CHARACTERIZING SCORING FUNCTION STABILITY UNDER OPEN-WORLD TAXONOMY GROWTH

Anonymous authors
Paper under double-blind review

ABSTRACT

Taxonomy expansion and completion represent a query and the available taxonomy, score candidate placements, and select parent links. Knowledge growth can change this attachment problem after training. A parent can become a grandparent, a sibling can become a parent, or another parent can become necessary without changing the query's meaning. Evaluation against a fixed seed taxonomy leaves these consequences untested. In this paper, we develop `TaxoAtlas` as a theoretical framework to analytically study the stability of scoring and selection families under such growth. It records taxonomy relationships and places concepts using calibrated semantic similarities. Related concepts contribute support to an evidence field, a map showing where evidence for a query is concentrated. Under explicit conditions, we establish which structural and semantic distinctions this construction preserves and bound how growth changes its fields. To isolate scorer limitations from representation error, every rule receives exactly the current query and candidate inputs prescribed by the evolving reference. We call this *oracle reference encoding*. The theory characterizes eight recurring scoring-operation classes as mathematical objects, independent of how well an encoder or complete model is trained. The resulting conditions cover retaining valid parents, acquiring newly required parents, rejecting false links, and exact parent-set recovery. They separate the effects of changing evidence, new competitors, normalization, structural constraints, and final selection. Growth elsewhere can suppress an unchanged local match, while a correct ranking can still yield missing or false links. Conversely, query and parent changes can cancel and preserve their score. In a controlled construction, a parent-set score that penalizes larger sets misses a new parent in all forty later states. A Gaussian score ranks both true parents highest but admits false links in thirty-nine. In paired conditional ranking with a parent-complete candidate neighborhood, a structurally gated angular scorer's Hit@1 falls from 49.0% to 21.6% on Environment and from 37.0% to 31.5% on Science after growth. A separate Environment replay and descriptive field profiles for Environment, Science, and Food provide further grounding. Our theoretical and empirical results provide detailed insights into the stability of scoring and selection rules under taxonomy growth.

1 INTRODUCTION

Taxonomies organize concepts in product catalogs, biomedical ontologies, scientific knowledge bases, and other systems where retrieval and reasoning depend on hierarchical structure. Taxonomy expansion and completion infer missing direct parent-child relations from a concept's semantic description and the observed taxonomy. The task is difficult even when the taxonomy is fixed because semantic relatedness does not uniquely determine structural position. A sibling may resemble the query more strongly than its true parent, a broad ancestor may be semantically appropriate without being the correct direct parent, and multiple parents may be valid simultaneously. Most systems therefore represent the query and candidate concepts, score candidate-parent compatibility, and convert these scores into one or more predicted links. Taxonomy growth adds a distinct difficulty because the query description can remain unchanged while its correct parent set changes. For example, *electric bus* may initially have *bus* as its only available parent. When *electric vehicle* is introduced, *bus* remains valid and an additional parent becomes necessary. An intermediate inser-

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.

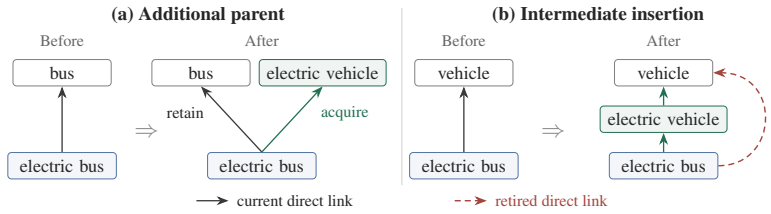

Figure 1: Two independent taxonomy-growth examples with an unchanged query. (a) *Bus* remains a parent when *electric vehicle* becomes an additional parent. (b) Inserting *electric vehicle* retires the direct link to *vehicle*, which remains an ancestor along the solid two-edge path. Green marks new concepts and links. Arrows point from child to parent.

tion can instead turn a former direct parent into an ancestor. A scoring rule used over an evolving taxonomy must therefore retain valid parents, acquire newly required parents, remove retired direct links, and reject false links. We ask: *under what conditions does a scoring and selection rule that is correct before taxonomy growth remain correct afterward?*

Existing taxonomy expansion and completion work primarily evaluates whether an end-to-end method can recover useful attachments in a supplied hierarchy. Recent approaches use fuzzy-overlap, distributional, structural, and generative formulations (Xu et al., 2024; Mishra et al., 2026a;b; Shi et al., 2024). Complementary work studies insertion order, sequential revision, temporal construction, and adaptation as the hierarchy evolves (Song et al., 2021; Xia et al., 2023; Tuan et al., 2016; Kargupta et al., 2025). These lines of work address placement quality. The stability of the scoring and selection mechanism itself under taxonomy growth remains unexplored. Since a prediction can change because the representation shifts, a new candidate becomes more competitive, score rescaling changes when the candidate set grows, a structural rule blocks an otherwise suitable parent, or the final selector returns the wrong set, end-to-end accuracy hides these effects.

In this paper, we characterize the stability of scoring and selection rules under taxonomy growth - *which rules remain correct, which conditions preserve correctness, and which failures persist even with exact current inputs*. To isolate scorer behavior, every analyzed rule starts from a correct pre-growth decision and, after each update, receives the exact quantities prescribed by its reference representation. We call this *oracle reference encoding*. Comparing heterogeneous scoring operations also requires a common evolving reference, since growth can alter direct parenthood, ancestry, sibling context, semantic competition, and evidence at multiple structural scales. We propose `TaxoAtlas`, which compiles each taxonomy state into a structured semantic universe with preserved parent-child ownership and ancestry, bounded semantic-allocation distortion, and bounded multiscale field response under admitted growth. Thus, when *electric vehicle* appears, the relation between *electric bus* and *bus* remains represented while the surrounding structural and semantic evidence changes in a controlled way. Different scorers read different projections of this same updated reference. We propagate these guarantees through eight recurring operation classes: distance, bilinear or angular similarity, distributional or Gaussian scoring, boxes or regions, nonlinear neural maps, normalized logits, joint parent-set objectives, and structurally gated angular scores.

We find that across these classes, a common stability principle emerges - *growth changes the effective evidence, and correctness survives only while the relevant decision margin or feasibility slack exceeds that change*. Exact recovery additionally requires the selector to return all and only the valid parents. In the running example, *bus* must retain its lead, *electric vehicle* needs a separate acquisition condition, and false candidates must remain outside the selected set. This explains why normalization can change a decision despite an unchanged pairwise preference, structural gates can block a semantically preferred parent, set penalties can reject a newly valid parent, and cutoffs can admit false links even when all true parents rank highest. Joint query-candidate movement can also cancel, giving much tighter guarantees than separate worst-case bounds. We test these consequences with an all-eight-class controlled history, matched fixed/open-world studies on Environment and Science, and sequential replays on six observed taxonomies. The joint-set objective misses the new parent in all 40 post-growth states despite retaining the old one, while the Gaussian rule ranks both true parents highest but admits false links in 39/40 states. Joint-response analysis certifies 200/200 distance and 195/200 Gaussian transitions, and all 192 before-reveal source-law checks match the predicted

changes. Under matched open-world competition, the structurally gated angular scorer's Hit@1 falls from 49.0% to 21.6% on Environment and from 37.0% to 31.5% on Science. Across 6 taxonomies, cosine retains all 72 initially correct links but acquires only 17/89 newly required parents.

