# WHEN RETRIEVAL ENDPOINT SCORES MISS THE VALUE OF FEEDBACK

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Equal blind and known-condition retrieval scores do not certify equal partial-feedback value. An exact average-precision construction matches dimension, action count, and endpoints while reversing the preferred question. A natural-image submenu census finds disjoint optimal one-question sets in 13 of 1,920 cells; six also favor a paid question over stopping, and one native-rank rational-AP case verifies the reversal. This evaluator-only existence result does not establish a deployable selector. Across the same learned libraries, endpoint-preserving deletion preserves every optimal tree in 1,887 cells, including 318 of 351 with positive allowed-probe risk. Publicly predicted endpoint cores show no established replanning advantage over matched menus. At exactly two answers, SOURCE-calibrated planning improves single-ranker AP over SOURCE-selected controls, while fusion gains remain unestablished. Endpoint insufficiency, retained-action inadequacy, and deployed-policy failure are distinct diagnoses. These exposed-population results support evaluation under explicit information and cost assumptions, not a universal repair or serving speedup.

## 1 INTRODUCTION

A retrieval system may have several useful notions of similarity. A shoe can match a query by material, category, or appearance; a bird can match by bill shape or color. Conditional retrieval makes these choices explicit (Veit et al., 2017; Tan et al., 2019; Vaze et al., 2023). When the condition is unknown, a system can ask a question before choosing a ranker. The central question is then not simply which representation retrieves well, but which representation remains useful after partial feedback.

A common evaluation tests the best blind ranker and the best ranker for each known condition. These scores are attractive summaries: the first measures performance without information, and the second measures performance with all relevant information. Equal values at both endpoints, however, need not imply equal values between them. A ranker that wins at neither endpoint can be optimal after one answer. Conversely, a ranker that is useful under some posterior can be irrelevant to every question worth asking.

This distinction motivates an evaluation with three levels. *Diagnostic risk* measures the loss caused by deleting actions under possible observations. *Policy relevance* asks whether that loss occurs on good question paths. *Operational value* compares frozen actors, including question costs and strong terminal alternatives. These levels separately test possible deletion loss, its relevance to useful question paths, and improvement of a deployed policy. Figure 1 shows why partial answers can change decisions despite matched endpoints.

The analysis separates endpoint insufficiency from policy relevance. State and erasure analyses explain preservation under unresolved uncertainty (Table 1). Response correction improves utility under adaptive stopping and single-ranker AP at exactly two answers, but fusion and stronger-actor controls limit the benefit.

The contribution is a retrieval-specific preservation test and a controlled separation of these outcomes. The generic value of information, dynamic programming, convex pruning, and rank fusion are established tools. No universal neural-dimension law, new general planner, or end-to-end serving speedup is asserted.



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.

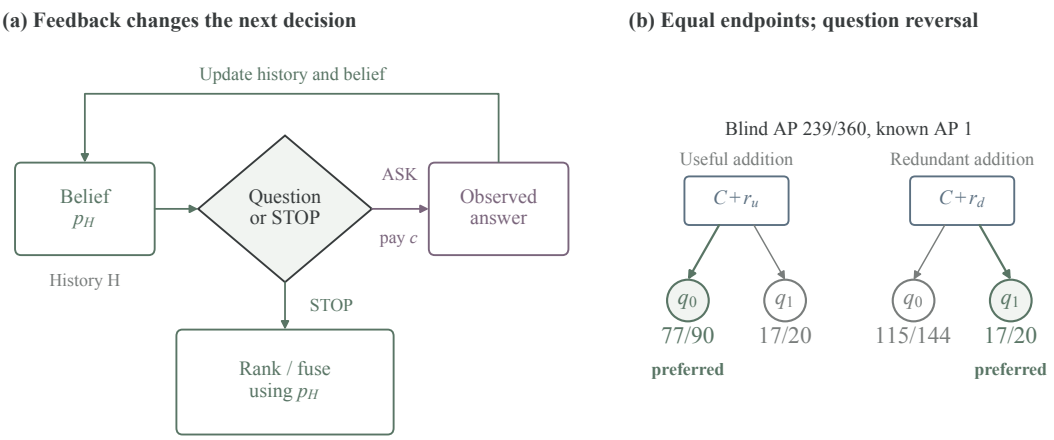


Figure 1: **Feedback changes the readout and the next question.** (a) Answers update the belief used to ASK again or STOP and rank/fuse. Each answer returns to the belief before reconsideration. Deployment uses public images and purchased answers; true payoffs and the response law remain evaluator information. (b) Equal blind AP $239/360$ and known-condition AP 1 conceal a question reversal. Highlighted branches mark preferred questions; fractions give forced noiseless one-answer AP without a tariff (Table 2). Geometry is schematic, not proportional to AP or paid-policy gains.

Table 1: **Nontrivial endpoint-preserving deletion often preserves optimal paths.** $N$ counts query components; $K \to |B|$ gives full and retained action counts (range). Every cell deletes actions. $L_A = \ell_B(\pi_A)$ is deletion loss on a full-menu optimal tree $\pi_A$. Means use $10^{-3}$ AP for deletion losses and $10^{-3}$ AP-minus-price for $G$; cap two, price .005. Preserved counts include three libraries per query, not independent samples. Stored-payoff endpoints are exactly preserved (Section 5); optimal-tree losses agree across ties (Appendix A).

| Library | $N$ | $K \to \lvert B\rvert$ | $\mathcal{D}_{\text{all}}$ | $\mathcal{D}_{\mathcal{Q},2}$ | $L_A$ | $G$ | Preserved |
|---|---|---|---|---|---|---|---|
| Zappos / DINOv2 | 64 | $8 \to 3$–5 | 7.0731 | 1.7226 | .1004 | .0811 | 185/192 |
| Zappos / CLIP | 64 | $8 \to 2$–5 | 6.8507 | 1.6021 | .1309 | .1257 | 183/192 |
| CUB / D32 | 256 | $6 \to 1$–5 | .5578 | .1407 | .0011 | .0011 | 765/768 |
| CUB / D128 | 256 | $6 \to 1$–5 | .6115 | .1821 | .0160 | .0105 | 754/768 |

## 2 Related work and scope

**Information for decisions.** The value of an observation depends on the decision and loss, not uncertainty alone (Blackwell, 1953; DeGroot, 1962). Decision-region identification, noisy active learning, and loss-driven acquisition make this dependence operational (Javdani et al., 2014; Golovin et al., 2010; Huang et al., 2026). Adaptive submodularity provides guarantees under additional structure (Golovin & Krause, 2011); such structure is not assumed here. Informational substitutes characterize diminishing decision value (Chen & Waggoner, 2016). The present question concerns a different intervention: reducing the available retrieval actions while keeping the observation process fixed.

**Finite decision geometry.** Belief-state planning and pruning of value vectors are standard in partially observable control (Kaelbling et al., 1998; Cassandra et al., 1997). Bayes-plausible posterior splitting is also central to information design (Kamenica & Gentzkow, 2011). The deletion diagnostic below specializes these tools to AP payoffs and tests whether the resulting risk is reached by useful image questions. The finite-policy identity serves as an accounting device, not a claim of a new general planning theorem.

**Representations, calibration, and fusion.** Conditional similarity changes the relevance relation (Veit et al., 2017; Tan et al., 2019; Vaze et al., 2023); nested representations change the available dimensionality (Kusupati et al., 2022). Retrieval-dimension bounds concern the attainable ranking

family and margin (Weller et al., 2026; Bangachev et al., 2026), whereas equal-dimensional menus can still differ in feedback value. Calibration of probabilities (Guo et al., 2017) and calibration for decisions (Zhao et al., 2021) motivate separate response and terminal-payoff checks. Reciprocal-rank fusion (Cormack et al., 2009) supplies a strong existing-head alternative that must be tested before attributing a gap to missing representations.

## 3 A RETRIEVAL DECISION BETWEEN TWO ENDPOINTS

A session contains public query and gallery images $x$, a hidden relevance condition $s$ with prior $p_0$, and a finite menu $A$ of image rankers. The payoff $R_{s,a} \in [0,1]$ is full-gallery AP of action $a$ under condition $s$. Every ranker is a fixed image function; the hidden condition and target labels are not actor inputs. A history $H$ consists of purchased probe IDs and answers. A policy can ask a permitted question or STOP, then choose a terminal ranker. The reported objective is

$$U = \mathbb{E}[\mathrm{AP} - cQ], \qquad \Phi_A(p) = \max_{a \in A} p^\top R_{\cdot,a}, \tag{1}$$

where $Q$ counts purchased answers. The tariff $c$ is an explicit modeling assumption, not measured annotation time or GPU cost.

The blind endpoint is $\Phi_A(p_0)$; known-condition endpoints are $\Phi_A(e_s)$. These are summaries, not the complete payoff matrix. The full matrix, prior, observation law, and feasible policies determine the finite problem. Dimension $D$, action count $K$, gallery caches, and executed rankers remain distinct resources.

**Three losses.** For nonempty $B \subseteq A$, let $d_B(p) = \Phi_A(p) - \Phi_B(p)$. The loss under a fixed experiment is $L_B(H) = \mathbb{E}d_B(p_H)$, where $p_H$ is the true-law posterior. Define $\mathcal{D}_{\mathrm{all}}$ as the largest loss over finite stochastic experiments with prior $p_0$, and $\mathcal{D}_{\mathcal{Q},q}$ as the largest loss over permitted question trees with at most $q$ questions. Let $V_A^\star$ and $V_B^\star$ independently optimize paid value over this same tree class. Then

$$0 \le L_B(H) \le \mathcal{D}_{\mathcal{Q},q} \le \mathcal{D}_{\mathrm{all}}, \qquad 0 \le G := V_A^\star - V_B^\star \le \mathcal{D}_{\mathcal{Q},q}. \tag{2}$$

The first bound concerns an allowed cap-$q$ history. The second follows by executing an $A$-optimal tree with $B$'s terminal actions; question costs cancel, and replanning in $B$ cannot reduce value. The common observation law, feasibility constraints, and prices are essential. Action-dependent feedback would require another model.

**A finite diagnostic.** Let $M = A \setminus B$. Unrestricted deletion risk equals the linear program

$$\begin{aligned} \mathcal{D}_{\mathrm{all}} = \max_{z,t \ge 0} \quad & \sum_{a \in M} t_a \\ \text{subject to} \quad & \sum_{a \in M} z_{a,s} \le p_{0,s}, \\ & t_a \le \sum_s z_{a,s}(R_{s,a} - R_{s,b}) \quad (a \in M, b \in B). \end{aligned} \tag{3}$$

The variables allocate prior mass to signals on which a deleted action can be useful. Appendix A proves equality. An optimizing signal can be unavailable to actual image probes, and numerical LP outputs are floating-point diagnostics rather than exact proofs.

**Preserving good paths.** For a question tree $\pi$, let $J_A(\pi)$ include optimal terminal actions and the common question cost. Define regret $r_A(\pi) = V_A^\star - J_A(\pi)$ and same-tree deletion loss $\ell_B(\pi) = J_A(\pi) - J_B(\pi)$. A direct rearrangement gives

$$G = \min_\pi \{r_A(\pi) + \ell_B(\pi)\}. \tag{4}$$

Thus a small loss on a near-optimal tree suffices, even when a bad tree suffers a large loss. For $\delta \ge 0$, writing $b_B(\delta) = \min_{r_A(\pi) \le \delta} \ell_B(\pi)$ gives

$$\min\{\delta, b_B(\delta)\} \le G \le \delta + b_B(\delta). \tag{5}$$

In a finite tree class, $G = 0$ exactly when at least one $A$-optimal tree has zero deletion loss. This condition is more decision-relevant than a large worst-observation diagnostic.

Table 2: **The same endpoints can hide a question reversal.** Exact full-gallery AP; equal $D = 6$, $K = 5$, common gallery and prior. Values for $q_0, q_1$ force one noiseless answer without charging. The final rows instead optimize paid policies independently.

| Decision | $C + r_u$ | $C + r_d$ | Difference |
|---|---|---|---|
| Blind | $239/360$ | $239/360$ | 0 |
| Known condition | 1 | 1 | 0 |
| Question $q_0$ | $77/90$ | $115/144$ | $41/720$ |
| Question $q_1$ | $17/20$ | $17/20$ | 0 |
| Preferred question | $q_0$ | $q_1$ | reversal |
| Replanned cap one, $c = .1$ | — | — | $1/180$ |
| Replanned cap two, $c = .005$ | .99 | .99 | 0 |

## 4 AN EXACT SEPARATION, AND ITS LIMITS

An AP-realizable construction isolates the endpoint issue without optimization error. Four equally likely conditions define positive sets on six gallery items. A common four-action menu $C$ is extended by either a useful ranking $r_u$ or a redundant ranking $r_d$. Unit dot-product embeddings realize both five-action menus at $D = 6$. Both retain blind value $239/360$ and known-condition value one, yet their preferred one-answer questions differ (Table 2). The questions have identical answer entropy but different partitions of the conditions.

