TALC: TARGET-AWARE LOSS COMPARISON FOR COST-BASED DECISIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

When learned costs feed a decision solver, a lower prediction loss can leave the decision unchanged. We introduce TALC (Target-Aware Loss Comparison), a framework that separates target choice, scoring rule, training procedure and decision cost. Our analysis decomposes squared threshold excess risk into decoded-mean error and a component invisible to the decoder. For arbitrary finite linear decision sets and finite cost alphabets, expected-cost regret bounds depend only on the decoded-mean component, by specializing existing squared-risk bounds. A cost-preserving projection can improve threshold Brier loss and binary cross-entropy while preserving the argmin set in exact arithmetic. We compare these two scores with decoded-mean MSE using a shared head, decoder and solver. On 300 saved models, projection reduces Brier scores with no observed change in true-cost error or regret; floating-point path identities can differ. Brier's extra residual constraint does not uniformly reduce cost error or regret. Across shortest-path tasks, corrected comparisons support threshold supervision in several settings. On fresh synthetic cost rules, specified learning-rate and duration changes significantly shift the BCE–Brier error difference, with observed mean orderings crossing. On Warcraft, most exact route errors add less than 0.1% stored cost, and their incidence depends on the cost precision used. A predicted training-duration reversal fails to transfer to battery scheduling. TALC connects what a loss constrains to what a decision actually costs.

1 INTRODUCTION

Learned costs often determine a downstream decision: a model predicts edge costs, and a solver chooses the cheapest route. Improving prediction loss is useful to that solver only through the costs it receives. Yet a comparison between classification-style and regression losses typically changes both the prediction targets and the rule used to score them. Classification-style value prediction has improved deep reinforcement learning (Farebrother et al., 2024), and distributional regression has motivated comparisons between component losses and decoded-mean losses (Imani et al., 2026). These findings raise a precise question for cost-based decisions: which part of the supervised error affects the decision, and which part can change without affecting it?

A cost predictor can assign a probability to each binary question "is the cost at least this level?" Adding these threshold probabilities to a base cost decodes a numerical cost. If two threshold predictions have equal and opposite errors, their contributions to that cost cancel. A loss on the decoded cost can then see no error even though both threshold outputs miss their labels. The decoded cost remains correct: supervision of each threshold imposes a constraint beyond supervision of their sum. This distinction is the starting point of TALC (Target-Aware Loss Comparison).

Decoded-mean squared error (Mean-MSE) scores the reconstructed cost; Brier scores each threshold probability with squared error; binary cross-entropy (BCE) scores those same thresholds with log loss. Brier (Brier, 1950; Gneiting & Raftery, 2007) lets us change one choice at a time: Mean-MSE versus Brier changes the supervised quantity, while Brier versus BCE changes the penalty on identical threshold targets.

Our analysis makes this distinction exact. For any finite linear decision set and finite, possibly unevenly spaced cost alphabet, weighted Brier excess risk decomposes into decoded-mean error and a residual component invisible to the decoder. The decoded-mean component controls expected

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.

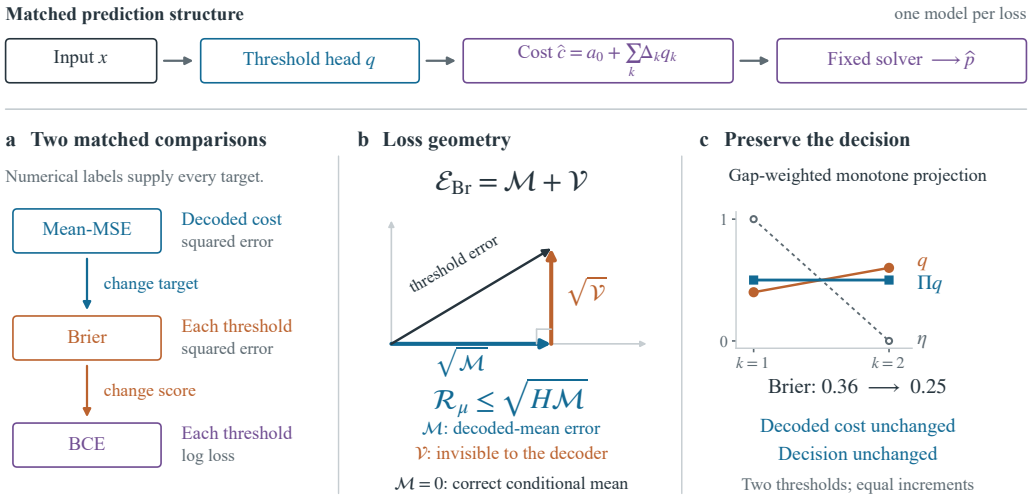

Figure 1: **TALC separates threshold supervision from decoder-visible error.** Each loss trains a separate model with a matched prediction head, decoder and solver (top). **(a)** Brier provides the squared-loss control between decoded-mean MSE and threshold BCE, separating target choice from scoring rule. **(b)** Threshold Brier excess risk decomposes into decoded-mean error $\mathcal{M}$ and a component $\mathcal{V}$ invisible to the decoder; the regret bound depends on $\mathcal{M}$. **(c)** Gap-weighted monotone projection can improve both scores while preserving decoded costs and decisions. With unit increments and target $(1, 0)$, $q = (.4, .6)$ projects to $(.5, .5)$, yielding the illustrated Brier decrease.

cost regret through an existing geometric risk bound (Ho-Nguyen & Kılınç-Karzan, 2022). A complementary projection result shows that both threshold scores can decrease while every decoded cost and decision stays fixed. Together, these results identify what threshold supervision adds and which part of its error reaches the solver. They motivate measuring decision cost alongside prediction loss; neither the variance identity nor the underlying isotonic algorithm is new.

TALC turns this analysis into a matched experimental protocol. A shared prediction head, decoder and solver isolate the two loss comparisons; paired training changes test the stability of their rankings; realized regret measures the cost of the resulting decisions. A portable audit of 300 models connects sample residuals to cost error and regret, and quantifies score reductions under decision-preserving projection. Across three shortest-path studies, corrected comparisons support threshold supervision in several settings, while the same-target scoring comparison changes direction between procedures. Tests on fresh synthetic cost rules isolate two significant shifts in the BCE–Brier error difference under specified learning-rate and duration changes, with observed mean orderings crossing. On Warcraft, exact route error is sensitive to cost precision even though most counted errors add little cost. Figure 1 summarizes the analysis and diagnostic protocol.

2 Predicting Costs and Choosing a Path

Each example contains an input X, a known graph and true edge or vertex costs C. A network predicts costs from numerical contexts or pixels; a common solver chooses a minimum-predicted-cost path. We condition on the known graph, including it in the input when it varies across examples.

Formally, let $\mathcal{P} \subset \{0, 1\}^d$ be the nonempty finite set of feasible paths, represented by their edge or vertex indicators. For positive true costs $C \in (0, \infty)^d$, let

$$F(C) = \arg\min_{p \in \mathcal{P}} C^\top p, \qquad \widehat{p} \in \arg\min_{p \in \mathcal{P}} \widehat{c}(X)^\top p.$$

A common deterministic rule breaks predicted ties. Any path in $F(C)$ counts as correct, so the model need not reproduce one designated route or recover every tied optimum. We measure both the fraction

of wrong path choices and the extra true cost incurred:

$$\mathbf{1}\{\widehat{p} \notin F(C)\}, \qquad C^\top \widehat{p} - \min_{p \in \mathcal{P}} C^\top p.$$

The second quantity is *realized regret*, the extra true path cost. The two measures can rank models differently: one model may make fewer mistakes but occasionally choose a much more expensive path.

The three central objectives all receive full numerical costs during training. Their threshold labels are computed from those costs, not obtained as additional observations. True costs are used again to evaluate decisions, but are unavailable to the predictor and solver at inference. We do not sort threshold outputs, round predicted costs or change the path solver in the original comparisons.

3 TALC: A MATCHED LOSS-TO-DECISION AUDIT

3.1 PROTOCOL AND SHARED PREDICTION STRUCTURE

TALC trains one model per loss with matched data, prediction structure, decoder and paired training streams. It records Brier–Mean-MSE and BCE–Brier contrasts, their changes under specified training choices, and both decision error and regret. Confirmatory comparisons use prespecified independent units and correction families. Figure 1 shows the shared *cumulative head*: each output is a probability for one threshold question, and their sum supplies a numerical cost. All three losses train this same output structure.

More generally, write one cost as $C = a_0 + sY$, where $Y \in \{0, \dots, m\}$, $m \geq 1$, and adjacent costs differ by $s > 0$. The three central tasks use $s = 1$. From the observed cost we form binary labels $t_k = \mathbf{1}\{Y \geq k\}$ for $k = 1, \dots, m$. The network outputs separate logits z_k, without enforcing their threshold order, and predicts

$$q_k = \sigma(z_k), \qquad \widehat{c} = a_0 + s \sum_{k=1}^{m} q_k. \tag{1}$$

Each objective uses these same outputs. Let $e_k = q_k - t_k$ and $\bar{e} = m^{-1} \sum_k e_k$. We compare

$$L_{\text{mean}} = \left(\frac{\widehat{c} - C}{sm} \right)^2 = \bar{e}^2, \tag{2}$$

$$L_{\text{Brier}} = \frac{1}{m} \sum_{k=1}^{m} e_k^2, \tag{3}$$

$$L_{\text{BCE}} = \frac{1}{m} \sum_{k=1}^{m} \left[\log(1 + e^{z_k}) - t_k z_k \right]. \tag{4}$$

Mean-MSE scores the decoded cost after dividing its error by the range sm. Brier scores each threshold with a squared error, without a factor of two. BCE scores those same binary labels with log loss, evaluated stably from logits. All three also average over cost coordinates and examples.

Table 1: **What changes between the three objectives.** The output head, numerical labels, cost decoder, path solver, initialization and batch streams are shared within each study. The training and selection procedures are described in Section 5.

Objective	What is compared with its label?	Per-cost normalization
Mean-MSE	The summed cost	Divide squared error by $(sm)^2$
Brier	Each threshold probability	Average over m thresholds
BCE	Each threshold probability	Average over m thresholds

Mean-MSE versus Brier isolates supervision of the decoded sum from supervision of each threshold. Brier versus BCE isolates the scoring rule on those same thresholds. The direct Mean-MSE/BCE comparison changes both.

3.2 WHAT THE ADDITIONAL THRESHOLD CONSTRAINT DOES

The two squared losses satisfy

$$\boxed{L_{\text{Brier}} = L_{\text{mean}} + \frac{1}{m}\sum_{k=1}^{m}(e_k - \bar{e})^2.} \tag{5}$$

Positive and negative threshold residuals can cancel: $-\delta$ and $+\delta$ give zero decoded error despite nonzero threshold errors. Mean-MSE correctly assigns zero cost error; Brier also penalizes their dispersion. This additional constraint need not improve learned decisions, which is the question tested by the matched experiments.

At the population level, all three unrestricted objectives recover the conditional mean cost required for a Bayes expected-cost decision. Brier and BCE identify each conditional threshold probability; Mean-MSE identifies their decoded sum. Section 4 separates the part of threshold error that changes decoded costs from the part that cancels, and derives regret bounds for general finite linear decisions.

All objectives use the same separate sigmoid outputs and cost decoder. Because these outputs are not forced to decrease with the threshold, they may cross, with $q_{k+1} > q_k$. Original comparisons evaluate them directly; our post-hoc audit also applies a cost-preserving monotone projection (Appendix C). Hyperparameters are selected separately under equal within-study search budgets.

The Warcraft extension has unevenly spaced costs. It replaces the unit increments by the observed cost gaps and applies the same normalized gap weights to both threshold losses. The weighted variance identity preserves the two comparisons; Appendix D gives the exact formulas.

Auditing a fitted model. Save threshold probabilities, alphabet gaps, original decoded costs and true costs for each independent unit. Report the sample mean-error and residual terms together with paired decision error and regret; apply gap-weighted isotonic projection to quantify score changes at fixed costs. Check floating-point decoding separately, and repeat cost comparisons at fixed and validation-selected checkpoints. These diagnostics characterize existing comparisons; they neither select a universally best loss nor replace confirmatory tests.

4 FROM THRESHOLD ERROR TO DECISION REGRET

The decoder retains one weighted mean from each threshold vector. We separate its error from the remaining threshold error, show how both scores can improve at fixed decoded costs, and transfer the retained error to regret using existing squared-risk geometry.

Condition on $X = x$, including the known graph, and allow any nonempty finite action set $\mathcal{P} \subset \mathbb{R}^d$. Coordinate e has alphabet $a_{e,0} < \cdots < a_{e,m_e}$, $m_e \geq 1$, increments $\Delta_{e,k} = a_{e,k} - a_{e,k-1}$, range $R_e = a_{e,m_e} - a_{e,0}$, and weights $w_{e,k} = \Delta_{e,k}/R_e$. Set $\eta_{e,k} = \Pr(C_e \geq a_{e,k} \mid x)$ and $\widehat{c}_e = a_{e,0} + \sum_k \Delta_{e,k} q_{e,k}$ for $q_{e,k} \in [0,1]$. Write $\mu = \mathbb{E}[C \mid x]$, $p_\mu \in \arg\min_{p\in\mathcal{P}} \mu^\top p$, $\widehat{p} \in \arg\min_{p\in\mathcal{P}} \widehat{c}^\top p$, and $\mathcal{R}_\mu = \mu^\top(\widehat{p} - p_\mu)$. For $\delta_{e,k} = q_{e,k} - \eta_{e,k}$, define

$$h_e = \sum_k w_{e,k}\delta_{e,k} = (\widehat{c}_e - \mu_e)/R_e, \qquad \mathcal{M} = \sum_e h_e^2,$$

$$\mathcal{E}_{\text{Br}} = \sum_{e,k} w_{e,k}\delta_{e,k}^2, \qquad \mathcal{V} = \sum_{e,k} w_{e,k}(\delta_{e,k} - h_e)^2,$$

$$\mathcal{E}_{\text{BCE}} = \sum_{e,k} w_{e,k} D_{\text{KL}}(\text{Ber}(\eta_{e,k}) \| \,\text{Ber}(q_{e,k}))\,.$$

These conditional excess risks subtract unrestricted minima and sum over cost coordinates. Here $\mathcal{M}$ is range-normalized Mean-MSE excess; the threshold risks use gap-proportional weights. Divide by d for coordinate averages; Appendix A.4 treats other weights.

Proposition 1 (Weighted decoder decomposition)**.**

$$\mathcal{E}_{\text{Br}} = \mathcal{M} + \mathcal{V}. \tag{6}$$

This is the weighted variance identity specialized to the decoder. The residual measured by $\mathcal{V}$ decodes to zero. The decoder-visible $\mathcal{M}$ is sufficient to control regret, but is not an exact measure of decision-relevant error: even decoded errors can leave all action-cost differences unchanged.

Proposition 2 (Improved threshold scores with unchanged decisions)**.** *For one coordinate, let r minimize $\sum_k w_k(r_k - q_k)^2$ over $r_1 \geq \cdots \geq r_m$. Then $r \in [0,1]^m$, $\sum_k \Delta_k r_k = \sum_k \Delta_k q_k$, and for every threshold label vector t,*

$$\sum_k w_k(q_k - t_k)^2 \geq \sum_k w_k(r_k - t_k)^2 + \sum_k w_k(q_k - r_k)^2.$$

If $q \in (0,1)^m$ violates the threshold order, this projection strictly reduces both weighted Brier and weighted BCE for every possible cost label.

This standard weighted isotonic projection (de Leeuw et al., 2009) preserves all action costs and argmin sets, hence the selected decision under fixed tie-breaking in exact arithmetic. Appendix A.3 proves the score improvement. Training and the original loss comparisons use unprojected predictions.