Contributions. We formulate taxonomy growth as a scorer-characterization problem that separates retention, acquisition, retired-link removal, and false-link control. We introduce `TaxoAtlas` as an evolving structural-semantic reference with preservation and perturbation guarantees, and derive a unified characterization of eight recurring scoring operations and their selectors under growth. Controlled and observed-taxonomy studies expose the predicted failure mechanisms, including cases where initially correct rules fail despite exact current reference inputs.

2 Related Work

Taxonomy expansion and completion. Taxonomy expansion and completion place concepts in an existing hierarchy using semantic and structural evidence. Recent methods span fuzzy overlap (Xu et al., 2024), Gaussian distributional scoring (Mishra et al., 2026a), angular scores with structural statistics (Mishra et al., 2026b), and implicit-concept insertion (Shi et al., 2024). These works primarily evaluate end-to-end link recovery. We instead characterize when an already-correct scoring and selection rule remains correct as the taxonomy changes. The eight classes analyzed here are recurring mathematical operations, not eight end-to-end systems. Appendix F gives selected formula-level correspondences, with broader operation-level coverage in Appendix H.

Evolving and sequential taxonomies. Prior work also studies non-static construction. TaxoOrder learns insertion order within an expansion batch (Song et al., 2021); ATTEMPT predicts parent and child relations for non-leaf insertion (Xia et al., 2023); temporal construction updates relations as evidence changes (Tuan et al., 2016); and TaxoAdapt adjusts taxonomy width and depth as the corpus evolves (Kargupta et al., 2025). These methods study how the taxonomy itself should be updated. We condition on the revealed states and characterize their effect on a fixed scoring and selection mechanism, separating retention, acquisition, retired-link removal, and exact recovery.

Ranking stability and perturbation analysis. Ranking-region methods study sensitivity to scoring-weight changes (Asudeh et al., 2018; Chen & Jagadish, 2026), margin-based certification relates decision stability to bounded perturbations (Tsuzuku et al., 2018), and graphon transfer analyzes changing graph size under shared structural assumptions (Ruiz et al., 2020). Our perturbation is induced by taxonomy growth itself, which can alter representations, competitors, normalization, structural constraints, feasible outputs, and selection. We derive this reference change first and then propagate it through the complete scoring and selection rule.

3 Problem Formulation

In this section, we formalize the taxonomy state, the complete decision chain, and the four obligations that define correctness after one update. At stage t, the available taxonomy is a finite directed acyclic is-a graph with concept set V_t. Each concept is represented by the semantic evidence available to the task, such as its name and textual description, together with its observed taxonomy context. For a query q, the target is the set of its *direct parents*, $P_t(q) \subseteq V_t$, and the system returns $\widehat{P}_t(q)$. Thus $P_{t-1}(\textit{electric bus}) = \{\textit{bus}\}$ can become $P_t(\textit{electric bus}) = \{\textit{bus}, \textit{electric vehicle}\}$ when the second parent becomes available. We analyze a fixed decision chain across taxonomy states. A declared adapter maps the current reference evidence to a scoring operation; the *complete score* includes any normalization, interaction, penalty, or structural gate; and a selector returns $\widehat{P}_t(q)$. Under *oracle reference encoding*, each analyzed operation receives exactly the current inputs prescribed by its reference map. The gold set $P_t(q)$ is never supplied, so exact reference inputs imply neither a correct score nor a correct parent set. For one transition, we define $R_t = P_{t-1} \cap P_t,\ N_t = P_t \setminus P_{t-1},\ D_t = P_{t-1} \setminus P_t$, where dependence on q is suppressed. A correct update must retain R_t, acquire N_t, remove D_t, and exclude every other false link; exact recovery is $\widehat{P}_t = P_t$. In the running example, *bus* must be retained and *electric vehicle* acquired. If an intermediate concept later makes an old direct parent only an ancestor, that link belongs to D_t and must disappear. These obligations distinguish changes in the taxonomy from failures of

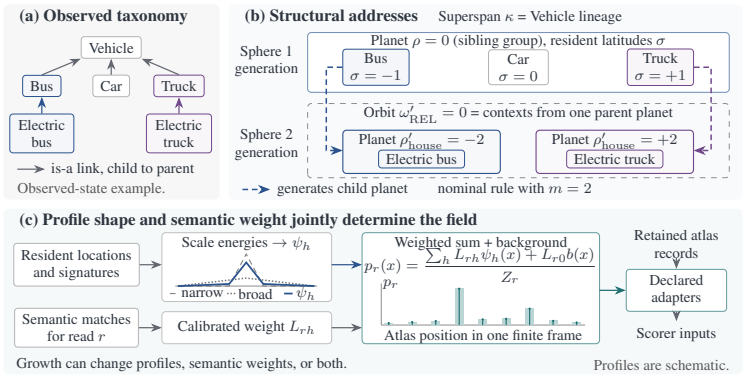

Figure 2: `TaxoAtlas` construction. (a,b) Labelled residents share a sibling planet and generate distinct child planets within one next-generation orbit. Coordinates illustrate $m = 2$. Links are observed construction inputs, not held-out query targets. (c) Scale energies determine emission profiles, while semantic calibration supplies their read-dependent weights. Normalizing the weighted sum with background gives a field over atlas positions, not parent-correctness probabilities. The labelled atlas retains observed ownership and ancestry. Semantic distortion and field changes are bounded under the stated allocation and update assumptions.

the representation, score, or selector. Characterizing them requires a reference whose structural and semantic state remains comparable as the taxonomy evolves.

4 `TaxoAtlas`: An Evolving Reference

The formulation in Section 3 requires a reference separating structural role, semantic position, and query-specific evidence. When *electric vehicle* arrives, *bus* can remain a child of *vehicle* while evidence for *electric bus* changes. `TaxoAtlas` makes these changes separately identifiable before scoring.

Reference construction. An *occurrence* is a concept's instance in an inherited structural context. Its labelled address records generation (sphere s), lineage (superspan κ), sibling-context family (orbit ω), sibling group (planet ρ), and position within that group (latitude σ). For a common positive refinement factor m, the nominal rule generates

$$s' = s + 1, \qquad \kappa' = \kappa, \qquad \omega'_{\mathrm{REL}} = m\rho, \qquad \rho'_{\mathrm{house}} = m\sigma. \tag{1}$$

Primes identify generated coordinates, with ω'_{REL} the child orbit and ρ'_{house} the child planet. Thus sibling residents *bus* and *truck* generate distinct child planets within one orbit (Figure 2). Ownership records preserve these distinctions under the corrections analyzed in Appendix A.1.

A *frame* F contains resident indices I_F and available integer positions X_F. In a sibling frame, resident i's position $z_i \in X_F$ is its allocated latitude tick. Semantic dissimilarities determine order and preferred positions before rounding and collision resolution. Every current resident receives unit raw structural mass, so a frame with $|I_F|$ residents uses equal structural weights $a_i = 1/|I_F|$. The same rule applies when a completed lower-level spectrum becomes a resident of a planet, orbit, or top-lineage frame. Growth rebuilds each affected frame from the current atlas, so new residents enter with unit mass and the equal shares are recomputed. The structural-summary vector v_i is its *signature*. Atomic signatures start at (1), while a compiled lower-level context supplies its normalized spectrum, aligned to the receiving scale components, as the signature used at the next level. Thus an atomic frame containing *bus*, *car*, and *truck* places three unit signatures with weights $1/3$ at their respective ticks. The measure $\mu_F = \sum_{i \in I_F} a_i v_i \delta_{z_i}$ places these weighted signatures at their positions, with δ_{z_i} a unit point mass.