The useful action is strictly worse at every declared endpoint. A common-action mixture bounds its unrestricted deletion loss by $41/720$, attained by $q_0$. The redundant action is dominated by a common mixture at every belief. Such mixtures certify expected whole-ranking payoffs; a mixture is not a fused ranking. The complete rankings and certificate appear in Appendix B.

The example also prevents an overinterpretation. The $41/720$ diagnostic is not a paid gain. Replanning reduces the cap-one gap to $1/180$ at $c = .1$, and cap two at the natural-study tariff gives exactly zero. Exact enumeration identifies a bounded price interval in which deletion matters (Figure 3a). Endpoint insufficiency therefore does not imply that adding a ranker will improve every operational policy.

The auxiliary width result restricts scoring to nonnegative diagonal weights on at most $D$ coordinates and uniformly averages AP within exact-score ties (McSherry & Najork, 2008). Identity-specific lookup and ID-based tie breaking lie outside this class; the separate six-item witness has no such restriction (Appendix C).

## 5 WHAT HAPPENS IN LEARNED IMAGE LIBRARIES?

**Data and actions.** UT-Zappos (Yu & Grauman, 2014) supplies condition-specific shoe retrieval. DINOv2 and CLIP features (Oquab et al., 2024; Radford et al., 2021) support separate CSN-derived D128 libraries: four condition masks at two learning rates, repeated for three training seeds. CUB-200-2011 (Wah et al., 2011) supplies four attribute relations—bill shape, wing color, breast color, and back color—with six same-width actions at D32 or D128. Present-attribute intersections define relevance; shared absence and bird class are not substitutes. Every ranking is computed by an image function, not a label-aware teacher permutation.

**A reproducible deletion rule.** For each query, $B$ is the union of the first maximizing action at the blind endpoint and at each of four known-condition endpoints, in frozen action-index order. Duplicate winners are retained once. Thus $|B| \leq 5$: all 1,920 cells delete at least one action from $K = 8$ or 6. Every endpoint has a unique winner in all 1,920 cells. Their union is consequently the unique minimum feasible core, although the rule performs no cardinality optimization; target payoffs still prevent SOURCE-only deployment. Endpoint equality and zero tree loss are exact for rational conversions of the stored binary64 payoffs; these stored-value equalities do not establish symbolic exactness of the underlying computed image AP. Appendix G records cardinalities, numerical screening, and conditional denominators.

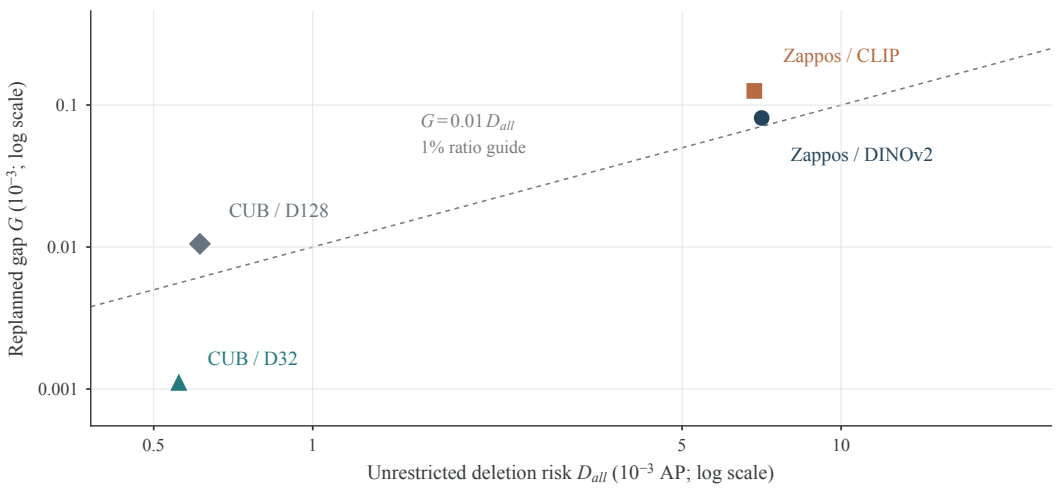


Figure 2: **Large diagnostic risk can coexist with a small policy gap.** Four cohort means compare unrestricted risk $\mathcal{D}_{\text{all}}$ and independently replanned gap $G$ at cap two, price .005. Both axes are logarithmic, in $10^{-3}$ AP for risk and AP-minus-price for $G$. The dashed 1% ratio guide is neither a fit nor a bound. Points have no uncertainty intervals. Means first average three libraries within each query. Preserved cells incur zero loss on every full-menu optimal tree. Table 1 and Appendix G retain diagnostics and preservation counts; three libraries share each query.

**Population and uncertainty.** All evaluated populations were historically exposed. Zappos selection uses 256 FIT, 128 TUNE, and 128 READOUT queries and a 1,736-image SOURCE gallery. A separate 206-query, 1,760-gallery transport cohort is disjoint from selection and model-training identities but was already used for final evaluation. CUB uses 256 eligible TEST-image queries, 128 probes, and 3,181 gallery images; image disjointness does not certify distinct individual birds. Intervals first average the three seed libraries within each query, then bootstrap query components (Efron, 1979). CUB also receives a species-cluster sensitivity analysis. These intervals are conditional on the common gallery, probe pool, and fitted models. A zero-containing interval is not an equivalence result.

**Positive risk, mostly preserved decisions.** Table 1 retains the full cap-two/.005 census. Of the cells with $\mathcal{D}_{\mathcal{Q},2} > 0$, every optimal tree is loss-free in 51/58 (DINOv2), 58/67 (CLIP), 102/105 (CUB D32), and 107/121 (D128): 318/351, rather than the unconditional 1,887/1,920. The other 1,569 cells have no positive allowed-probe loss. No optimal tie has differing deletion losses. Preservation therefore is not explained by no deletion or favorable optimal-tree selection, but the pooled count includes many cells without reachable risk.

Replanning compensation is small: Zappos recovers only $.0000194/.0000052$ from optimal-path losses $.0001004/.0001309$. Useful paths already avoid most diagnostic loss. Seed-averaged $G > .002$ occurs in 1/64 components per Zappos encoder and 0/256 per CUB width; small means do not rule out those tails.

**Preservation is not only identification.** For retained menu $B$, define perfect-condition opportunity $\Gamma_B(p) = \sum_s p_s \max_{b \in B} R_{s,b} - \Phi_B(p)$. Then $d_B(p) = \Gamma_B(p) - \Gamma_A(p)$ when known-condition endpoints are preserved. Terminal histories are classified as PRIOR, KNOWN, or intermediate; intermediate mass is split into common optimum ($\Gamma_B = 0$), preserved compromise ($\Gamma_B > 0, d_B = 0$), and lost action ($d_B > 0$). Table 3 shows substantial preserved-compromise mass. Perfect-condition opportunity is not the value of an available paid probe.

Erasing the second answer on the fixed tree raises deletion loss to $.0009668/.0010798$ on DINOv2/CLIP. This terminal-AP intervention does not replan or measure a new paid gap (Appendix H). A fixed-menu sensitivity analysis of answer price and question budget shows that, at .005, allowing one question increases mean $G$ to $.000696/.000827$, with 9/64 components per encoder above .002. At cap two, prices $0, .001, .005, .02$ keep both Zappos means below .000171

Table 3: **Uncertainty can remain while a compromise is preserved.** Mean terminal probability mass in percent on the original optimal trees. The five columns exhaust mass up to rounding; branch rows are not independent samples.

| Library | Prior | Known | Common optimum | Compromise | Lost action |
|---|---|---|---|---|---|
| Zappos / DINOv2 | 1.562 | 64.193 | 5.990 | 26.432 | 1.823 |
| Zappos / CLIP | .521 | 67.708 | 2.865 | 26.172 | 2.734 |
| CUB / D32 | 54.036 | 13.184 | 12.663 | 19.889 | .228 |
| CUB / D128 | 52.995 | 13.770 | 12.988 | 19.076 | 1.172 |

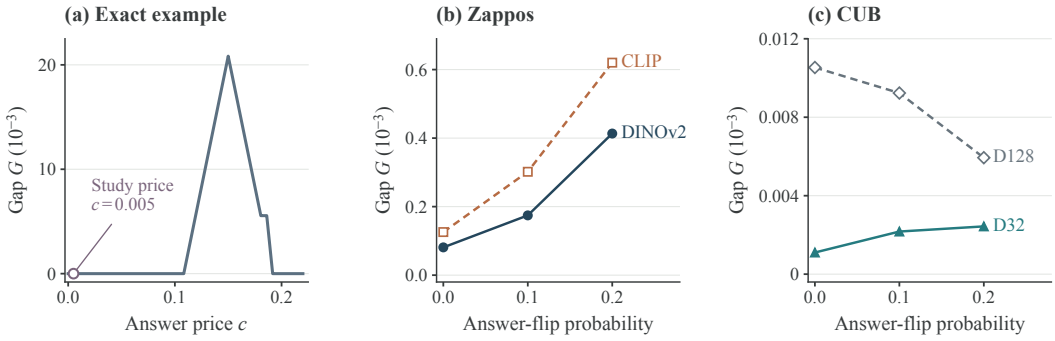


Figure 3: **Policy relevance depends on price and channel.** (a) Exact rational tariff segments for the six-item cap-two example. The gap is positive only for $13/120 < c < 23/120$; the study price $c = .005$ lies outside. (b–c) Retained numerical means under controlled answer noise on exposed image banks. Noise increases Zappos gaps but not consistently CUB gaps; separate scales reveal smaller CUB changes. Gaps use $10^{-3}$ AP-minus-price. Points are means, not uncertainty intervals; paired intervals appear in Appendix D.

(Appendix G). Unique endpoint winners require every endpoint-feasible subset to contain $B$. Thus all such supersets preserve optimal value in the original 1,887 zero-gap cells; the 33 positive-gap cells receive only an upper bound (Appendix G).

**A heterogeneous-library check.** Replacing the two-rate CSN menu by four CSN and four GeneCIS actions retains the query, gallery, relevance, and question law. Each library passes its own endpoint-preserving deletion check. Mean $G$ is $.00009152/.00033356$ for DINOv2/CLIP; seed-averaged $G > .002$ occurs in only 1/64 and 5/64 components. The paired increase has a positive conditional interval only for CLIP. Architecture, training, and width change together (CSN128/GeneCIS512), and shared GeneCIS actions are not independent encoder replications. A same-512D body-adaptation check changes $G$ in opposite directions, $+.00014479/-.00019364$. Thus neither pure diversity nor body adaptation alone explains the gap.

**Controlled noise tests a possible fragility.** An independent answer-flip channel with probability .1 or .2 is applied to the existing joint response law. The true-law policy is reoptimized separately for each menu. At noise .2, mean gaps rise to $.0004133/.0006203$ on Zappos but remain $.00000243/.00000594$ on CUB (Figure 3b–c). Zappos paired gap changes have positive conditional intervals; the response is not uniform across datasets. All four mean gaps remain below the declared .002 materiality threshold. That threshold is a decision criterion, not an equivalence margin established by a test.

Keeping an old noiseless question tree under the noisy law incurs a substantially larger regret than the deletion gap. At noise .2, mean regret ranges over noiseless optimal ties from .00418 to .00818 for DINOv2 and .00539 to .00968 for CLIP. Terminal actions are reoptimized under the new posterior in this diagnostic. The result localizes sensitivity to the question path; the experiment is not a frozen deployable actor under distribution shift.

Table 4: **Questioning helps; uncertainty-based menu selection shows no established advantage.** Paired utility differences on 206 exposed transport queries. Values are $10^{-3}$ AP-minus-price, with conditional pointwise 95% intervals. Each own-STOP comparison uses that method's terminal rule.

| Controlled comparison | DINOv2 | CLIP |
|---|---|---|
| Uncertainty menu − strong endpoint | $-2.35$ $[-5.12, .33]$ | $-.57$ $[-3.17, 1.86]$ |
| Endpoint DP − own STOP | $75.81$ $[66.92, 84.92]$ | $103.82$ $[93.93, 113.55]$ |
| Endpoint DP − entropy | $7.62$ $[3.14, 12.40]$ | $7.97$ $[2.48, 13.61]$ |
| Endpoint DP − one-step VOI | $2.99$ $[.57, 5.47]$ | $6.18$ $[2.86, 9.77]$ |
| Posterior fusion − own STOP | $54.24$ $[47.58, 61.07]$ | $76.31$ $[67.81, 84.81]$ |
| Posterior − prior fusion, same history | $63.59$ $[56.91, 70.45]$ | $86.14$ $[77.66, 94.63]$ |

## 6 USEFUL FEEDBACK DOES NOT ESTABLISH A BETTER SELECTOR

**A strong matched menu baseline.** The learned actor estimates world payoffs and a joint response model using SOURCE data, then uses cap-two dynamic programming. STOP, entropy, one-step value of information, and fixed-question policies share the same predicted payoffs. Target labels remain evaluator-only. The strong endpoint menu selects among blind-only, known-condition, and equally weighted endpoint candidates using TUNE utility. This is an explicitly disclosed development amendment, not independent confirmation.