Squared prediction excess controls decision excess through feasible-set diameter (Ho-Nguyen & Kılınç-Karzan, 2022). Range rescaling yields the following corollary; the last step is Bernoulli Pinsker.

Corollary 1 (Regret transfer)**.** *Let $H = \max_{p,r \in \mathcal{P}} \sum_e R_e^2 (p_e - r_e)^2$. If $H > 0$, using natural logarithms,*

$$\mathcal{R}_\mu \leq \sqrt{H\mathcal{M}} \leq \sqrt{H\mathcal{E}_{\mathrm{Br}}} \leq \sqrt{H\mathcal{E}_{\mathrm{BCE}}/2}. \tag{7}$$

For $H = 0$, $\mathcal{R}_\mu = 0$. Constants 1 and $1/2$ are sharp, even for two parallel edges with binary positive costs.

For deterministic costs, $\mathcal{R}_\mu$ is realized regret; otherwise it is excess expected cost over the Bayes action. Appendix A.4 gives the prior-risk mapping, sharpness and positive-gap guarantees. These are fixed-prediction relations; they do not determine how training changes $\mathcal{M}$.

5 Experiments

We apply TALC to shortest paths so that every decision can be checked against the true costs. Its two matched comparisons are summarized in Appendix Table 6: Brier versus Mean-MSE changes which quantities receive squared-error supervision; BCE versus Brier changes the penalty on the same binary threshold targets. BCE versus Mean-MSE changes both. A *tied context* has several paths with the same lowest true cost; choosing any of them counts as correct. A synthetic *teacher* is a sampled rule that generates true edge costs from an input, not a trained predictor. An image *partition* is a split of source digit images. Study A uses a fixed short budget, Study B a longer validation-selected procedure, and Study C fixed Warcraft maps. Each study corrects its own planned comparisons; Appendix Table 6 does not pool studies. "Registered" denotes locally recorded protocols; hashes bind contents, not independently certified timestamps.

5.1 Study A: fixed short-budget comparison

Tasks and analysis units. Two synthetic tasks predict positive integer edge costs on 11-node, 27-edge DAGs from eight-dimensional contexts. Their teachers use random ReLU features or quadratic products. Each split contains 512/512/2,048 training/validation/evaluation contexts. Primary error uses naturally tied contexts (about 61% and 64%); selecting any true optimum counts as correct. Digits predicts $1 + d$ costs from 4×4 boards of digit images and chooses among 20 right/down paths. It uses all evaluation boards and disjoint source-image pools from UCI Optical Digits (Alpaydin & Kaynak, 1998; scikit-learn developers, n.d.). Its 20 seeds vary optimization on fixed images and boards; synthetic seeds also vary graph, teacher and contexts. Appendix B.1 gives data construction and locates the unique-optimum results in the supplement.

A common training and selection budget. The network has two 64-unit ReLU hidden layers and the same cumulative head. Methods share initialization, contexts and minibatch order within a seed. Adam runs for 600 updates, with batch size 32 for synthetic tasks and 16 for images; we evaluate

Table 2: **Path choices after 600 updates.** Primary error and true regret are equal-seed means $\pm$ sample SD over 20 evaluation seeds. Synthetic errors use naturally tied contexts; Digits uses all boards. Error comparisons share one nine-test Holm family; regret is descriptive. The same 12-configuration search is available to each objective.

Objective	ReLU (tied)	Quadratic (tied)	Digits (all)
Primary error (%)			
BCE	$\mathbf{8.364} \pm 1.028$	$\mathbf{12.553} \pm 1.292$	$\mathbf{9.429} \pm 1.313$
Brier	9.940 ± 1.222	13.576 ± 1.811	11.484 ± 2.323
Mean-MSE	9.799 ± 1.139	13.386 ± 0.961	14.653 ± 0.818
True regret			
BCE	0.087 ± 0.012	0.137 ± 0.015	0.226 ± 0.045
Brier	0.104 ± 0.013	0.148 ± 0.021	0.248 ± 0.070
Mean-MSE	0.101 ± 0.012	0.141 ± 0.011	0.323 ± 0.034

only the final checkpoint. Each loss receives 12 configurations: learning rate $\{.0005, .002, .006, .02\}$ crossed with weight decay $\{0, .0001, .001\}$. Five development seeds select a setting by primary validation error, then regret, then deterministic parameter tie breaks. All selections are fixed before evaluation on 20 new seeds per task.

Paired comparisons. For each task we test all three pairs, with one Holm correction across the nine comparisons. The seed is the unit of analysis; synthetic seeds vary the graph, teacher and contexts, while image seeds vary optimization. Exact paired sign-flip tests enumerate 2^{20} assignments and assume a sign-symmetric paired null. Reported 95% bootstrap intervals use 20,000 paired-seed draws and are marginal, not simultaneous. Appendix B.1 lists every test and selected setting.

Results. BCE makes fewer nonoptimal choices than both squared objectives on all three tasks (Table 2). Each of those six planned comparisons passes the nine-test Holm correction. In particular, BCE beats Brier when both train the same individual thresholds. Thus the BCE–Mean-MSE gap in this study cannot be explained solely by which quantities receive supervision; the scoring rule matters empirically here. On all synthetic contexts, the same Study A models retain BCE's mean ordering (Appendix Table 13); the tied-context endpoint remains the planned primary comparison.

Brier improves on Mean-MSE on Digits, but neither synthetic task establishes that advantage. Several development selections reach the learning-rate search boundary, motivating the next procedure.

5.2 Study B: an extended training and selection procedure

We keep the head and decoder, train every model for 6,000 updates, and select checkpoints by validation error, regret, then earliest step. Each objective receives the same 18 settings: six learning rates from .0001 to .06 and the previous three weight decays. Three development teacher seeds per synthetic task and three new Digits partitions (two initializations each) choose the settings. Confirmation uses ten new teachers per synthetic task and ten new Digits partitions, again with two initializations each. All image partitions reuse the same 1,797-image pool.

The primary endpoint is nonoptimal-path error on *all* contexts, so its synthetic percentages are not directly comparable with A's tied-only percentages. The unit is a teacher seed or a partition mean over two initializations. The nine comparisons receive a within-study Holm correction at level .025; Appendix B.2 gives the allocation plan, complete contrasts and fixed-step checks.

Brier improves over Mean-MSE on both synthetic tasks and over BCE on ReLU after correction; BCE improves over Mean-MSE on quadratic. No Digits comparison passes correction. Brier leads on ReLU in this procedure, whereas BCE led in A. Appendix Figure 4 shows the paired-unit patterns.

Table 3: **Path choices after extended training and validation selection.** All-context error (%), mean ± SD across ten teacher or partition units; true regret is the mean across those same units. Every model trains for 6,000 updates; its checkpoint is selected on validation. This is a separate study from Table 2.

Objective	ReLU (all)	Quadratic (all)	Digits (all)
Primary error (%)			
BCE	10.762 ± 1.250	14.585 ± 1.767	7.632 ± 2.994
Brier	10.103 ± 1.372	13.584 ± 1.515	8.442 ± 2.814
Mean-MSE	12.627 ± 1.002	17.197 ± 1.882	8.501 ± 2.821
Mean true regret			
BCE	0.116	0.162	0.163
Brier	0.106	0.147	0.163
Mean-MSE	0.132	0.183	0.190

5.3 What improves between threshold scores and decisions?

Table 4: **Reduced residual dispersion need not reduce decision cost.** Brier minus Mean-MSE, using all contexts from the same fitted models. M_{samp} is range-normalized sample cost MSE and V_{samp} the sample residual term, both averaged over coordinates and units; neither is a population excess-risk estimate. Negative values favor Brier. Regret intervals are descriptive marginal 95% paired-unit bootstrap intervals (20,000 draws), not new corrected tests.

Study	Task	ΔV_{samp}	ΔM_{samp}	Error (pp)	Regret	Marginal 95% CI
A	ReLU	-0.0873	+0.0004	+0.41	+0.0070	[+0.0008, +0.0143]
A	Quadratic	-0.1200	+0.0005	+0.28	+0.0113	[+0.0026, +0.0200]
A	Digits	-0.1428	-0.0016	-3.17	-0.0748	[-0.1046, -0.0416]
B	ReLU	-0.0980	-0.0037	-2.52	-0.0257	[-0.0326, -0.0189]
B	Quadratic	-0.1268	-0.0044	-3.61	-0.0361	[-0.0402, -0.0316]
B	Digits	-0.1864	-0.0007	-0.06	-0.0277	[-0.0736, +0.0144]

The post-hoc probability audit links scores to costs (Table 4). Brier's residual dispersion is lower in all six task–procedure combinations. In A's synthetic tasks, sample mean error and regret instead increase; in B's synthetic tasks both decrease. Thus the extra threshold constraint alone does not establish a decision benefit. These paired associations do not test mediation, and A's original synthetic primary endpoint remains tied-only.

Improving scores without observed true-cost changes. We apply monotone projection to all 300 saved A/B models. Across 491,520 context–model evaluations, replayed true-cost error and regret are unchanged, while Brier and finite-subset BCE scores decrease. Exact arithmetic preserves argmin sets; floating-point roundoff can switch the selected path near a tie. For example, Study B Digits Mean-MSE has a 37.9% reduction in mean Brier score. Some threshold-trained models cross frequently yet have much smaller reductions. Appendix C reports all 18 groups, the finite-loss BCE subset and roundoff checks. This measures the extent of the proven separation in fitted models, without changing the original comparisons.

5.4 How training and selection change the comparison

Within B, fixed-step and validation-selected readings also differ: ReLU finishes nearly tied but selection favors Brier; Digits favors Brier at the final step and BCE after selection, without a corrected primary difference (Table 12).

Training choices change relative error. Prior Study B results guided a learning-rate contrast on ReLU and a training-duration contrast on quadratic (Appendix B.4). We evaluated these fixed contrasts on 20 new teachers per task, pairing the three losses on the same teacher, data and minibatch stream. Figure 2 shows the observed mean BCE–Brier difference crossing zero in both comparisons; the two tests of the paired change pass their planned Holm correction. Mean regret changes in the same directions descriptively. Appendix B.5 gives the complete protocol and tests.

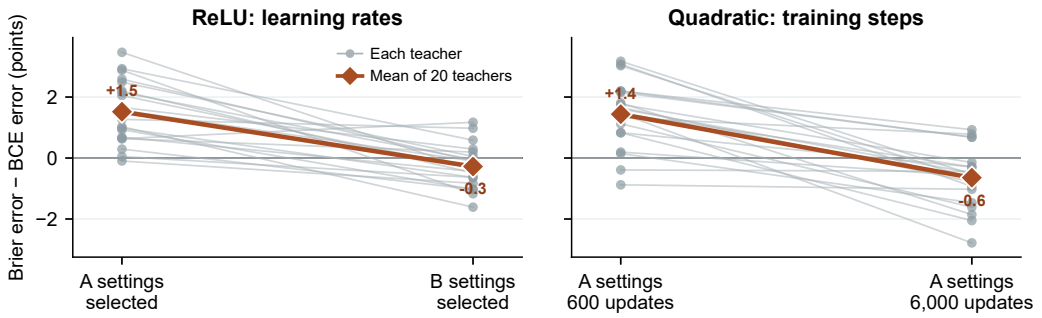

Figure 2: **TALC's training check detects two changes in relative error on fresh teachers.** Grey lines pair 20 new teachers; orange diamonds show their means, labeled in percentage points. Above zero favors BCE; below zero favors Brier. ReLU changes BCE/Brier learning rates from .006/.02 to .002/.006 at a common 6,000-update budget and validation rule; quadratic changes reading time on one A-setting trajectory. Both paired changes pass the planned two-test Holm correction; the observed means cross zero. Neither endpoint ranking has its own significance test.

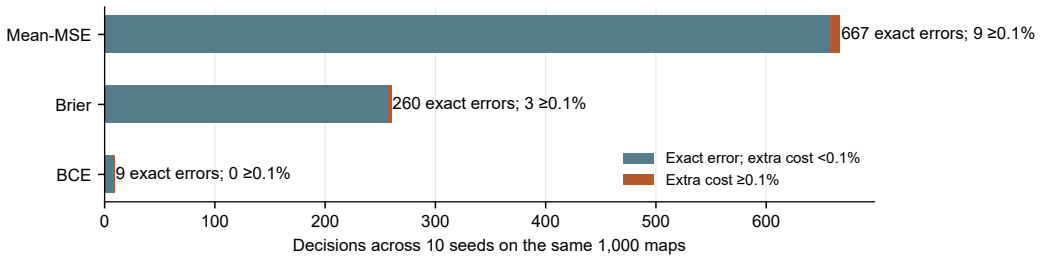

Figure 3: **TALC's cost audit: most exact path errors add less than 0.1% stored cost.** Bars and text count nonoptimal decisions across ten training seeds on the same fixed set of 1,000 maps. The orange portion meets the exploratory 0.1% relative-regret threshold; the blue portion does not. The same 1,000 maps recur over ten seeds.

5.5 Study C: Warcraft and the cost of an error

On standard 12×12 Warcraft maps (Vlastelica et al., 2020; Vlastelica, 2021), a shared ResNet18 feature extractor and cumulative head predict terrain costs; an eight-neighbor solver chooses routes. Both threshold losses use the same gap weights for the five uneven cost levels. We use the original 10,000/1,000/1,000 split, three development seeds and learning rates per loss, then ten new paired seeds with common early stopping and selection (Appendix D). Exact error and regret use costs stored in the released files; relative regret divides by optimal stored cost. A separate post-hoc nominal interpretation maps costs to one decimal.

BCE has lower exact error than Brier, which has lower exact error than Mean-MSE; both planned pairwise comparisons favor the same method in all ten paired seeds and pass their two-test Holm correction. Most errors add less than 0.1% optimal cost (Figure 3); nearly all disappear under nominal costs

Table 5: **Exact path error depends on the cost interpretation.** Stored-cost exact error is mean ± SD across ten paired seeds. Nominal exact error recomputes path costs at one-decimal levels as a post-hoc sensitivity check. Mean regret is extra stored cost per map.

Objective	Stored exact error (%)	Nominal exact error (%)	Mean stored regret
Mean-MSE	6.67 ± 1.84	0.09	1.96×10^{-4}
Brier	2.60 ± 2.00	0.03	5.87×10^{-5}
BCE	0.09 ± 0.07	0.00	8.79×10^{-7}

(Table 5). Even archived route labels are stored-cost suboptimal on 636/61/61 train/validation/test maps, yet optimal after nominal remapping. Thus exact-error rankings depend on cost interpretation. Appendix D gives the audit and tests; Figure 7 shows the full cost-tolerance curve.

6 Related work

Relation to classification-based prediction. Farebrother et al. (2024) report gains from categorical training of bootstrapped RL values, motivating our question about target choice and scoring. Our observed edge costs and separately predicted thresholds differ from that setting. Binary threshold outputs already appear in ordinal prediction (Niu et al., 2016); distributional regression can improve optimization (Imani & White, 2018), and Imani et al. (2026, Sections 4.4–4.5) compare ℓ_2+Softmax and ℓ_2+Softmax+Mean: squared losses on individual bin-weighted contributions versus their sum. Their softmax outputs and scalar targets differ from our independent binary thresholds. TALC pairs target controls with a Brier–BCE contrast and regret analysis. Batch-RL analyses also link value-learning losses to policy costs (Ayoub et al., 2024; 2025).

Scoring the same threshold targets. Brier and binary log loss are probability scores whose expected value is minimized by the true binary probability (Brier, 1950; Gneiting & Raftery, 2007). When threshold probabilities decrease in the required order, their squared-error sum is related to the ranked probability score (Epstein, 1969). Our separately predicted outputs need not obey that order, so we call both losses *threshold-wise*. CORAL enforces an order using a shared score (Cao et al., 2020), while CORN uses conditional probabilities (Shi et al., 2023); both change the output structure studied here.