Scales $j = 0, \ldots, M$ use radii 2^j, with M the smallest nonnegative integer for which 2^M exceeds the largest absolute tick offset from the frame's reference. The smoothing operator T_j spreads each signature centered at z with weights proportional to $\max\{2^j - d_F(x, z), 0\}$ over $x \in X_F$, normalized at each center to sum to one. Here d_F is interval or shortest cyclic tick distance. Thus

T_0 preserves integer positions, while broader views mix nearby *bus*, *car*, and *truck* signatures. Set $A_j = T_j - T_{j+1}$ for $j < M$ and $A_M = T_M$. These operators retain differences between adjacent views and the coarsest view. Nonempty frames with the unit-mass signatures above have positive total energy, giving

$$e_{Fj} = \|A_j \mu_F\|_F^2, \qquad \theta_{Fj} = \frac{e_{Fj}}{\sum_{k=0}^{M} e_{Fk}}. \tag{2}$$

The squared Frobenius norm sums squared entries over positions and signature components. The fractions θ_{Fj} form the frame's *spectrum*, recording how its energy is distributed across scales. Appendix A.1 gives the aggregation, operator, and interaction identities. *Signed detail responses can reinforce or cancel before squaring, so sibling contributions need not add independently.*

A *contributor* h supplies one spatial field component, using its lineage's sphere-zero orbit spectrum in the lineage-level construction. Relative-scale alignment to a common receiving chart X gives nonnegative weights β_{hj} summing to one. Its emission is $\psi_h(x) = \sum_j \beta_{hj} K_{hj}(x)$, where K_{hj} is the normalized tent at receiving scale j, centered at the contributor's allocated position z_h. A query, candidate, or calibration occurrence *read* r has a tick ξ_r, obtained from semantic distances by the frame's read rule without collision reassignment. Reads can coincide despite distinct allocated centers. Calibration uses a separate weighting rule: occurrence i of concept c_i receives residual-calibration weight $\widetilde{w}_i = 1/\nu(c_i)$, where $\nu(c_i)$ is its current visible occurrence multiplicity. For contributor h, let c_{rh} be the total calibration mass whose signed residual matches $\xi_r - z_h$, and let n_h be the total calibration mass in h's visible population. With B fixed residual bins, half-count smoothing gives $L_{rh} = (c_{rh} + 1/2)/(n_h + B/2)$. Growth recomputes these calibration weights from the current visible occurrence multiplicities. Held queries and their parent labels are excluded from the calibration populations. Appendix B.1 gives the residual-count definition and update law.

With a background probability profile b and nonnegative read-specific weight L_{r0}, the raw evidence T_r, total mass Z_r, and normalized field p_r are

$$T_r(x) = \sum_h L_{rh}\psi_h(x) + L_{r0}b(x), \qquad p_r(x) = \frac{T_r(x)}{Z_r}, \quad Z_r = \sum_h L_{rh} + L_{r0} > 0. \tag{3}$$

The sums cover contributors on X. Controlled constructions use uniform background, calibrated from all visible occurrences relative to the chart's reference tick. A lineage's profile can remain fixed while added members change its weight for *electric bus*. Unlike the individual component ψ_h, p_r combines all contributions for read r. *The field describes evidence over atlas positions, not probabilities of correct parenthood.*

Guarantees under growth. Three properties support scorer analysis. First, labelled occurrences and ownership records preserve parent-child ownership and ancestry, although spectra and fields are lossy (Appendix A.1). Second, semantic allocation has bounded distortion. For ordered normalized source coordinates x_i with dissimilarity $|x_i - x_k|$, reference coordinate x_0, and scale $R > 0$, let $t_i = R(x_i - x_0)$ be the ideal position. If $\delta = \max_i |z_i - t_i|$ bounds allocation error, then

$$\left| \frac{|z_i - z_k|}{R} - |x_i - x_k| \right| \le \frac{2\delta}{R}. \tag{4}$$

Appendix A.2 derives δ and the resulting spectral error for the analyzed allocation branch. The two endpoint errors bound spacing distortion, but a smaller source advantage need not survive quantization.

A surviving distance advantage yields a positive emission advantage under shared-profile conditions. On a common cyclic chart with distance d, let $z_q = \xi_q$ be query q's read tick, and let z_h, z_c be the centers associated with its independently labelled true parent h and rival c. Write $\Gamma = d(z_q, z_c) - d(z_q, z_h) > 0$ for the remaining distance gap. Assume the emissions share an aligned spectrum derived from nonnegative unit-mass source smoothings with N_s positions, D_s signature components, and M_s detail scales. Alignment retains the full broadest source weight at receiving scale M, whose tent radius exceeds the compared distances. Its normalizer Z_M is center-independent because the chart is cyclic.

Proposition 1 (Semantic preference)**.** *Under these conditions,*

$$\psi_h(z_q) - \psi_c(z_q) \ge \frac{\Gamma}{N_s D_s (2M_s + 1) Z_M} > 0.$$

The proof is in Appendix A.3.

Unit mass guarantees a minimum broad-scale weight, whose tent favors the closer *bus* over a rival. *Broad-scale support preserves a strict preference even when narrower profiles do not reach the query.* Unequal emissions require the allowances in Appendix A.3. Calibration and contextual contributions additionally affect the full field comparison (Appendix B.3).

Third, growth has a bounded effect on the completed field. For a fixed read r, primes now denote post-growth quantities. Entire profiles are aligned on a common chart before comparison. Let $\mathcal{S}, \mathcal{B}, \mathcal{D}$ contain persistent, born, and removed contributors. Total variation between normalized profiles is $\mathrm{TV}(u, v) = \frac{1}{2}\sum_x |u(x) - v(x)|$. With $\rho_h = \mathrm{TV}(\psi'_h, \psi_h)$, the raw-evidence allowance is

$$\begin{aligned} A_r = &\sum_{h\in\mathcal{S}} \big(|L'_{rh} - L_{rh}| + 2L'_{rh}\rho_h\big) + \sum_{h\in\mathcal{B}} L'_{rh} + \sum_{h\in\mathcal{D}} L_{rh} \\ &+ |L'_{r0} - L_{r0}| + 2L'_{r0}\,\mathrm{TV}(b', b), \end{aligned} \tag{5}$$

so $\|T'_r - T_r\|_1 \le A_r$, where the norm sums absolute changes over chart positions. The terms account for reweighting, profile movement, births, removals, and background change. Normalization additionally depends on both endpoint masses.

Theorem 2 (Normalized field response)**.** *For nonnegative weights and positive masses Z_r, Z'_r on aligned finite charts,*

$$\mathrm{TV}(p'_r, p_r) \le e_r := \min\left\{1, \frac{A_r + |Z'_r - Z_r|}{2\max\{Z_r, Z'_r\}}\right\}. \tag{6}$$

The proof is in Appendix B.2.

