} = \frac{\delta^2(10 - F_k)}{80}, \quad D^{(k)}_{\mathrm{ind}} = \frac{\delta^2(F_k - G_k)}{80}. \end{aligned} \tag{188}$$

The residual R_{irr} is the risk with known indicator and parameter. The terms $D^{(k)}_{\mathrm{param}}$ and $D^{(k)}_{\mathrm{ind}}$ are expected parameter and indicator uncertainty, whose differences simplify to

$$\begin{aligned} D^{(4)}_{\mathrm{param}} - D^{(6)}_{\mathrm{param}} &= \frac{z^4(1 - z)(2z^3 + 4z^2 + z + 1)}{20(1 + z^2)(1 + z^3)} > 0, \\ D^{(6)}_{\mathrm{ind}} - D^{(4)}_{\mathrm{ind}} &= \frac{z^4(1 - z)\Pi_{\mathrm{ind}}(z)}{80(1 + z^2)(1 + z^3)} > 0, \\ \Pi_{\mathrm{ind}}(z) &= -z^6 - z^5 - z^4 + 4z^3 + 11z^2 + 2z + 2. \end{aligned} \tag{189}$$

For $0 < z < 1$, the inequality $z^4 + z^5 + z^6 \le 3z^3$ bounds Π_{ind} below by $z^3 + 11z^2 + 2z + 2 > 0$. Subtracting the increase in indicator uncertainty from the reduction in parameter uncertainty proves Equation 183. Both risks approach the same zero-noise limit, and complete acquisition fixes their values across acquisition classes. □

The ordering also holds for full-population loss, which multiplies both risks by $5/8$.

N.2 Common-Target Information Orderings

Use the nine common targets in each region, that is, $\{1, 2\} \times \mathcal{V}_\cap$ from Appendix M. Write $R^{(k)}_\cap(12, \eta)$ for expected Brier on these targets after complete acquisition from the length-k pool in both regions.

Corollary 1 (Model-evidence ordering and common-target prediction)**.** *For every $0 < \eta < 1/2$, the length-four observations can be randomized into the length-six observation distribution by a channel independent of the indicator vector. Their optimal equal-prior model-testing error is strictly smaller. For prediction on common targets, the risks obey Equation 182. The length-six pool has a smaller parameter component and a larger prediction-weighted model-uncertainty component, with a net reduction in prediction risk.*

Proof. The parity $S_k \in \{-1, 1\}$ is sufficient for the regional indicator because the likelihood is $1 + \delta^k S_k$. It is fair under independence and has mean δ^k under the affine model. Take an independent sign $E \in \{-1, 1\}$ with mean δ^2, so $\mathbb{P}(E = -1) = (1 - \delta^2)/2$. Then $S_4 E$ follows the length-six parity law under either model. Sampling a six-bit assignment uniformly conditional on this sign reconstructs the full length-six observation law. Apply this channel independently in both regions to compare the exclusive indicator states. The equal-prior testing error $1/2 - \delta^k/4$ is strictly smaller for length four.

For prediction, remove from each unqueried set the coordinate observed only by the other pool. The remaining common targets have six representation pairs of weights $(3, 3)$ and three of weights $(3, 5)$ under the length-four pool. Under the length-six pool, all nine pairs have weights $(3, 3)$. Their energy coefficients are

$$\begin{aligned} A_{4,\cap} &= 15\delta^6 + 3\delta^{10}, \quad B_{4,\cap} = 12\delta^6 + 6\delta^8, \\ A_{6,\cap} &= B_{6,\cap} = 18\delta^6, \end{aligned} \tag{190}$$

where $A_{k,\cap}$ and $B_{k,\cap}$ collect squared and parity-dependent energy on $\mathcal{V}_\cap$. Averaging the regional parity signs as in Proposition 5 gives

$$\begin{aligned} F_{k,\cap} &= \frac{A_{k,\cap} - B_{k,\cap}\delta^k}{1 - \delta^{2k}}, \quad G_{k,\cap} = \frac{A_{k,\cap} + F_{k,\cap}}{4}, \\ R_\cap^{(k)}(12, \eta) &= \frac{1}{4} - \frac{\delta^2 G_{k,\cap}}{72}. \end{aligned} \tag{191}$$

The sums $F_{k,\cap}$ and $G_{k,\cap}$ are the indicator-informed and marginal energies on $\mathcal{V}_\cap$. Substitution yields the Brier gap and the component differences

$$\begin{aligned} D^{(4)}_{\mathrm{param},\cap} - D^{(6)}_{\mathrm{param},\cap} &= \frac{z^4(1-z)(2z^3 + 5z^2 + z + 1)}{24(1 + z^2)(1 + z^3)} > 0, \\ D^{(6)}_{\mathrm{ind},\cap} - D^{(4)}_{\mathrm{ind},\cap} &= \frac{z^4(1-z)\Pi_{\mathrm{ind},\cap}(z)}{96(1 + z^2)(1 + z^3)} > 0, \\ \Pi_{\mathrm{ind},\cap}(z) &= -z^6 - z^5 - z^4 + 4z^3 + 14z^2 + 2z + 2. \end{aligned} \tag{192}$$

The expected components are $D^{(k)}_{\mathrm{param},\cap} = \delta^2(9 - F_{k,\cap})/72$ and $D^{(k)}_{\mathrm{ind},\cap} = \delta^2(F_{k,\cap} - G_{k,\cap})/72$. The inequality $z^4 + z^5 + z^6 \le 3z^3$ bounds the final polynomial below by $z^3 + 14z^2 + 2z + 2 > 0$, and subtracting the component differences gives the positive common-target risk gap. □

The opposite orderings for these two decisions imply Blackwell incomparability when the state includes both regional indicators and common target labels.

N.3 Public-Prior Dependence

Keep the pools, twelve queries, and common targets fixed, and let $p = \mathbb{P}(Z_1 = 1 \mid W)$ be the public prior. Proposition 4 establishes the strict Brier ordering for every $p \in [0, 1]$ and $0 < \eta < 1/2$.

Set $v = (2p - 1)^2$ for squared prior imbalance and $u_k = \delta^k$ for the check signal. Using $A_{k,\cap}$, $B_{k,\cap}$, and $F_{k,\cap}$ from Appendix N.2, the parity pair (s_1, s_2) has density $1 + u_k[ps_1 + (1 - p)s_2]$ relative to independent fair signs. Summing the four outcomes gives

$$\begin{aligned} G_{k,\cap}(p) &= \frac{1 + v}{4} F_{k,\cap} + \frac{(1 + v)A_{k,\cap} - 2vB_{k,\cap}u_k}{4(1 - vu_k^2)}, \\ R_\cap^{(k)}(12, \eta, p) &= \frac{1}{4} - \frac{\delta^2 G_{k,\cap}(p)}{72}. \end{aligned} \tag{193}$$

The unnormalized energy $G_{k,\cap}(p)$ weights squared affine-sign means on common targets by squared posterior model probability. At $p = 1/2$, it reduces to $G_{k,\cap}$. At $p \in \{0, 1\}$, the indicators are known, and the expression equals $F_{k,\cap}$, so strict ordering at the endpoints isolates the effect of target representations. By continuity, risk is $1/8 + p(1 - p)/8$ at zero noise and $1/4$ at noise $1/2$.

Dependence on $(2p - 1)^2$ makes the expression symmetric in p. We extend the ordering to symmetric prediction losses in Appendix N.7.

N.4 Noisy Acquisition Recursions

A regional history has six entries, each recording an unobserved label, zero, or one. This gives 3^6 regional histories and 3^{12} joint histories. Summing over the 32 affine parameters yields regional likelihoods and posterior signs, which Equation 4 combines. Query-selection factors cancel, so the partially labeled set is sufficient for subsequent decisions.