Prediction loss and downstream decisions. SPO+ uses optimization structure to define a decision-oriented surrogate (Elmachtoub & Grigas, 2022); Liu & Grigas (2021) establish its risk bounds and calibration. Squared prediction risk already controls decision risk through feasible-set geometry (Ho-Nguyen & Kılınç-Karzan, 2022); our regret corollary specializes this result to the threshold decoder. Our analysis combines the weighted variance decomposition with cost-preserving isotonic post-processing to distinguish threshold-score improvement from decision improvement. Black-box differentiation propagates through a solver (Vlastelica et al., 2020). Appendix E compares complete decision-focused training recipes; TALC holds the head and decoder fixed to isolate the supervised loss.

Scope of the evidence. Loss rankings depend on the tested procedures; A and B vary several factors together, while fresh teachers isolate specified changes on two generators. Some B selections reach the training cap. Digits reuses one image pool without writer separation; Warcraft uses fixed maps and an alternative nominal-cost interpretation. The predicted duration reversal did not transfer to battery scheduling: the registered interaction was 0.022 GBP/day (95% interval $[-0.181, 0.203]$; Appendix F). Its unequal cost gaps and uniform threshold weights also differ from our gap-weighted score identity. TALC supplies an audit, not a validated cross-task loss-selection rule.

7 Conclusion

TALC separates threshold supervision, decoder-visible error and decision cost. Its decomposition and cost-preserving projection motivate matched controls; the 300-model audit measures their implications in fitted models. Fresh-teacher experiments confirm specified changes in relative loss performance, while the predicted battery reversal fails to transfer. Warcraft shows that large exact-error gaps can have small cost consequences. The resulting audit identifies what improves, under which training choices, and with what decision benefit.

REPRODUCIBILITY STATEMENT

The anonymous supplement includes manuscript sources and eight evidence/source folders under `research_artifact/`. `review_evidence/` contains frozen Study A/B protocols; `revision_evidence/` contains per-unit diagnostics, Warcraft threshold counts and the exploratory bridge. `independent_confirmation/` contains the new-teacher protocol fixed before evaluation, all unit summaries, 240 training receipt hashes, and a portable exact-test verifier. `decision_chain/` records the exploratory residual-to-decision audit. `prediction_replay/` holds saved A/B and Warcraft predictions and a standalone verifier for path choices, stored-cost regret and the nominal-cost sensitivity. `battery_evidence/` supplies the transfer protocol, sealed receipts, per-day outcomes, the frozen implementation and deterministic replay of the 20,000-resample block bootstrap. `probability_audit/` supplies all 300 A/B threshold probability arrays, original decoded costs, labels and paths, plus a standalone NumPy verifier for the residual and projection analyses. Extra fixed-checkpoint regret summaries are reaggregated from saved per-run metrics; their per-sample predictions are omitted to meet the upload limit. `training_source/` provides the A/B and fresh-teacher data generators, models, training/selection controllers, dependencies, small Digits source arrays and isolated-workspace instructions. Appendices B.1–E give the complete comparison families and designs. Saved-output replay, source-hash checks, imports and no-update smoke checks were performed for this packaging audit; the complete experimental training matrix was not rerun. Earlier engineering replays are described separately. Battery raw-price data must be acquired separately; the package does not supply all historical Warcraft or baseline training dependencies. Historical, exploratory and new-teacher evidence are labeled separately.

ETHICS STATEMENT

Digits uses public UCI images with attribution and disjoint source identities within each split, but reuses the image pool and does not establish writer separation. Warcraft uses the attributed released terrain-map benchmark. Both are constructed routing evaluations, not operational deployments.

AI USE STATEMENT

Generative AI tools assisted with study design, implementation of experimental and verification code, analysis and interpretation of results, literature identification and checking, and drafting and revising manuscript text. They also assisted with mathematical derivations, proof development and explanations, scientific figure design and revision, and pre-submission feedback. The authors are responsible for the final text, claims, code and results.

REFERENCES

E. Alpaydin and C. Kaynak. Optical recognition of handwritten digits. UCI Machine Learning Repository, 1998. URL https://doi.org/10.24432/C50P49.

Alex Ayoub, Kaiwen Wang, Vincent Liu, Samuel Robertson, James McInerney, Dawen Liang, Nathan Kallus, and Csaba Szepesvári. Switching the loss reduces the cost in batch reinforcement learning. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 2135–2158, 2024. URL https://proceedings.mlr.press/v235/ayoub24a.html.

Alex Ayoub, David Szepesvári, Alireza Bakhtiari, Csaba Szepesvári, and Dale Schuurmans. Rectifying regression in reinforcement learning. *Reinforcement Learning Journal*, 6:1440–1454, 2025. URL https://rlj.cs.umass.edu/2025/papers/Paper154.html.

Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950. doi: 10.1175/1520-0493(1950)078⟨0001:VOFEIT⟩2.0.CO;2.

Wenzhi Cao, Vahid Mirjalili, and Sebastian Raschka. Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognition Letters*, 140:325–331, 2020. doi: 10.1016/j.patrec.2020.11.008.

Jan de Leeuw, Kurt Hornik, and Patrick Mair. Isotone optimization in R: Pool-adjacent-violators algorithm (PAVA) and active set methods. *Journal of Statistical Software*, 32(5):1–24, 2009. doi: 10.18637/jss.v032.i05.

Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1): 9–26, 2022. doi: 10.1287/mnsc.2020.3922.

Edward S. Epstein. A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, 8(6):985–987, 1969. doi: 10.1175/1520-0450(1969)008⟨0985:ASSFPF⟩2.0.CO;2.

Jesse Farebrother, Jordi Orbay, Quan Vuong, Adrien Ali Taiga, Yevgen Chebotar, Ted Xiao, Alex Irpan, Sergey Levine, Pablo Samuel Castro, Aleksandra Faust, Aviral Kumar, and Rishabh Agarwal. Stop regressing: Training value functions via classification for scalable deep RL. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 13049–13071, 2024. URL https://proceedings.mlr.press/v235/farebrother24a.html.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437. URL https://sites.stat.washington.edu/people/raftery/Research/PDF/Gneiting2007jasa.pdf.

Nam Ho-Nguyen and Fatma Kılınç-Karzan. Risk guarantees for end-to-end prediction and optimization processes. *Management Science*, 68(12):8680–8698, 2022. doi: 10.1287/mnsc.2022.4321. URL https://pubsonline.informs.org/doi/10.1287/mnsc.2022.4321.

Ehsan Imani and Martha White. Improving regression performance with distributional losses. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2157–2166, 2018. URL https://proceedings.mlr.press/v80/imani18a.html.

Ehsan Imani, Kai Luedemann, Sam Scholnick-Hughes, Esraa Elelimy, and Martha White. Investigating the histogram loss in regression. *Journal of Machine Learning Research*, 27(66):1–54, 2026. URL https://www.jmlr.org/papers/v27/24-0260.html.

Alexander I. Jordan, Anja Mühlemann, and Johanna F. Ziegel. Characterizing the optimal solutions to the isotonic regression problem for identifiable functionals. *Annals of the Institute of Statistical Mathematics*, 74(3):489–514, 2022. doi: 10.1007/s10463-021-00808-0. URL https://link.springer.com/article/10.1007/s10463-021-00808-0.

Heyuan Liu and Paul Grigas. Risk bounds and calibration for a smart predict-then-optimize method. In *Advances in Neural Information Processing Systems*, volume 34, pp. 22083–22094, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/b943325cc7b7422d2871b345bf9b067f-Abstract.html.

Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. Ordinal regression with multiple output cnn for age estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4920–4928, 2016. URL https://openaccess.thecvf.com/content_cvpr_2016/html/Niu_Ordinal_Regression_With_CVPR_2016_paper.html.

Open Power System Data. Data package: Time series (2020-10-06), 2020. URL https://data.open-power-system-data.org/time_series/2020-10-06/. GB_GBN_price_day_ahead field; accessed 2026-09-25.

scikit-learn developers. load_digits: Handwritten digits dataset. scikit-learn documentation, n.d. URL https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_digits.html. Accessed 2026-09-25.

Xintong Shi, Wenzhi Cao, and Sebastian Raschka. Deep neural networks for rank-consistent ordinal regression based on conditional probabilities. *Pattern Analysis and Applications*, 26:941–955, 2023. doi: 10.1007/s10044-023-01181-9. URL https://link.springer.com/article/10.1007/s10044-023-01181-9.

Marin Vlastelica. Differentiation of blackbox combinatorial solvers. Edmond research dataset, version 1, 2021. https://doi.org/10.17617/3.YJCQ5S.

Marin Vlastelica, Anselm Paulus, Vít Musil, Georg Martius, and Michal Rolínek. Differentiation of blackbox combinatorial solvers. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=BkevoJSYPB.

A MATHEMATICAL DETAILS

A.1 MATCHED OBJECTIVES AND POPULATION TARGETS

For the equally spaced alphabet in Section 3, $Y = \sum_{k=1}^{m} t_k$ pointwise, so $(\widehat{c} - C)/(sm) = m^{-1}\sum_k (q_k - t_k) = \bar{e}$. Expanding the squares yields

$$\frac{1}{m}\sum_k (e_k - \bar{e})^2 = \frac{1}{m}\sum_k e_k^2 - \bar{e}^2,$$

which proves Equation (5). For arbitrary positive increments, $C_e = a_{e,0} + \sum_k \Delta_{e,k} t_{e,k}$ pointwise. With $w_{e,k} = \Delta_{e,k}/R_e$ and $\bar{e}_e = \sum_k w_{e,k} e_{e,k}$, the same expansion gives $\sum_k w_{e,k} e_{e,k}^2 = \bar{e}_e^2 + \sum_k w_{e,k}(e_{e,k} - \bar{e}_e)^2$. This includes Warcraft's objectives exactly.

For one threshold with conditional success probability η,

$$\mathbb{E}[(q - t)^2 \mid x] = (q - \eta)^2 + \eta(1 - \eta).$$

Its conditional excess is $(q - \eta)^2$. Binary log-loss minus binary entropy is $D_{\mathrm{KL}}(\mathrm{Ber}(\eta) \| \mathrm{Ber}(q))$. Log-loss uses natural logarithms and extended values at zero and one, with $0 \log 0 = 0$. Summing these identities with positive weights proves the threshold excess-risk formulas. Dependence among threshold labels or cost coordinates does not affect linearity of expectation. Each proper score has the unique unrestricted optimum $q_{e,k} = \eta_{e,k}$. A boundary optimum is approached by finite logits when it is not attained by them.

Taking conditional expectation of the alphabet identity gives $\mu_e = a_{e,0} + \sum_k \Delta_{e,k}\eta_{e,k}$. The variance identity

$$\mathbb{E}\left[\frac{(\widehat{c}_e - C_e)^2}{R_e^2}\middle| x\right] = \frac{(\widehat{c}_e - \mu_e)^2}{R_e^2} + \frac{\mathrm{Var}(C_e \mid x)}{R_e^2}$$

therefore identifies $\mathcal{M}$ with Mean-MSE excess summed over coordinates. Its unrestricted minimizers are exactly the probability vectors satisfying $\sum_k \Delta_{e,k} q_{e,k} = \sum_k \Delta_{e,k}\eta_{e,k}$ for every e. The closed-cube solution can be nonunique; it is unique at the two extreme means and when $m_e = 1$.

A.2 WEIGHTED DECODER DECOMPOSITION

Proof of Proposition 1. For each coordinate, $\sum_k w_{e,k} = 1$ and $\sum_k w_{e,k}(\delta_{e,k} - h_e) = 0$. Expanding around h_e gives

$$\sum_k w_{e,k}\delta_{e,k}^2 = h_e^2 + \sum_k w_{e,k}(\delta_{e,k} - h_e)^2.$$

Summing over e proves Equation (6). In particular the residual vector $\delta_{e,k} - h_e$ contributes zero to the decoded error because $\Delta_{e,k} = R_e w_{e,k}$. □

Decoder-visible does not mean decision-sensitive. Decoded errors orthogonal to $\mathrm{span}(\mathcal{P} - \mathcal{P})$ leave all action-cost differences unchanged; more generally, an error can preserve the minimizing action without preserving every difference. For example, take $\mathcal{P} = \{(1, 0), (0, 1)\}$, alphabet $\{1, 3\}$ on each coordinate, $\eta = (.2, .5)$ and $q = (.4, .7)$. Then $\mu = (1.4, 2)$ and $\widehat{c} = (1.8, 2.4)$ have the same unique minimizer and the same cost difference, although $\mathcal{M} = .08$. Thus $\mathcal{M}$ is a sufficient control quantity, not a necessary measure of decision change. This does not weaken its upper bound.

A.3 COST-PRESERVING IMPROVEMENT OF BOTH THRESHOLD SCORES

Weighted isotonic regression and pool-adjacent-violators are standard (de Leeuw et al., 2009). More generally, isotonic fits can simultaneously minimize loss families eliciting the same functional (Jordan et al., 2022). Here we use the gap weights to preserve a particular decoder and prove score improvement against every threshold label vector at fixed predictions.

Proof of Proposition 2. Fix one coordinate and suppress its index. The order cone $\mathcal{C} = \{r \in \mathbb{R}^m : r_1 \geq \cdots \geq r_m\}$ is closed, convex and nonempty. Positive weights make the squared objective coercive and strictly convex, so its minimizer r exists and is unique. Clipping every coordinate to

$[0,1]$ preserves order and reduces squared distance to $q \in [0,1]^m$ if any coordinate is outside that interval. Thus $r \in [0,1]^m$. Every translate $r + c\mathbf{1}$ is in $\mathcal{C}$; differentiating the objective along this line at $c = 0$ gives $\sum_k w_k(r_k - q_k) = 0$. Since $\Delta_k = Rw_k$, decoding is preserved.

For every $t \in \mathcal{C}$, the line segment $r + u(t - r)$, $0 \le u \le 1$, is feasible. Nonnegativity of its objective's right derivative at $u = 0$ gives $\sum_k w_k(q_k - r_k)(t_k - r_k) \le 0$. Expanding the square in $q - t = (q - r) + (r - t)$ now yields

$$\sum_k w_k(q_k - t_k)^2 \ge \sum_k w_k(r_k - t_k)^2 + \sum_k w_k(q_k - r_k)^2.$$

Every threshold label vector is decreasing, so this proves the samplewise Brier assertion, with strict improvement whenever $q \notin \mathcal{C}$. The same inequality applies to the decreasing population vector η.

For BCE, use the weighted pool-adjacent-violators algorithm, which computes this unique projection for positive weights on a total order (de Leeuw et al., 2009). At a merge, adjacent blocks have fitted values $a < b$ and weights $u, v > 0$, and their new common value is $c = (ua + vb)/(u + v) \in (a, b)$. Let α and β be the weighted average labels on the left and right blocks. Decreasing labels imply $1 \ge \alpha \ge \beta \ge 0$. Write $\ell(z, t) = -t \log z - (1 - t) \log(1 - z)$. The loss decrease

$$D(\alpha, \beta) = u[\ell(a, \alpha) - \ell(c, \alpha)] + v[\ell(b, \beta) - \ell(c, \beta)]$$

is affine in (α, β), with

$$D(\alpha, \beta) = (1 - \alpha)D(0, 0) + (\alpha - \beta)D(1, 0) + \beta D(1, 1).$$

For $0 < a < b < 1$, strict convexity of $-\log(1 - z)$ and $-\log z$ gives $D(0, 0) > 0$ and $D(1, 1) > 0$ by weighted Jensen. Also $D(1, 0) = u \log(c/a) + v \log((1 - c)/(1 - b)) > 0$. The three coefficients are nonnegative and sum to one; each violating merge therefore strictly decreases BCE. For $q \in (0, 1)^m$, all pooled values remain interior. The finite sequence of merges gives the final projection and proves strict BCE improvement whenever a merge is needed. This argument also applies to decreasing conditional probabilities in place of realized labels, establishing the population version. □

These are changes to threshold scores at fixed decoded costs. Every linear action cost and its entire minimizing set is identical before and after projection; fixed tie-breaking therefore selects the same action in exact arithmetic. For unequal gaps, an unweighted projection need not preserve the decoder: $q = (.2, .8)$ with gaps $(1, 2)$ decodes an increment 1.8, whereas its unweighted projection $(.5, .5)$ decodes 1.5. Gap-weighted pooling gives $(.6, .6)$ and retains 1.8. The result concerns post-processing, not training with a constrained head. Training and the original loss comparisons use unprojected predictions. Floating-point re-summation can alter the last bits of near-tied path costs, so a numerical projection audit must distinguish these changes from the exact-arithmetic invariance.