The budget e_r controls the normalized field, including redistribution caused by evidence elsewhere. *An unchanged local contribution need not retain its relative weight when the taxonomy grows.* Learned parameters, structural counts, feasible sets, and permitted outputs require additional records or response bounds when p_r does not determine them (Appendices B.4 and C.1).

Repeated arrivals separate geometric preservation from local field protection. Appendix A.4 controls allocation after every prefix of the declared catalogue and forecasts fields before the next identity is revealed. In Appendix A.6, protection means that an anchored position has greater field evidence than its neighbors. Supporting and competing arrivals can preserve or erase this advantage, yielding different intermediate protection histories despite identical endpoints. The scorer analysis must still determine when these reference changes preserve the returned parent set.

5 Scoring and Parent Recovery

A bounded change in `TaxoAtlas` does not by itself guarantee a stable parent decision. A scorer may read the field together with quantities such as structural counts or fixed parameters, while the selector acts only after normalization, interactions, penalties, gates, and output constraints. We call the value finally presented to the selector the *complete score*. The analysis therefore follows the chain from field change to score change and then to parent-set recovery.

From field change to score change. Fix query q, and let larger complete scores be preferred. For candidate h, write $S_0(h)$ for its pre-growth score and $S_t(h)$ for its score at current state t. A score allowance E_h satisfies $|S_t(h) - S_0(h)| \le E_h$. The field budget e_r from Theorem 2 yields such an allowance only after passing through the scorer's declared readout. For example, let $u_r = \sqrt{p_r}$ elementwise. Since p_r has unit mass, u_r has unit Euclidean norm, and $\mathrm{TV}(p'_r, p_r) \le e_r$ implies $\|u'_r - u_r\|_2 \le \sqrt{2e_r}$. Hence the negative-distance score $S(h) = -\|u_q - u_h\|_2$ has the valid allowance $E_h = \sqrt{2e_q} + \sqrt{2e_h}$. Appendix C.1 derives the corresponding transfers for all eight classes, including additional records when p_r alone does not determine the score.

Now let y be a still-valid parent such as *bus*, and let h be an incorrect candidate already present before growth. If their initial lead is $m_h = S_0(y) - S_0(h) > 0$, then

$$S_t(y) - S_t(h) \ge m_h - E_y - E_h. \tag{7}$$

A positive remainder certifies that y still outranks this persistent rival. Retaining y as an output additionally requires the corresponding comparisons against all relevant rivals and the conditions

Table 1: Eight scoring-operation classes as specializations of the same stability problem.

Class	Reference readout	Characteristic stability issue
Distance/angular	relative geometry	only relative query-candidate motion determines the pairwise score change
Bilinear/fuzzy	paired features and comparison weights	features, weights, or a normalization denominator can change
Gaussian	field-derived means and variances	spread can change a decision even when centers remain stable
Box/region	region boundaries and overlap	containment can also hold for a broader ancestor
Fixed neural	derived features and fixed network parameters	bounded input motion needs bounded network response; identical inputs remain indistinguishable
Normalized token	pre-normalization scores and supported outputs	adding outputs can change normalized admission without changing old logits
Joint set	unary scores, interactions, and feasible sets	the complete objective can prefer an incomplete parent set
Structural gate	semantic score and graph statistics	missing structural evidence can suppress a strong semantic match

imposed by the actual selector. Along a growth trajectory, satisfying these sufficient conditions at every prefix certifies uninterrupted preservation, while failure of a bound is merely inconclusive. Appendix C.2 proves Equation 7. *An old margin protects an ordering only against the relative score movement that growth can induce.*

Acquisition is different because a newly required relation has no inherited correctness margin, even when the node itself existed earlier. Let V_t be the current candidate universe, $A_t \subseteq V_t$ the candidates accessible to the selector, and $P_t \subseteq V_t$ the true current parent set. For threshold τ, define $\widehat{P}_t = \{h \in A_t : S_t(h) > \tau\}$. Suppose every available candidate has a valid score interval $\underline{S}_h \le S_t(h) \le \overline{S}_h$.

Proposition 3 (Exact parent-set certificate)**.** *If $P_t \subseteq A_t$, every $h \in P_t$ satisfies $\underline{S}_h > \tau$, and every $r \in A_t \setminus P_t$ satisfies $\overline{S}_r \le \tau$, then $\widehat{P}_t = P_t$. (**Proof:** Appendix C.3.)*

For *electric bus*, both *bus* and *electric vehicle* must lie above the cutoff while every false or retired parent lies at or below it. Other selectors impose analogous access, capacity, feasibility, or decoding conditions (Appendices C.3 and C.5). *Retention preserves old required links, acquisition establishes newly required links, and exact recovery additionally excludes every false or retired link.*

Eight scoring-operation classes. The same field change can affect different scoring operations through different retained quantities. Table 1 summarizes these interfaces, with complete response bounds in Appendix C.1 and the information condition in Appendix B.4. Separate endpoint bounds can be conservative because they discard coordinated movement. Since u_q and u_h share an aligned chart,

$$\big|\|u'_q - u'_h\|_2 - \|u_q - u_h\|_2\big| \le \|(u'_q - u'_h) - (u_q - u_h)\|_2. \tag{8}$$

The right-hand side measures change in the relative query-candidate representation. If *electric bus* and *bus* move together, their individual movements can be large while this quantity is small or zero. Appendix C.1 proves this refinement and its Gaussian counterpart. Prospective certification additionally requires this joint response to be bounded before the update. *Decision-relevant change can be much smaller than the separate movement of the objects being compared.* Two obstructions show why favorable unary evidence still need not yield the correct parent set. First, let $\mathcal{F}_t$ be a downward-closed feasible family, meaning every subset of a feasible set is feasible, and consider

$$J(P) = \sum_{h \in P} s_h - \lambda \frac{|P|(|P|-1)}{2}, \qquad 0 \le s_h \le 1, \qquad \lambda > 1.$$

Here s_h is unary evidence and λ penalizes each selected pair. For any $|P| \ge 2$, deleting one member increases $J(P)$ because the removed unary gain is at most one while the removed pair penalty exceeds one. Thus no multi-parent set can maximize this objective, even when *bus* and *electric vehicle* both receive maximal unary evidence (Appendix C.4). *Perfect evidence for each parent can still be incompatible with the objective used to select them jointly.*

Structural gating creates a different obstruction. Let $s_h \in [-1, 1]$ be semantic compatibility, n_h the number of visible distinct strict descendants, and $\beta > 0$ the structural weight. With

$$g_h = s_h - 1 + \beta[\log_2(1 + n_h/2)]^2, \qquad S_h^{\text{gate}} = s_h \mathbf{1}\{g_h > 0\},$$

a leaf has $n_h = 0$, hence $g_h = s_h - 1 \le 0$ even when $s_h = 1$. A newly valid leaf can therefore fail the gate despite maximal semantic compatibility. Its gated score becomes zero, while final exclusion

still depends on the selector. Appendices C.4 and C.5 give the complete construction and decoder analysis. *Semantic evidence can remain favorable while interaction terms, structural eligibility, or selection alter the returned parent set.* The characterization therefore separates four interfaces: growth can erase an old ordering, a newly required parent can fail its current acquisition conditions, the complete score can change despite favorable unary evidence, and the selector can still return an incomplete or overinclusive set. The evaluation below tests these mechanisms separately.