Table 4 separates useful questioning from menu selection. The uncertainty-aware four-action menu has no established advantage over the strong four-action endpoint menu. Eight directional contrasts cross two encoders, forced-one/optional-two policies, and two recipient menus. All pointwise intervals include zero; a strict reversal between these selected menus remains unestablished (Appendix D). Different terminal rules have different own-STOP values; the magnitudes of their gains cannot rank final methods.

**Separating candidate generation from validation.** A later matched factorial uses endpoint or feedback-aware candidate generation on FIT and endpoint or utility validation on TUNE. Each family receives three candidates and the same four-of-eight selection budget. All four arms choose the same menu in each of the six Zappos libraries; both primary contrasts are exactly zero on READOUT and the old transport cohort. This is an identity of selected actions, not statistical equivalence between arbitrary selectors.

The corresponding CUB comparison uses 367 SOURCE FIT and 367 TUNE queries and the existing 256-query evaluation. D128 again selects the same menu in all four arms. D32 utility validation changes value by $-.00000664$, and feedback-aware generation changes value by $+.00004376$ relative to its matched comparator; both intervals include zero. Independent dynamic-programming recomputation agrees within $2.22 \times 10^{-16}$. These tests close a selection-budget confound without establishing a new selector advantage.

**Fusion is a substantive control.** Posterior score fusion uses SOURCE/TUNE-fixed normalization and world-to-action weights. Comparing posterior with prior weights on identical purchased histories isolates use of the answers; comparing with its own STOP includes the tariff. Both effects are positive on Zappos (Table 4). Reciprocal-rank fusion also gains $.058160/.079991$ over its own STOP. The question policy remains optimized for single-action predictions, so these results do not claim fusion-optimal acquisition.

On CUB, posterior weighting improves over prior weighting by $.000111/.000096$ at D32/D128, yet paid fusion loses $.000400/.000373$ against its own STOP, with negative intervals. Better use of an answer and a worthwhile purchase are different claims. The available evaluator headroom does not imply that the learned actor can recover that headroom.

**Correcting the channel, not only the terminal.** A crossed response-noise test fixes the latent actor, predicted payoffs, global four-action menu, tariff, and joint response law. $C_{ij}$ uses old or corrected question planning ($i$) and terminal posterior ($j$). After the known-rate control, 2,048 synthetic SOURCE answer pairs give $\widehat{\eta} = .1557617$; evaluation retains the frozen flip rate $\eta = .15$. Thus

Table 5: **Question correction does not transfer uniformly across terminal rules.** Four actions, 206 exposed transport components, cap two, price .005. Top rows report $\Delta U$ in $10^{-3}$ with pointwise conditional 95% intervals. Bottom rows subtract the single-ranker correction effect from each fusion effect; nominal family-four 98.75% intervals are post-observation diagnostics.

| Terminal / contrast | DINOv2 | CLIP |
|---|---|---|
| Single: $C_{11} - C_{01}$ | $4.651\ [2.589, 6.827]$ | $2.488\ [.536, 4.519]$ |
| RRF: new − old question | $2.234\ [.911, 3.603]$ | $1.079\ [-.256, 2.406]$ |
| Z-score: new − old question | $.813\ [-.451, 2.109]$ | $-.566\ [-1.939, .774]$ |
| RRF − single interaction | $-2.417\ [-4.194, -.783]$ | $-1.409\ [-3.046, .133]$ |
| Z-score − single interaction | $-3.838\ [-6.070, -1.939]$ | $-3.053\ [-4.888, -1.305]$ |

$C_{11} - C_{01}$ changes questioning while holding the corrected terminal fixed. Neither target labels nor target utility select the correction.

Single-ranker utility improves (Table 5), but $\Delta\,\mathrm{AP} = .001890/.000428$ has zero-containing intervals, while $\Delta Q = -.55214/-.41189$. The tariff term accounts for approximately 59%/83% of the utility point estimates: accounting, not causal mediation or GPU savings. A SOURCE-selected old-order stopping control reduces but does not remove the utility gain: $C_{11}$ adds $.003303/.002409$, with family-two 97.5% intervals $[.001082, .005535]/[.000152, .004737]$; both adjusted AP intervals include zero.

At exactly two answers, corrected planning instead improves single-ranker AP by $.002517/.003380$ over terminal-specific SOURCE-selected controls; both adjusted intervals are positive. Fewer purchases cannot explain this comparison. The signs persist across eight synthetic calibration draws, but all four corresponding fusion intervals include zero. CUB does not establish transport of the gain (Appendix K).

Applying both question policies to the same posterior-fusion terminal weakens the effect. Z-score fusion loses AP by $.001947/.002625$, with negative intervals. The paired question-by-terminal interaction has $\Delta Q = 0$, hence identical AP and utility contrasts; three of four adjusted intervals are negative. The actor was not separately optimized for fusion, so this diagnoses limited portability of the tested correction, not impossibility of fusion-aware acquisition. Fusion remains outside the original single-ranker oracle action class.

On CUB, the analogous four-action utility changes are $+.0002520/+.0001966$, but AP falls and own-STOP superiority remains unresolved. Corrected myopic and dynamic policies coincide on this finite panel. Avoiding harmful purchases is not new feedback headroom. Earlier complete-label oracle and calibration controls remain in Appendix F.

## 7 Can a better target or more training close the gap?

A history-conditioned readout adds a residual constrained to zero at the blind and known-condition endpoints. With body adaptation, this endpoint-null readout gains .004087 over the world-conditioned readout and .036330 over its own STOP, yet loses .017911 against posterior fusion (95% interval $[-.023954, -.012352]$). Body adaptation versus a frozen history readout remains unresolved. Appendix E retains the training and comparator contracts.

Rollout supervision also improves over STOP and direct-Q, but has utility differences $-.010232/-.014274$ against the existing DINOv2/CLIP latent actors, with negative family-adjusted intervals (Figure 4). Rollout and value-target readouts remain unresolved. SOURCE selection retains the old actor in every library: selected-policy change is exactly zero. Improvement over weaker comparators does not establish a stronger final policy.

**Further repairs remain inconclusive.** Larger-support and terminal-matched repairs remain inconclusive. A fixed-budget fusion learner has four negative target means and one negative adjusted interval; two SOURCE response-model repairs also fail against strong controls (Appendices E and L). These results concern the tested protocols.

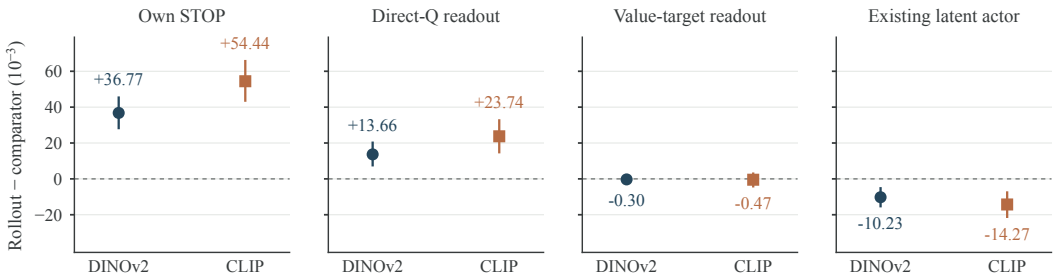


Figure 4: **A gain over STOP can coexist with a loss against a strong actor.** Paired rollout-target utility contrasts ($10^{-3}$ AP-minus-price) on 128 exposed READOUT queries. Thin uncapped lines show nominal 99.1667% intervals; the primary six-contrast family contains the value-target and existing-actor comparisons, whereas STOP and direct-Q comparisons are secondary. Both fitted policies replace only the root decision: value-target supervision assumes optimal child continuation, while rollout-target supervision uses the frozen existing child. The fitted value-target comparator is not a performance bound.

**The tested shared image functions did not realize the reversal.** The six trained directional gaps remain negative, from $-.007229$ to $-.003587$ (Appendix E). A separate evaluator-only census finds endpoint-preserving submenu reversals in 13/1,920 cells, six with positive paid value at cap one. A fixed native-rank rational-AP case reverses the optimal question with regrets $.004560/.002970$ (Appendix N). These natural submenu examples do not establish successful shared-representation training.

## 8 Scope and Practical Implications

The optimal-value and actor tests answer different questions. On CUB, evaluator-derived restriction improves average actor utility, yet individual zero-gap cells can lose utility (Appendix J). Publicly predicted cores give four unresolved primary contrasts against the full menu. Against equal-size menus containing the same public blind winner, both replanned utility means are negative (Appendix M).

All populations are exposed. The 1,527 unaccounted images in 783 components fail the four-relation complete-label requirement; global non-exposure remains unverified (Appendix D). True-law frontiers remain evaluator diagnostics. The tariff excludes encoding, storage, fusion, and fitting; evaluation computes the full library, and CLIP also uses a DINO response channel. Direct execution failed cached action/channel agreement in all 87 cases, so its latency cannot support the cached utility claim.

## 9 Conclusion

Endpoint scores do not determine partial-feedback value, as exact and natural-image submenu examples show. Such existence does not establish a deployable selector: publicly predicted endpoint cores retain no matched replanning advantage. Preserved compromises explain part of the small natural deletion gaps. Fixed-budget correction improves single-ranker question quality without establishing uniform fusion gains. Representation adequacy and policy portability therefore require separate tests under explicit information, action, and cost assumptions.

## AI USE STATEMENT

Generative AI tools assisted hypothesis refinement, mathematical formulation and proof development, experimental design and implementation, numerical analysis, literature search and reference verification, figure preparation, and manuscript drafting and editing. The authors are responsible for the final content.

## ETHICS STATEMENT

The experiments use existing image datasets and pretrained encoders for conditional retrieval, rather than identification of individuals. Dataset and model terms remain applicable to redistributed artifacts. Question prices are stipulated evaluation costs, and cached utility does not establish measured annotation costs or end-to-end serving savings.

## REPRODUCIBILITY STATEMENT

Appendices A, B, and C give the preservation proofs, exact six-item construction, and restricted width result. Appendices D–F specify protocols and adverse controls; Appendices G–I retain deletion denominators, stored price/cap results, state occupancy, and response-cost accounting. Appendices J–L detail the actor, mandatory-budget, and response-model controls. Appendix M gives public-menu controls and Appendix N gives the natural-image submenu census.

## REFERENCES


Kiril Bangachev, Guy Bresler, Jonathan Kogan, and Yury Polyanskiy. Is dimensionality a barrier for retrieval models?, 2026. URL https://arxiv.org/abs/2605.23556v1. arXiv:2605.23556.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953. URL https://doi.org/10.1214/aoms/1177729032.

Anthony R. Cassandra, Michael L. Littman, and Nevin Lianwen Zhang. Incremental pruning: A simple, fast, exact method for partially observable Markov decision processes. In *UAI*, pp. 54–61, 1997. URL https://arxiv.org/abs/1302.1525.

Yiling Chen and Bo Waggoner. Informational substitutes. In *FOCS*, 2016. URL https://arxiv.org/abs/1703.08636. Full version, arXiv:1703.08636.

Gordon V. Cormack, Charles L. A. Clarke, and Stefan Büttcher. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In *SIGIR*, 2009. URL https://plg.uwaterloo.ca/~gvcormac/cormacksigir09-rrf.pdf.

Morris H. DeGroot. Uncertainty, information, and sequential experiments. *The Annals of Mathematical Statistics*, 33(2):404–419, 1962. URL https://doi.org/10.1214/aoms/1177704567.

Bradley Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1): 1–26, 1979. URL https://doi.org/10.1214/aos/1176344552.

Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011. URL https://arxiv.org/abs/1003.3967. Revised proof in arXiv version 5, 2017.

Daniel Golovin, Andreas Krause, and Debajyoti Ray. Near-optimal Bayesian active learning with noisy observations. In *NeurIPS*, 2010. URL https://arxiv.org/abs/1010.3091.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *ICML*, volume 70, pp. 1321–1330, 2017. URL https://proceedings.mlr.press/v70/guo17a.html.