For partial history H, let P_H be its probability under the target mixture and $\mathcal{L}(H)$ its full-population posterior Brier loss. Define the history-weighted terminal cost $\mathcal{B}(H) = P_H \mathcal{L}(H)$, assigning zero to zero-probability histories. At total budget b, the best fixed continuation has value

$$F_b(H) = \min_{C \subseteq \mathcal{I}_\mathcal{P} \setminus Q(H),\ |C| = b - |Q(H)|} \sum_{y_C} \mathcal{B}\big(H \cup (C, y_C)\big), \tag{194}$$

where $\mathcal{I}_\mathcal{P}$ is the available regional query union from Section 3, $Q(H)$ records acquired identities, and C is the fixed continuation. Each terminal cost already includes its history probability. Marginalizing over continuation outcomes yields

$$R^*_{\text{fix}}(b) = F_b(\varnothing), \quad R^*_2(b) = \min_{S \subseteq \mathcal{I}_\mathcal{P},\ |S| \le b} \sum_{y_S} F_b\big((S, y_S)\big). \tag{195}$$

Choose the first set S before outcomes, and use the best fixed continuation for each resulting history. Either batch may be empty.

For adaptive acquisition, initialize $A_b(H) = \mathcal{B}(H)$ whenever $|Q(H)| = b$ and recurse backward using

$$A_b(H) = \min_{v \in \mathcal{I}_\mathcal{P} \setminus Q(H)} \left[A_b\big(H \cup (v, 0)\big) + A_b\big(H \cup (v, 1)\big) \right]. \tag{196}$$

The initial value $A_b(\varnothing)$ is the optimal adaptive population risk. Enumeration covers every available action and continuation. Multiplying by $32/(32 - b)$ gives unqueried risk.

N.5 Common-Target Acquisition Results

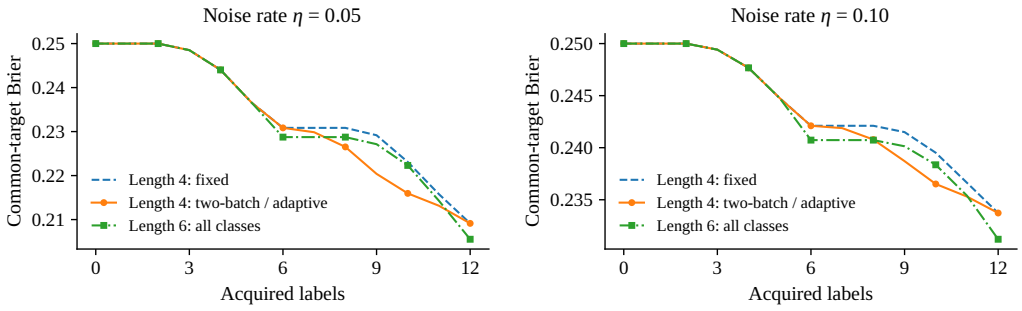


Figure 7: Exact acquisition-class optima for two six-point query pools of equal augmented rank on the same eighteen targets. Their unique dependencies have length four and six.

For $0 \le b \le 12$, common-target optimization substitutes the fixed-population terminal loss

$$\mathcal{L}_\cap(H) = \frac{1}{4} - \frac{\delta^2}{8|\mathcal{V}_\cap|} \sum_{j=1}^{2} \gamma_j(H)^2 \sum_{x \in \mathcal{V}_\cap} M_{j,x}(H)^2. \tag{197}$$

We then use $\mathcal{B}_\cap(H) = P_H \mathcal{L}_\cap(H)$ in place of $\mathcal{B}$ in each acquisition recursion.

Figure 7 shows the common-target optima at noise rates .05 and .10; the full grid is in Tables 18 to 21. Two batches attain the adaptive optimum. The short-check pool has lower risk at several intermediate budgets. Complete acquisition at positive noise favors the long-check pool, as established by Equation 182. Expected risk decreases or stays constant as labels are added on the fixed target population.

	Common targets			Unqueried targets		
b	Length 4 fixed	Length 4 two-batch/adaptive	Length 6 all classes	Length 4 fixed	Length 4 two-batch/adaptive	Length 6 all classes
0	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
1	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
2	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
3	0.246528	0.246528	0.246528	0.247845	0.247845	0.247845
4	0.236111	0.236111	0.236111	0.241071	0.241071	0.241071
5	0.218750	0.218750	0.218750	0.224537	0.224537	0.224537
6	0.208333	0.208333	0.208333	0.217949	0.217949	0.217949
7	0.208333	0.204861	0.208333	0.216667	0.214167	0.216667
8	0.208333	0.194444	0.208333	0.215278	0.204861	0.215278
9	0.203704	0.175926	0.203704	0.209239	0.182971	0.209239
10	0.187500	0.171875	0.187500	0.187500	0.176136	0.187500
11	0.166667	0.166667	0.166667	0.166667	0.166667	0.166667
12	0.156250	0.156250	0.156250	0.156250	0.156250	0.156250

Table 18: Exact optimal Brier at zero noise. Common-target risk averages the same eighteen targets across pools and budgets. Unqueried risk averages each history's $32-b$ remaining targets. Coincident classes share a column.

	Common targets			Unqueried targets		
b	Length 4 fixed	Length 4 two-batch/adaptive	Length 6 all classes	Length 4 fixed	Length 4 two-batch/adaptive	Length 6 all classes
0	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
1	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
2	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
3	0.248505	0.248505	0.248505	0.249072	0.249072	0.249072
4	0.244021	0.244021	0.244021	0.246157	0.246157	0.246157
5	0.236548	0.236548	0.236548	0.239382	0.239382	0.239382
6	0.230843	0.230843	0.228744	0.235433	0.235433	0.233649
7	0.230843	0.229851	0.228744	0.234851	0.234166	0.232995
8	0.230843	0.226515	0.228744	0.234220	0.231003	0.232287
9	0.229168	0.220391	0.227136	0.232223	0.223944	0.230258
10	0.223096	0.215954	0.222311	0.223937	0.219602	0.223937
11	0.215769	0.213141	0.214270	0.216667	0.214944	0.215067
12	0.209137	0.209137	0.205528	0.209594	0.209594	0.205528

Table 19: Exact optimal Brier at $\eta = .05$. Target populations and shared acquisition columns follow Table 18.

N.6 Dependency-Support Validation

We enumerate every binary linear code of lengths one through seven via reduced row-echelon generator matrices. For a nonzero code, the largest feasible packing count is

$$\kappa_{\max} = \min_{S:\dim \mathcal{C}_S > 0} \left\lfloor \frac{|S|}{\dim \mathcal{C}_S} \right\rfloor . \tag{198}$$

We compute $\kappa_{\max}$ from the shortened-subcode dimensions and independently find the maximum number of disjoint information sets by subset dynamic programming. Their agreement validates the packing criterion in Lemma 4.

For rational attenuation $\delta = p/q$ in $\{0, 1/4, 1/2, 3/4, 1\}$, let a_w count codewords of weight w. With $k_{\mathcal{C}} = \dim \mathcal{C}$ denoting code dimension, we verify Equation 73 through the integer comparison

$$\sum_{w=0}^{n} a_w p^w q^{n-w} \le (q^{\kappa_{\max}} + p^{\kappa_{\max}})^{k_{\mathcal{C}}} q^{n-\kappa_{\max} k_{\mathcal{C}}} . \tag{199}$$

The support condition implies $\kappa_{\max} k_{\mathcal{C}} \le n$, so the exponents on the right are nonnegative. Zero-dimensional codes satisfy equality.

	Common targets			Unqueried targets		
b	Length 4 fixed	Length 4 two-batch/adaptive	Length 6 all classes	Length 4 fixed	Length 4 two-batch/adaptive	Length 6 all classes
0	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
1	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
2	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
3	0.249417	0.249417	0.249417	0.249638	0.249638	0.249638
4	0.247670	0.247670	0.247670	0.248502	0.248502	0.248502
5	0.244757	0.244757	0.244757	0.245957	0.245957	0.245957
6	0.242105	0.242105	0.240729	0.244051	0.244051	0.242869
7	0.242105	0.241881	0.240729	0.243813	0.243686	0.242583
8	0.242105	0.240782	0.240729	0.243555	0.242599	0.242274
9	0.241497	0.238709	0.240137	0.242799	0.240250	0.241475
10	0.239514	0.236514	0.238358	0.240077	0.238087	0.239632
11	0.236632	0.235305	0.235395	0.237209	0.236192	0.235882
12	0.233728	0.233728	0.231206	0.234058	0.234058	0.231206

Table 20: Exact optimal Brier at $\eta = .10$. Target populations and shared acquisition columns follow Table 18.