6 Evaluation Setup

The evaluation asks two complementary questions. Controlled studies test the predicted failure mechanisms under exact reference encoding, while observed-taxonomy studies test whether the same distinctions arise under real taxonomy structure with fixed supplied semantic vectors.

Controlled studies. The parent-acquisition study evaluates the eight scoring-operation classes characterized above on the same change in the correct parent set, from $\{B\}$ to $\{B, C\}$. Once C becomes valid, the query and both true parents receive equally favorable profiles, $p_q = p_B = p_C$. Every rule is correct before growth, after which its calibration or selector remains fixed for 40 states. We measure retention of B, acquisition of C, false-link exclusion, and exact recovery of $\{B, C\}$ (Appendix D.3). A second study keeps the correct parent fixed across 201 states while the reference changes. Since every realized decision remains correct, it isolates actual preservation from what the sufficient certificates can prove. Separate worst-case bounds are compared with joint-response refinements, and before-reveal field predictions are tested under the declared arrival processes (Appendices D.5 and A.4).

Observed taxonomy growth. The matched study uses Environment and Science from SemEval-2016 Task 13 (Bordea et al., 2016), evaluating the same held queries and scoring formulas before and after growth with every current correct parent included before additional candidates. This parent-complete design isolates growth-induced candidate competition from retrieval failure. We evaluate distance, bilinear or angular similarity, Gaussian scoring, boxes or regions, nonlinear neural scoring, normalized logits, and structurally gated angular scoring; joint parent-set objectives are excluded because they score complete sets rather than individual candidates. The main paper reports the five-choice setting, with larger neighborhoods, MRR, and paired transitions in Appendix G.2. A separate Environment study follows 82 held single-parent queries under fixed supplied embeddings as 41 concepts arrive, testing candidate competition, structural gating, and arrival-order effects (Appendix G.3). A six-source study covers Environment, Science, and Food (Bordea et al., 2016), MeSH 2026 (National Library of Medicine, 2026), and WordNet Noun and WordNet Verb (Miller, 1992), measuring retention of initially correct links and acquisition of newly required parents under fixed supplied vectors and initial-state calibration. It covers cosine similarity and structural gating and not all eight classes, testing the predicted retention-acquisition separation (Appendix G.5). The controlled studies isolate the characterized mechanisms under exact inputs, while these observed studies test whether they recur under real taxonomy structure; results are therefore interpreted within each protocol rather than pooled across settings.

7 Results

Exact Inputs Can Fail. The controlled history changes the correct set from $\{B\}$ to $\{B, C\}$ while assigning identical reference profiles $p_q = p_B = p_C$ to the query and both true parents. The failures arise at distinct interfaces: the joint-set objective retains B but rejects the newly required C despite maximal unary evidence, the Gaussian cutoff admits false links despite ranking both true parents highest, normalization changes decisions that remain exact under the corresponding unnormalized logits, and structural gating can delay an otherwise valid parent. *Retention, acquisition, and exact recovery are distinct properties even with exact current reference inputs.*

Table 2: Parent recovery after C becomes required alongside B. Exact recovery returns only $\{B, C\}$.

Construction	Retains B	Acquires C	Exact set	False-link states
Distance	40	40	40	0
Bilinear	40	40	40	0
Moment Gaussian	40	40	1	39
Box overlap	40	40	40	0
Fixed neural score	40	40	40	0
Normalized token score	22	22	22	0
Penalized joint set	40	0	0	0
Structural gate	40	39	39	0

Joint Response Sharpens Certification. All eight rules remain correct throughout a separate 201-state history, while separate worst-case bounds certify none of the 200 distance or Gaussian transitions. Joint response raises coverage to 200/200 and 195/200, leaving only five correct Gaussian transitions uncertified (Appendix D.5). *Much of the apparent looseness disappears when changes that occur together are analyzed together.*

Observed Growth Shows the Same Effects. Observed-taxonomy studies test these mechanisms under real taxonomy structure and fixed supplied semantic vectors; full sweeps and controls appear in Appendices G.2, G.3, and G.5.

The matched studies include every current correct parent, ruling out retrieval omission as the source of turnover. In Environment, angular similarity loses correct decisions even when the correct-parent score is unchanged; for *plant resources*, arriving *animal resources* overtakes *natural resources* by only 2.80×10^{-4}. Structural gating further shows that the same final scores can follow different transient failures under different arrival orders. Across six taxonomies, the fixed-vector studies reproduce the retention-acquisition gap rather than estimate trained-system accuracy. Full-source `TaxoAtlas` constructions on Environment, Science, and Food also agree with exact readouts to numerical precision (Appendix G.4). *Growth can therefore change competition, structural eligibility, and selection without first corrupting the underlying semantic evidence.*

Table 3: Observed-taxonomy summary. Metrics are protocol-specific and should be read row-wise.

Study	Primary result	What changes under growth
Matched Environment, F8	Hit@1 $49.0\% \rightarrow 21.6\%$	20 paired losses, 6 gains; tie-safe $37.3\% \rightarrow 17.6\%$
Matched Science, F8	Hit@1 $37.0\% \rightarrow 31.5\%$	13 paired losses, 10 gains; tie-safe $29.6\% \rightarrow 27.8\%$
Environment angular replay	20/24 initial successes retained	4 are overtaken although their old correct-parent scores do not change
Environment structural gate	8/8 correct at endpoint, 7/8 uninterrupted	one initially correct decision is lost and later recovered; order changes the trajectory
Six-source cosine	72/72 retained, 17/89 acquired	perfect retention coexists with 19.1% acquisition
Six-source structural gate	36/37 retained, 13/89 acquired	97.3% retention coexists with 14.6% acquisition

8 Discussion and Limitations

The characterization exposes two evaluation issues that are easy to miss in growing taxonomies. Final-state accuracy does not reveal whether a decision failed and later recovered along the growth path, and separate worst-case perturbation bounds can hide stability that is visible when coupled changes are analyzed jointly. Consequently, growth-aware evaluation should distinguish endpoint from trajectory behavior and preserve dependencies among the quantities that move together. The guarantees apply to scoring and selection rules whose representations and updates satisfy the stated correspondence to `TaxoAtlas`. The observed studies use fixed, supplied semantic vectors and declared scoring rules, rather than trained end-to-end systems. The six-source study covers cosine and structural gating, while the matched study evaluates parent-complete conditional ranking instead of retrieval. The target is direct-parent recovery; child selection, unrestricted graph repair, and recovery of missing internal concepts require additional decision objects and guarantees.

9 Conclusion

In this paper, we characterize stability of scoring and selection rules under taxonomy growth using `TaxoAtlas`, a common evolving reference with structural, semantic, and perturbation guarantees. Across eight recurring scoring-operation classes, retention depends on surviving margin or feasibility slack, acquisition requires a separate current-state condition, and exact recovery additionally depends on the selector. Controlled and observed studies show that exact reference inputs can still produce incomplete or overinclusive parent sets, strong retention can coexist with weak acquisition, and identical endpoints can conceal different preservation histories. *Growth-aware taxonomy models should therefore separate semantic compatibility, structural eligibility, and set selection, with an explicit stability margin at each interface.*

REPRODUCIBILITY STATEMENT

The appendix specifies the field laws, input maps, scoring and selection rules, proof assumptions, constructed histories, and dataset protocols used in this paper. The released artifacts reconstruct the controlled examples, arrival forecasts, matched ranking aggregates, and declared formula replays at the scope stated in Appendix I. These calculations evaluate declared mathematical operations and not trained end-to-end systems.