Zhuoyue Huang, Freddie Bickford Smith, and Tom Rainforth. Loss-driven Bayesian active learning. In *AISTATS*, volume 300, pp. 5140–5148, 2026. URL https://proceedings.mlr.press/v300/huang26b.html.

Shervin Javdani, Yuxin Chen, Amin Karbasi, Andreas Krause, Drew Bagnell, and Siddhartha Srinivasa. Near optimal Bayesian active learning for decision making. In *AISTATS*, volume 33, pp. 430–438, 2014. URL https://proceedings.mlr.press/v33/javdani14.html.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1–2):99–134, 1998. URL https://doi.org/10.1016/S0004-3702(98)00023-X.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101 (6):2590–2615, 2011. URL https://doi.org/10.1257/aer.101.6.2590.

Aditya Kusupati, Gantavya Bhatt, Aniket Rege, Matthew Wallingford, Aditya Sinha, Vivek Ramanujan, William Howard-Snyder, Kaifeng Chen, Sham Kakade, Prateek Jain, and Ali Farhadi. Matryoshka representation learning. In *NeurIPS*, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/c32319f4868da7613d78af9993100e42-Abstract-Conference.html.

Frank McSherry and Marc Najork. Computing information retrieval performance measures efficiently in the presence of tied scores. In *ECIR*, 2008. URL https://marc.najork.org/papers/ecir2008.pdf.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. URL https://arxiv.org/abs/2304.07193.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *ICML*, volume 139, pp. 8748–8763, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Reuben Tan, Mariya I. Vasileva, Kate Saenko, and Bryan A. Plummer. Learning similarity conditions without explicit supervision. In *ICCV*, 2019. URL https://arxiv.org/abs/1908.08589.

Sagar Vaze, Nicolas Carion, and Ishan Misra. GeneCIS: A benchmark for general conditional image similarity. In *CVPR*, 2023. URL https://sgvaze.github.io/genecis/.

Andreas Veit, Serge Belongie, and Theofanis Karaletsos. Conditional similarity networks. In *CVPR*, 2017. URL https://arxiv.org/abs/1603.07810.

Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The Caltech-UCSD Birds-200-2011 dataset. Technical report, California Institute of Technology, 2011. URL https://authors.library.caltech.edu/records/cvm3y-5hh21.

Orion Weller, Michael Boratko, Iftekhar Naim, and Jinhyuk Lee. On the theoretical limitations of embedding-based retrieval. In *ICLR*, 2026. URL https://arxiv.org/abs/2508.21038v2.

Aron Yu and Kristen Grauman. Fine-grained visual comparisons with local learning. In *CVPR*, 2014. URL https://www.cs.utexas.edu/~grauman/papers/yu-cvpr2014.pdf.

Shengjia Zhao, Michael P. Kim, Roshni Sahoo, Tengyu Ma, and Stefano Ermon. Calibrating predictions to decisions: A novel approach to multi-class calibration. In *NeurIPS*, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/bbc92a647199b832ec90d7cf57074e9e-Abstract.html.

## A DELETION AND PRESERVATION PROOFS

The state space and menus are finite, and menus are nonempty; payoffs are bounded; priors assign nonnegative mass summing to one. The question-tree class is finite and common to both menus. Terminal actions do not affect the observation law. Randomization cannot improve the maximum of the finite linear policy objective.

**Proof of Equation 3.** For any signal, collect each positive-loss posterior atom under an action $a \in A \setminus B$ maximizing the full-menu payoff at that atom. Write its unnormalized state mass as $z_{a,s}$. Pooling atoms assigned to the same $a$ cannot decrease the minimum over $b \in B$ of their linear differences. Their total loss is therefore no larger than a feasible LP value. Zero-loss atoms can be omitted, giving the componentwise mass inequality.

Conversely, for a feasible $z, t$, create one signal atom for every $a$ with mass $m_a = \sum_s z_{a,s} > 0$ and posterior $z_a/m_a$. Add the residual prior mass as another atom. At the $a$ atom,

$$m_a\{\Phi_A(z_a/m_a) - \Phi_B(z_a/m_a)\} \geq \min_{b \in B} \sum_s z_{a,s}(R_{s,a} - R_{s,b}) \geq t_a.$$

The residual contributes nonnegative loss. Thus the largest signal loss is at least every feasible LP objective. The two inequalities establish equality. A zero-mass atom contributes zero and requires no posterior definition.

**Same-tree bound and frontier identity.** For an $A$-optimal tree $\pi_A$, $V_B^\star \geq J_B(\pi_A)$, hence $G \leq J_A(\pi_A) - J_B(\pi_A) \leq \mathcal{D}_{\mathcal{Q},q}$. Also,

$$V_B^\star = \max_\pi [J_A(\pi) - \ell_B(\pi)] = V_A^\star - \min_\pi [r_A(\pi) + \ell_B(\pi)].$$

For trees with regret at most $\delta$, loss is at least $b_B(\delta)$; for the remaining trees, regret exceeds $\delta$. This proves the lower bound by splitting the minimization. A tree attaining $b_B(\delta)$ proves the upper bound. Nonnegativity of both terms gives the zero-gap equivalence.

**Loss on optimal ties.** Let $b_B(0)$ and $u_B(0)$ denote the smallest and largest deletion loss among full-menu optimal trees. A positive $u_B(0)$ alone does not imply a paid gap because another optimal tree may have zero loss. The reported census explicitly evaluates both values. Equality $b_B(0) = u_B(0)$ holds in every tested query–library cell, eliminating this tie-selection ambiguity in the reported analysis.

**Numerical checks.** Exact AP values are checked using two independent formulas: positive-rank precision summation and complete relevance-vector accumulation. Image-derived floating-point payoffs receive an independent dynamic-programming check; this check does not establish symbolic exactness. Controlled-noise policies are checked by explicit rational tree calculations against Bellman recursion, with maximum discrepancy $3.33 \times 10^{-16}$. The CUB matched-menu check has discrepancy $2.22 \times 10^{-16}$; the earlier condition-specific results reproduce within $3.33 \times 10^{-16}$. The entropy-only gap-forecast check preserves the frozen forecast predictions exactly and reproduces the control statistics to floating precision.

## B THE EXACT SIX-ITEM WITNESS

The positive sets in worlds $s_0, \ldots, s_3$ are $\{0,1,2\}$, $\{3,4,5\}$, $\{0,2,4\}$, and $\{1,3,5\}$. The prior is uniform. Common menu $C$ contains the rankings

$$c_0 = (1,2,0,5,3,4), \qquad c_1 = (4,3,5,0,2,1),$$
$$c_2 = (4,2,0,5,3,1), \qquad c_3 = (1,3,5,4,2,0).$$

The candidate additions are $r_u = (0,2,3,1,4,5)$ and $r_d = (0,3,2,5,1,4)$, with payoff vectors

$$R_u = (11/12, 37/90, 13/15, 4/9), \qquad R_d = (34/45, 1/2, 13/18, 8/15).$$

For gallery vectors $g_i = e_i \in \mathbb{R}^6$, a query action assigns coordinates $6, 5, \ldots, 1$ in the desired order and divides by $\sqrt{91}$. All vectors have unit norm, rankings are strict, and both menus have the same dimension and common gallery. This is a row-specific construction, not a shared learned kernel.

Question $q_i$ asks whether item $i$ is relevant. The useful action's blind score is $95/144$, below the common optimum by $1/240$, and each known-condition score is below one. The common-action mixture $(13/16, 0, 3/16, 0)$ bounds its positive coordinatewise residual by $(0, 0, 41/180, 0)$. Prior averaging gives $\mathcal{D}_{\text{all}} \le 41/720$, and $q_0$ attains the bound. The redundant action is dominated by $(R_{c_0} + R_{c_2})/2$, so its loss is zero at every posterior.

Exact cap-two enumeration yields the following paid gap:

$$G(c) = \begin{cases} 0, & 0 \le c \le 13/120, \\ c/2 - 13/240, & 13/120 \le c \le 3/20, \\ 23/240 - c/2, & 3/20 \le c \le 13/72, \\ 1/180, & 13/72 \le c \le 67/360, \\ 23/120 - c, & 67/360 \le c \le 23/120, \\ 0, & 23/120 \le c \le 1. \end{cases}$$

The maximum is $1/48$ at $c = 3/20$. The operational gap is therefore smaller than the worst-signal diagnostic even at its best tariff. The natural experiments retain their own fixed .005 tariff; the toy price window does not retrospectively change that choice.

## C A RESTRICTED WIDTH–QUALITY LAW

For integers $1 \le A \le L$ giving the positive count and tie-block size, define

$$F(A, L) = \frac{A}{L} + \left(1 - \frac{A}{L}\right) \frac{H_L - 1}{L - 1},$$

where $H_L$ is the harmonic number. The following spectrum concerns the stated diagonal-support class only. The gallery consists of all flip masks $M \subseteq [m]$. A hidden set $S$, $1 \le |S| = s < m$, declares $M$ relevant exactly when $M \cap S = \varnothing$, giving $A_+ = 2^{m-s}$ positives. For integer $D \ge 1$, an admissible action has $w_i \ge 0$, $\sum_i w_i = 1$, support at most $D$, and score $1 - 2\sum_{i \in M} w_i$. AP averages uniformly over permutations inside exact-score ties. The convention $F(1, 1) = 1$ handles the singleton case.

**Theorem 1.** *For $1 \le D \le s$, the maximal tie-averaged AP is $F(A_+, 2^{m-D})$. Equality holds exactly when the support contains $D$ coordinates, all in $S$. For $D \ge s$, the optimum is one.*

*Proof.* Let $t$ support coordinates lie in $S$. If $t < s$, fixing all support bits partitions the gallery into equal-score fibers. Fibers with any retained relevant bit flipped are wholly negative and can be deleted for an upper bound. Each remaining fiber has positive fraction $p = 2^{-(s-t)} \in (0, 1)$. Merging equal-density blocks cannot lower expected AP, by the argument below. The merged block has $2^{m-t}$ items and all $A_+$ positives, yielding $F(A_+, 2^{m-t})$. For fixed $A_+$, inserting an extra negative into a uniformly permuted block strictly lowers expected AP. Since $t \le D$, the bound follows. A support of $D$ relevant coordinates attains the bound in its top tie. If $t < D$, monotonicity is strict; if $t = D$, the width constraint leaves no irrelevant support. When $t = s$, AP is at most one and relevant-only support attains one. □

**Equal-density merge.** Suppose ordered tie blocks have sizes $n_j$ and exactly $pn_j$ positives, $0 < p < 1$. Let $K = \sum_j n_j$ and $A = pK$. At rank $r = B + i + 1$ of a size-$n$ block, after $B$ items, the expected AP contribution is

$$\frac{p}{Ar}\left[pr + (1 - p)\left(1 - \frac{i}{n - 1}\right)\right].$$

Only the correction sequence depends on the partition. For a prefix of length $q = B + \ell$, put $y = n - \ell$ and let $C = K - B - n$. The merged-minus-separated correction prefix sum is

$$D_q = \frac{\ell(\ell - 1)C + y(y - 1)B + BC(n - 1)}{2(K - 1)(n - 1)} \ge 0.$$

The full sum is zero. Summation by parts gives

$$\mathbb{E}\,\mathrm{AP}_{\mathrm{merged}} - \mathbb{E}\,\mathrm{AP}_{\mathrm{separated}} = \frac{p(1-p)}{A} \sum_{q=1}^{K-1} D_q \left( \frac{1}{q} - \frac{1}{q+1} \right) \geq 0.$$

Multiple blocks give a strictly positive prefix difference at the first boundary. The case $p = 1$ is handled directly: every remaining item is positive. Unequal-density blocks need not obey this lemma. The restrictions are indispensable; an unrestricted nonlinear encoder can escape a diagonal support constraint.

## D PROTOCOLS AND UNCERTAINTY

**Data roles.**

**What the Zappos questions reveal.** The four hidden conditions are SubCategory, Gender, HeelHeight, and Closure. Each metadata value is the sorted tuple of its distinct, stripped semicolon-separated entries; missing values are excluded by the complete-case rule. A probe is a candidate shoe image, and a queried pair consists of the query image and that probe. The binary answer states whether the pair has the same categorical tuple under the hidden condition. The evaluator uses these labels to generate the true joint answer law; noisy experiments additionally flip each answer independently at the stated rate. The policy does not observe the condition or those labels. SOURCE-trained categorical predictions supply a shared-query latent response model, preserving dependence between answers about the same query rather than multiplying independent question marginals. Both DINOv2 and CLIP ranker comparisons retain the DINO response channel. This learned law and the evaluator's true response law are distinct.