	Common targets			Unqueried targets		
b	Length 4 fixed	Length 4 two-batch/adaptive	Length 6 all classes	Length 4 fixed	Length 4 two-batch/adaptive	Length 6 all classes
0	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
1	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
2	0.250000	0.250000	0.250000	0.250000	0.250000	0.250000
3	0.249942	0.249942	0.249942	0.249964	0.249964	0.249964
4	0.249767	0.249767	0.249767	0.249850	0.249850	0.249850
5	0.249475	0.249475	0.249475	0.249606	0.249606	0.249606
6	0.249152	0.249152	0.248974	0.249369	0.249369	0.249211
7	0.249152	0.249152	0.248974	0.249344	0.249344	0.249179
8	0.249152	0.249066	0.248974	0.249317	0.249262	0.249145
9	0.249094	0.248883	0.248916	0.249241	0.249058	0.249062
10	0.248918	0.248630	0.248741	0.249033	0.248810	0.248876
11	0.248625	0.248493	0.248449	0.248711	0.248609	0.248516
12	0.248299	0.248299	0.247947	0.248354	0.248354	0.247947

Table 21: Exact optimal Brier at $\eta = .20$, with target sets and acquisition columns as in Table 18.

N.7 Proof of the Confidence Ordering

Acquire every available label from the two six-point pools and evaluate their eighteen common targets. The proof first compares target representations at a fixed check signal, then compares check signals for a fixed representation type. Both steps allow us to use every public regional prior.

Proof of Proposition 4. Write $\delta = 1 - 2\eta$, $z = \delta^2$, and $a = \delta^4$. Set $p_1 = p$ and $p_2 = 1 - p$, the two regional model probabilities. For check signal $u \in (0, 1)$, the regional parity signs $s_1, s_2 \in \{-1, 1\}$ have probability $D_u(s_1, s_2)/4$, where $D_u(s_1, s_2) = 1 + u[ps_1 + (1 - p)s_2]$. This positive density is relative to independent fair signs. Equation 5 gives confidence $p_j a|1 + s_j|/D_u$ for a type-$(3, 3)$ target. A type-$(3, 5)$ target in the short-check pool instead has confidence $p_j a|1 + zs_j|/D_{z^2}$.

Let $A_u(p)$ denote the first confidence distribution with a uniformly selected target region, and let $B_z(p)$ denote the second. In each region, six of the short pool's nine common targets have type $(3, 3)$ and three have type $(3, 5)$. All nine long-pool targets have type $(3, 3)$. Therefore,

$$\mathcal{L}(C_4) = \tfrac{2}{3}A_{z^2}(p) + \tfrac{1}{3}B_z(p), \quad \mathcal{L}(C_6) = A_{z^3}(p). \tag{200}$$

Comparing the two target representation types. Condition on target region j and the other region's sign t. Put $w = p_j$, $u = z^2$, and $d_0 = 1 + u(1 - w)t$. Conditional probabilities of $s_j = s$

are $(d_0 + uws)/(2d_0)$. For type $(3, 3)$, the confidence takes the endpoints zero and

$$M = \frac{2wa}{d_0 + uw}. \tag{201}$$

For type $(3, 5)$, its two values are

$$\frac{wa(1+z)}{d_0 + uw}, \quad \frac{wa(1-z)}{d_0 - uw}. \tag{202}$$

Both conditional distributions have mean wa/d_0. When $w = 0$, each is concentrated at zero. For $w > 0$, the first type-$(3, 5)$ value is strictly inside $[0, M]$. The second is also strictly inside, because $(1+z)d_0 - (3-z)uw \geq (1-z)\big[(1+z)^2 - 2z^2 w\big] > 0$. For a fixed mean, a distribution supported at the endpoints of $[0, M]$ maximizes every convex expectation. Thus, conditioning and averaging establish $B_z(p) \preceq_{\mathrm{cx}} A_{z^2}(p)$. Strict convexity makes the inequality strict because at least one region has positive model probability. Averaging over the other sign t, whose marginal mass is $d_0/2$, gives regional mean wa and overall mean $a/2$ under equal target weights.

Comparing the two check signals. For a threshold $t \geq 0$, define $\mathcal{C}_u(t) = \mathbb{E}(A_u(p) - t)_+$, taking the expectation under confidence distribution $A_u(p)$. By symmetry, assume $p \geq 1/2$ and put $q = 1 - p$ and $v = 2p - 1$. Summing over the parity pair yields

$$\begin{aligned} 8\mathcal{C}_u(t) = &\big(2pa - t(1+u)\big)_+ + \big(2qa - t(1+u)\big)_+ \\ &+ \big(2pa - t(1+uv)\big)_+ + \big(2qa - t(1-uv)\big)_+. \end{aligned} \tag{203}$$

The first two summands are nonincreasing in u. The final pair has the same property because

$$\frac{2pa}{1+uv} \geq \frac{2qa}{1-uv}, \tag{204}$$

which follows from $p - q = v \geq uv$. Whenever the final summand is positive, the preceding one is also positive, and their slopes in u cancel. With only the preceding summand positive, the slope is nonpositive. Continuity covers the breakpoints. Thus, Equation 203 is nonincreasing in u. For $t < 0$, all these expectations equal $a/2 - t$. The equal means and positive-part inequalities establish $A_{z^2}(p) \preceq_{\mathrm{cx}} A_{z^3}(p)$. Combining the signal comparison with the strict representation-type comparison in Equation 200 proves the ordering for each p.

For a continuous symmetric concave Bayes entropy H_ℓ, the function $c \mapsto -H_\ell((1+c)/2)$ is convex. Expected Bayes risk is $\mathbb{E}H_\ell((1 + C_k)/2)$, so the confidence order implies weakly lower risk under the length-six pool, and strictly lower risk for strictly concave entropy. Classification depends only on mean confidence. Since both posterior predictors are calibrated under the model,

$$R^{(4)}_{01,\cap}(12, \eta, p) = R^{(6)}_{01,\cap}(12, \eta, p) = \frac{1}{2} - \frac{(1-2\eta)^4}{4}. \tag{205}$$

Canceling D_u in the expected absolute posterior target sign produces the same classification formula. Continuity extends the result to the noise endpoints. □

At $p = 1/2$, the type-$(3, 3)$ confidence distribution has three atoms:

confidence	0	$a/(1+u)$	a
probability	$(2-u)/4$	$(1+u)/4$	$1/4.$

(206)

As u decreases, the middle positive atom rises and the mean stays fixed, so the confidence distribution becomes more dispersed. Table 22 gives the resulting Brier, log, and classification risks, and the conditional argument establishes the ordering for unequal priors.

N.8 Higher-Dimensional Construction

We extend the pair to a larger affine parameter by adding independent observable constraints. Let $d_{\mathrm{ext}} \geq 0$ be the number of added coordinates, with $d = 4 + d_{\mathrm{ext}}$ and $r = 5 + d_{\mathrm{ext}}$. Embed each base point as $(x, 0_{d_{\mathrm{ext}}})$ and add the standard basis vectors $e_5, \ldots, e_d$ in the additional directions: $\mathcal{P}^{(k)}_{d_{\mathrm{ext}}} = (\mathcal{P}^{(k)} \times \{0_{d_{\mathrm{ext}}}\}) \cup \{e_5, \ldots, e_d\}, \quad k \in \{4, 6\}$, where $0_{d_{\mathrm{ext}}}$ is the zero vector in the added

	Brier		Log loss (bits)		Classification
η	Length 4	Length 6	Length 4	Length 6	Both pools
0.05	0.209137	0.205528	0.874908	0.863169	0.335975
0.10	0.233728	0.231206	0.952052	0.944425	0.397600
0.20	0.248299	0.247947	0.995080	0.994061	0.467600

Table 22: Bayes risks computed by exact summation over outcomes at $p = 1/2$, with twelve acquired labels and identical targets.

coordinates. The added set is empty at $d_{\mathrm{ext}} = 0$. Each regional pool has $r + 1$ points, augmented rank r, and the original unique dependency. Define $E_{d_{\mathrm{ext}}} = \{z \in \mathbb{F}_2^{d_{\mathrm{ext}}} : |z| \text{ is even}\}$, $\mathcal{V}_{d_{\mathrm{ext}}} = \mathcal{V}_{\cap} \times E_{d_{\mathrm{ext}}}$. The set E_0 consists of the empty zero vector, and $\mathcal{V}_{d_{\mathrm{ext}}}$ is disjoint from both pools. Both regions use the same pool and have equal target weight. For $0 \le t \le 1$, put

$$\varphi_0(t) = 1, \quad \varphi_{d_{\mathrm{ext}}}(t) = \frac{(1+t)^{d_{\mathrm{ext}}} + (1-t)^{d_{\mathrm{ext}}}}{2^{d_{\mathrm{ext}}}} \quad (d_{\mathrm{ext}} \ge 1). \tag{207}$$

This factor equals $\mathbb{E}t^{|U_{\mathrm{ext}}|}$ for U_{ext} uniform on $E_{d_{\mathrm{ext}}}$. At fixed positive noise, $\varphi_{d_{\mathrm{ext}}}(\delta^2)$ tends to zero with dimension, determining the magnitude of the strict finite-dimensional ordering.