A.4 Regret transfer and sharpness

All statements condition on an input containing the known graph and action set. The alphabets are finite and strictly increasing, so each $R_e > 0$. A constant-cost coordinate can be kept as a known offset in the action objective and omitted from the error sums. The finite, nonempty action set ensures that predicted and Bayes minima exist.

Relation to existing squared-risk guarantees. Ho-Nguyen & Kılınç-Karzan (2022) bound excess decision risk by squared prediction excess and the feasible-set diameter. Their squared loss is $\ell(d, c) = \|d - c\|_2^2$, without a factor $1/2$. To specialize their result, set

$$D_R = \operatorname{diag}(R_e), \quad \widetilde{C} = D_R^{-1} C, \quad \widetilde{c} = D_R^{-1} \widehat{c}, \quad \mathcal{X} = \operatorname{conv}(D_R \mathcal{P}).$$

Use zero deterministic objective term, so $\widetilde{c}^\top D_R p = \widehat{c}^\top p$. The convex hull is compact, preserves linear minima and has squared diameter H. Its selector may use the original finite-action tie rule; convexification does not require fractional solver outputs. Finite alphabets ensure bounded support and square integrability. With the covariate fixed at x, squared-loss excess is precisely $\|\widetilde{c} - D_R^{-1}\mu\|_2^2 = \mathcal{M}$. Their decision loss is relative to the realized oracle, but its excess over unrestricted optimal risk is

$$\left(\mu^\top \widehat{p} - \mathbb{E}[\min_p C^\top p \mid x]\right) - \left(\min_p \mu^\top p - \mathbb{E}[\min_p C^\top p \mid x]\right) = \mathcal{R}_\mu.$$

Hence, for $H > 0$, their squared-risk guarantee gives $\mathcal{R}_\mu^2 \le H\mathcal{M}$, with no additional factor. This mapping credits the regret-transfer step, rather than the threshold-space decomposition or projection. The proof below also verifies sharpness for positive binary costs.

Proof of Corollary 1. Set $d_e = \widehat{p}_e - p_{\mu,e}$. Predicted optimality gives $\widehat{c}^\top d \le 0$, while Bayes optimality gives $\mu^\top d \ge 0$. Consequently,

$$0 \le \mathcal{R}_\mu \le (\mu - \widehat{c})^\top d = -\sum_e R_e d_e h_e$$
$$\le \left(\sum_e R_e^2 d_e^2\right)^{1/2} \left(\sum_e h_e^2\right)^{1/2} \le \sqrt{H\mathcal{M}}.$$

The middle inequality is Cauchy–Schwarz. Nonnegativity of $\mathcal{V}$ gives the next bound. Bernoulli Pinsker with natural logarithms states $2(q-\eta)^2 \le D_{\mathrm{KL}}(\mathrm{Ber}(\eta) \| \mathrm{Ber}(q))$. Apply it coordinatewise with the positive weights and sum to obtain $2\mathcal{E}_{\mathrm{Br}} \le \mathcal{E}_{\mathrm{BCE}}$. Infinite KL gives an infinite upper bound when $H > 0$. If $H = 0$, $R_e > 0$ implies that all feasible actions are the same vector, so regret is zero directly, without forming $0 \cdot \infty$.

For sharpness, take two parallel one-edge routes with actions $(1,0)$ and $(0,1)$ and cost alphabets $\{1,2\}$. Let their conditional upper-cost probabilities be $1/2 - \varepsilon$ and $1/2 + \varepsilon$, where $0 < \varepsilon < 1/2$. Predict $q_1 = q_2 = 1/2$, assigning the higher true mean to the route selected by the fixed tie rule. Then

$$H = 2, \quad \mathcal{M} = \mathcal{E}_{\mathrm{Br}} = 2\varepsilon^2, \quad \mathcal{R}_\mu = 2\varepsilon = \sqrt{H\mathcal{M}}.$$

Thus the unit constant in the squared-error bounds is attained. Writing $f(\varepsilon) = D_{\mathrm{KL}}(\mathrm{Ber}(1/2 + \varepsilon) \| \mathrm{Ber}(1/2))$, differentiation gives $f(0) = f'(0) = 0$ and $f''(0) = 4$. Hence $\mathcal{E}_{\mathrm{BCE}} = 2f(\varepsilon) = 4\varepsilon^2 + o(\varepsilon^2)$, and $\mathcal{R}_\mu^2/(H\mathcal{E}_{\mathrm{BCE}}) \to 1/2$. No smaller universal coefficient can hold. A prediction perturbation of order $o(\varepsilon)$ toward the worse route makes its predicted optimum unique and retains both limiting ratios. □

Other positive threshold weights. The result also has an exact form when the training weights are not proportional to the gaps. For arbitrary $\omega_{e,k} > 0$, put

$$S_e = \sum_k \frac{\Delta_{e,k}^2}{\omega_{e,k}}, \quad b_e = \widehat{c}_e - \mu_e, \quad v_{e,k} = \delta_{e,k} - \frac{b_e \Delta_{e,k}}{S_e \omega_{e,k}}.$$

Then $\sum_k \Delta_{e,k} v_{e,k} = 0$ and the cross term vanishes in the weighted square expansion, giving the exact orthogonal decomposition

$$\sum_{e,k} \omega_{e,k} \delta_{e,k}^2 = \sum_e \frac{b_e^2}{S_e} + \sum_{e,k} \omega_{e,k} v_{e,k}^2.$$

Cauchy–Schwarz applied to $\sum_e d_e b_e$ gives

$$\mathcal{R}_\mu \le \sqrt{H_\omega \sum_e b_e^2/S_e}, \qquad H_\omega = \max_{p,r \in \mathcal{P}} \sum_e S_e (p_e - r_e)^2.$$

This is bounded by the corresponding weighted Brier bound and by $\sqrt{H_\omega \mathcal{E}_{\mathrm{BCE},\omega}/2}$ when $H_\omega > 0$. For $\omega_{e,k} = \Delta_{e,k}/R_e$, $S_e = R_e^2$ and this gives Proposition 1 and Corollary 1. For $\omega_{e,k} = 1$, it recovers the unweighted sum-risk constant $D = \max_{p,r} \sum_{e,k} (p_e - r_e)^2 \Delta_{e,k}^2$. When an objective averages the main-text risks over d cost coordinates, the corresponding bound is $\sqrt{Hd\,\overline{\mathcal{E}}_{\mathrm{Br}}}$ or $\sqrt{Hd\,\overline{\mathcal{E}}_{\mathrm{BCE}}/2}$. For equal threshold averages $\omega_{e,k} = 1/m_e$, $S_e = m_e \sum_k \Delta_{e,k}^2$. If Mean-MSE instead uses a scale $\sigma_e > 0$, its conditional excess is b_e^2/σ_e^2 and the decoded term in this decomposition is $(\sigma_e^2/S_e)(b_e^2/\sigma_e^2)$. Thus an unequal-gap decoder or a different Mean-MSE scale need not give a unit coefficient.

Positive gaps and averaging over inputs. Let $F(\mu) = \arg\min_{p\in\mathcal{P}} \mu^\top p$. If a nonoptimal action exists, finiteness gives $\gamma(x) = \min_{p\notin F(\mu)} \mu^\top(p - p_\mu) > 0$. On a nonoptimal choice, $\mathcal{R}_\mu \geq \gamma(x)$, so

$$\mathbf{1}\{\widehat{p} \notin F(\mu)\} \leq \frac{H\mathcal{M}}{\gamma(x)^2}, \qquad \mathcal{R}_\mu \leq \frac{H\mathcal{M}}{\gamma(x)}.$$

The first follows by squaring the regret bound; the second uses $\mathcal{R}_\mu \leq \mathcal{R}_\mu^2/\gamma(x)$ on a nonoptimal choice. Both have zero left sides on optimal choices. If all actions are optimal, the left sides are zero and no gap is needed. The same expressions with $\mathcal{E}_{\mathrm{Br}}$ or $\mathcal{E}_{\mathrm{BCE}}/2$ in place of $\mathcal{M}$ also hold. For deterministic costs these concern true-cost optimality; for stochastic costs they concern Bayes mean-cost optimality. They allow tied optima and assert membership in the optimal set.

For a fixed graph, Jensen's inequality gives $\mathbb{E}\mathcal{R}_\mu(X) \leq \sqrt{H\mathbb{E}\mathcal{M}(X)}$. If graphs or alphabets vary, retain $\mathbb{E}\sqrt{H(X)\mathcal{M}(X)}$, or use a uniform finite $H_{\max}$ to obtain $\sqrt{H_{\max}\mathbb{E}\mathcal{M}(X)}$. A uniform positive gap similarly yields the linear bound $\mathbb{E}\mathcal{R}_\mu \leq H_{\max}\mathbb{E}\mathcal{M}/\gamma_{\min}$. These are population excess-risk relations; no generalization or optimization bound is assumed for the fitted networks.

Continuous costs and finite encoding. The same argument separates prediction error from encoding error. Let C have finite conditional first moments, and fix coordinatewise encodings $g_e(C_e)$ into the finite alphabets above. Define $\eta^g_{e,k} = \Pr(g_e(C_e) \geq a_{e,k} \mid x)$, $\mu^g = \mathbb{E}[g(C) \mid x]$, $b = \mu - \mu^g$, and $\mathcal{M}_g = \sum_e (\sum_k w_{e,k}(q_{e,k} - \eta^g_{e,k}))^2$. For p_μ optimal under the original mean costs μ and $\widehat{p}$ optimal under $\widehat{c}$, predicted optimality gives

$$\mu^\top(\widehat{p} - p_\mu) \leq \sqrt{H\mathcal{M}_g} + \mathrm{osc}_{\mathcal{P}}(b), \qquad \mathrm{osc}_{\mathcal{P}}(b) = \max_{p,r\in\mathcal{P}} b^\top(p - r).$$

Indeed, split $\mu - \widehat{c} = (\mu^g - \widehat{c}) + b$; the first term has the previous Cauchy–Schwarz bound for the action difference $\widehat{p} - p_\mu$, and the second is at most the displayed oscillation. If $\mathbb{E}[|C_e - g_e(C_e)| \mid x] \leq \rho_e(x)$, this oscillation is at most $\max_{p,r} \sum_e |p_e - r_e|\rho_e(x)$. The decomposition and threshold-score bounds apply to the encoded labels with η^g. Original-cost regret additionally includes this encoding term; the finite grid alone does not bound clipping tails.

Expected cost and realized error. When costs remain stochastic after observing x, expected realized regret is $\mathcal{R}_\mu$ plus $\min_p \mu^\top p - \mathbb{E}[\min_p C^\top p \mid x]$, a nonnegative term independent of the predictor. Thus mean-cost optimality need not minimize the probability of a realized error. For example, a route costing one with probability .6 and 101 otherwise competes with a constant cost-two route. The latter is Bayes optimal with error .6 and expected realized regret .6; the former has error .4 and expected realized regret 39.6.

B Details of the two matched studies

Table 6: **TALC's two controlled axes and their combined comparison.** Each cell names the method with lower path-error rate after correction within that study; n.s. means no established direction, and — means untested. A and C use .05 families; B uses its prespecified .025 family. A's synthetic endpoint uses tied contexts; B uses all contexts, so cell signs summarize within-study tests rather than a pooled effect.

	A: 600 updates			B: validation-selected			C
Comparison	ReLU	Quad.	Digits	ReLU	Quad.	Digits	Warcraft
Target: Brier vs Mean-MSE	n.s.	n.s.	Brier	Brier	Brier	n.s.	Brier
Score: BCE vs Brier	BCE	BCE	BCE	Brier	n.s.	n.s.	BCE
Both: BCE vs Mean-MSE	BCE	BCE	BCE	n.s.	BCE	n.s.	—

B.1 Study A: fixed short budget

Two synthetic cost-prediction tasks. Each seed defines an 11-node, 27-edge directed acyclic graph (DAG) with shared subpaths and three skip edges. The network sees an eight-dimensional Gaussian context and predicts the edge costs. The labels are positive integers generated by a nonlinear function of that context. One teacher uses 32 centered and scaled random ReLU features; the

other uses products $(u_e^\top x)(v_e^\top x)$ of orthogonal projections. Both bin their scores at $-.5, .5$ into a three-level penalty added to the edge's layer increment. Train, validation and evaluation sets contain 512, 512 and 2,048 independent contexts. The primary error rate uses contexts with multiple true optimal paths. This set-valued decision case checks whether a model selects *any* valid optimum, rather than matching an arbitrary path label. Such ties arise naturally in about 61% of ReLU and 64% of quadratic evaluation contexts across the 20 seeds; no context is resampled to create them. Unique-optimum and all-context results are recorded separately in `revision_evidence/study_a_all_context_diagnostics.csv`, using the `unique` and `all` strata; the PDF does not contain a separate unique-optimum table.

A path task built from digit images. A board contains 4×4 handwritten digits, and a vertex containing digit d has true cost $1 + d$. A shared network predicts those costs from normalized pixels. The solver chooses among 20 right/down paths, each containing seven vertices. We split the 1,797 images in scikit-learn's `load_digits` subset of UCI Optical Digits (Alpaydin & Kaynak, 1998; scikit-learn developers, n.d.) into disjoint pools of 1,078/359/360 source images, then sample with replacement to make 1,024/256/1,024 training, validation and evaluation boards. The primary endpoint uses *all* evaluation boards, unlike the tied-only synthetic endpoints.

The image pools and sampled boards stay fixed across its 20 evaluation seeds, which vary initialization and minibatch order. Study B instead uses new image partitions from the same source pool.

Registered analysis and recipes. Study A comprises 540 development runs and 180 evaluation runs. The learning-rate grid was expanded to include .02 before new outcomes were observed; it differs from the earlier 15-method study. Selection uses mean validation error across development seeds, then regret, then the smaller learning rate and weight decay. All objectives share initial parameters for a task and seed. Brier averages squared binary residuals without a factor of two; Mean-MSE divides the cost residual by the number of thresholds before squaring; BCE averages binary log loss over those same thresholds. No threshold sorting, monotonic adjustment, rounding, or alternative path decoder is used in the original comparisons. The separate post-hoc projection audit is reported in Appendix C.

Table 7: Selected learning rate / weight decay in Study A, using only five development seeds.

Objective	ReLU	Quadratic	Digits
BCE	0.006 / 0.0001	0.006 / 0.0001	0.02 / 0.0001
Brier	0.02 / 0.0001	0.006 / 0.0001	0.02 / 0.0001
Mean-MSE	0.006 / 0.0001	0.006 / 0.0001	0.02 / 0.0001

Table 8: All nine prespecified error contrasts in Study A. Negative first-minus-second differences favor the first method. Differences and intervals are percentage points. Marginal bootstrap intervals are not adjusted; Holm correction is applied jointly to the nine exact paired sign-flip tests.