Zappos libraries contain two learning rates crossed with four condition masks, analyzed separately for DINOv2 and CLIP and repeated for three seeds. Product IDs and identical raw-image SHA-256 values define connected components before selection. Among 4,512 complete-case SOURCE development images, eligible queries have a positive example outside their component under each of the four relations. A fixed salted SHA-256 ordering of image IDs selects representatives from the first 64 distinct components. Each gallery includes all bank images outside its query's component; eight probes follow a separate fixed query–candidate ID hash order. Relation labels determine eligibility, but query and probe ordering use no retrieval payoffs. Hash ordering does not define the components. Natural-ranking ties use a separate image-ID hash, distinct from the restricted theorem's uniform tie averaging. Selection instead uses 256 FIT, 128 TUNE, and 128 READOUT queries and a 1,736-image gallery. The old 206-query transport set uses a 1,760-image gallery. Disjoint selection identities do not erase historical exposure of that set.

CUB relations use present attributes with certainty at least three. Official TRAIN supplies 2,935 FIT and 734 DEV images. Eligible TEST-image components supply 256 queries, 128 probes, and 3,181 gallery images. Four 32-image probe pools are assigned to query subsets, with four candidate pairs per session. Each width has six actions, a frozen DINOv2 backbone, 1,024 head updates and 512 observer updates per seed. CUB is a task-specific fitted evaluation, not zero-shot transfer from Zappos.

**Question-policy transfer.** The eight contrasts cross DINOv2/CLIP, exactly one forced question/optional STOP with a two-question cap, and strong-endpoint/uncertainty-selected recipient menus. At price .005, a contrast subtracts the other menu's question/STOP policy from the recipient's own policy while reselecting terminal actions from the recipient menu at every reached history. Donor terminal actions are not copied. The strong-endpoint menu is selected by TUNE paid utility among blind, known, and balanced candidates. Both encoders retain the existing DINO categorical response channel. Worlds and three seeds are averaged within each of 206 exposed query components before 4,000 paired bootstrap draws. All eight descriptive pointwise 95% intervals contain zero; no simultaneous family or independent menu-pair replication is implied.

**Eligibility of previously unaccounted images.** An identity census of the 7,553-image historical FINAL partition excludes whole product/exact-duplicate components used by known evaluation consumers, leaving 1,527 images in 783 components. None overlaps the 4,390-image four-relation

complete-case roster by ID or component. These candidates therefore cannot supply confirmation under the original label requirement, despite cached feature IDs. Global non-exposure and an actual actor/channel evaluation on these candidates remain unverified. No feature values, new payoffs, or missing-label replacements enter this identity check.

**Near-optimal-path sensitivity.** The mean minimum deletion losses $b_B(\delta)$ at $\delta = 0, .0005, .002, .005$, in $10^{-3}$ AP units, are DINOv2 $(.100438, .079807, .079807, .079807)$, CLIP $(.130897, .130897, .130897, .130897)$, CUB D32 $(.001104, .000528, .000528, 0)$, and CUB D128 $(.016009, .013486, .004757, .000081)$. These evaluator frontiers do not alone establish better policies: the sum of regret and deletion loss determines the replanned gap.

**State and response-correction checks.** The original 1,920 cells yield 3,709 terminal rows and 7,680 erasure rows. Original same-tree deletion losses reproduce exactly. The heterogeneous follow-up has 384 cells but only 64 query components; each seed's GeneCIS actions are shared across encoders. Full-gallery AP checks preserve within-library endpoints, not endpoints across the homogeneous and heterogeneous menus. SOURCE-rate and fusion results use one fixed synthetic calibration sample; rates are not natural-user estimates. The 16 question-by-terminal comparisons retain four primary four-action interactions on the original transport cohort. All intervals remain conditional on exposed queries, fitted actors, and shared galleries.

**Joint responses and terminal payoffs.** Zappos response modeling retains joint latent response histories rather than multiplying independent per-question marginals. True-law evaluation preserves the joint distribution before adding independent answer flips. The initial actor treats world payoffs as fixed conditional on the world and public features, leaving within-world history dependence to the later matched readout tests. Evaluator access to true AP must not be counted as deployable information.

**Intervals and independence.** World outcomes are first averaged under the fixed prior; library seeds are then averaged within each query. Query components, not world rows or query–seed pairs, are resampled. Species clustering is an additional CUB sensitivity analysis. The intervals do not account for repeated adaptive reuse of these populations or variability over new galleries. Main selection intervals are descriptive and pointwise. The rollout-target primary family contains six comparisons and uses nominal 99.1667% intervals; secondary controls are labeled separately. No practical-equivalence claim follows from an interval that includes zero.

At noise .2, paired $\Delta G$ intervals are $[.0000979, .0006253]$ for Zappos DINOv2 and $[.0002092, .0008252]$ for CLIP. CUB D32's tiny increase and D128's decrease do not supply a universal noise law. CUB noiseless-tree regret ranges across optimal ties are .0009471–.0009747 and .0010558–.0010865, respectively; terminal actions are reoptimized, as in the Zappos diagnostic.

## E SECONDARY REPAIR AND FORECASTING RESULTS

**Endpoint-neutral and rollout-target readouts.** A history-conditioned readout gains from expanding the finite terminal catalogue without establishing superiority over continuous fusion. The endpoint-null residual uses 20,372 images, eight epochs, and three seeds, comparing a frozen DINO body with adaptation of the final two blocks. The residual is exactly zero at blind and known-world endpoints. On the old 206-query cohort, the body-adapted history readout gains .004087 over the world-conditioned readout and .036330 over its own STOP, but loses .017911 against posterior fusion, with conditional 95% interval $[-.023954, -.012352]$. The body-adaptation contrast remains unresolved.

The matched MLP comparison changes the root question-advantage target. Value-target supervision averages future answers under the fixed public response model and assumes an optimal child continuation; rollout-target supervision instead fixes that continuation to the existing actor. Both fitted policies replace only the root decision and execute the same frozen child thereafter. Targets are expected continuation minus STOP value and price .005, not scalar state values or full-information bounds. Rollout gains $.036774/.054443$ over STOP and $.013658/.023741$ over direct-Q, but has utility differences $-.010232/-.014274$ against the existing actor. The six primary contrasts cross

the two encoders with rollout − value-target, rollout − existing actor, and value-target − existing actor. Seed-averaged resampling of 128 exposed READOUT queries uses 10,000 draws and nominal 99.1667% intervals. STOP, direct-Q, and AP-ridge comparisons are secondary. All four encoder-by-teacher TUNE choices retain the old actor; selected-policy change is exactly zero.

Both new teachers use 2,048 queries (the original 256 FIT plus 1,792 earlier backbone/channel TRAIN queries), identical 634-dimensional public inputs, and a $634 \to 256 \to 128 \to 8$ GELU network. Inputs pool hypothetical worlds under the public posterior; no actual hidden world is supplied. The 634 coordinates comprise 522 base features, 28 relevance-free ranking-geometry features, and 84 geometry-by-world-contrast features. AdamW uses learning rate .0003, weight decay .0001, 512 updates, and batches of 2,048 sampled states. Targets are divided by .05 during fitting and predictions rescaled. Root states receive half the sampling mass; one-answer states share the other half according to public history mass. Loss divides summed squared error by the count of valid outputs, so eight root versus seven child outputs preclude an exactly half-and-half loss interpretation.

The older direct-Q ridge predicts question advantages at both decision depths; the AP ridge predicts fixed-fusion AP changes from root to history. Both use 256 FIT queries, the 634-feature blocks, and penalty .01. Only the new teacher pair matches support, architecture, inputs, and child execution. The existing actor instead predicts $4 \times 8$ native-action AP from 256 query-embedding coordinates, choosing penalty from $.1, 1, 10$ by TUNE MSE; its terminal uses four-action Z-score fusion. TUNE128, READOUT128, and the 1,736-image gallery have disjoint component roles but are historically exposed. Utility excludes geometry computation, fitting, and fusion costs.

**Unrealized natural reversal.** At D32, the baseline and endpoint-constrained image-training rules each receive 4,096 updates in three contexts with shared initialization, draws, common actions, and full-gallery AP forward computation. All six directional gaps remain negative ($-.007229$ to $-.003587$), while endpoint excess ranges from $.165049$ to $.295079$. Endpoint-preserving feasible-path alternatives keep the original question preference. Full-gallery AP recomputation differs by at most $2.56 \times 10^{-15}$. The failed recipes do not establish information impossibility or justify changing the criterion after observing results.

**More support does not yet close the strong-baseline gap.** A nested support experiment expands the AP learner from 256 to 512, 1,024, and 2,048 queries, retaining a fixed public-feature map, regularization, and question tariff. TUNE selects 512 for DINOv2 and 1,024 for CLIP. On the fixed READOUT cohort, selected − existing-actor utility is $+.001211$ and $+.000301$, with conditional 97.5% intervals $[-.001886, .004393]$ and $[-.001624, .002289]$. The additional support comes from earlier backbone/channel TRAIN images, so sample count and TRAIN mixture both change; no new population is introduced.

**Matching the target to the executed terminal rule.** A separate ridge experiment holds the 634 public inputs, regularization $\lambda = .01$, FIT weights, and TUNE utility criterion fixed while replacing the single-action target with the actual fusion-terminal target. Both targets retain the same fusion terminal at evaluation. The 37,888 augmented rows represent only 256 distinct FIT queries, not a larger query population. Fusion-target − single-action-target utility on READOUT is $-.0002745$ for DINOv2 and $+.0006364$ for CLIP; conditional 97.5% intervals are $[-.0017856, .0010751]$ and $[-.0025735, .0037451]$. The matched comparison removes an input-feature explanation but does not establish a target-alignment advantage. Thus neither larger support nor terminal-target alignment supplies a confirmed strong-baseline improvement.

**Predicting a material deletion gap.** A SOURCE-frozen ridge forecast predicts operational gap from endpoint features, unrestricted risk, frontier descriptors, or a zero baseline. This target concerns a SOURCE-selected global four-of-six menu, not the query-specific endpoint-preserving deletion in Table 1. Predictions using public inputs are fixed before evaluation payoffs are examined. A separate entropy-only feature control uses the same fit and tariff/cap inputs. At cap two and .005, the frontier forecast's mean absolute error is $.000923/.000784$ on CUB D32/D128, compared with $.000595/.000516$ for zero. The extra diagnostic structure does not improve this operational forecast. The choice rule uses a fixed .002 threshold to select the four-action menu or the full six actions;

Table 6: **Unrecovered value is not all feedback headroom.** Complete-label protocols at cap two and answer price .005. Utility and gaps are in $10^{-3}$ AP-minus-price; $\mathbb{E}Q$ is the mean number of answers. Comparisons within a row share a problem; differences across rows are not causal effects of width or dataset. These exposed historical protocols are distinct from the matched four-action menu test in Table 4.

| Protocol | Learned $U$ | Own-STOP gain | $\mathbb{E}Q$ | Feedback headroom $V_2^\star - V_0^\star$ | Actor shortfall $V_2^\star - U$ |
|---|---|---|---|---|---|
| Zappos D32 | 575.404 | +36.471 | 1.6185 | 64.877 | 37.886 |
| Zappos D128 | 577.483 | +38.378 | 1.6677 | 66.596 | 38.136 |
| CUB D32 | 492.679 | -.048 | .0830 | 2.255 | 6.070 |
| CUB D128 | 494.632 | -.166 | .0791 | 2.102 | 5.924 |

cached evaluation of that rule does not measure execution savings. Reused images and labels remain exposed even though the forecast was fixed before evaluation.

## F Earlier positive evidence and adverse controls

This section reports earlier controls that constrain the central claim.

**Missingness is not a null result.** An earlier incomplete-label Zappos transport left all 18 primary contrasts unresolved. Requiring all four annotations before observing utility gives a separate complete-case cohort of 154 queries and 3,465 gallery images. Its nine-action menu mixes conditional widths and frozen blind rankers, unlike the later same-width menu comparison. Replacing each purchased answer by a coherent independent world's answer reduces utility to $.514376/.514913$, below the historical blind baseline $.520059$. Matching marginal answer statistics does not preserve relation-specific action value.

**Readout training against stronger baselines.** Earlier history-readout tests find gains from expanding the terminal catalogue but no reliable gain over continuous fusion. The endpoint-null history readout with body adaptation loses to that stronger fusion. The later rollout-target readout likewise loses to the existing latent actor. SOURCE selection retains the old actor, giving zero selected-policy change. These comparisons leave the natural reversal unestablished, and reuse of the exposed population does not provide independent validation.