Corollary 2 (Confidence ordering for the embedded construction)**.** *For every integer $d \ge 4$, every $p \in [0, 1]$, and $0 < \eta < 1/2$, query all $2(r + 1)$ available labels in the two regions. Let $C_{k,d_{\mathrm{ext}}}$ denote posterior confidence at a uniform target in their copies of $\mathcal{V}_{d_{\mathrm{ext}}}$. Then $C_{4,d_{\mathrm{ext}}} \preceq_{\mathrm{cx}} C_{6,d_{\mathrm{ext}}}$. Every continuous symmetric concave Bayes entropy has weakly lower risk for the length-six pool, strictly lower for a strictly concave entropy. Its model-testing experiment remains a randomization of the length-four experiment. The two classification risks agree, and the Brier difference satisfies*

$$\begin{aligned} R^{(k)}_{01,d_{\mathrm{ext}}} &= \frac{1}{2} - \frac{\delta^4}{4}\varphi_{d_{\mathrm{ext}}}(\delta), \\ R^{(4)}_{\mathrm{Br},d_{\mathrm{ext}}} - R^{(6)}_{\mathrm{Br},d_{\mathrm{ext}}} &= \varphi_{d_{\mathrm{ext}}}(\delta^2)[R^{(4)}_{\cap}(12,\eta,p) - R^{(6)}_{\cap}(12,\eta,p)]. \end{aligned} \tag{208}$$

The risks use the extended common target set, where Br *denotes the Brier loss.*

Proof. Under the affine model, the added basis signs are independent fair signs, independent of the base five-dimensional parameter. To see this, replacing each extra slope by its sum with the intercept is an invertible change of variables. Thus, noisy observations at the added coordinates are independent fair bits under both models and contribute unit likelihood for the indicator vector.

For a target $(x, z) \in \mathcal{V}_{d_{\mathrm{ext}}}$, the even weight of z implies

$$f_\theta(x, z) = f_\theta(x, 0_{d_{\mathrm{ext}}}) \prod_{i:z_i=1} f_\theta(e_{4+i}). \tag{209}$$

The even weight of z makes the intercept appear an odd number of times on both sides. Each added factor has conditional posterior mean $\delta(-1)^{Y_{j,e_{4+i}}}$. Posterior independence multiplies the base signed prediction by these factors, and taking absolute values yields

$$C_{k,d_{\mathrm{ext}}} \overset{\mathcal{L}}{=} \delta^{|U_{\mathrm{ext}}|} C_k, \tag{210}$$

where U_{ext} is independent of the base history and target choice. Composing with a positive scaling factor preserves convexity and strict convexity. By Proposition 4, we obtain the comparison after averaging over U_{ext}. The first two confidence moments determine classification and Brier risk, so the displayed formulas follow. We can generate extra fair observations independently using the same channel that randomizes the six-point pair's model-state observations. □

Enumeration of regional affine parameters and queried outcomes at $d \in \{4, 5, 6, 7, 8\}$ and $\eta \in \{.05, .10, .20\}$ validates the extension. Target-sign numerators agree with Equation 210, and integer subset sums confirm the unique dependency. At $d \in \{4, 5, 6\}$, independent summation over joint outcomes checks the first four confidence moments under five priors, as well as the evaluated losses, rational inequalities, and numerical residuals.

O ACQUISITION-OBJECTIVE COMPARISONS

O.1 OBJECTIVES AND BATCH CONSTRUCTION

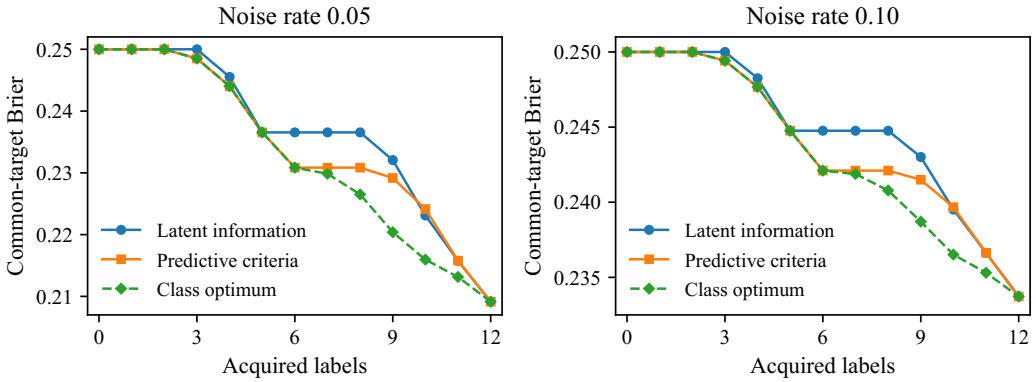


Figure 8: Exact common-target Brier for the length-four pool. Latent information denotes the coincident BALD and BatchBALD risks. Predictive criteria denote the coincident EPIG and one-step Brier risks. Two-batch and adaptive optima agree. Curves use ties resolved by the public order, with region one first. Table 24 gives ranges over the four sequential tie conventions.

We compare BALD (Houlsby et al., 2011), BatchBALD (Kirsch et al., 2019), and expected predictive information gain (EPIG) (Smith et al., 2023) on the two six-point pools. The model uses $d = 4$, a uniform exclusive regional prior, and eighteen common targets in $\{1, 2\} \times \mathcal{V}_\cap$. Each method has the exact posterior and the same available identities as the exhaustive acquisition experiment. We evaluate $b \in \{0, \dots, 12\}$ at noise rates 0, .05, .10, and .20.

Let $\Theta_\star = (J_\star, \theta_{J_\star})$ contain the affine-region index and its parameter. Let X be an independent uniform target in the fixed common set, and let $\mathcal{H}$ record acquired identities and labels. Sequential BALD selects an available unobserved identity v maximizing $I(Y_v, \Theta_\star \mid \mathcal{H}) = h_2\left(\mathbb{P}(Y_v = 1 \mid \mathcal{H})\right) - \gamma_j h_2(\eta) - (1-\gamma_j), \quad v \in \mathcal{P}_j$, where $I(A, B \mid H)$ denotes conditional mutual information between A and B given H. Conditional label entropy is $h_2(\eta)$ in the affine region and one bit in the independent region.

BatchBALD greedily constructs S to maximize the joint objective $I(Y_S, \Theta_\star \mid \mathcal{H})$. Conditional independence given $\Theta_\star$ makes the conditional entropy additive, and we integrate joint predictive entropy over all binary assignments. The first batch contains $\lfloor b/2 \rfloor$ labels. After observing it, the policy selects the entire remaining batch before using any second-batch outcome.

EPIG selects v maximizing target-averaged predictive information,

$$\mathbb{E}_X I(Y_v, Y_X \mid \mathcal{H}, X) = \mathbb{E}_X \left[h_2\left(\mathbb{P}(Y_X = 1 \mid \mathcal{H}, X)\right) - \mathbb{E}_{Y_v \mid \mathcal{H}} h_2\left(\mathbb{P}(Y_X = 1 \mid \mathcal{H}, Y_v, X)\right) \right]. \tag{211}$$

The one-step Brier control replaces entropy with posterior target variance, measuring the immediate expected reduction in common-target Brier (Cohn et al., 1996; Huang et al., 2026). Sequential rules update their acquisition score after each label. For ties within numerical tolerance, select the smallest public query index, with region-one indices preceding region-two indices.