AI USE STATEMENT

We used AI to identify relevant literature and edit and polish portions and figures of the manuscript. AI-assisted feedback also supported checks of clarity. The authors take full responsibility for all mathematical arguments, experimental protocols, numerical results, and bibliographic references against the available research materials and primary sources.

ETHICS STATEMENT

`TaxoAtlas` analyzes taxonomy structure and supplied semantic evidence rather than collecting new personal or behavioral data. The reported studies use existing benchmark and public taxonomy resources and do not introduce new human-subject data collection. Taxonomy errors can nevertheless propagate to downstream retrieval, recommendation, or knowledge systems, and this risk may be more consequential in domains such as biomedical knowledge organization. The stability guarantees characterize declared scoring and selection rules under stated assumptions; they do not establish that a predicted parent is factually correct, socially appropriate, or suitable for consequential deployment. Any deployment should therefore preserve domain-specific review, provenance, and correction mechanisms.

REFERENCES

Abolfazl Asudeh, H. V. Jagadish, Gerome Miklau, and Julia Stoyanovich. On obtaining stable rankings. *Proceedings of the VLDB Endowment*, 12(3):237–250, 2018. doi: 10.14778/3291264.3291269. URL https://www.vldb.org/pvldb/vol12/p237-asudeh.pdf.

Georgeta Bordea, Els Lefever, and Paul Buitelaar. SemEval-2016 task 13: Taxonomy extraction evaluation (TExEval-2). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pp. 1081–1091, San Diego, California, June 2016. Association for Computational Linguistics. doi: 10.18653/v1/S16-1168. URL https://aclanthology.org/S16-1168/.

Houming Chen and H. V. Jagadish. The general stability of ranking. arXiv:2607.01546v1, 2026. URL https://arxiv.org/abs/2607.01546v1.

Priyanka Kargupta, Nan Zhang, Yunyi Zhang, Rui Zhang, Prasenjit Mitra, and Jiawei Han. TaxoAdapt: Aligning LLM-based multidimensional taxonomy construction to evolving research corpora. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 29834–29850. Association for Computational Linguistics, 2025. doi: 10.18653/v1/2025.acl-long.1442. URL https://aclanthology.org/2025.acl-long.1442/.

George A. Miller. WordNet: A lexical database for English. In *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, February 23-26, 1992*, 1992. URL https://aclanthology.org/H92-1116/.

Sahil Mishra, Srinitish Srinivasan, Srikanta Bedathur, and Tanmoy Chakraborty. Taxobell: Gaussian box embeddings for self-supervised taxonomy expansion. In *Proceedings of the ACM Web Conference 2026*, pp. 4326–4337, 2026a.

Sahil Mishra, Srinitish Srinivasan, Sourish Dasgupta, and Tanmoy Chakraborty. Polaris: Coupled orbital polar embeddings for hierarchical concept learning. In *Forty-third International*

Conference on Machine Learning, 2026b. URL https://openreview.net/forum?id=FQNYtjPBxv.

National Library of Medicine. Release of 2026 MeSH XML, RDF datasets. NLM Technical Bulletin, Jan-Feb 2026, No. 468, b2, 2026. URL https://www.nlm.nih.gov/pubs/techbull/jf26/brief/jf26_MeSH_RDF_2026_DSrelease.html.

Luana Ruiz, Luiz F. O. Chamon, and Alejandro Ribeiro. Graphon neural networks and the transferability of graph neural networks. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/12bcd658ef0a540cabc36cdf2b1046fd-Abstract.html.

Jingchuan Shi, Hang Dong, Jiaoyan Chen, Zhe Wu, and Ian Horrocks. Taxonomy completion via implicit concept insertion. In *Proceedings of the ACM on Web Conference 2024*, pp. 2159–2169, 2024. doi: 10.1145/3589334.3645584. URL https://www.cs.ox.ac.uk/people/ian.horrocks/Publications/download/2024/Shi0CW024.pdf.

Xiangchen Song, Jiaming Shen, Jieyu Zhang, and Jiawei Han. Who should go first? a self-supervised concept sorting model for improving taxonomy expansion. arXiv:2104.03682v2, 2021. URL https://arxiv.org/abs/2104.03682v2.

Yusuke Tsuzuku, Issei Sato, and Masashi Sugiyama. Lipschitz-margin training: Scalable certification of perturbation invariance for deep neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/485843481a7edacbfce101ecb1e4d2a8-Abstract.html.

Luu Anh Tuan, Siu Cheung Hui, and See Kiong Ng. Utilizing temporal information for taxonomy construction. *Transactions of the Association for Computational Linguistics*, 4:551–564, 2016. doi: 10.1162/tacl_a_00117. URL https://aclanthology.org/Q16-1039/.

Fei Xia, Yixuan Weng, Shizhu He, Kang Liu, and Jun Zhao. Find parent then label children: A two-stage taxonomy completion method with pre-trained language model. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 1032–1042, Dubrovnik, Croatia, 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.eacl-main.73. URL https://aclanthology.org/2023.eacl-main.73/.

Fred Xu, Song Jiang, Zijie Huang, Xiao Luo, Shichang Zhang, Yuanzhou Chen, and Yizhou Sun. FUSE: Measure-theoretic compact fuzzy set representation for taxonomy expansion. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 2707–2720, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.158. URL https://aclanthology.org/2024.findings-acl.158/.