**How much opportunity can a learner recover?** An earlier complete-label protocol provides a useful calibration of these scales (Table 6). Zappos uses 154 queries, a 3,465-image gallery, and four conditional rankers plus five frozen blind rankers; the mixed-width menu is not the later same-width test. CUB retains the six-action libraries and 256-query evaluation described above. Within each row, the learned actor and the true-law oracles share the relevance problem, question supply, cap, and tariff. The difference $V_2^\star - V_0^\star$ measures the value of permitted feedback to an optimal policy; $V_2^\star - U$ measures the learned actor's shortfall from that policy. The latter also includes errors possible without buying an answer. Dividing a learned gain over its own STOP by oracle headroom therefore need not measure a fraction of information recovered.

The CUB actor shortfall is larger than the entire true-law feedback headroom. The learned-DP own-STOP intervals are $[-.000289, .000205]$ and $[-.000368, .000019]$, so neither positive value nor practical equivalence is established. On unchanged histories, substituting true payoffs recovers more value than substituting the true posterior alone. This comparison identifies terminal prediction as an important error source in the tested actor, not an intrinsic lack of observable information. The intervention remains evaluator-only: a deployed actor cannot read the true AP matrix.

Two further controls prevent an easy but unsupported repair claim. A TRAIN-calibrated payoff-by-response factorial retains negative mean paid gains in all eight cells. Increasing evaluator probe supply from four to 64 pairs raises headroom to $.003098/.003020$, without demonstrating an actor improvement. Better calibration, a larger question catalogue, and stronger terminal prediction are distinct interventions; their effects cannot be added across experiments. The appropriate operational test remains a paired comparison against a strong actor with the same information and cost contract.

Table 7: **All cells delete actions; only a subset has reachable deletion risk.** Counts summarize the same query–library cells evaluated in Table 1. $Z$ means every full-menu optimal tree has zero loss; $Q^+$ means $\mathcal{D}_{\mathcal{Q},2} > 0$. All values use cap two and price .005.

| Library | Cells with deletion | $\mathcal{D}_{\text{all}} > \tau_{\text{LP}}$ | $Q^+$ | $Z \cap Q^+$ | All $Z$ |
|---|---|---|---|---|---|
| Zappos / DINOv2 | 192/192 | 154 | 58 | 51 | 185 |
| Zappos / CLIP | 192/192 | 161 | 67 | 58 | 183 |
| CUB / D32 | 768/768 | 403 | 105 | 102 | 765 |
| CUB / D128 | 768/768 | 405 | 121 | 107 | 754 |

**Execution accounting.** The cached studies compute the entire candidate library. A selected four-action menu does not prove that only four rankers were executed or stored. A direct-execution audit across 29 SOURCE banks and three seeds measured 11.935ms resident-gallery query time in BF16 and 13.767ms in FP32, but neither precision preserved the cached action/channel contract in any of the 87 cases. Re-evaluation with precomputed features did preserve the action and response specification. Consequently, those timings are not combined with cached utility to claim an end-to-end speedup.

## G Deletion denominators and the stored price–cap boundary

**Rule, ordering, and numerical meaning.** For each stored four-condition payoff matrix, the original pruning code takes the union of `argmax` of the blind payoff and `argmax` in each condition row. Equal floating values choose the first action in the frozen index order. Sorting and deduplication determine $B$; no iterative pruning, cardinality optimization, or target-selected threshold follows. Later enumeration converts each stored binary64 entry to its exact rational value and verifies blind and known-condition endpoint equality without an equality tolerance. The paid-tree catalogue uses those rational payoffs, a uniform four-condition prior, and the original deterministic joint answers. Optimality ties and zero tree losses are exact in that stored-value problem. These statements do not make the underlying floating image scores or full-gallery AP symbolic quantities.

The unrestricted-risk LP remains numerical. Table 7 screens LP values using $\tau_{\text{LP}} = 2 \times 10^{-7}$, the original diagnostic's numerical consistency allowance, rather than treating every positive floating residue as scientific risk. Applying $10^{-12}$ or $10^{-8}$ instead gives counts $(154, 161, 404, 405)$; the $2 \times 10^{-7}$ screen gives $(154, 161, 403, 405)$. Neither screen certifies exact zero unrestricted risk. Positive allowed-probe counts are unchanged at thresholds $0$, $10^{-12}$, and $10^{-8}$. Numerical screening is separate from the development materiality criterion .002 AP.

The conditional count 318/351 addresses the central distinction directly: reachable harmful observations need not occur on an optimal paid tree. The unconditional 1,887/1,920 count additionally includes 1,569 cells without positive allowed-probe risk. Cells with no unrestricted risk, or no reachable risk, must not be described as evidence that large risk was avoided. The retained-size census in Table 8 rules out a no-deletion explanation. Exact endpoint comparisons find unique maximizers at all five endpoints in all 1,920 cells. An independent enumeration checks $384 \cdot 255 + 1,536 \cdot 63 = 194,688$ nonempty subsets. Every endpoint-feasible $B'$ must contain the winner union $B$, so no distinct same-cardinality alternative exists. This is a set-inclusion certificate, not additional policy simulation. Under the same prior, response law, price, and feasible policies, $V^\star_B \le V^\star_{B'} \le V^\star_A$, hence $0 \le G(B') \le G(B)$. All feasible subsets have zero gap in the original 1,887 zero-gap cells (185/183/765/754 in table order); the other 33 (7/9/3/14) receive only the bound. This does not guarantee unique endpoint winners in other libraries or actor preservation.

**Sensitivity to question budget and price.** Policies for $A$ and the same retained menu $B$ are independently replanned at prices $0, .001, .005, .02$ and question budgets of zero, one, or two. Table 9 reports all nonzero-cap settings; cap zero has zero gap because the blind endpoint is preserved. These are exposed-population sensitivities under the original response law, not a realistic-price estimate, a frozen-policy accounting curve, or a new confirmation. Prices .0025 and .01 have not been evaluated. The primary remains cap two and .005.

Table 8: **Retained cardinality is query-specific.** Each entry counts cells with the indicated $|B|$; three seed libraries share each query. All columns, including singleton retained menus, remain in the denominator.

| Library | Full $K$ | $\lvert B\rvert = 1$ | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Zappos / DINOv2 | 8 | 0 | 0 | 30 | 123 | 39 |
| Zappos / CLIP | 8 | 0 | 3 | 24 | 107 | 58 |
| CUB / D32 | 6 | 24 | 246 | 353 | 134 | 11 |
| CUB / D128 | 6 | 27 | 240 | 345 | 144 | 12 |

Table 9: **The answer budget changes policy relevance.** Mean independently replanned $G$ in $10^{-3}$ AP-minus-price units. Every row reuses the same per-cell deletion; no price-dependent menu selection occurs. Query means first average the three libraries.

| Cap | Price | Zappos / DINOv2 | Zappos / CLIP | CUB / D32 | CUB / D128 |
|---|---|---|---|---|---|
| 1 | 0 | .695960 | .854892 | .025063 | .051213 |
| 1 | .001 | .695960 | .849684 | .021835 | .042032 |
| 1 | .005 | .695960 | .826570 | .004921 | .020014 |
| 1 | .02 | .695960 | .799853 | 0 | 0 |
| 2 | 0 | .079807 | .170544 | .011948 | .038461 |
| 2 | .001 | .079807 | .158826 | .009092 | .029208 |
| 2 | .005 | .081062 | .125741 | .001103 | .010535 |
| 2 | .02 | .100438 | .099025 | 0 | 0 |

At the primary setting, seed-averaged maximum query gaps are .0025803, .0044300, .0001351, and .0006897 in table order. At cap one and .005, the Zappos maxima rise to .0076950/.0116946, and the fractions above .002 rise from 1/64 per encoder to 9/64. CUB has no query mean above .002 at any stored setting. Zero CUB gaps at price .02 are conditional finite-policy results, not evidence that all feedback is useless. Figure 2 compares primary cohort means; its ratio guide does not summarize the distribution across prices and budgets.

## H State occupancy and answer erasure

Each terminal row carries its probability under an optimal full-menu tree. The 1,920 cells contain 3,709 terminal rows and 7,680 erasure rows, with terminal probabilities summing to one within each cell. The five categories in Table 3 are disjoint. PRIOR mass exceeds half of the CUB total and cannot be interpreted as successful condition identification. Among intermediate states, a common optimum has $\Gamma_B = 0$, a preserved compromise has $\Gamma_B > 0$ and $d_B = 0$, and a lost action has $d_B > 0$. Their reported masses are distinct from counts of independent branches or queries.

Stored minimum and maximum category occupancy over optimal ties agree in every cell. Thus selection of a favorable tied tree does not explain the state census. The check remains restricted to the declared actions, probes, and stored payoffs. A compromise can remain optimal without identifying the condition; that property does not establish that another paid probe would have positive net value.

The erasure intervention holds the original question tree fixed, hides only the second answer, and reoptimizes the terminal action from the remaining information. No actor is refitted or question tree replanned. Table 10 therefore concerns terminal-AP deletion loss, not a new paid gap $G$. Complete erasure also differs from the independent-flip channel. CUB's additional species sensitivity uses 136 clusters; changing the resampling unit does not change the recorded point estimate.

## I Response correction: quality, purchases, and terminal scope

The four cells $C_{ij}$ use old or corrected planning in index $i$ and old or corrected terminal posterior in index $j$. The primary contrast $C_{11} - C_{01}$ holds the corrected terminal fixed. A fixed sample of 2,048 synthetic SOURCE answer pairs gives $\widehat{\eta} = .15576171875$; the evaluation channel remains $\eta = .15$. Existing labels and a seeded independent flip process generate the calibration pairs, not natural-user

Table 10: **Hiding the second answer increases fixed-tree deletion loss.** Original loss is $L_A = \ell_B(\pi_A)$ on a full-menu optimal tree. Losses use $10^{-3}$ AP units. The interval concerns the paired increase under query-component resampling, conditional on the fixed libraries and trees. Positive counts are components after seed averaging.

| Library | Original $L_A$ | Full erasure | Increase | Conditional 95% interval | Positive / $N$ |
|---|---|---|---|---|---|
| Zappos / DINOv2 | .1004 | .9668 | .8663 | [.4511,1.3380] | 23/64 |
| Zappos / CLIP | .1309 | 1.0798 | .9489 | [.4816,1.4911] | 23/64 |
| CUB / D32 | .0011 | .0103 | .0092 | [.0006,.0217] | 5/256 |
| CUB / D128 | .0160 | .0371 | .0211 | [.0044,.0436] | 9/256 |

Table 11: **Utility improvement need not imply retrieval improvement.** Values use $10^{-3}$ AP or AP-minus-price units. Brackets are marginal, pointwise conditional 95% intervals over 206 exposed components, not joint confidence regions or the family-adjusted interaction intervals in Table 5.

| Encoder | Terminal | $\Delta$ AP [interval] | $\Delta U$ [interval] |
|---|---|---|---|
| DINOv2 | Single | $+1.890$ $[-.140, 4.057]$ | $+4.651$ $[2.589, 6.827]$ |
| | RRF | $-.527$ $[-1.868, .850]$ | $+2.234$ $[.911, 3.603]$ |
| | Z-score | $-1.947$ $[-3.271, -.622]$ | $+.813$ $[-.451, 2.109]$ |
| CLIP | Single | $+.428$ $[-1.556, 2.517]$ | $+2.488$ $[.536, 4.519]$ |
| | RRF | $-.981$ $[-2.346, .379]$ | $+1.079$ $[-.256, 2.406]$ |
| | Z-score | $-2.625$ $[-4.018, -1.268]$ | $-.566$ $[-1.939, .774]$ |

observations. Neither target utility nor target labels choose the rate estimate. The sample is one calibration realization, not 2,048 independent image populations.

The three terminals share the same changed question paths within each encoder. For DINOv2, $\Delta Q = -.552144$ with conditional 95% interval $[-.620841, -.488389]$; for CLIP, $\Delta Q = -.411893$ with $[-.465008, -.361489]$. Multiplication by the stipulated tariff gives savings $.002760720/.002059466$. Therefore $\Delta U = \Delta\,\mathrm{AP} - c\Delta Q$ accounts for all paired means; the approximately 59%/83% tariff fractions are not causal mediation, user time, or GPU savings. The single-ranker AP intervals contain zero, and both Z-score AP intervals are negative.

Subtracting the single-ranker correction effect from each fusion effect cancels $\Delta Q$. Three of the four declared interaction intervals are negative; CLIP RRF remains unresolved. The question actor was not optimized separately for fusion. Accordingly, the result limits portability of the tested correction, not all possible fusion-aware acquisition. Posterior fusion can output rankings outside the single-action menu, so the finite single-ranker oracle is not a fusion performance bound. The SOURCE-selected old-order stopping comparison retains residual single-ranker utility gains, with unresolved AP gains (Appendix K); fewer purchases alone do not establish a need for complex replanning. Absolute cross-method rankings also cannot be reconstructed by subtracting gains from different own-STOP baselines.