Sequential policies are compared with R^*_{ad}, and two-batch BatchBALD is compared with R^*_2. The latter optimizes first-batch size and the acquired sets, so its comparison with BatchBALD includes the acquisition objective, greedy set construction, and batch split.

Integrating reachable histories, regional affine parameters, and candidate-batch outcomes gives the policy risks across the full grid. In Figure 8, BALD and BatchBALD coincide, as do EPIG and one-step Brier. Predictive criteria give lower risk at several budgets; latent information gives lower risk at a later budget. The class optimum accounts for subsequent regional allocation, and every rule reaches the same posterior at complete acquisition.

Check length	η	BALD	EPIG	One-step Brier
4	0.00	0.031250	0.017361	0.017361
4	0.05	0.013452	0.007473	0.007473
4	0.10	0.005243	0.002913	0.002913
4	0.20	0.000525	0.000323	0.000323
6	0.00	0.031250	0.017361	0.017361
6	0.05	0.013452	0.007804	0.007804
6	0.10	0.005243	0.004028	0.004028
6	0.20	0.000525	0.000501	0.000501

Table 23: Largest Brier variation across four tie conventions, maximized over all budgets $0 \le b \le 12$ on the same eighteen common targets. Table 24 gives the complete budget ranges for the length-four pool at the two positive noise rates in Figure 8.

O.2 Acquisition-Score Ties

We evaluate sequential BALD, EPIG, and one-step Brier under four tie conventions: the public order with region one first, reversed indices within each region, interleaved regional indices, and uniform selection among maximizers. We treat scores within 10^{-12} of the maximum as ties. The pools, prior, noise, targets, and posterior are identical across conventions.

For objective a and tie convention k_{tie}, let $R_{a,k_{\mathrm{tie}}}(b)$ be the expected Brier score on the eighteen common targets. Define the largest variation across budgets as

$$\Delta_{\mathrm{tie}}(a) = \max_{0 \le b \le 12} \left[\max_{k_{\mathrm{tie}} \in \mathcal{K}} R_{a,k_{\mathrm{tie}}}(b) - \min_{k_{\mathrm{tie}} \in \mathcal{K}} R_{a,k_{\mathrm{tie}}}(b) \right]. \tag{212}$$

The set $\mathcal{K}$ contains the four conventions, and $\Delta_{\mathrm{tie}}(a)$ is the largest within-budget risk range.

Labels tied in their current acquisition score can lead to different later allocations. This accounts for the risk variation in Table 23. EPIG and one-step Brier agree under every convention. Table 24 gives the budget-wise ranges, which coincide at complete acquisition.

We integrate over every reachable history, including uniformly randomized ties. At history H, label y at coordinate v has probability $P_{H \cup (v,y)}/P_H$. Multiplying by the tie-selection probability propagates the probability of each policy history. Summing posterior target variances then gives expected risk, with all 32 regional affine parameters represented in the posterior.

P Affine-Flat Controls and Dimension-Dependent Bounds

P.1 Affine-Flat Search and Paired Prediction

An eight-point affine flat has the form $a + \mathrm{span}\{u, v, w\}$, with anchor $a \in \mathbb{F}_2^d$ and independent directions $u, v, w \in \mathbb{F}_2^d$. Its augmented rank is four. A uniform ambient affine parameter restricts to a uniform affine function on $\mathbb{F}_2^3$, so independent label errors produce exactly the testing law of an eight-point subcube.

Before acquiring labels, group all pairs of public pool coordinates by their binary sum, in order of first public-index occurrence. Two pairs sharing a sum form an affine plane. For directions u, v, the sorted triple $\{u, v, u+v\}$ identifies the direction subspace, and the smallest coordinate in the plane identifies its coset. Two distinct cosets with this direction form an eight-point affine flat. Select the first such pair of planes and order its coordinates by public index. If the candidate set is empty, apply the two-check rule in Appendix L.3. We validate every selected flat for distinctness, affine closure, and augmented rank four.

Every three-dimensional affine flat decomposes into parallel affine planes, so the search is exhaustive. The selected-flat, two-check, and prescribed-subcube designs exclude their designated sets from the regional continuation orders and follow Algorithm 1. Each first batch is in region one, with threshold-one selection and region-one ties at the tolerance in Appendix O.2.

The study uses $d \in \{8, 12\}$, $q = (d+1)^2$, $\eta \in \{.01, .025, .05\}$, and a uniform exclusive prior. Budgets are $(13, 15, 17)$ at $d = 8$ and $(19, 22, 24)$ at $d = 12$. For each dimension and noise pair, 512

η	b	BALD risk range	EPIG and one-step Brier risk range
0.05	0	[0.250000, 0.250000]	[0.250000, 0.250000]
0.05	1	[0.250000, 0.250000]	[0.250000, 0.250000]
0.05	2	[0.250000, 0.250000]	[0.250000, 0.250000]
0.05	3	[0.249796, 0.250000]	[0.248505, 0.250000]
0.05	4	[0.245516, 0.250000]	[0.244021, 0.248505]
0.05	5	[0.236548, 0.250000]	[0.236548, 0.244021]
0.05	6	[0.236548, 0.250000]	[0.230843, 0.236548]
0.05	7	[0.236548, 0.245516]	[0.230843, 0.230843]
0.05	8	[0.236548, 0.241032]	[0.230843, 0.230843]
0.05	9	[0.232064, 0.232578]	[0.229168, 0.229168]
0.05	10	[0.223096, 0.224124]	[0.224144, 0.224869]
0.05	11	[0.215769, 0.216345]	[0.215769, 0.216018]
0.05	12	[0.209137, 0.209137]	[0.209137, 0.209137]
0.10	0	[0.250000, 0.250000]	[0.250000, 0.250000]
0.10	1	[0.250000, 0.250000]	[0.250000, 0.250000]
0.10	2	[0.250000, 0.250000]	[0.250000, 0.250000]
0.10	3	[0.249921, 0.250000]	[0.249417, 0.250000]
0.10	4	[0.248252, 0.250000]	[0.247670, 0.249417]
0.10	5	[0.244757, 0.250000]	[0.244757, 0.247670]
0.10	6	[0.244757, 0.250000]	[0.242105, 0.244757]
0.10	7	[0.244757, 0.248252]	[0.242105, 0.242105]
0.10	8	[0.244757, 0.246505]	[0.242105, 0.242105]
0.10	9	[0.243009, 0.243353]	[0.241497, 0.241497]
0.10	10	[0.239514, 0.240202]	[0.239672, 0.239936]
0.10	11	[0.236632, 0.236991]	[0.236632, 0.236788]
0.10	12	[0.233728, 0.233728]	[0.233728, 0.233728]

Table 24: Common-target Brier ranges over all four sequential tie conventions for the length-four pool at every budget. Each interval gives the minimum and maximum expected risk at that budget. The $b = 5$ rows attain the EPIG and one-step Brier maxima in Table 23 at these two noise rates.

independent samples share target labels, pools, and common evaluation coordinates across policies and budgets. Each sample contains the affine-region indicator, a uniform affine parameter, and labels generated once per coordinate in both regions.

Independent study streams follow Appendix L.2, and the pools use the coupling in Appendix L.2. Tables 25 and 26 report conditional and realized Brier for every budget and sample, including the two-check fallback, with paired Student intervals on 511 degrees of freedom. Table 27 records flat availability and regional selection.

Joint paired comparisons. For each row and risk definition, we compute 512 within-sample differences between the affine-flat control and the subcube. Two-sided paired Student tests use 511 degrees of freedom, with Holm adjustment applied jointly to 36 contrasts: 18 for conditional risk and 18 for realized risk. For ordered unadjusted values $p_{(1)} \le \cdots \le p_{(36)}$, the adjusted value at position k is

$$p_{(k)}^{\mathrm{Holm}} = \min\left\{1, \max_{1 \le j \le k} (37 - j) p_{(j)}\right\}. \tag{213}$$

We assign each adjusted value to its original contrast. Table intervals remain pointwise. Under the paired-test assumptions, the joint Holm procedure controls family-wise error and allows dependence across budgets.

The terminal-risk contrasts in Tables 25 and 26 have both signs, and all jointly Holm-adjusted p-values exceed the significance threshold. An available flat has the same verification law as the subcube. Prediction also depends on the continuation coordinates available in each pool.