Task	First − second	Difference (pp)	Marginal 95% CI	Exact p	Holm p
ReLU	Brier − BCE	1.576	[1.061, 2.163]	0.0000038	0.0000305
ReLU	Mean-MSE − BCE	1.435	[1.021, 1.824]	0.0000153	0.0001068
ReLU	Mean-MSE − Brier	-0.142	[-0.618, 0.290]	0.5785599	1.0000000
Quadratic	Brier − BCE	1.024	[0.361, 1.669]	0.0087547	0.0262642
Quadratic	Mean-MSE − BCE	0.833	[0.306, 1.332]	0.0061531	0.0246124
Quadratic	Mean-MSE − Brier	-0.191	[-0.952, 0.578]	0.6386070	1.0000000
Digits	Brier − BCE	2.056	[1.172, 3.018]	0.0001831	0.0009155
Digits	Mean-MSE − BCE	5.225	[4.697, 5.762]	0.0000019	0.0000172
Digits	Mean-MSE − Brier	3.169	[2.144, 4.023]	0.0000496	0.0002975

The statistical random seed is 9349026, with a separate bootstrap stream for each contrast in the fixed task/contrast order. Synthetic inference averages seed-specific tied error rates equally rather than pooling their different denominators. The image board sample is fixed across the 20 evaluation

seeds, making its uncertainty conditional on those images and boards. No new inference uses the residual-variance diagnostic.

Table 9: Study A's exploratory threshold-residual diagnostic on the primary synthetic subsets, averaged equally over 20 seeds. Dispersion is the additional term in Equation (5), not total threshold error. Error differences and Holm p retain the planned nine-test family; these diagnostics are neither mediation nor equivalence tests.

	Residual dispersion		Error difference	Holm
Task	Mean-MSE	Brier	Mean-MSE − Brier (pp)	p
ReLU	0.114648	0.027062	-0.142	1.000
Quadratic	0.153007	0.029890	-0.191	1.000

Residual diagnostic and search endpoints. The additional residual component is smaller for Brier, but neither task establishes a decision advantage over Mean-MSE. All nine selected recipes use weight decay .0001; the learning-rate ceiling .02 is selected by ReLU Brier and all three Digits objectives. No convergence claim follows.

Figure 4 reports both comparison axes; all nine tests above remain the inferential record.

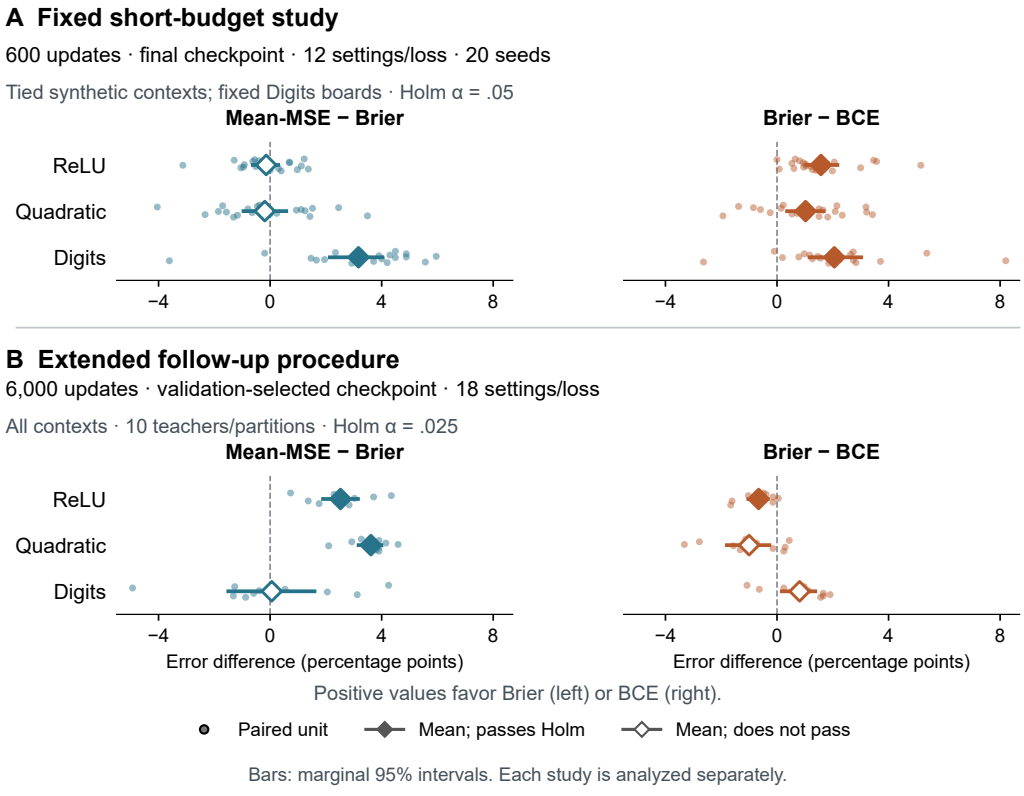

Figure 4: **Loss rankings are not preserved across the two study protocols.** Positive values favor Brier on the left and BCE on the right. Each circle is a paired analysis-unit difference; vertical offsets only separate points. Study A uses 20 seeds, tied synthetic contexts and fixed Digits boards. Study B uses all contexts and ten teachers or image partitions, averaging its two Digits initializations within each partition. Diamonds show means and bars are marginal 95% bootstrap intervals. Filled diamonds pass that study's nine-test Holm correction; open diamonds do not, even if a marginal interval excludes zero. The different endpoints, samples and selection procedures preclude interpreting the panels as an isolated effect of adding updates. All nine tests for each study are in Appendices B.1 and B.2.

B.2 STUDY B: EXTENDED TRAINING AND SELECTION PROCEDURE

The frozen follow-up contains 648 development and 120 confirmation fits. Its model, data-generating tasks, losses and ordinary shortest-path decoder match the earlier implementation. Synthetic splits remain 512/512/2,048 contexts. Each Digits partition has disjoint pools of 1,078/359/360 source images and 1,024/256/1,024 sampled boards. New random partitions support inference conditional on the fixed image pool; they are not new datasets or evidence of writer-separated generalization.

Selection and statistical family. Adam runs for 6,000 updates at a constant learning rate. The learning-rate grid is $\{.0001, .0005, .002, .006, .02, .06\}$, crossed with weight decay $\{0, .0001, .001\}$. Checkpoints at 600, 900, . . ., 6,000 are eligible; the minimum validation error, then regret, then earliest step determines each run's checkpoint. Development averages units equally and selects error, then regret, then smaller learning rate and weight decay. Selected ReLU learning rates are $.002/.006/.006$ for BCE/Brier/Mean-MSE; quadratic uses $.002$ for all three; Digits uses $.02/.02/.006$. Weight decay is $.0001$ except Digits Mean-MSE, which uses zero.

The family contains all three pairs on all three tasks. Exact two-sided sign-flip tests enumerate 2^{10} assignments with inclusive tails using integer error fractions; their null requires independent sign-symmetric paired differences. The Digits unit first averages its two paired optimization seeds. These nine comparisons use Holm correction at the prospectively allocated level $.025$. The supplied protocol records this long-budget family as round $r = 0$ of a predeclared sequence, with familywise level $.05/2^{r+1}$ for round r; unused levels are not recycled. The full rule is retained in the supplied protocol. The 20,000-draw unit bootstrap uses statistical seed 9469001 with a fixed contrast offset; its 95% intervals are marginal, not simultaneous.

Table 10: All nine longer-budget comparisons. Differences are first minus second in percentage points; negative favors the first. Decisions use Holm $p < .025$, not whether a marginal interval excludes zero.

Task	First minus second	Difference (pp)	95% CI (pp)	Holm p
ReLU	BCE – Brier	+0.659	$[+0.332, +1.030]$	0.023438
ReLU	BCE – Mean-MSE	-1.865	$[-2.607, -1.084]$	0.029297
ReLU	Brier – Mean-MSE	-2.524	$[-3.149, -1.899]$	0.017578
Quadratic	BCE – Brier	+1.001	$[+0.273, +1.797]$	0.140625
Quadratic	BCE – Mean-MSE	-2.612	$[-3.486, -1.684]$	0.017578
Quadratic	Brier – Mean-MSE	-3.613	$[-3.989, -3.174]$	0.017578
Digits	BCE – Brier	-0.811	$[-1.372, -0.176]$	0.140625
Digits	BCE – Mean-MSE	-0.869	$[-2.329, +0.723]$	0.640625
Digits	Brier – Mean-MSE	-0.059	$[-1.592, +1.499]$	0.951172

Table 11: Longer-budget true regret, averaged over the same ten units. Regret is descriptive. A lower nonoptimal rate need not imply lower regret.

Task	Mean-MSE	Brier	BCE
ReLU	0.132129	0.106396	0.115967
Quadratic	0.183496	0.147412	0.161621
Digits	0.190381	0.162646	0.162695

The registered strong-BCE criterion required BCE to beat both controls on at least two tasks by at least one percentage point and 10% relative error, with Holm $p < .025$ and no mean-regret increase on those tasks. It also required no adverse error difference greater than one point on the remaining task. Zero tasks qualified, so this criterion was not met. The narrower pairwise findings above retain their original interpretation.

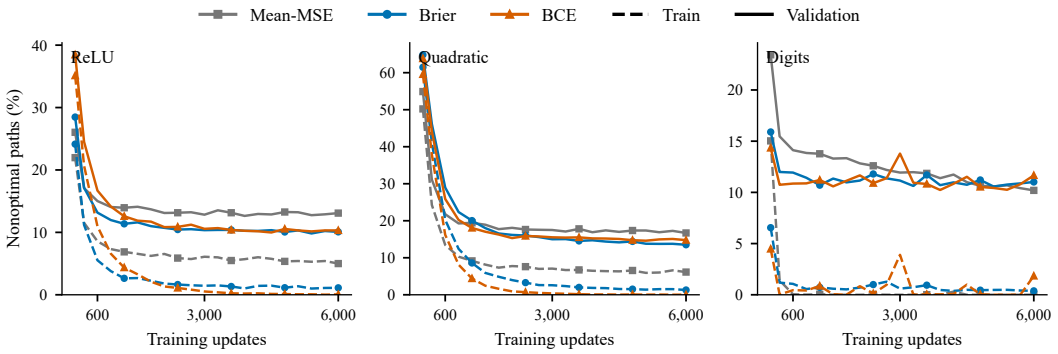

Figure 5: Error trajectories of the frozen confirmation recipes, averaged equally over ten teacher or partition units. Solid lines are validation; dashed lines are training. Digits first averages two initializations per partition. These are descriptive means, not test-set trajectories or proof of convergence.

Table 12: Predeclared secondary fixed-step test errors (%). These snapshots use the long-run selected recipes, not separate tuning for each budget. The validation-selected checkpoint remains the primary endpoint.

Task	Objective	600	1200	3000	6000	Selected
ReLU	Mean-MSE	14.48	14.13	13.14	12.95	12.63
ReLU	Brier	13.48	11.87	10.64	10.28	10.10
ReLU	BCE	16.59	12.84	10.89	10.27	10.76
Quadratic	Mean-MSE	21.53	18.59	17.44	17.09	17.20
Quadratic	Brier	28.61	19.68	14.97	13.55	13.58
Quadratic	BCE	25.70	17.64	14.79	14.12	14.58
Digits	Mean-MSE	12.05	11.65	10.59	8.62	8.50
Digits	Brier	9.59	9.09	8.85	8.67	8.44
Digits	BCE	8.13	7.97	10.59	8.93	7.63

Optimization context. Development decisions nearly fit the training data while validation errors remain nonzero. Quadratic Brier selects steps 6,000/6,000/5,700, so the cap does not establish convergence. The fixed-checkpoint table uses the same selected recipes, not a new search for each step budget. In particular, ReLU error at the fixed final step is 10.269% for BCE and 10.283% for Brier; the validation-selected primary results are 10.762% and 10.103%. This distinction concerns the entire selection procedure and does not justify substituting the final step after observing test outcomes.

B.3 Auxiliary views on common endpoints

Table 13 reports Study A on all evaluation contexts, recomputed from its saved final-model predictions. This supplements its tied-only synthetic primary endpoint without changing the original tests. It aligns the reported population with Study B's all-context endpoint; samples, configurations and selection procedures still differ.

Table 13: Study A all-context diagnostics over its original 20 seeds. Error is mean ± SD; regret is an equal-seed mean. These auxiliary results do not replace the registered primary comparisons.

Task	Objective	All-context error (%)	Mean regret
ReLU	BCE	12.351 ± 1.298	0.133057
ReLU	Brier	14.419 ± 1.424	0.154785
ReLU	Mean-MSE	14.014 ± 1.424	0.147754
Quadratic	BCE	17.976 ± 1.198	0.202246
Quadratic	Brier	19.661 ± 1.872	0.220532
Quadratic	Mean-MSE	19.380 ± 1.361	0.209204
Digits	BCE	9.429 ± 1.313	0.226318
Digits	Brier	11.484 ± 2.323	0.248047
Digits	Mean-MSE	14.653 ± 0.818	0.322803

Table 14: Mean regret for all three Study B objectives at the original fixed checkpoints and validation-selected endpoint. Units and recipes match the fixed-step error table; Digits averages two initializations within each partition before averaging partitions.

Task	Objective	600	1,200	3,000	6,000	Selected
ReLU	BCE	0.18198	0.13896	0.11743	0.10942	0.11597
ReLU	Brier	0.14355	0.12554	0.11187	0.10796	0.10640
ReLU	Mean-MSE	0.15347	0.14834	0.13750	0.13608	0.13213
Quadratic	BCE	0.31240	0.19829	0.16450	0.15635	0.16162
Quadratic	Brier	0.35464	0.22383	0.16479	0.14668	0.14741
Quadratic	Mean-MSE	0.23818	0.19897	0.18560	0.18125	0.18350
Digits	BCE	0.18306	0.17568	0.38545	0.20723	0.16270
Digits	Brier	0.18774	0.17549	0.17065	0.16177	0.16265
Digits	Mean-MSE	0.23833	0.23906	0.22080	0.19346	0.19038

Table 15: Exploratory Brier-minus-Mean-MSE differences from saved evaluation predictions, averaged equally over teacher, partition or optimization units. All rows use all contexts; Study A's registered synthetic endpoint remains tied-only. Dispersion is the extra threshold term in Equation (5); pathwise spread is the maximum minus minimum across feasible paths of their summed decoded-cost errors. Smaller values favor Brier in every column. These descriptive differences do not test mediation or replace the original comparison families.

Study	Task	Threshold dispersion	Decoded cost MSE	Pathwise spread	Path error (pp)
A	ReLU	-0.087	+0.002	+0.010	+0.41
A	Quadratic	-0.120	+0.002	-0.010	+0.28
A	Digits	-0.143	-0.131	-0.803	-3.17
B	ReLU	-0.098	-0.015	-0.172	-2.52
B	Quadratic	-0.127	-0.017	-0.209	-3.61
B	Digits	-0.186	-0.053	+0.091	-0.06

The pathwise spread for an example is a sufficient stability check: if it is below the true cost gap to the cheapest nonoptimal path, every route minimizing predicted cost is truly optimal. Its average in the table is not itself a certificate for every example. The complete unit-level values and certified fractions are in the analysis artifact.

These tables were regenerated from bound saved arrays using exact integer comparisons of the binary32 predicted path costs. The supplement's `revision_evidence/` records each source receipt, array hash, checkpoint, stratum and analysis unit. No model was retrained or reselected for these views. The original fixed-step endpoints retain their secondary status; this cross-study interpretation is descriptive.

B.4 EXPLORATORY BRIDGE ON COMMON SYNTHETIC UNITS

This post-hoc diagnostic reuses the ten Study B evaluation teachers per synthetic task. It is not independent confirmation. Before running the new fits, we fixed the manifest of tasks, units, objectives, settings and four readings. Each A-selected learning rate and weight decay is applied to the same B graph, teacher, train/validation/evaluation contexts, initialization and pre-generated minibatch sequence. The inherited two-layer, 64-unit model, loss normalization, constant-rate Adam and batch size 32 are unchanged.