## J Question budgets and frozen-actor behavior

**Paired budget changes.** At price .005, Table 12 compares independently replanned deletion gaps for the same retained menu. Three libraries are averaged within query before 4,000 paired resamples. Zappos has 64 query components; CUB has 256 queries in 136 species. Whole-species resampling preserves the query-weighted estimand through query sums and counts. Intervals are conditional, pointwise post-exposure summaries; query-level increases in CUB preclude a universal monotonicity claim. These are changes in deletion gaps, not the AP benefit of buying another answer.

**Optimal value and a fixed actor.** A CUB intervention uses the frozen predicted-payoff actor with strict ASK decisions. AA plans and terminates in all six actions; AB keeps AA's ordered questions, answers, STOP decisions, and purchase count but restricts terminal selection to the evaluator-derived endpoint core $B$; BB replans and terminates in $B$. All 256 queries, two widths, and three seeds enter: 1,536 cells and 18,432 world–arm outcomes. Worlds and seeds are averaged within query. Four

Table 12: Two-question − one-question deletion gap at price .005, in $10^{-3}$ AP-minus-price. Brackets are 95% percentile intervals. Counts classify query-mean increases, decreases, and unchanged values.

| Cohort | $\Delta G$ | Query interval | Species interval | $+/-/0$ |
|---|---|---|---|---|
| Zappos DINOv2 | $-.614898$ | $[-1.057330, -.260736]$ | — | 0/16/48 |
| Zappos CLIP | $-.700829$ | $[-1.258132, -.265620]$ | — | 0/14/50 |
| CUB D32 | $-.003817$ | $[-.010521, .000919]$ | $[-.010472, .000901]$ | 1/3/252 |
| CUB D128 | $-.009479$ | $[-.020109, -.001831]$ | $[-.019687, -.001984]$ | 1/7/248 |

utility contrasts receive 4,000 paired query resamples and nominal family-four 98.75% intervals. Unlike true-payoff deletion loss, AA−AB can be negative because terminal decisions use predicted payoffs. Restriction improves all four average utilities, also under a 136-species sensitivity analysis.

Table 13: Full-menu actor − restricted actor. AP and utility use $10^{-3}$ units; purchases use questions. Negative means favor restriction. Intervals apply only to utility.

| Width | Contrast | $\Delta$ AP | $\Delta Q$ | $\Delta U$ | 98.75% interval |
|---|---|---|---|---|---|
| 32 | AA−AB | $-1.390$ | .00000 | $-1.390$ | $[-2.064, -.849]$ |
| 32 | AA−BB | $-1.370$ | .01725 | $-1.457$ | $[-2.152, -.905]$ |
| 128 | AA−AB | $-1.298$ | .00000 | $-1.298$ | $[-1.822, -.850]$ |
| 128 | AA−BB | $-1.258$ | .01725 | $-1.345$ | $[-1.882, -.877]$ |

Nevertheless, among the original zero-$G$ cells (765/754 at D32/D128), AA−AB exceeds .002 in 16 cells per width, with maxima $.020161/.020578$. Averaging included zero-gap seeds within query then queries equally gives negative means $-.001397/-.001294$. These individual losses do not establish average harm from pruning. The evaluator-derived restriction is not a deployable selector, and the experiment does not isolate all prediction and decision errors. Independent backward induction checks 101,376 states and preserves AA/AB transcripts.

## K Stopping controls and mandatory answer budgets

**A fixed old question order.** The single-ranker stopping control selects on the original 128-query SOURCE/TUNE cohort among $C_{01}$, four margin thresholds $0, .001, .005, .01$, and always-STOP. At an old ASK node, a margin below threshold stops; otherwise the old question remains. The margin is the old planner's predicted paid continuation minus its STOP value, without recomputing continuation for the modified policy. Old STOP cannot become ASK. Corrected terminal inference, the categorical joint response law, unsupported-history fallback, and costs remain fixed. Exact ties use the stated candidate order. DINOv2 selects .01 for each seed; CLIP selects $.005, .001, .005$. $C_{11}$ is a comparator, not a selection candidate.

A separate RRF extension selects by RRF utility, using rank constant 60 and the same corrected posterior and public hash ties. DINOv2 again selects .01 throughout; CLIP selects $.001, .01, C_{01}$. Both studies evaluate the same exposed 206-query cohort after SOURCE selection and average conditions and seeds within query. Acquisition remains single-ranker optimized. The remaining

Table 14: $C_{11}$ − terminal-specific SOURCE-selected stopping. AP and utility use $10^{-3}$ units. All intervals use 4,000 paired query resamples. Single-ranker intervals are nominal family-two 97.5%; RRF utility uses that family, while RRF AP and purchases use descriptive pointwise 95% intervals.

| Terminal | Encoder | $\Delta$ AP [interval] | $\Delta Q$ [interval] | $\Delta U$ [interval] |
|---|---|---|---|---|
| Single | DINOv2 | $2.100\ [-.191, 4.382]$ | $-.2406\ [-.2978, -.1864]$ | $3.303\ [1.082, 5.535]$ |
| Single | CLIP | $.506\ [-1.769, 2.806]$ | $-.3805\ [-.4361, -.3243]$ | $2.409\ [.152, 4.737]$ |
| RRF | DINOv2 | $.400\ [-.924, 1.680]$ | $-.2406\ [-.2907, -.1938]$ | $1.603\ [.169, 3.017]$ |
| RRF | CLIP | $-.876\ [-2.279, .439]$ | $-.3628\ [-.4095, -.3172]$ | $.938\ [-.634, 2.420]$ |

utility gain is positive for both single-ranker encoders but unresolved for CLIP RRF. Every AP

interval contains zero. These controls do not exhaust inexpensive controllers or establish a need for complex replanning. The synthetic calibration sample, images, and models are fixed; the tariff excludes computation and human time.

**Exactly one or two answers.** A separate test holds the four-action menu, predicted AP, probes, and corrected terminal posterior fixed. Corrected DP is compared with a SOURCE-selected control drawn from old DP, corrected entropy, and eight cyclic fixed-question orders. Each library, budget, and terminal receives its own TUNE128 utility selection before evaluation on the exposed 206 queries. The synthetic flip rate is .15 and the SOURCE estimate is .15576171875. Worlds and three seeds are averaged within query. Single-ranker comparisons use 4,000 paired resamples for four encoder–budget primaries; two-answer fusion uses 10,000 for a separate four encoder–terminal family. Neither family integrates calibration or model uncertainty. For two-answer single-ranker

Table 15: Corrected DP − terminal-specific SOURCE-selected control, in $10^{-3}$ AP. Exact purchase counts give $\Delta U = \Delta\,\mathrm{AP}$. Single-ranker and two-answer fusion brackets are nominal 98.75% conditional intervals from separate families; one-answer fusion brackets are descriptive 95% intervals.

| Answers | Terminal | DINOv2 $\Delta$ AP [interval] | CLIP $\Delta$ AP [interval] |
|---|---|---|---|
| 1 | Single | $.605\ [-.627, 1.932]$ | $1.801\ [.050, 3.721]$ |
| 2 | Single | $2.517\ [.272, 4.889]$ | $3.380\ [.469, 6.407]$ |
| 1 | RRF | $-.257\ [-.875, .311]$ | $.554\ [-.231, 1.399]$ |
| 1 | Z-score | $-.630\ [-1.310, -.047]$ | $.806\ [-.364, 2.030]$ |
| 2 | RRF | $-.711\ [-2.291, .846]$ | $.891\ [-1.621, 3.374]$ |
| 2 | Z-score | $-1.336\ [-2.753, .037]$ | $-.148\ [-2.854, 2.611]$ |

evaluation, DINOv2 selects old DP in all seeds; CLIP selects entropy/old DP/entropy. The CLIP contrast against old DP alone is .002203, with nominal 98.75% interval $[-.000351, .004702]$. Both selected-control lower bounds remain below the .002 reference effect size. All two-answer fusion libraries select entropy; RRF retains constant 60, Z-score a $10^{-9}$ standard-deviation floor, and both posterior-to-head temperature .05. All four primary fusion intervals contain zero. DINOv2 Z-score's negative pointwise 95% interval $[-.002419, -.000257]$ does not replace its primary interval. Different terminal-specific controls prevent subtracting these rows as a same-control causal interaction. The fixed-budget comparison excludes fewer purchases as an explanation of the single-ranker gain, without decomposing the adaptive-stopping result.

**Calibration draws and CUB transport.** Eight prechosen seeds regenerate independent flips on the same 2,048 SOURCE pairs, giving estimated rates from .1357421875 to .16064453125. Previously chosen controls stay fixed. Each two-answer comparison uses a common corrected terminal, but changing the rate changes both planning and the terminal between draws. These repeated signs are conditional sensitivity results, not new populations or significance tests. CUB instead uses

Table 16: Two-answer differences across eight synthetic calibration draws, in $10^{-3}$ AP. Ranges are observed minima and maxima, not confidence intervals. All draws reuse the 206 queries and fixed controls.

| Encoder | Terminal | Mean | Observed range | Positive draws |
|---|---|---|---|---|
| DINOv2 | Single | 2.780 | $[2.517, 3.272]$ | 8/8 |
| CLIP | Single | 3.157 | $[2.913, 3.500]$ | 8/8 |
| DINOv2 | RRF | $-.477$ | $[-.711, -.095]$ | 0/8 |
| CLIP | RRF | .739 | $[.431, .891]$ | 8/8 |
| DINOv2 | Z-score | $-1.160$ | $[-1.336, -.905]$ | 0/8 |
| CLIP | Z-score | $-.328$ | $[-.561, -.148]$ | 0/8 |

its attribute-overlap Bernoulli response law with stipulated flip rate .15, not a SOURCE-estimated Zappos rate. Exactly two distinct questions precede a common corrected single-action terminal. SOURCE-TUNE367 selects among old DP, corrected entropy, corrected myopic, and six fixed probe pairs for each width, seed, and four/full-six-action menu. All histories have support. On the existing 256-query cohort, all four primary query intervals, four species-cluster 95% sensitivities, and

four full-six secondary intervals include zero. Forced questioning improves AP only about .0802–.0886pp over STOP while reducing utility by about .9114–.9198pp-equivalent after two answer charges; that separate comparison does not replace the matched-budget test.

Table 17: CUB four-action fixed-budget AP differences in percentage points. Nominal family-four 98.75% paired query intervals follow world and seed averaging; exactly two answers give $\Delta U = \Delta\,\mathrm{AP}$.

| Width | Corrected DP − | $\Delta$ AP (pp) | Interval (pp) |
|---|---|---|---|
| 32 | Old DP | $-.001099$ | $[-.014032, .010681]$ |
| 32 | SOURCE-selected | $-.003201$ | $[-.017739, .010899]$ |
| 128 | Old DP | $-.001742$ | $[-.016333, .012406]$ |
| 128 | SOURCE-selected | .008695 | $[-.013448, .044454]$ |

## L FUSION-TARGET ACQUISITION AND RESPONSE-MODEL CONTROLS

A fixed-budget fusion learner targets the executed RRF or Z-score terminal. Fusion and single-action targets share 336 public features, weighted ridge penalty .01, FIT256, and TUNE128 selection in six libraries. Every non-STOP policy buys exactly two answers. Among 14 controls, SOURCE selects entropy in eleven library–terminal cells and the existing single-action DP in one CLIP/RRF cell. This experiment is distinct from the 634-feature ridge and root-only MLP tests. All four primary target means are negative, as are the matched single-action-target differences. Positive gains over myopic or own-STOP controls do not establish superiority over the selected stronger policy.

Table 18: Fusion DP − SOURCE-selected control, AP percentage points, on the exposed 206-query cohort. Nominal family-four 98.75% intervals use paired resampling after seed averaging.

| Encoder | Terminal | $\Delta$ AP (pp) | Interval (pp) |
|---|---|---|---|
| DINOv2 | RRF | $-.13643$ | $[-.31343, .03828]$ |
| DINOv2 | Z-score | $-.28568$ | $[-.45487, -.12328]$ |
| CLIP | RRF | $-.00488$ | $[-.18275, .18050]$ |
| CLIP | Z-score | $-.08211$ | $[-.34178, .18817]$ |

An exactly-two-answer SOURCE oracle uses actual AP, clean-answer patterns, and the true .15 flip law on 128 queries. Its first question is shared across hidden worlds; the second depends only on observed answers. In DINOv2/RRF, DINOv2/Z-score, CLIP/RRF, CLIP/Z-score order, oracle − selected AP is $.9801, 1.0655, 1.5318, 1.6272$pp; oracle − fusion-learner is $1.1817, 1.3412, 1.7764, 1.8357$pp. Replacing only learned payoffs adds $.3776, .5483, .5572, .6521$pp; replacing only the response law adds $.5875, .6316, 1.2675, 1.1554$pp. Individual queries can worsen, and nonzero interactions preclude additive causal attribution. The true law supplies query/world-specific transitions, not merely a noise-rate correction.