APPENDIX CONTENTS

A Construction and Correspondence Proofs **14**

A.1 Structural Organization and Energy 14

A.2 Semantic Ordering and Allocation Distortion 15

A.3 Semantic Preference and Its Limits 16

A.4 Inductive Correspondence and Next-Arrival Forecasts 17

A.5 Potential Protection and New Contributors 19

A.6 Successive Arrivals and Shared-Query Survival 20

B Taxonomy Fields **22**

B.1 Source Construction 22

B.2 Field Response 24

B.3 Calibrated Semantic Order 26

B.4 When Does the Reference Contain Enough Information? 29

C Family Responses and Decisions **32**

C.1 From Field Change to Score Change 32

C.2 Preservation from Complete Scores 40

C.3 Retention, Acquisition and Exact Parent Sets 42

C.4 Family-Specific Obstructions 43

C.5 Composition and the Actual Decoder 48

D Constructed Comparisons **52**

D.1 A Fixed Target Under Remote Structural Growth 52

D.2 Experimental Evidence Map 56

D.3 Shared Acquisition History 56

D.4 Output Obstructions and Controls 57

D.5 Preservation Certificates: Correctness versus Certifiability 58

E Gated Scorer Compatibility **59**

E.1 Semantic Compatibility Meets Structural Eligibility 60

E.2 When a Gate Forces Error Across a Scorer Class 60

E.3 How Far Must a Representation Move to Become Eligible? 61

E.4 Binding the Gate to the Constructed Field 62

E.5 Removing Remote-Range Drift Without Removing the Gate 63

E.6 Can One Global Gate Be Selective Enough? 64

E.7 The Limit of Compensation 65

F Specific Scorer Formula Correspondence **65**

G Dataset Protocols and Evidence **67**

G.1 Theory-to-Evidence Coverage 67

G.2 Matched Conditional Ranking Protocol 67

G.3 Environment Formula Replay 68

G.4 Full-Source Descriptive Profiles 69

G.5 Parent Acquisition on Six Source Taxonomies 70

H Related Methods and Scoring Operations **71**

I Reproducibility and Claim Scope **72**

A Construction and Correspondence Proofs

A.1 Structural Organization and Energy

The reference construction requires structural separation. A concept occurrence belongs to the sibling context generated by its parent, while its own children occupy a newly generated context. Under the nominal rule in Equation 1, two siblings at different latitudes necessarily generate different child-planet coordinates because multiplication by the positive refinement factor m is injective. The address therefore preserves ownership distinctions needed by the later analysis. This correspondence is an association of labeled addresses and not an isometry between the bounded and cyclic metrics. Corrected routes retain the same distinction through their recorded ownership and correction witnesses.

Structural dependence is kept distinct from semantic calibration. In the analyzed compiler, unit occurrence signatures enter a latitude summary and then the corresponding planet and orbit summaries. A lineage-level emitter reads its sphere-zero orbit signature. A deeper child context therefore affects an ancestral spectral object only when it lies in that object's declared dependency graph. Calibration membership is separate: deeper occurrences may change a lineage's calibration population without changing its sphere-zero emission shape. Claims of unchanged inputs below always refer to this complete typed dependency graph.

Proposition 4 (Structural correlation and interaction)**.** *For a fixed frame F with aligned vector-valued measure $\mu = \mu_F = \sum_{i \in I_F} a_i v_i \delta_{z_i}$, define $G_j = A_j^\top A_j$. Then*

$$e_{Fj} = \sum_{i,k \in I_F} a_i a_k \langle v_i, v_k \rangle G_j(z_i, z_k). \tag{9}$$

For raw additive contributions u, v and $Q_j(V) = \|A_j V\|_F^2$,

$$Q_j(V+u+v) - Q_j(V+u) - Q_j(V+v) + Q_j(V) = 2\langle A_j u, A_j v \rangle_F.$$

Proof. Linearity gives

$$A_j \mu = \sum_i a_i (A_j \delta_{z_i}) v_i.$$

Expanding the squared Frobenius norm separates the interaction into spatial and signature terms. The spatial term is $(A_j^\top A_j)(z_i, z_k)$, yielding Equation 9. For two additive changes, expanding the four squared norms cancels the individual and baseline terms, leaving exactly $2\langle A_j u, A_j v \rangle_F$. □

The interaction term gives the first important consequence of the multiscale construction. Two additions need not contribute independent positive amounts of detail energy. An off-diagonal entry of $A_j^\top A_j$ may be negative even though the full matrix is positive semidefinite, so two additions can partially cancel at a particular scale. Normalizing the energies introduces a second coupling through the changing denominator. The raw identity therefore isolates an actual source of interaction without introducing an extra force or heuristic coupling term.

The signed archive retains more information than its normalized spectrum. With $D_j = (T_j - T_{j+1})\mu$ and $S_M = T_M \mu$,

$$S_M + \sum_{j<M} D_j = T_0 \mu = \mu,$$

when the finest smoothing is the identity. Recovering the labelled taxonomy state additionally requires the labelled atoms, known nonzero weights, and typed incidence records. The normalized energy vector alone is intentionally lossy. For a single-child measure,

$$e_j = \|v\|_2^2 \|A_j \delta_z\|_2^2,$$

so the signature magnitude cancels from every normalized scale weight. A child's signature can therefore change while its parent's normalized spectrum remains unchanged.

Lemma 5 (Dependence and consistent relabelling)**.** *On a finite fixed compiler dependency graph, unchanged inputs outside the downstream closure of changed inputs retain their computed values. A relabelling that transports all identities, positions, operators, references and incidences consistently preserves scalar energies and transported fields.*

Proof. Process the compiler graph in topological order. A node outside the downstream closure of the changed inputs receives exactly the same inputs and applies the same prescribed function, so its output is unchanged. Induction completes the first claim.

For a consistent coordinate permutation P and channel permutation Q, a transformed wave is $P(A_j\mu)Q$. Permutations preserve the Frobenius norm, hence preserve scalar energies. The normalized coefficients, kernels, weights, identities, and incidences are transported consistently. Finite summation then commutes with the same transport, proving the field claim. □

The lemma fixes the relevant notion of locality. A changed reference or a shared calibration population belongs to the changed-input closure even when it is taxonomically distant. Conversely, pure coordinate relabelling does not alter the construction when every dependent object is transported consistently. Taxonomic graph distance alone is therefore insufficient to determine which reference quantities can change.

A.2 Semantic Ordering and Allocation Distortion

Structural ownership determines which concepts share a context. The declared source family then quantifies how semantic arrangement changes when it is converted to finite atlas coordinates.

Consider $n = 2k + 1$ distinct sibling coordinates

$$0 \le x_1 < \cdots < x_n \le 1$$

with distance $|x_i - x_j|$. Unit vectors $(\cos(\pi x_i), \sin(\pi x_i))$ realize this normalized angular distance on a semicircle. The bounded latitude frame is $[-R, R]$, with $R \ge k$, median reference $x_0 = x_{k+1}$, and fixed orientation. The preferred integer tick is

$$p_i = \operatorname{sign}(x_i - x_0)\, \lfloor R \operatorname{round}_{\text{even}}(M|x_i - x_0|)/M \rfloor, \qquad M = 2^{32} - 1.$$

Zero is reserved for the reference. The declared branch first resolves same-ray collisions by nearest vacant slots and then assigns Hamiltonian-ordered residents to the sorted resolved slots on each ray. The guarantee below applies to this exact branch rather than to arbitrary Hamiltonian heuristics.

Proposition 6 (Order and distance distortion)**.** *Increasing or decreasing source order minimizes the open Hamiltonian length. The oriented sorted-slot assignment preserves source order and minimizes maximum error relative to ordered ideals among assignments to those same slots. If $C = \max_i |c_i - p_i|$ is the provisional slot displacement, then*

$$\delta = \max_i |z_i - R(x_i - x_0)| \le C + 1 + R/(2M).$$

Equation 4 follows.

Proof. Every path visiting both source extremes has length at least $x_n - x_1$. Increasing order attains exactly this value by telescoping, whereas any change of direction among distinct line positions introduces additional length. For the allocation claim, let ordered ideals satisfy $t_i \le t_j$ and suppose assigned slots cross, with $u > v$. Swapping the assignments cannot increase $\max(|u - t_i|, |v - t_j|)$. Repeated uncrossing therefore yields the order-preserving assignment with no larger maximum error.

Quantization contributes at most $R/(2M)$ and the floor contributes less than one to $|p_i - t_i|$. Preferred positions are ordered on each ray, so the same uncrossing argument bounds the final displacement relative to the preferred positions by C. Triangle inequality gives the stated bound on δ, and applying it at both endpoints yields Equation 4. □

The proposition turns finite atlas allocation into a controlled semantic distortion. For the empty-ray nearest-vacancy construction with at most k residents per ray and at least k available slots, $C \le k$. Other slot landscapes require their actual displacement certificate. For pairs separated by at least γ, a positive multiplicative guarantee requires

$$2\delta < R\gamma.$$

No uniform multiplicative guarantee is possible for arbitrarily close source points because finite quantization imposes a nonzero additive error.