We validate exhaustive search by independently enumerating anchors and three independent directions on small pools. In dimensions two through six, direct affine-parameter summation agrees with transform likelihoods and posteriors to maximum absolute discrepancy below 10^{-12}. Supplemen-

d	η	b	Flats	Two checks	Affine flat	Subcube	Flat minus subcube
8	0.010	13	512/512	0.201143	0.218523	0.217798	0.000725 ± 0.002760
8	0.010	15	512/512	0.190423	0.193812	0.194116	-0.000304 ± 0.002970
8	0.010	17	512/512	0.173772	0.168795	0.169902	-0.001107 ± 0.002907
8	0.025	13	512/512	0.214372	0.228411	0.225943	0.002467 ± 0.002260
8	0.025	15	512/512	0.205343	0.208484	0.208292	0.000192 ± 0.003041
8	0.025	17	512/512	0.191059	0.188831	0.189984	-0.001153 ± 0.003207
8	0.050	13	512/512	0.227961	0.238149	0.236963	0.001186 ± 0.001743
8	0.050	15	512/512	0.220428	0.226527	0.227263	-0.000736 ± 0.002783
8	0.050	17	512/512	0.212609	0.213817	0.215487	-0.001670 ± 0.003071
12	0.010	19	399/512	0.196559	0.198952	0.199203	-0.000251 ± 0.002988
12	0.010	22	399/512	0.173869	0.166448	0.163941	0.002507 ± 0.003144
12	0.010	24	399/512	0.162019	0.154773	0.153360	0.001413 ± 0.003618
12	0.025	19	395/512	0.208833	0.212343	0.214752	-0.002410 ± 0.002861
12	0.025	22	395/512	0.188798	0.186951	0.185423	0.001528 ± 0.003302
12	0.025	24	395/512	0.179121	0.174259	0.172646	0.001613 ± 0.003805
12	0.050	19	404/512	0.226754	0.230789	0.231067	-0.000278 ± 0.002384
12	0.050	22	404/512	0.213677	0.214751	0.214234	0.000517 ± 0.003016
12	0.050	24	404/512	0.206553	0.205976	0.204695	0.001282 ± 0.003413

Table 25: Conditional Brier in the independent affine-flat study. Flats counts region-one pools containing a flat among 512 samples. The contrast is flat control minus subcube with a pointwise 95% paired Student half-width. Multiplicity results are summarized in the text.

d	η	b	Flats	Two checks	Affine flat	Subcube	Flat minus subcube
8	0.010	13	512/512	0.201725	0.218583	0.217340	0.001243 ± 0.003686
8	0.010	15	512/512	0.189927	0.195500	0.194319	0.001181 ± 0.004197
8	0.010	17	512/512	0.172541	0.168706	0.171348	-0.002642 ± 0.005447
8	0.025	13	512/512	0.212648	0.226889	0.227145	-0.000256 ± 0.003345
8	0.025	15	512/512	0.203552	0.208605	0.208345	0.000260 ± 0.004336
8	0.025	17	512/512	0.191306	0.188053	0.189892	-0.001839 ± 0.005665
8	0.050	13	512/512	0.228006	0.236932	0.237706	-0.000774 ± 0.002972
8	0.050	15	512/512	0.221088	0.225046	0.226012	-0.000966 ± 0.004133
8	0.050	17	512/512	0.209973	0.212280	0.212873	-0.000592 ± 0.005158
12	0.010	19	399/512	0.198807	0.201888	0.203758	-0.001870 ± 0.004054
12	0.010	22	399/512	0.173124	0.169368	0.169482	-0.000113 ± 0.005557
12	0.010	24	399/512	0.162964	0.155221	0.154750	0.000471 ± 0.005282
12	0.025	19	395/512	0.208200	0.211687	0.214135	-0.002448 ± 0.003893
12	0.025	22	395/512	0.188355	0.186575	0.187330	-0.000756 ± 0.005437
12	0.025	24	395/512	0.177935	0.175085	0.173020	0.002064 ± 0.005452
12	0.050	19	404/512	0.227917	0.231240	0.232067	-0.000826 ± 0.003425
12	0.050	22	404/512	0.214518	0.218277	0.216755	0.001522 ± 0.004501
12	0.050	24	404/512	0.205309	0.207903	0.204932	0.002971 ± 0.004900

Table 26: Realized Brier on the same targets and samples as Table 25. Contrasts and interval conventions are unchanged. The paired tests for these contrasts are included in the joint Holm adjustment summarized above.

tary results include per-sample risks, queried identities, affine-region indicators, and selected-flat coordinates.

P.2 Availability of Eight-Point Affine Flats

Let A_d be the number of three-dimensional affine flats in $\mathbb{F}_2^d$. Counting such flats and their inclusion in a uniform pool gives

$$A_d = 2^{d-3} \prod_{i=0}^{2} \frac{2^d - 2^i}{8 - 2^i}, \quad \mathbb{E}N_3(\mathcal{P}) = A_d \frac{(q)_8}{(2^d)_8}, \tag{214}$$

d	η	Pools with a flat	Two-check localization	Affine-flat localization	Subcube localization
8	0.010	512/512	0.851562	0.925781	0.919922
8	0.025	512/512	0.789062	0.898438	0.871094
8	0.050	512/512	0.716797	0.814453	0.791016
12	0.010	399/512	0.814453	0.896484	0.927734
12	0.025	395/512	0.787109	0.855469	0.884766
12	0.050	404/512	0.736328	0.779297	0.812500

Table 27: Regional selection on the affine-flat study's independent samples from Table 25. First sets and localization decisions are shared across budgets. An available flat has the subcube's testing law.

where $N_3(\mathcal{P})$ counts contained flats and $(a)_8 = a(a-1)\cdots(a-7)$ is the falling factorial. The product counts three-dimensional linear subspaces, the leading factor counts their translates, and the final ratio is the inclusion probability of a specified flat. Thus, $\mathbb{P}(N_3 \geq 1) \leq \min\{1, \mathbb{E}N_3\}$. Table 28 evaluates the bound at $q = (d+1)^2$.

d	q	Expected contained affine flats	Existence probability upper bound
8	81	2.44×10^{2}	1
12	169	1.49×10^{0}	1
16	289	1.78×10^{-3}	1.79×10^{-3}
20	441	8.26×10^{-7}	8.27×10^{-7}
24	625	2.09×10^{-10}	2.10×10^{-10}
32	1089	4.21×10^{-18}	4.22×10^{-18}

Table 28: Eight-point affine-flat availability at $q = (d+1)^2$.

Along the dimension grid, eight-point flats become rarer, and the prescribed subcube remains available. At $d = 8, 12$, the pair-count argument guarantees two disjoint checks. The random-pool converse controls query sets selected by either search procedure.

P.3 Finite Bounds with the Analytic Cutoff

We evaluate the finite inequalities for two equally weighted regions with public prior $1/2$, fixed noise $\eta = .05$, and budget $b_r = \lfloor (c + 1/2)r \rfloor$. Each pool has $q = r^2$ coordinates. The structured pool contains the prescribed subcube of size t_r, and we use $n = b_r - t_r$ coordinates for continuation. Take

$$T = \lfloor 3r/4 \rfloor, \quad T_{\mathrm{rec}} = b_r - T, \quad \kappa = \left\lfloor \frac{r - T - 1}{2 \log_2(2q)} \right\rfloor. \tag{215}$$

These cutoffs satisfy $T < r$, $T \leq T_{\mathrm{rec}} < cr$, and $\kappa \geq 1$. The support-density argument bounds the two-pool failure probability by $2/(2q-1)$ and gives the uniform likelihood envelope

$$e_r^{\mathrm{ev}} = (1 + \delta^{\kappa})^{\lfloor T/\kappa \rfloor} - 1. \tag{216}$$

On the support event in Appendix D.4, this envelope bounds $E_T(\mathcal{P}, \eta)$.

We evaluate $\Psi_\eta(T_{\mathrm{rec}}, r)$ by minimizing the likelihood split over binomial support points, bounding tails below the display range with a Chernoff inequality. For random pools, we take the larger of Equation 15 and the lower bound with the indicator and parameter known. For the structured upper bound, we optimize the testing and decoding thresholds over $(\eta, 1/2)$. A union bound limits the reciprocal distinct-coordinate correction to $1 + 10^{-12}$ at these dimensions.