**Public response-model repairs.** A categorical inverse-temperature fit on FIT256 shares one coefficient per classifier seed across encoders and terminals. Range $[.25, 4]$ and 96 bisection steps yield $1.069815315, 1.068139112, 1.054348755$. Terminal rankings, learned rewards, rate estimate $319/2048$, true flip rate .15, answer budget, and selected control identities stay fixed. TUNE categorical NLL falls from .664629113 to .662019040, while joint pair-response NLL rises from 1.190308400 to 1.190536198. Fusion DP changes by $-.004373, -.006814, -.011122, -.020758$pp in the four panels; matched corrected controls remain .234353–.318257pp higher. These descriptive means have no new intervals.

A second fit adds two unary or four full residual features to the original shared-query categorical law over 256 answer patterns. Both models preserve dependence. Fitting minimizes regularized clean eight-response NLL; noisy pair cross entropy is a separate evaluation loss. Per seed, FIT256 plus 1,212 distinct probe images gives 1,468 labeled images and 8,192 derived relation bits, not independent observations. Penalty .01, features, likelihood, noise, and two-answer terminals are fixed. TUNE128 queries are distinct from FIT, but 554 probe images overlap and all SOURCE roles are exposed. Additional pair supervision makes this more than a change of objective relative to temperature fitting. Both fitted policies remain below same-law entropy and selected controls in all four panels.

Table 19: Coherent response fits on SOURCE/TUNE. The lower panel reports AP differences in percentage points at exactly two answers. All entries are descriptive means without new intervals. UNARY and FULL denote two- and four-feature residuals. The last column applies FULL to previously selected control identities without reselection.

| Prediction loss | Original | UNARY | FULL |
|---|---|---|---|
| Clean eight-response NLL | 1.980121 | 1.980865 | 1.983413 |
| Noisy pair cross entropy | 1.190308 | 1.190304 | 1.190431 |

| Panel | UNARY−old | FULL−old | FULL−entropy | Selected−old selected |
|---|---|---|---|---|
| DINOv2 / RRF | −.022762 | −.006143 | −.237417 | +.029599 |
| DINOv2 / Z-score | −.005283 | −.002814 | −.309243 | +.030761 |
| CLIP / RRF | −.014064 | −.000568 | −.301175 | +.005301 |
| CLIP / Z-score | +.040563 | +.031218 | −.233483 | +.056257 |

FULL worsens both proper losses for every seed; UNARY's tiny mean pair-loss reduction includes two worsening seeds. Roundoff of at most $5.55 \times 10^{-16}$ changes eight near-zero sign/count summaries by one; original-array aggregation recovers the reported counts. These SOURCE-only results do not replace the adverse target comparisons or establish a strong-policy gain.

## M PUBLICLY PREDICTED ENDPOINT CORES AND MATCHED MENUS

These CUB controls use the same 256 exposed queries, three fixed seeds, and two widths, D32 and D128 (1,536 query–seed–width cells). Four relevance conditions have uniform prior. The clean joint response law, two-question cap, and price .005 are fixed. The public planner asks only when its continuation value strictly exceeds STOP; ties use the existing action and question order. All 33 public histories have support. These controls are distinct from the noisy mandatory-two-answer experiments.

Let $\widehat{B}$ be the union of the public predicted-AP blind winner and the four predicted known-condition winners, taking the first maximizer at ties. True AP and the evaluator-derived core $B$ are not policy inputs. AA plans and terminates in all six actions. AB keeps AA's questions, answers, STOP decisions, and purchase count, while restricting terminal selection to $\widehat{B}$. BB independently replans and terminates in $\widehat{B}$. Worlds and the three seeds are averaged within each query before 4,000 paired-query bootstrap resamples. Each experiment's four primary contrasts use nominal 98.75% intervals; the two experiments do not form one simultaneous family. Whole-species sum/count resampling over 136 species supplies separate pointwise 95% sensitivities, preserving query weighting.

Table 20: Public core − full-six-action policy. AP and utility differences are percentage points; question differences are unscaled. All four primary query intervals include zero. A lower bound of exactly zero is not a strictly positive bound.

| Width | Contrast | $\Delta$ AP | $\Delta Q$ | $\Delta U$ | Query 98.75% interval |
|---|---|---|---|---|---|
| 32 | AB − AA | .000243 | 0 | .000243 | $[-.000036, .001008]$ |
| 32 | BB − AA | .000243 | 0 | .000243 | $[-.000036, .001008]$ |
| 128 | AB − AA | .000887 | 0 | .000887 | $[0, .002588]$ |
| 128 | BB − AA | −.001903 | −.001302 | −.001252 | $[-.006737, .001533]$ |

Species intervals in table order are $[-.000026, .000778]$, $[-.000026, .000778]$, $[0, .002173]$, and $[-.005230, .001078]$pp. The mean retained sizes are 2.971354 and 3.026042. Actual blind endpoints change in 161/768 and 149/768 cells, and at least one known-condition endpoint changes in 456/768 and 440/768. Predicted endpoint identity therefore does not certify actual endpoint preservation. AA asks at the root in only 75/768 and 67/768 cells. One D128 query retains mean utility loss $-.435357$pp, including a $-1.306070$pp seed-specific mean. Small cohort averages do not rule out these losses.

**Matching menu size and the public blind winner.** For each cell, all full-six subsets with the same size as $\widehat{B}$ and containing the same public blind winner are enumerated, including $\widehat{B}$ itself (at most ten menus). The comparator uniformly samples one complete menu before the conversation; the reported expectation neither selects the best menu nor resamples at later histories. AB retains AA's transcript; BB replans separately for each menu. The comparison contains 26,224 complete policies. Public policies are fixed before evaluating their outcomes.

Table 21: Public core − uniformly sampled matched menu. Units and primary query intervals follow Table 20. Positive AP need not imply positive utility because the replanned policy purchases more answers.

| Width | Policy | $\Delta$ AP | $\Delta Q$ | $\Delta U$ | Query 98.75% | Species 95% |
|---|---|---|---|---|---|---|
| 32 | AB | .023197 | 0 | .023197 | $[.005878, .045955]$ | $[.009029, .040417]$ |
| 32 | BB | .026848 | .062695 | −.004499 | $[-.025701, .018611]$ | $[-.020114, .013144]$ |
| 128 | AB | .010389 | 0 | .010389 | $[-.001284, .023167]$ | $[.002665, .019329]$ |
| 128 | BB | .011515 | .044303 | −.010637 | $[-.027453, .004814]$ | $[-.023396, .000836]$ |

All four mean utility contrasts are below the .002 raw-utility (.2pp) practical reference. D32 AB has a small positive terminal-selection contrast, but neither replanned comparison establishes a utility advantage. A positive species interval for D128 AB does not replace the zero-containing primary query interval. Query counts with utility gain above .2pp / loss below −.2pp are 14/3, 7/11, 10/4,

and 5/9 in table order. Against the older global SOURCE-selected four-action menu, BB has secondary pointwise 95% gains of .025189pp $[.010492, .039327]$ and .021756pp $[.003522, .039968]$, with losses on 39 and 40 queries. That comparison changes global fixed size to query-adaptive size and does not override the matched-menu result.

Table 22: Absolute means for the public-menu comparisons. AP and utility are raw values; $U = \mathrm{AP} - .005Q$. All arms use the same fixed evaluator and exposed cohort.

| Width | Arm | AP | Questions | Utility |
|---|---|---|---|---|
| 32 | Full6 AA | .4931278018 | .1048177083 | .4926037132 |
| 32 | Public core AB | .4931302333 | .1048177083 | .4926061448 |
| 32 | Public core BB | .4931302333 | .1048177083 | .4926061448 |
| 32 | Uniform AB | .4928982666 | .1048177083 | .4923741781 |
| 32 | Uniform BB | .4928617484 | .0421223958 | .4926511364 |
| 32 | SOURCE K4 | .4928653224 | .1022135417 | .4923542547 |
| 128 | Full6 AA | .4950841554 | .0930989583 | .4946186606 |
| 128 | Public core AB | .4950930294 | .0930989583 | .4946275346 |
| 128 | Public core BB | .4950651230 | .0917968750 | .4946061387 |
| 128 | Uniform AB | .4949891345 | .0930989583 | .4945236398 |
| 128 | Uniform BB | .4949499756 | .0474934896 | .4947125081 |
| 128 | SOURCE K4 | .4948573297 | .0937500000 | .4943885797 |

Taking the union of public full-menu terminal winners over all 33 histories preserves all actions and questions of the public policy, with maximum value discrepancy $2.22 \times 10^{-16}$. This is a finite-policy compression identity, not another target-performance or speed comparison. Full-library predictions are computed before any menu restriction; storage or latency savings are not measured.

## N ENDPOINT-PRESERVING SUBMENUS IN NATURAL-IMAGE LIBRARIES

The 1,920 existing query–library cells support a separate evaluator-only existence test. For each cell, all submenus containing the union $B$ of the unique blind and known-condition winners are enumerated. Each pair has equal cardinality, width, endpoints, and probe law. The census contains different cardinalities and does not define one deployable library. Unlike the public core in Appendix M, menu construction here uses true AP.

A forced-one reversal requires the complete sets of globally optimal questions for the two menus to be disjoint; a shared third optimum disqualifies the pair. The paid condition additionally requires both menus to strictly prefer ASK over STOP at cap one and cost .005. A separate materiality condition requires both directional question regrets to exceed .002 raw AP. The paid condition is not the original cap-two actor event. Census arithmetic treats each stored binary64 AP as an exact rational number; this alone does not certify native symbolic AP.

Table 23: Natural submenu census. Reversal columns give cells / equal-cardinality menu pairs. Cells and pairs are dependent descriptive counts without confidence intervals or population-frequency claims.

| Cohort | Cells | Eligible | Menus | Pairs | Forced | Paid cap1 | Regret $>$ .002 |
|---|---|---|---|---|---|---|---|
| Zappos DINOv2 | 192 | 192 | 3,240 | 6,855 | 3/86 | 3/86 | 1/56 |
| Zappos CLIP | 192 | 192 | 3,136 | 7,167 | 1/4 | 1/4 | 1/4 |
| CUB D32 | 768 | 757 | 8,086 | 11,534 | 2/18 | 1/10 | 1/7 |
| CUB D128 | 768 | 756 | 8,064 | 11,664 | 7/45 | 1/15 | 0/0 |
| Total | 1,920 | 1,897 | 22,526 | 37,220 | 13/153 | 6/115 | 3/67 |

Twenty-three cells have no comparable pair, and 1,907 cells have no forced reversal. No query exhibits any of the three events across all three seeds. The positive cases establish existence on the exposed libraries, not a frequent or seed-stable effect.

**A fixed native-rank rational-AP example.** A deterministic salted-hash ordering selects one positive case, rather than maximizing effect size. The case uses a 128-dimensional DINOv2 library and six actions in each menu: $\{0, 1, 2, 4, 6, 7\}$ versus $\{0, 1, 4, 5, 6, 7\}$. Actions 0–3 and 4–7 are the four condition masks from two separately learned embedding geometries; the replacement swaps mask 2 from the first geometry for mask 1 from the second. Thus the example uses learned image-function rankers, but not one shared embedding or a newly trained matched representation.

The original float64 rank computation is repeated for this fixed case with normalization epsilon $10^{-6}$ and score-then-image-hash tie ordering. Excluding the query's seven product images from 4,512 source images leaves 4,505 gallery images, with 571, 3,159, 1,543, and 425 positives under the four conditions. The eight rankings and all 32 original AP values are reproduced. Computing AP directly as rational sums over positive ranks changes float values by at most $1.11 \times 10^{-16}$ and preserves all endpoints, question optima, and strict paid comparisons.

Table 24: Native-rank rational-AP case. Decimals display exact rational comparisons. Regrets use the other menu's optimal question. Both menus favor one paid answer over STOP.

| Quantity | Menu 1 | Menu 2 |
|---|---|---|
| Blind AP | .527567851965 | .527567851965 |
| Forced $q_0$ AP | .585176991488 | .585176991488 |
| Forced $q_3$ AP | .589736525211 | .582207287292 |
| Globally optimal question | $q_3$ | $q_0$ |
| Other-question regret (pp) | .455953372308 | .296970419595 |
| Best ASK $-$ STOP, cost .005 | .057168673246 | .052609139523 |

The example is selected after observing the census and remains an exposed-data illustration. Native verification covers this case from frozen embeddings, not new encoder training or all 1,920 cells. The natural submenu result does not overturn the failed shared-representation training experiment, establish a deployable selector, or imply cap-two policy improvement.