The four readings are A settings at step 600, those same settings at step 6,000, the same long trajectory with validation selection, and B settings with the same selection rule. Eligible steps are 600, 900, . . ., 6,000; selection minimizes validation all-context error, then regret, then step. The A-settings/600 reading therefore does not reproduce A's original tied-only endpoint or original evaluation samples. Changing A to B settings compares previously selected configurations; their searches used different development data and procedures, so it does not isolate search-space size.

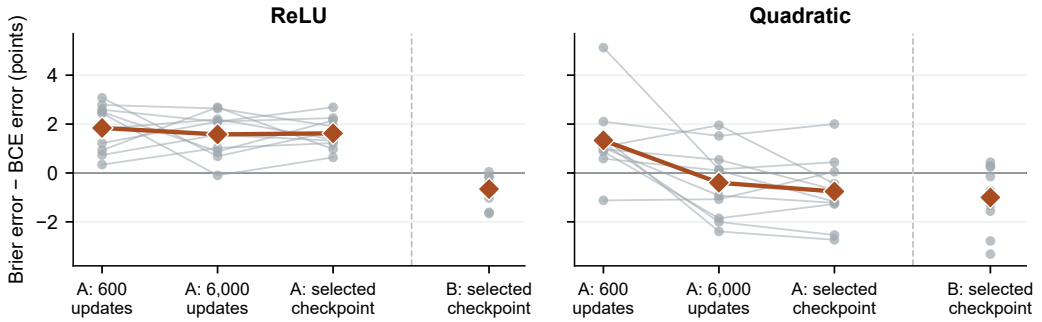

Figure 6: Exploratory bridge on the previously inspected B teachers. Thin lines pair the same ten teachers under A settings; diamonds are means. Positive Brier-minus-BCE error favors BCE. The separated final reading uses B settings. All readings use all contexts. No new significance tests are attached to this previously inspected sample.

Table 16: Settings used in the bridge. Entries are learning rate / weight decay from the original selections. All trajectories run for 6,000 steps.

Task	Objective	A settings	B settings
ReLU	BCE	.006/.0001	.002/.0001
ReLU	Brier	.02/.0001	.006/.0001
ReLU	Mean-MSE	.006/.0001	.006/.0001
Quadratic	BCE	.006/.0001	.002/.0001
Quadratic	Brier	.006/.0001	.002/.0001
Quadratic	Mean-MSE	.006/.0001	.002/.0001

Table 17: All-context error and regret across the four bridge readings. Each entry averages the same ten teachers equally. MSE denotes decoded-mean MSE. These are descriptive results, with no added significance tests.

		All-context error (%)			Mean regret		
Task	Procedure	BCE	Brier	MSE	BCE	Brier	MSE
ReLU	A step 600	12.402	14.243	14.482	0.13340	0.15273	0.15347
ReLU	A step 6,000	10.649	12.227	12.954	0.11387	0.13003	0.13608
ReLU	A selected	10.718	12.339	12.627	0.11460	0.13110	0.13213
ReLU	B selected	10.762	10.103	12.627	0.11597	0.10640	0.13213
Quadratic	A step 600	17.886	19.219	19.194	0.20312	0.21812	0.20840
Quadratic	A step 6,000	14.512	14.111	17.910	0.16123	0.15317	0.19121
Quadratic	A selected	14.575	13.818	18.047	0.16289	0.14976	0.19321
Quadratic	B selected	14.585	13.584	17.197	0.16162	0.14741	0.18350

The ten ReLU Mean-MSE trajectories have identical A/B configurations and are reused; the remaining 50 A-settings trajectories are new. All 60 B-settings trajectories are reused. A full 6,000-step engineering replay of one reused ReLU Mean-MSE configuration reproduced the historical model, Adam state, RNG state, loss history, validation curves, selection and fixed snapshots bit for bit. A separate interruption at step 600 and exact-state resume reproduced the continuous trajectory. These two engineering runs are distinct from the 50 new scientific fits. Calibration and new training occupied 275.0 seconds on the local CPU, with four workers and one computation thread per worker; analysis and manuscript preparation are excluded.

All 50 new fits completed. Evaluation followed sealing the complete new training manifest, although the B test units had been inspected previously. Saved predictions were evaluated with independent integer path comparisons; all original B reference arrays and receipts remain unchanged. The supplement records every teacher, setting, checkpoint, stratum and paired contrast. Error and regret means share the reported BCE/Brier orderings. The ordering is not uniform across teachers: at quadratic A/6,000, Brier has lower error than BCE on five of ten teachers despite its lower mean. This sequential diagnostic does not remove interactions or establish a unique causal attribution of the historical A/B difference. No run count, recipe, endpoint or stopping criterion was adapted to its results.

B.5 Independent confirmation on fresh synthetic teachers

Before generating new evaluation predictions, we fixed the protocol, seeds, two tests and failure rule in the supplied preregistration file. Twenty new teachers per task use data seeds 9470001–9470020 for ReLU and 9470101–9470120 for quadratic. They do not overlap A, B or the exploratory bridge. Each teacher receives all three objectives under both historical A- and B-selected settings: 240 complete 6,000-update trajectories. Within a teacher, losses and settings share training, validation and evaluation contexts, initial weights and minibatch streams. The settings are frozen from prior studies; none is selected on these new teachers. All 2,048 evaluation contexts per teacher form the primary endpoint. Fixed 600- and 6,000-update snapshots and validation-selected checkpoints are read from the registered trajectories. All fits completed before the evaluation split was opened; no teacher was replaced or added.

Let $d_u(r)$ be Brier's nonoptimal-path error minus BCE's error on teacher u at reading r, in percentage points. Positive values favor BCE. The predeclared ReLU interaction compares $d_u(\text{A selected})$ with $d_u(\text{B selected})$: both use 6,000 updates and the same validation rule; within each objective only the historical learning rate changes. The quadratic interaction compares $d_u(\text{A at 600})$ with $d_u(\text{A at 6,000})$ on the same trajectory. Both contrasts are first-minus-second. Two-sided exact paired sign flips enumerate all 2^{20} assignments for each interaction under independent, sign-symmetric teacher differences. The two tests form a new Holm family at .05; their marginal 95% bootstrap intervals are descriptive. We required both a corrected interaction and a crossing of the two mean d values to call each change confirmed. These tests do not assess the significance of either reading by itself.

Table 18: Fresh-teacher readings used by the two registered interactions. Errors are mean ± SD across 20 teachers. Regret is the descriptive mean on the same true costs. A/B denote frozen historical learning-rate settings applied to the new units; "selected" uses the common validation rule. Full per-teacher readings include both settings at all checkpoints.

Task	Reading	BCE	Brier	Mean-MSE
All-context nonoptimal-path error (%): mean ± SD				
ReLU	A selected	10.38 ± 1.18	11.90 ± 1.27	12.84 ± 1.37
ReLU	B selected	10.16 ± 0.94	9.88 ± 1.05	12.84 ± 1.37
Quadratic	A at 600	18.47 ± 1.52	19.91 ± 1.53	19.83 ± 1.34
Quadratic	A at 6,000	15.05 ± 1.27	14.40 ± 1.27	18.70 ± 0.94
Mean true-cost regret: descriptive				
ReLU	A selected	0.111	0.127	0.134
ReLU	B selected	0.108	0.104	0.134
Quadratic	A at 600	0.210	0.225	0.214
Quadratic	A at 6,000	0.167	0.155	0.200

Table 19: The two registered teacher-paired interactions. Positive first-minus-second differences indicate a shift toward Brier. Exact two-sided p values use sign flips; Holm values correct this two-test family. Bootstrap intervals are marginal, not simultaneous.

Interaction	Difference (pp)	Marginal 95% CI	Exact p	Holm p
ReLU: settings	1.797	[1.296, 2.295]	0.0000076	0.0000076
Quadratic: budget	2.083	[1.528, 2.637]	0.0000019	0.0000038

Both mean contrasts cross zero in the preregistered direction, and both interactions pass Holm correction. ReLU's B setting has lower Brier mean error than BCE on 14 of 20 teachers; quadratic's long A reading does so on 16 of 20. True-cost regret follows both mean error orderings, but no new regret test was registered. This supports the two specific changes on these synthetic generators. It does not prove a universal Brier or BCE advantage, convergence by step 6,000, or that learning rate alone explains every difference between the original A and B studies.

C PORTABLE PROBABILITY-TO-DECISION AUDIT

This post-hoc analysis uses every saved evaluation probability array from Studies A and B: 180 final A models and 120 validation-selected B models. No model is retrained, selected again, or omitted. The supplement's `probability_audit/` contains these probabilities, original binary32 decoded costs, true costs, feasible paths, bases, unit gaps, unit mappings and checksums. Its standalone NumPy verifier reproduces the earlier decision-chain metrics and the projection analysis from these arrays. For B's extra fixed checkpoints (600, 1,200, 3,000 and 6,000 updates), the size-limited supplement retains per-run metrics and recomputes paired regret summaries from them; it omits those extra per-sample predictions. The analysis protocol was recorded before this new computation, after the original outcomes were known; it is not a prospective confirmation.

Sample quantities and independent units. We compute $M_{\mathrm{samp}} = \mathrm{avg}_{i,e}(m^{-1}\sum_k(q_{i,e,k} - t_{i,e,k}))^2$ and $V_{\mathrm{samp}} = \mathrm{avg}_{i,e}\, m^{-1}\sum_k(q_{i,e,k} - t_{i,e,k} - \bar{e}_{i,e})^2$ using observed threshold labels. Their sum is sample Brier loss. They are coordinate averages of sample residuals, not estimates of conditional excess risks without additional assumptions. The maximum numerical identity error is 3.1×10^{-16}. All audit endpoints use all contexts. We average equally over A's 20 seeds and B's ten teachers or ten image partitions, averaging the two initializations within each B Digits partition first. Bootstrap intervals use 20,000 paired unit resamples, seed 26092607, and are marginal and descriptive. No new significance family is introduced. Shared selected endpoints reuse the same interval across tables. Original primary tests and their correction families remain unchanged.

Projection and floating-point decisions. Weighted pool-adjacent-violators (PAVA) produces a decreasing threshold vector with weights proportional to cost gaps (uniform here). A crossing is any $q_{k+1} > q_k$ in one cost coordinate; percentages average these coordinate-level indicators, not the fraction of entire contexts with a crossing. We report all task–procedure–loss groups in Tables 20 and 21. The example 37.9% reduction is the difference of group-mean Brier scores divided by the unprojected group mean, not the mean of individual ratios. High crossing incidence can accompany tiny score changes.

Projection preserves decoded costs and argmin sets in exact arithmetic. Binary64 summation can change the selected representative near a tie; path-identity counts are environment-dependent. Our original local replay recorded zero projection path changes and 26 changes when re-decoding saved probabilities instead of using archived binary32 costs. A supplied external replay report recorded one and 30, respectively, with no change in true-cost error or regret. Its identified projection change selects between two true-cost-26 optima at a 3.55×10^{-15} predicted-cost difference. The report's environment was not supplied, and its counts were not reproduced locally. The original logs and a provenance-labeled account are retained in `probability_audit/NUMERICAL_REPLAY_NOTES.md`.

The original maximum decoder/action-cost discrepancies were $5.4 \times 10^{-15}/2.2 \times 10^{-14}$. The revised verifier records its runtime, checks sample identities, score nonincrease and explicit decoder/action-cost tolerances, and tests true-cost error/regret equality. It separately reports path identity and labels a projection path change `PASS_WITH_PATH_CHANGES` when the other checks pass. A change in true-cost outcomes fails the audit. These statuses concern numerical replay, not a floating-point proof of Proposition 2. Original decision metrics retain archived binary32 costs with exact integer path-sum comparisons; earlier decision-chain metrics reproduce within 10^{-10}.

BCE at stored probability endpoints. Some binary32 probabilities equal zero or one. We evaluate probability log loss with its true extended-real values, without clipping, and report BCE decreases only on cost-coordinate vectors whose before and after losses are both finite. Excluded fractions are shown explicitly; the largest group mean is 0.02747%. These are probability-level summaries, not a reconstruction of numerically stable logit training losses. Of 100 initially infinite vectors, 90 become finite and ten remain infinite; no finite vector becomes infinite. Finite endpoint vectors remain included. There are no Brier or finite-subset BCE increases beyond the stated numerical tolerances. The full CSV records endpoint rates and crossing intervals as well as the displayed summaries. The projection analysis covers unit-gap A/B only: the Warcraft replay lacks original probabilities, and battery's uniform scores do not match its unequal decoder gaps.

Table 20: Full post-hoc projection audit. Crossings and excluded fractions are percentages of cost-coordinate vectors, averaged over independent units. Decreases are unprojected minus projected score. † BCE uses the finite subset defined above. No true-cost error or regret change is observed in these replays; floating-point path identity can change near ties.

Study	Task	Trained loss	Crossings (%)	Brier decrease	BCE decrease†	Excluded (%)
A	Digits	BCE	98.376	0.000202	0.001216	0.00000
A	Digits	Brier	99.383	0.000139	0.000356	0.00000
A	Digits	Mean-MSE	100.000	0.000304	0.000706	0.00000
A	Quadratic	BCE	2.057	0.000364	0.001604	0.00009
A	Quadratic	Brier	2.152	0.000321	0.000957	0.00000
A	Quadratic	Mean-MSE	50.536	0.013318	0.029510	0.00000
A	ReLU	BCE	0.065	0.000004	0.000020	0.00018
A	ReLU	Brier	0.204	0.000011	0.000049	0.00000
A	ReLU	Mean-MSE	49.529	0.026534	0.067810	0.00000
B	Digits	BCE	99.154	0.000163	0.000698	0.00000
B	Digits	Brier	97.966	0.000131	0.000298	0.00000
B	Digits	Mean-MSE	99.973	0.074520	0.380256	0.02747
B	Quadratic	BCE	1.423	0.000289	0.001547	0.00054
B	Quadratic	Brier	1.603	0.000216	0.000675	0.00000
B	Quadratic	Mean-MSE	50.022	0.015598	0.034931	0.00000
B	ReLU	BCE	0.036	0.000002	0.000023	0.00072
B	ReLU	Brier	0.042	0.000001	0.000004	0.00000
B	ReLU	Mean-MSE	50.114	0.036715	0.096857	0.00000

Table 21: Marginal paired-unit 95% bootstrap intervals for the score decreases in Table 20. † BCE uses the same finite subset. These intervals describe magnitude across the original units; they are not tests of the projection theorem.

Study	Task	Trained loss	Brier decrease: marginal 95% CI	BCE decrease†: marginal 95% CI
A	Digits	BCE	[0.000171, 0.000234]	[0.000998, 0.001448]
A	Digits	Brier	[0.000117, 0.000168]	[0.000263, 0.000486]
A	Digits	Mean-MSE	[0.000223, 0.000388]	[0.000527, 0.000893]
A	Quadratic	BCE	[0.000332, 0.000398]	[0.001445, 0.001772]
A	Quadratic	Brier	[0.000293, 0.000351]	[0.000864, 0.001056]
A	Quadratic	Mean-MSE	[0.012197, 0.014496]	[0.026978, 0.032198]
A	ReLU	BCE	[0.000001, 0.000007]	[0.000008, 0.000035]
A	ReLU	Brier	[0.000007, 0.000016]	[0.000032, 0.000068]
A	ReLU	Mean-MSE	[0.023795, 0.029405]	[0.060541, 0.075351]
B	Digits	BCE	[0.000118, 0.000209]	[0.000414, 0.001018]
B	Digits	Brier	[0.000111, 0.000150]	[0.000213, 0.000381]
B	Digits	Mean-MSE	[0.061799, 0.087129]	[0.266422, 0.501053]
B	Quadratic	BCE	[0.000241, 0.000334]	[0.001314, 0.001772]
B	Quadratic	Brier	[0.000178, 0.000255]	[0.000556, 0.000801]
B	Quadratic	Mean-MSE	[0.014043, 0.017239]	[0.031218, 0.038822]
B	ReLU	BCE	[0.000000, 0.000007]	[0.000003, 0.000054]
B	ReLU	Brier	[0.000000, 0.000002]	[0.000002, 0.000007]
B	ReLU	Mean-MSE	[0.032168, 0.040726]	[0.084325, 0.108045]

Table 22: Post-hoc paired true-regret differences. Each interval is marginal, based on 20,000 paired unit resamples; no new confirmatory tests. A uses its final 600-update model. B uses the same selected recipe at each reading.