At $T = \lfloor 3r/4 \rfloor$ and $q = r^2$, the support parameter satisfies $\kappa \sim r/(16 \log_2 r)$. As $\lfloor T/\kappa \rfloor \delta^{\kappa}$ vanishes, the random and structured bounds approach $1/4 - g/2$ and $1/4 - g$, respectively, where $g = \delta^2/8$. Figure 5 plots these analytic-cutoff curves alongside the direct finite-sum calculation developed next.

P.4 Direct Evaluation of the Finite Support Sum

By evaluating Equation 75 directly, we can use larger κ than in the analytic construction. We evaluate the risk inequalities at the integer cutoffs listed below.

r	t_r	t_r/r	Testing error bound	Random adaptive lower bound	Structured two-batch upper bound
256	32	0.12500	3.91×10^{-2}	0.148749	0.199426
1024	64	0.06250	2.79×10^{-4}	0.148749	0.149029
4096	128	0.03125	8.78×10^{-9}	0.148749	0.148751
16384	128	0.00781	8.78×10^{-9}	0.193510	0.148751
65536	128	0.00195	8.78×10^{-9}	0.199374	0.148751
262144	256	0.00098	5.53×10^{-18}	0.199374	0.148751

Table 29: Finite-bound grid using the analytic cutoff at $\eta = .05$, $p = 1/2$, and $b_r = \lfloor (c + 1/2)r \rfloor$. The first three rows use the geometry-independent lower bound with indicator and parameter revealed. The last three establish a strict optimal-risk separation. Error and risk bounds are rounded outward.

r	T	T_{rec}	κ	e_r^{ev}	$\Psi_\eta(T_{\mathrm{rec}}, r)$ upper bound
256	192	294	1	3.32×10^{53}	4.54×10^{-3}
1024	768	1178	6	4.93×10^{23}	1.59×10^{-8}
4096	3072	4715	20	4.21×10^{7}	3.52×10^{-30}
16384	12288	18863	70	1.16×10^{-1}	1.27×10^{-115}
65536	49152	75454	248	8.89×10^{-10}	$\leq 10^{-250}$
262144	196608	301816	885	7.10×10^{-39}	$\leq 10^{-250}$

Table 30: All cutoffs and error components used in Table 29. For large early-evidence envelopes, the active lower bound is the one with indicator and parameter revealed. The support parameter uses the stated $T = \lfloor 3r/4 \rfloor$ throughout.

Fix $\eta = 1/20$, $p = 1/2$, $q = r^2$, and $b = \lfloor (c + 1/2)r \rfloor$. For $T < r$ and $\kappa \geq 1$, set $T_{\mathrm{rec}} = b - T$ and require $1 \leq T \leq T_{\mathrm{rec}}$. Write $\Gamma = \Gamma_{r,q,T}(\kappa)$ for the one-pool failure bound in Equation 75 and $e = (1 + \delta^\kappa)^{\lfloor T/\kappa \rfloor} - 1$ for the evidence bound. Proposition 1, with total affine-region energy at most one, yields

$$L_r(T, \kappa) = \frac{1}{4} - \frac{b}{4\,2^r} - g \min\left\{1, \frac{1}{2} + \frac{e}{2} + 2\Gamma + \Psi_\eta(T_{\mathrm{rec}}, r)\right\} \leq R_{\mathrm{ad}}^{*,\mathrm{rand}}(b). \tag{217}$$

The minimum includes the bound with indicator and parameter revealed, so the expression also holds for large evidence terms. Table 31 lists feasible choices of T and κ.

Arithmetic for the lower bound. At $\eta = 1/20$, the terms in Γ and the evidence value e are rational. For the likelihood calculation, write $n = T_{\mathrm{rec}}$ and, for integer $a \in \{0, \ldots, n\}$, define

$$P_{n,r}(a) = \frac{19^{n-a}}{10^n 2^r} + \frac{1}{20^n} \sum_{i=0}^{a-1} \binom{n}{i} 19^{n-i}. \tag{218}$$

The first term equals $2^{S_n(a)-r}$, where $S_n(a)$ is the supplied-parameter log-likelihood with a errors. The second equals $\mathbb{P}(B_n < a)$, taking an empty sum to be zero. Substituting $\zeta = S_n(a)$ in Equation 10 yields $\Psi_\eta(n, r) \leq \min\{1, P_{n,r}(a)\}$ at the thresholds in Table 31.

We compute binomial coefficients exactly and accumulate the nonnegative rational terms with upward rounding. We use these bounds to establish separation before rounding for display.

Arithmetic for the upper bound. Let $n = b - t_r$ be the continuation length. Write $\varepsilon_{\mathrm{sel}}$ for regional selection-error probability and $\varepsilon_{\mathrm{dec}}$ for the parameter-uniform decoding-error bound. Under the equal prior, the bounds in Appendix E give

$$\begin{aligned} \varepsilon_{\mathrm{sel}} &\leq \tfrac{1}{2} e^{-t_r \,\mathrm{kl}(\tau_{\mathrm{test}}, \eta)} + t_r 2^{-t_r[1 - h_2(\tau_{\mathrm{test}})]}, \\ \varepsilon_{\mathrm{dec}} &\leq e^{-n \,\mathrm{kl}(\tau_{\mathrm{dec}}, \eta)} + \frac{2^r - 2}{p_{\mathrm{dist}}(m, n, t_r)} 2^{-n[1 - h_2(\tau_{\mathrm{dec}})]}, \\ U_r &= \min\{1/4,\, 1/4 - g + \varepsilon_{\mathrm{sel}} + \varepsilon_{\mathrm{dec}}\}. \end{aligned} \tag{219}$$

r	b	t_r	T	T_{rec}	κ	a	τ_{test}	τ_{dec}
768	1460	64	433	1027	30	45	.1925	.0882
1024	1946	64	602	1344	36	54	.1925	.0897
2048	3893	64	1242	2651	64	99	.1925	.0919
4096	7787	128	2452	5335	120	203	.2037	.0927

Table 31: Parameters for direct finite evaluation with a subcapacity recovery cutoff. The binomial threshold uses error count a. The final two columns give the disagreement thresholds for the subcube test and parameter decoder.

The threshold-one likelihood test has an error sum at most that of the disagreement test. Averaging the two model errors under the equal prior introduces the factor $1/2$. The distinct-coordinate correction satisfies

$$p_{\mathrm{dist}}(m, n, t_r) \geq 1 - \frac{nt_r + \binom{n}{2}}{m} > 0 \tag{220}$$

at every row of the table. We bound the logarithms and exponentials by directed interval arithmetic, so $R_2^{*,\mathrm{str}}(b) \leq U_r$ holds for the prescribed policy and its Bayes predictor.

r	Γ upper	e upper	$\Psi_\eta(T_{\mathrm{rec}}, r)$ upper	$\varepsilon_{\mathrm{sel}}$ upper	$\varepsilon_{\mathrm{dec}}$ upper
768	1.822×10^{-5}	.788258	.166912	2.790×10^{-4}	2.861×10^{-8}
1024	4.805×10^{-4}	.428257	.041753	2.790×10^{-4}	1.083×10^{-11}
2048	6.451×10^{-8}	.022641	.000866	2.790×10^{-4}	1.422×10^{-25}
4096	3.532×10^{-8}	.000065	.000017	8.771×10^{-9}	2.409×10^{-46}

Table 32: Error components for the explicit cutoffs, each rounded upward. Final risks use the higher-precision bounds before the coarser display rounding in this table.

r	t_r/r	Random adaptive lower	Structured two-batch upper	Strict certificate
768	.08333	.148749	.149029	$L_r \leq U_r$
1024	.06250	.173369	.149029	$L_r > U_r$
2048	.03125	.198141	.149029	$L_r > U_r$
4096	.03125	.199370	.148751	$L_r > U_r$

Table 33: Finite risk certificates from Equation 217. Lower risks are rounded downward and upper risks upward. Rows with $L_r > U_r$ establish strict separation between acquisition classes.