Study / task	Reading	Contrast	Difference	Marginal 95% CI
A / ReLU	final, 600 updates	Brier – Mean-MSE	+0.0070	[+0.0008, +0.0143]
A / ReLU	final, 600 updates	BCE – Brier	-0.0217	[-0.0299, -0.0149]
A / Quadratic	final, 600 updates	Brier – Mean-MSE	+0.0113	[+0.0026, +0.0200]
A / Quadratic	final, 600 updates	BCE – Brier	-0.0183	[-0.0260, -0.0104]
A / Digits	final, 600 updates	Brier – Mean-MSE	-0.0748	[-0.1046, -0.0416]
A / Digits	final, 600 updates	BCE – Brier	-0.0217	[-0.0516, +0.0069]
B / ReLU	600	Brier – Mean-MSE	-0.0099	[-0.0149, -0.0054]
B / ReLU	600	BCE – Brier	+0.0384	[+0.0319, +0.0454]
B / ReLU	1200	Brier – Mean-MSE	-0.0228	[-0.0267, -0.0187]
B / ReLU	1200	BCE – Brier	+0.0134	[+0.0083, +0.0182]
B / ReLU	3000	Brier – Mean-MSE	-0.0256	[-0.0315, -0.0207]
B / ReLU	3000	BCE – Brier	+0.0056	[+0.0006, +0.0104]
B / ReLU	6000	Brier – Mean-MSE	-0.0281	[-0.0345, -0.0212]
B / ReLU	6000	BCE – Brier	+0.0015	[-0.0029, +0.0063]
B / ReLU	selected	Brier – Mean-MSE	-0.0257	[-0.0326, -0.0189]
B / ReLU	selected	BCE – Brier	+0.0096	[+0.0055, +0.0142]
B / Quadratic	600	Brier – Mean-MSE	+0.1165	[+0.1012, +0.1333]
B / Quadratic	600	BCE – Brier	-0.0422	[-0.0503, -0.0335]
B / Quadratic	1200	Brier – Mean-MSE	+0.0249	[+0.0154, +0.0363]
B / Quadratic	1200	BCE – Brier	-0.0255	[-0.0318, -0.0196]
B / Quadratic	3000	Brier – Mean-MSE	-0.0208	[-0.0259, -0.0152]
B / Quadratic	3000	BCE – Brier	-0.0003	[-0.0043, +0.0036]
B / Quadratic	6000	Brier – Mean-MSE	-0.0346	[-0.0391, -0.0304]
B / Quadratic	6000	BCE – Brier	+0.0097	[+0.0055, +0.0163]
B / Quadratic	selected	Brier – Mean-MSE	-0.0361	[-0.0402, -0.0316]
B / Quadratic	selected	BCE – Brier	+0.0142	[+0.0058, +0.0233]
B / Digits	600	Brier – Mean-MSE	-0.0506	[-0.0838, -0.0157]
B / Digits	600	BCE – Brier	-0.0047	[-0.0318, +0.0266]
B / Digits	1200	Brier – Mean-MSE	-0.0636	[-0.1125, -0.0175]
B / Digits	1200	BCE – Brier	+0.0002	[-0.0392, +0.0526]
B / Digits	3000	Brier – Mean-MSE	-0.0501	[-0.0879, -0.0172]
B / Digits	3000	BCE – Brier	+0.2148	[-0.0330, +0.6897]
B / Digits	6000	Brier – Mean-MSE	-0.0317	[-0.0732, +0.0063]
B / Digits	6000	BCE – Brier	+0.0455	[-0.0119, +0.1395]
B / Digits	selected	Brier – Mean-MSE	-0.0277	[-0.0736, +0.0144]
B / Digits	selected	BCE – Brier	+0.0000	[-0.0163, +0.0172]

Warcraft cost tolerance. The saved path-level arrays also support the full tolerance curve in Figure 7. We use exact error at zero and relative stored-cost regret at least the threshold otherwise. The grid covers 10^{-8} through 10^{-1}; the original 0.1% descriptive threshold is retained rather than selected from this curve. Each seed uses the same 1,000 maps, so seeds are not new independent map samples.

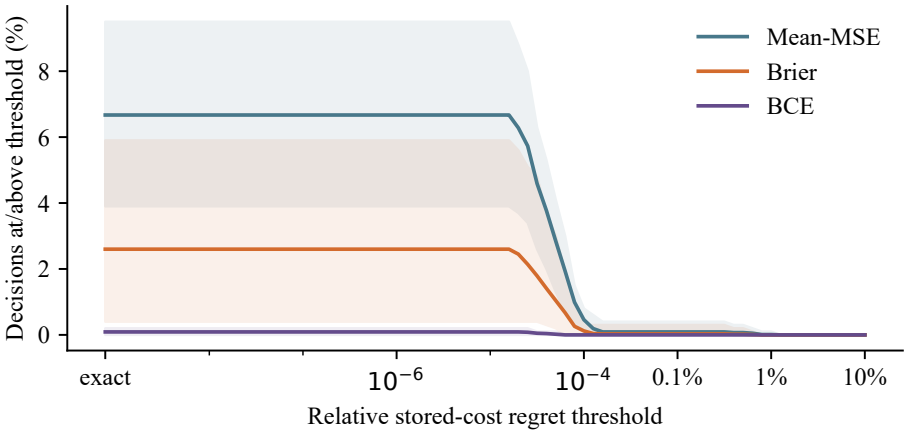

Figure 7: **Cost tolerance changes the size of the Warcraft error gap.** Lines average ten training seeds; shading is the seed minimum–maximum range, not a confidence interval. The horizontal scale is logarithmic above 10^{-8}, with a linear segment containing exact error at zero. The evaluation maps and stored-cost interpretation are fixed.

D WARCRAFT PROTOCOL, LEARNING CURVES AND NUMERICAL SENSITIVITY

Data and the decision rule. We downloaded the original Warcraft archive from Edmond, DOI `10.17617/3.YJCQ5S`, version 1.0, and checked its publisher MD5. Only the original 12×12 train and validation arrays were extracted before selection and confirmation training were sealed. The test split contains 1,000 fixed maps. Inference is conditional on these maps; the ten confirmation seeds vary initialization and minibatch order. Costs are positive vertex weights, moves use all eight neighbors, and both endpoints contribute to path cost. Dijkstra returns a deterministic minimum-cost simple path. Evaluation converts stored binary costs into integers with a shared power-of-two scale, avoiding tolerance-based ties. This is full-cost supervision, so these scores are not a like-for-like ranking against methods trained only on path labels.

Unequal cost levels. The training alphabet in the supplied float16 arrays is

$$(0.7998046875,\ 1.2001953125,\ 5.30078125,\ 7.69921875,\ 9.203125).$$

Let $a_0 < \cdots < a_m$ denote these levels, $\Delta_k = a_k - a_{k-1}$, $R = a_m - a_0$ and $w_k = \Delta_k / R$. With $t_k = \mathbf{1}\{C \geq a_k\}$, the shared decoder and objectives are

$$\widehat{c} = a_0 + \sum_k \Delta_k q_k, \qquad L_{\text{mean}} = \left(\sum_k w_k (q_k - t_k) \right)^2,$$

$$L_{\text{Brier}} = \sum_k w_k (q_k - t_k)^2, \qquad L_{\text{BCE}} = \sum_k w_k [-t_k \log q_k - (1 - t_k) \log(1 - q_k)].$$

Writing $e_k = q_k - t_k$ and $\bar{e}_w = \sum_k w_k e_k$ gives

$$L_{\text{Brier}} = L_{\text{mean}} + \sum_k w_k (e_k - \bar{e}_w)^2.$$

Indeed, expanding the last sum and using $\sum_k w_k = 1$ leaves $\sum_k w_k e_k^2 - \bar{e}_w^2$. Positive weights retain the componentwise proper-score optima, connecting the implemented objectives to Section 4. Both

threshold objectives use the same weights; using them on only one would change the comparison. No training labels or predicted costs are rounded to the nominal values.

Training and selection. The fresh network uses ResNet18's stem and first two residual blocks, adaptive pooling to 12×12 and a four-channel 1×1 cumulative head, without pretrained weights. On the exact 24×24 feature map, ordinary 2×2 max pooling has the same bins as adaptive pooling; this equivalent implementation permits deterministic CUDA backward. Input normalization uses training pixels only. Each loss receives learning rates $\{10^{-4}, 5 \times 10^{-4}, 2 \times 10^{-3}\}$, zero weight decay, and three development seeds. Adam uses batch size 70, up to 50 epochs, and learning-rate factors of .1 at epochs 30 and 40. Validation occurs every five epochs. Four checks without improvement stop a run, not before epoch 20; the best error, then regret, then earliest epoch selects its checkpoint. Mean-MSE and Brier select 5×10^{-4}; BCE selects 2×10^{-3}, the grid ceiling. All three BCE development seeds have zero selected validation error, but this does not establish an exhaustive search or asymptotic convergence. The 600-step validation is an auxiliary checkpoint and is excluded from selection.

There are 27 development and 30 confirmation runs. The prospective seed rule used the development paired standard deviations, inflated by 1.5, in a normal planning approximation for a one-percentage-point effect, 80% power and per-comparison $\alpha = .025$, constrained to 10–20 seeds. It selected the minimum of ten before any test access. Estimating variance from three development seeds is uncertain; this is not a power guarantee. The full study took 3.44 occupied hours on an RTX 4070 Ti.

Longer training. One full epoch is 143 updates; 600 updates expose 41,960 examples, or 4.196 epochs. Mean validation errors at step 600 are 8.61%, 5.53% and 2.22% for Mean-MSE, Brier and BCE. At epoch 20, which every run reaches, they are 6.64%, 2.83% and .02%. These are descriptive validation comparisons, not additional held-out test results. Selected checkpoints range from epochs 10 to 50; completed runs range from 30 to 50.

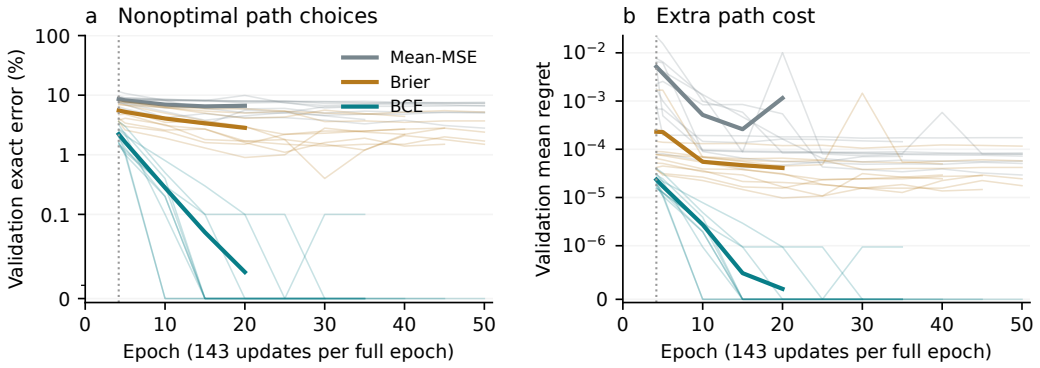

Figure 8: **Validation trajectories in the Warcraft extension.** Thin lines show all ten confirmation seeds per method and end at their actual stopping epochs. Thick lines are means only through epoch 20, where every seed is observed; later averages would mix different survivors. The dotted line marks the auxiliary 600-step checkpoint. The vertical axes use a symmetric log scale to show both zero and small positive errors. No curve was used to choose a new test endpoint after test access.

Registered tests and complete seed counts. The two comparisons form their own Holm family, separate from the central nine-test study and earlier 24-test study. Two-sided exact paired sign flips use all 2^{10} assignments under a sign-symmetric paired null; the statistic is the seed mean error difference. Marginal 95% intervals use 20,000 seed bootstrap draws. Both raw p values are $2/1024$.

Table 23: Warcraft's two predeclared comparisons. Negative differences favor the first method; intervals are marginal, not multiplicity-adjusted.

Comparison	Difference (pp)	Marginal 95% CI	Exact p	Holm p
BCE − Brier	-2.51	[-3.74, -1.37]	0.0019531	0.0039062
Brier − Mean-MSE	-4.07	[-4.78, -3.40]	0.0019531	0.0039062

Complete seed-level counts, including the nominal-cost diagnostics, are provided with the research artifact.

What numerical precision changes. We re-evaluated the saved paths at the nominal levels $(.8, 1.2, 5.3, 7.7, 9.2)$, using integer tenths for exact sums. Of the raw-cost errors, 98.65% for Mean-MSE, 98.85% for Brier and 100% for BCE are optimal under this diagnostic. The archived label masks are also suboptimal under stored costs on 636 training, 61 validation and 61 test maps; all these masks have optimal nominal cost. This pattern explains the precision sensitivity without changing the frozen endpoint. It does not identify the generator's historical floating-point execution. Casting the stored float16 values to float64 cannot recover pre-storage costs; the nominal-decimal remapping is a different diagnostic cost interpretation, not a higher-precision reconstruction of the original generator.

The relative-regret diagnostics in Figure 3 use $\mathbf{1}\{r/c^\star \geq \epsilon\}$. At .1% extra stored cost, the counts are 9, 3 and 0 of 10,000 seed–map evaluations per method; at 1%, they are 1, 0 and 0. Each seed evaluates the same 1,000 maps, so these are not 10,000 independent map samples. The empirical zeros are exact counts, not a claim of zero population risk. Stored costs have a common integer scale: the two checks use $1000r_{\text{int}} \geq c^\star_{\text{int}}$ and $100r_{\text{int}} \geq c^\star_{\text{int}}$, respectively. No observed case equals either threshold. These post-hoc thresholds were not used for training, selection or the primary tests. The original benchmark's code also reports tolerance-based cost ratios against archived masks, whereas our primary endpoint uses exact optima recomputed from stored costs. Together with the different supervision, this prevents presenting our .09% as a directly comparable benchmark record. The useful conclusion is that exact-hit differences can survive longer training while corresponding to very small extra path cost.

E Separate standard-method background study

This independent study used different seeds and boards, nine configurations per method, and its own 24-comparison Holm family. It is not an expansion of either current three-objective table. All 15 methods, including competitive ordinal and categorical alternatives, and every planned contrast are retained.

Table 24: Earlier 15-family study, primary error (%), equal-seed mean ± SD over 20 seeds. Synthetic columns use natural tied contexts; Digits uses all boards. Different heads and supervision define complete-recipe comparisons. These are separate-study evidence, not additional rows in Table 2.