Figure 5 plots the finite-sum grid $r \in \{768, 1024, 2048, 4096\}$ and the analytic-cutoff grid $r \in \{256, 1024, 4096, 16384, 65536, 262144\}$ for the expected-risk comparison in Section 6.3.

Q Subcube Verification under Positive Noise

We evaluate threshold-one verification on subcubes with 64, 128, and 256 labels and compare simulated errors with a bound for random access to the whole pool. Using affine symmetry, we integrate out the ambient parameter and evaluate the likelihood from the block labels. The finite terminal-risk bounds in Appendix P.4 combine testing and recovery bounds.

Q.1 Block Sizes and the Analytic Random-Pool Comparator

We use $r \in \{1024, 4096, 262144\}$, $q = r^2$, $\eta \in \{.05, .10, .20\}$, and an equal regional prior. The region-one first batch has $t_r = 2^{\lceil \log_2((\log r)^2) \rceil}$ labels, as in Theorem 1. At budget $b_r = \lfloor (c + 1/2)r \rfloor$, it leaves $n_r = b_r - t_r$ continuation labels. Table 36 gives their counts and capacity margins within the separation interval.

For the random-pool comparison, define

$$h_r = \max\left\{t_r, \left\lfloor \frac{r}{4\log_2 q} \right\rfloor\right\}, \qquad \rho_r^{\mathrm{short}} = \frac{2}{2^{r-1}-q} \sum_{\substack{4 \le w \le h_r \\ w \text{ even}}} \binom{q}{w}, \tag{221}$$

where h_r is the dependency-length cutoff and ρ_r^{short} bounds the probability that either pool contains a dependency of that length or shorter. After the first $w-1$ coordinates of a candidate relation are exposed, the last is determined and has conditional probability at most $(2^{r-1}-q)^{-1}$. Distinct augmented rows have dependencies of even weight at least four. A union bound over these weights and both pools proves Equation 221.

Outside this exceptional event, every set of at most $t_r \le h_r$ rows is independent, with $L_Q(y) = 1$ for all outcomes. This property holds uniformly, including for adaptive choices from the full pool. If $e^*_{\mathrm{rand}}(t_r)$ denotes the minimum equal-prior testing error with at most t_r labels, then

$$\frac{1}{2} - \frac{\rho_r^{\mathrm{short}}}{2} \le e^*_{\mathrm{rand}}(t_r) \le \frac{1}{2}. \tag{222}$$

For the lower bound, average fair-transcript testing error on the geometry event and use nonnegative error on its complement. A public decision achieves the upper bound. Exact integer comparisons establish the probability bounds in Table 34.

r	q	t_r	h_r	ρ_r^{short} upper
1024	1048576	64	64	4×10^{-12}
4096	16777216	128	128	10^{-523}
262144	68719476736	256	1820	10^{-64334}

Table 34: Analytic short-dependency bounds for both complete random pools. Outside the bounded event, every query set of size at most t_r has likelihood ratio exactly one.

Q.2 Exact Likelihood Evaluation and Sampling

On the prescribed subcube, let $s_r = \log_2 t_r$ be the binary dimension. The set of affine restrictions is

$$\mathcal{A}_{s_r} = \{(\beta + u^\top x)_{x \in \mathbb{F}_2^{s_r}} : \beta \in \mathbb{F}_2,\ u \in \mathbb{F}_2^{s_r}\}. \tag{223}$$

The family $\mathcal{A}_{s_r}$ contains $2t_r$ restrictions. For outcomes $y \in \mathbb{F}_2^{t_r}$, the exact affine-mixture likelihood is

$$L(y) = \frac{2^{t_r}}{2t_r} \sum_{a \in \mathcal{A}_{s_r}} (1-\eta)^{t_r - d_H(y,a)} \eta^{d_H(y,a)}, \tag{224}$$

where $d_H(y,a)$ is Hamming distance. We compute the likelihood from integer Walsh-Hadamard correlations and logarithmic normalization, as in Appendix I.1. The test selects region one when $L(y) \ge 1$ and makes the same decision on numerical ties.

For each dimension and noise pair, we simulate $M = 131072$ independent outcomes under each model. Adding an affine restriction to y leaves $L(y)$ unchanged, so Bernoulli(η) errors around the zero restriction produce the likelihood distribution for a uniform affine parameter. Under the independent model, every coordinate is a fair bit.

Let K_1 be the number of affine-model rejections and K_0 the number of independent-model acceptances. Their estimated equal-prior error is

$$\widehat{e}_{\mathrm{str}} = \frac{K_1 + K_0}{2M}. \tag{225}$$

The counts use independent samples. For a binomial count k from M trials, invert the exact binomial distribution to form a two-sided interval $[l(k), u(k)]$ with coverage at least .975. At an interior count, the lower endpoint is the .0125 quantile of $\mathrm{Beta}(k, M-k+1)$ and the upper endpoint is the .9875 quantile of $\mathrm{Beta}(k+1, M-k)$, where Beta denotes the beta distribution. Set $l(0) = 0$ and

$u(M) = 1$, using the same beta-quantile formulas for the other boundary endpoints. A union bound then gives the equal-prior error interval

$$\left[\frac{l(K_1) + l(K_0)}{2}, \frac{u(K_1) + u(K_0)}{2}\right] \tag{226}$$

with coverage at least .95 for each dimension and noise pair. Its upper endpoint remains positive even when both observed error counts are zero.

r	η	K_1	K_0	$\widehat{e}_{\mathrm{str}}$	95% error interval
1024	.05	3	1	1.526×10^{-5}	$[1.85 \times 10^{-6},\ 6.15 \times 10^{-5}]$
1024	.10	189	48	9.041×10^{-4}	$[7.376 \times 10^{-4},\ 1.100 \times 10^{-3}]$
1024	.20	5314	5206	.040131	[.038924, .041363]
4096	.05	0	0	0	$[0,\ 3.344 \times 10^{-5}]$
4096	.10	0	0	0	$[0,\ 3.344 \times 10^{-5}]$
4096	.20	216	204	.001602	[.001364, .001869]
262144	.05	0	0	0	$[0,\ 3.344 \times 10^{-5}]$
262144	.10	0	0	0	$[0,\ 3.344 \times 10^{-5}]$
262144	.20	0	0	0	$[0,\ 3.344 \times 10^{-5}]$

Table 35: Observed errors of the threshold-one subcube test. Each row uses 131072 outcomes under each regional model. Interval endpoints are rounded outward. The random-pool comparator is the whole-pool bound in Table 34.

Q.3 Verification and the Remaining Budget

Table 35 shows decreasing testing error as subcube size grows. The continuation's normalized capacity margin is $n_r C/r - 1$, which is positive throughout Table 36. We use these continuation counts in the decoding bound in Equation 17.

r	η	b_r	n_r	$n_r C/r - 1$
1024	.05	1946	1882	.311524
1024	.10	2440	2376	.232096
1024	.20	4194	4130	.121520
4096	.05	7787	7659	.334347
4096	.10	9761	9633	.248820
4096	.20	16778	16650	.130346
262144	.05	498424	498168	.356103
262144	.10	624747	624491	.264982
262144	.20	1073792	1073536	.138764

Table 36: Budgets and continuation counts for the verification experiment. The final column is the normalized capacity margin after the prescribed first batch.

Direct parameter summation agrees with transform likelihoods at subcube sizes eight, sixteen, and thirty-two. We also verify that adding an affine restriction leaves the likelihood unchanged. Sampling uses the independent streams in Appendix L.2.

R POTENTIAL LIMITATIONS

Our pool-separation results assume a known regional partition, known noise rates, uniform affine parameters, and conditionally independent label flips. Estimating the partition or noise rates adds uncertainty to regional testing and parameter recovery. Noise dependence can also change the parity evidence used in the likelihood bounds. Our finite experiments study the interaction between testing and estimation on binary affine populations with controlled query pools. Applying the analysis to learned regions and richer prediction families requires accounting for these additional sources of uncertainty.

We allow unrestricted computation when defining the statistical optima. Exact population prediction is near-linear in the explicit population size and exponential in coordinate dimension. Optimizing noisy acquisition also involves enumerating an exponentially growing history space (Appendix I.1). The guarantees for label efficiency and the computational cost of attaining them are therefore separate considerations. With approximate inference and acquisition, a further tradeoff exists between computational cost and prediction risk.