Method	ReLU	Quadratic	Digits routing
Scalar MSE	12.302 ± 1.889	18.412 ± 1.871	14.038 ± 1.061
Cumulative-head mean MSE	9.746 ± 1.492	13.651 ± 1.494	13.926 ± 1.684
Cumulative BCE	8.181 ± 0.896	12.672 ± 1.939	10.674 ± 0.929
Categorical CE	10.035 ± 1.225	14.835 ± 1.456	12.168 ± 1.761
Softmax-head mean MSE	9.938 ± 1.555	13.945 ± 1.568	12.827 ± 1.616
CORAL	8.380 ± 1.118	11.890 ± 1.173	14.048 ± 1.214
CORN	8.510 ± 0.888	14.243 ± 1.793	11.001 ± 1.401
Linear cumulative BCE	25.970 ± 3.199	60.370 ± 8.665	34.531 ± 1.009
Linear categorical CE	26.235 ± 3.010	60.428 ± 8.726	16.021 ± 0.969
Linear CORAL	27.562 ± 3.176	60.370 ± 8.749	53.887 ± 0.983
Uniform-path CE	16.474 ± 1.699	21.737 ± 2.336	18.955 ± 1.661
SPO+	19.495 ± 2.047	25.013 ± 2.356	17.817 ± 2.122
DBB	26.593 ± 2.312	34.997 ± 3.188	40.166 ± 2.903
Cumulative BCE + SPO+	8.557 ± 0.952	13.113 ± 1.402	10.186 ± 1.132
Categorical CE + SPO+	9.674 ± 1.017	15.265 ± 1.637	13.018 ± 0.885

Architecture and search. Nonlinear models use two 64-unit ReLU hidden layers. Synthetic models jointly predict 27 edge costs; image models share a $64 \to 64 \to 64$ feature extractor over grid locations. Cumulative and categorical heads have $K-1$ and K logits per cost, respectively. All matched-head pairs have identical parameter counts and initialization. The scalar head uses $b_e + (K-1)\,\mathrm{softplus}(z_e)$ and is not upper bounded. Batch sizes are 32 and 16 for synthetic and image tasks. Nonlinear learning rates are $\{.0005, .002, .006\}$; linear rates are $\{.02, .06, .2\}$. Most families pair these with weight decay $\{0, .0001, .001\}$; Uniform-path CE instead searches temperature $\{.25, 1, 4\}$, DBB perturbation $\{.1, 1, 10\}$, and hybrids SPO+ weight $\{.01, .1, 1\}$, with zero weight decay in these families. All receive nine configurations, which does not equate their effective search spaces.

Table 25: Selected learning rate / weight decay / auxiliary parameter. The auxiliary entry is unused for purely supervised objectives. Selections use development outcomes only.

Method	ReLU	Quadratic	Digits
Scalar MSE	0.006/0.0001/1	0.002/0.0001/1	0.006/0.0001/1
Cumulative-head mean MSE	0.006/0.0001/1	0.006/0.0001/1	0.006/0/1
Cumulative BCE	0.006/0.0001/1	0.006/0.0001/1	0.006/0.0001/1
Categorical CE	0.006/0.001/1	0.006/0.0001/1	0.006/0.001/1
Softmax-head mean MSE	0.006/0.0001/1	0.002/0.0001/1	0.006/0/1
CORAL	0.006/0/1	0.006/0/1	0.006/0.0001/1
CORN	0.006/0/1	0.006/0/1	0.006/0.0001/1
Linear cumulative BCE	0.06/0/1	0.2/0/1	0.06/0/1
Linear categorical CE	0.02/0.001/1	0.2/0.0001/1	0.2/0.0001/1
Linear CORAL	0.06/0/1	0.06/0.001/1	0.2/0.0001/1
Uniform-path CE	0.002/0/1	0.002/0/1	0.006/0/1
SPO+	0.002/0.001/1	0.002/0.001/1	0.006/0/1
DBB	0.006/0/10	0.006/0/10	0.002/0/0.1
Cumulative BCE + SPO+	0.006/0/0.1	0.006/0/0.1	0.006/0/0.01
Categorical CE + SPO+	0.006/0/0.1	0.006/0/0.1	0.002/0/1

Table 26: True regret at the primary endpoint and stratum, mean $\pm$ seed SD. These are secondary outcomes; no additional tests were introduced for this table.

Method	ReLU	Quadratic	Digits routing
Scalar MSE	0.130 ± 0.022	0.202 ± 0.022	0.271 ± 0.029
Cumulative-head mean MSE	0.101 ± 0.015	0.143 ± 0.017	0.276 ± 0.046
Cumulative BCE	0.086 ± 0.009	0.139 ± 0.023	0.237 ± 0.025
Categorical CE	0.105 ± 0.013	0.168 ± 0.018	0.279 ± 0.041
Softmax-head mean MSE	0.102 ± 0.015	0.148 ± 0.016	0.282 ± 0.046
CORAL	0.087 ± 0.012	0.126 ± 0.013	0.255 ± 0.030
CORN	0.090 ± 0.010	0.160 ± 0.023	0.266 ± 0.042
Linear cumulative BCE	0.313 ± 0.044	1.039 ± 0.199	0.979 ± 0.035
Linear categorical CE	0.312 ± 0.039	1.043 ± 0.200	0.356 ± 0.018
Linear CORAL	0.336 ± 0.044	1.040 ± 0.208	1.960 ± 0.059
Uniform-path CE	0.179 ± 0.020	0.253 ± 0.030	0.376 ± 0.046
SPO+	0.218 ± 0.023	0.292 ± 0.030	0.353 ± 0.056
DBB	0.331 ± 0.034	0.489 ± 0.056	1.092 ± 0.130
Cumulative BCE + SPO+	0.091 ± 0.010	0.145 ± 0.018	0.240 ± 0.030
Categorical CE + SPO+	0.103 ± 0.011	0.172 ± 0.021	0.320 ± 0.024

Baseline adaptation conventions. The CORAL implementation retains shared score weights but initializes threshold biases from 1 to -1 and averages BCE over threshold coordinates. These choices differ from the authors' reference initialization and loss normalization. DBB perturbs unnormalized costs by λC and then normalizes the scalar training objective. These conventions and the finite grids define the evaluated recipes; they are not claims to exhaustive tuning of the original methods.

All planned comparisons. Negative first-minus-second differences favor the first method. Synthetic tests condition on natural ties; digits tests use all boards. Exact count fractions determine the tests, with 1,048,576 sign assignments per comparison. All 24 comparisons share one Holm family. Marginal intervals use the protocol's bootstrap random stream and are not Holm-adjusted intervals.

Table 27: All 24 planned comparisons in the earlier standard-method study.

Task	First minus second	Δ (pp)	Marginal 95% CI	Holm p
ReLU	BCE − Mean MSE	-1.565	[-2.049, -1.093]	0.000206
ReLU	Cat. CE − Cat. MSE	0.098	[-0.480, 0.683]	0.916534
ReLU	BCE − Cat. CE	-1.854	[-2.285, -1.468]	0.000046
ReLU	BCE − CORAL	-0.199	[-0.543, 0.114]	0.814438
ReLU	BCE − CORN	-0.328	[-0.628, -0.041]	0.436485
ReLU	Linear BCE − BCE	17.788	[16.638, 19.086]	0.000046
ReLU	BCE+SPO+ − BCE	0.375	[0.045, 0.717]	0.436485
ReLU	Cat.+SPO+ − Cat. CE	-0.362	[-0.689, -0.059]	0.436485
Quadratic	BCE − Mean MSE	-0.980	[-1.562, -0.428]	0.044220
Quadratic	Cat. CE − Cat. MSE	0.890	[0.362, 1.398]	0.061121
Quadratic	BCE − Cat. CE	-2.164	[-2.657, -1.631]	0.000109
Quadratic	BCE − CORAL	0.781	[0.005, 1.521]	0.436485
Quadratic	BCE − CORN	-1.572	[-2.061, -1.053]	0.000610
Quadratic	Linear BCE − BCE	47.699	[44.059, 50.789]	0.000046
Quadratic	BCE+SPO+ − BCE	0.442	[-0.002, 0.922]	0.436485
Quadratic	Cat.+SPO+ − Cat. CE	0.429	[0.055, 0.822]	0.436485
Digits	BCE − Mean MSE	-3.252	[-4.097, -2.358]	0.000259
Digits	Cat. CE − Cat. MSE	-0.659	[-1.631, 0.293]	0.814438
Digits	BCE − Cat. CE	-1.494	[-2.378, -0.669]	0.036850
Digits	BCE − CORAL	-3.374	[-4.062, -2.720]	0.000046
Digits	BCE − CORN	-0.327	[-1.143, 0.518]	0.916534
Digits	Linear BCE − BCE	23.857	[23.203, 24.507]	0.000046
Digits	BCE+SPO+ − BCE	-0.488	[-0.928, -0.059]	0.436485
Digits	Cat.+SPO+ − Cat. CE	0.850	[0.186, 1.484]	0.275047

Earlier source-partition sensitivity. A separate descriptive check used five new partitions of the same 1,797 images, with three initializations per method and partition. Cumulative BCE beat its matched-head MSE control in five of five partition means but categorical CE in only three of five. It included no Brier runs and is not replaced by Study B's Brier comparison. All 75 method–partition means are provided in `research_artifact/earlier_partition_means.csv`; the full study records also retain every fit. No new significance claim is made.

F Prospective transfer to battery scheduling

Question and setup. The fresh-teacher synthetic study found two specific changes in the average BCE–Brier ordering when training settings changed. We asked whether one ordering change would predict a difference in a separate decision problem. Using the GB day-ahead price field from the Open Power System Data time-series package (Open Power System Data, 2020), we forecast the 24 hourly prices of each complete UTC day. Inputs were calendar features and same-hour price lags from two through eight days earlier. Each price input precedes its target hour by at least 48 hours; across a day's 24 forecasts, the latest input is 25 hours before the start of that UTC day. There was no same-day price input. The chronological split was 2015–2017 for training, 2018 for development, and 2019–September 2020 for confirmation. Five days lacked a complete target day and 22 lacked usable lag history, leaving 612 test days.

The decision was a daily schedule for a simulated 1-MWh battery. It started and ended at 0.5 MWh, changed stored energy by at most 0.25 MWh per hour, used a 0.05-MWh state grid and 90% round-trip efficiency. The same dynamic program chose a schedule from every model's predicted prices. We then calculated the resulting gross revenue using observed prices. Daily regret is the revenue lost relative to a schedule chosen with those observed prices. No-trade revenue is zero under the fixed initial and final state. These GBP amounts describe a price-taking simulation, not deployable net profit.

Table 28: **No registered BCE–Brier reversal in the independent battery task.** Mean daily regret in GBP per 1-MWh battery on 612 held-out days; lower means more realized revenue. The scalar-price MSE model has a different head and is a contextual baseline.

Training loss	600 updates	6,000 updates
BCE	4.175	4.402
Brier	4.089	4.293
Mean-MSE	3.959	4.393
Scalar-price MSE	5.100	5.341
Seven-day historical forecast	5.653	

Models and registered rule. The matched threshold models shared a two-layer, 64-unit ReLU trunk, a cumulative-threshold output, 64 training-price-quantile levels, the same decoded-price rule, initial weights, minibatch order and decision solver. Only the training loss differed: BCE, Brier, or Mean-MSE. A direct scalar price-MSE predictor used a different head and is a contextual complete-recipe baseline. The 600- and 6,000-update readings came from the same trajectory for each method and seed. For each loss, three development seeds selected a learning rate from the common grid $\{0.0003, 0.001, 0.003, 0.01, 0.03\}$ using 2018 decision regret averaged over both fixed readouts and the three development seeds. BCE, Brier and Mean-MSE selected 0.003; scalar MSE selected 0.03. Five new paired seeds per method were trained before confirmation prices were read.

Implemented encoding and loss weights. Let $a_0 < \cdots < a_{63}$ be the 64 training-price quantiles, $m = 63$, and $\Delta_k = a_k - a_{k-1}$. For an observed price y, the encoded label $g(y)$ is its nearest alphabet value (with endpoint clipping), and $t_k = \mathbf{1}\{g(y) \geq a_k\}$. The decoder is $\widehat{y} = a_0 + \sum_{k=1}^{m} \Delta_k q_k$, where $q_k = \sigma(z_k)$. The frozen implementation uses *uniform* threshold weights for both scores and the training-price standard deviation s_{tr} for Mean-MSE:

$$L_{\mathrm{Brier}} = \frac{1}{m}\sum_k (q_k - t_k)^2, \qquad L_{\mathrm{BCE}} = \frac{1}{m}\sum_k \big[\log(1+e^{z_k}) - t_k z_k\big], \qquad L_{\mathrm{mean}} = \frac{(\widehat{y} - g(y))^2}{s_{\mathrm{tr}}^2}.$$

Losses are also averaged over sampled hours. Scalar-price MSE instead scores the unencoded price, with the same s_{tr} normalization. Thus this transfer uses the general-positive-weight result in Appendix A.4, rather than the gap-weighted unit-coefficient decomposition used for Warcraft. Specifically, with $\omega_k = 1/m$, $S = \sum_k \Delta_k^2/\omega_k = m\sum_k \Delta_k^2$, $\eta_k = \Pr(g(Y) \geq a_k \mid x)$, $\delta_k = q_k - \eta_k$, and $b = \widehat{y} - \mathbb{E}[g(Y) \mid x]$, its per-hour Brier excess is

$$\mathcal{E}_{\mathrm{Br}} = \frac{b^2}{S} + \frac{1}{m}\sum_k \left(\delta_k - \frac{m\Delta_k b}{S}\right)^2.$$

The decoded component is b^2/S, whereas implemented Mean-MSE excess is b^2/s_{tr}^2: the former equals s_{tr}^2/S times the latter. The quantized-label comparison therefore does not have the unit-coefficient $L_{\mathrm{Brier}} = L_{\mathrm{mean}} +$ dispersion identity of the unit-gap studies. Decisions are always scored at the original prices; the encoding term in Appendix A.4 remains relevant.

We predicted BCE would have less regret than Brier at 600 updates, and Brier less than BCE at 6,000. Let $D_b = R_{\mathrm{Brier,b}} - R_{\mathrm{BCE,b}}$, where positive D_b favors BCE. Success required opposite signs of the two means, a positive interaction $D_{600} - D_{6000}$ with a confidence interval excluding zero and a Holm-corrected test, and a long-checkpoint Brier gain of at least both 10% of oracle-versus-no-trade revenue and GBP 5 per battery day, with a positive interval. The two registered effects used 20,000 resamples of circular blocks of 28 consecutive eligible-day observations, after averaging seeds within each day. Development already showed Brier slightly ahead at both steps; the prediction and confirmation rule were retained unchanged. The complete plan, source hashes, and stage order are archived with the experiment.

Confirmation and limits. Table 28 shows Brier's mean regret below BCE's at both steps: $D_{600} = -0.087$ and $D_{6000} = -0.109$ GBP/day. The interaction was 0.022 GBP/day, with 95% block-bootstrap interval $[-0.181, 0.203]$ and Holm-adjusted $p = 0.819$. Brier's long-checkpoint revenue benefit was 0.109 GBP/day, interval $[-0.003, 0.242]$, adjusted $p = 0.162$. The oracle's mean gross revenue was 22.021 GBP/day; the benefit was about 0.5% of that amount and far below the registered

GBP 5/day floor. A prespecified GBP 5 per MWh stored-energy-change charge reduced the mean benefit to 0.070 GBP/day. The registered joint success rule failed.

For transparency, the 2019 subset descriptively had $(D_{600}, D_{6000}) = (0.034, -0.049)$ GBP/day, whereas the 2020 subset had $(-0.245, -0.187)$; these were not separate confirmation tests. Every method had higher test regret at the later checkpoint. A price-grid-only oracle diagnostic incurred 0.546 GBP/day regret against the original-price oracle, and 0.52% of evaluated hours lay outside the training-derived price range. Neither diagnostic changes the registered result. The evidence shows that the particular synthetic short-to-long ordering prediction did not transfer to this fixed external task, and that the observed monetary gap was small. It does not establish a general ranking of the losses.