The same positional error propagates to the multiscale spectrum. Extend each normalized tent to a real-valued center t in the bounded frame:

$$K_j(v;t) = \frac{(2^j - |v-t|)_+}{\sum_w (2^j - |w-t|)_+}.$$

This analytical kernel agrees with the declared discrete kernel at integer centers. On each unit interval, its unnormalized vector w and denominator Z are affine, giving the global derivative bound

$$\lambda_j = \max_r \frac{\|\Delta w\, Z_0 - w_0 \Delta Z\|_1}{\min(Z_0, Z_1)^2}.$$

For identical child weights and signatures in the ideal and allocated configurations, define

$$\bar{d} = \sum_i a_i |z_i - t_i|, \qquad \eta_j = \min(2, \lambda_j \bar{d}).$$

Integrating the derivative bound gives

$$\|S_j^C - S_j^{\text{ideal}}\|_1 \le \eta_j.$$

Hence the detail-wave error is at most $b_j = \eta_j + \eta_{j+1}$, with $b_M = \eta_M$. The difference-of-squared-norms identity then bounds the energy error by

$$(\|W_j^C\|_1 + \|W_j^{\text{ideal}}\|_1) b_j.$$

The multiplier is at most four at detail scales and two at the coarse scale. Normalizing the energy vectors gives

$$\|\theta_C - \theta_{\text{ideal}}\|_1 \le \min\left(2, \frac{2\sum_j \text{energy-error}_j}{\max(E_C, E_{\text{ideal}})}\right).$$

A unit-mass coarse vector on NK entries has squared norm at least $1/(NK)$, providing an input-derived positive energy floor.

Thus the semantic-allocation bound propagates through the construction: bounded coordinate distortion implies bounded distortion of the multiscale summary. The conclusion concerns this declared summary. It does not claim that the normalized spectrum reconstructs all pairwise semantic distances.

A.3 Semantic Preference and Its Limits

Controlled distortion is useful only if an informative semantic preference can survive it. Let the true parent $h(q)$ be labelled independently of the field and assume, in a declared metric chart,

$$\Gamma(q) = \min_{b \ne h(q)} \left[d(q, c_b) - d(q, c_{h(q)}) \right] > 0.$$

One source family satisfying this condition places query reads within radius r of their true center and separates centers by $D > 2r$, yielding $\Gamma \ge D - 2r$ by triangle inequality. The positive gap is therefore an assumption on the semantic evidence, not a label induced by the field construction.

Proof of Proposition 1. The source coarse smoothing has $N_s D_s$ nonnegative entries summing to one. Cauchy-Schwarz gives

$$e_{s,M_s} \ge \frac{1}{N_s D_s}.$$

Each detail is the difference of two nonnegative unit-mass vectors and therefore has squared norm at most two. Coarse energy is at most one, so the total energy is at most $2M_s + 1$. Hence

$$\theta_{s,M_s} \ge \frac{1}{N_s D_s (2M_s + 1)}.$$

The declared relative-scale alignment maps source index j to jM/M_s when $M_s > 0$ and linearly divides its mass between adjacent receiving scales. Source index M_s maps exactly to M, retaining its full weight. All other contributions are nonnegative and total mass remains one. Therefore

$$\theta_M \ge \theta_{s,M_s}.$$

When $M_s = 0$, the single source weight is mapped to M and the same inequality holds.

On the cyclic receiving frame, every kernel $(2^j - d)_+/Z_j$ is nonincreasing in distance, and its normalizer is independent of the center. The coarse radius exceeds every receiving distance, so the coarse-scale preference between the true parent and a rival is at least

$$\frac{d_b - d_h}{Z_M} \geq \frac{\Gamma}{Z_M}.$$

Every other scale contributes a nonnegative difference. Combining this with the retained coarse-weight lower bound proves Proposition 1. □

The guarantee requires a quantitative amount of broad-scale support. If an alternative alignment retains only a certified fraction $\alpha > 0$ of the source coarse weight, the preference lower bound is multiplied by α. Positivity alone gives no uniform numerical floor.

The result also clarifies what may change without reversing the semantic preference. The spectra may vary at every step while the preference remains intact, provided the semantic gap stays positive and the common cyclic geometry remains valid. If the query and two candidate centers move by at most ϵ, s_h, and s_b, respectively, then the new semantic gap is at least

$$\Gamma_{\text{new}} \geq \Gamma_{\text{old}} - (2\epsilon + s_h + s_b).$$

Pure coordinate relabellings are aligned out before applying this bound.

Unequal growth introduces an additional response term. At fixed cyclic geometry, add and subtract the new spectrum evaluated at the old center:

$$|\psi_h^t(q) - \psi_h^0(q)| \leq \left| \sum_j (\theta_{hj}^t - \theta_{hj}^0) K_j(q, c_h^0) \right| + \sum_j \theta_{hj}^t |K_j(q, c_h^t) - K_j(q, c_h^0)| =: B_h(q).$$

The first term measures spectral redistribution; the second measures center motion. For a correct parent h and rival b, an initial emission margin m_{hb}^0 therefore remains positive whenever

$$m_{hb}^0 > B_h + B_b.$$

A nonpositive lower bound is inconclusive and not a certificate of reversal.

Full weighted contributions impose one further condition. From $L_h \psi_h > L_b \psi_b$,

$$\frac{L_h}{L_b} > \frac{\psi_b}{\psi_h}$$

is additionally required. Semantic emission order is therefore only one layer of the eventual scorer decision. Calibration, complete scoring, and selection remain separate objects.

A.4 Inductive Correspondence and Next-Arrival Forecasts

The preceding results control one compiled state. The following result extends correspondence to repeated insertion when the reference is rebuilt after every arrival.

Let

$$R = 64, \qquad \varepsilon = 2^{-28}, \qquad x_j = \frac{1}{2} + \frac{j}{R} + \varepsilon \operatorname{sign}(j)$$

for distinct integer indices $|j| \leq 12$. Every revealed prefix consists of siblings with the same ground-truth parent. At each prefix, rebuild the lower-median reference j_*, use the bounded frame $[-64, 64]$, and apply the preceding quantizer and sorted allocation. An arrival is any unrevealed catalogue identity.

Proposition 7 (Prefix closure)**.** *At every nonempty prefix, preferred and final ticks satisfy*

$$p_j = z_j = j - j_*.$$

Consequently,

$$|z_j - R(x_j - x_{j_*})| \leq 2R\varepsilon,$$

and the normalized pair-distance error is at most 4ε. The preceding spectral error bound therefore holds uniformly at each prefix.

Table 4: Forecasts of normalized mean field distance in two known source catalogues. Each starts with four unrevealed arrivals. All 24 orders are checked, yielding 96 transitions per catalogue. Values are descriptive calculations under the declared field laws. The minimum predicted increase is positive before every selected arrival. Repeated states are not independent forecasting trials.

Catalogue	**Initial spread**	**Final spread**	**Minimum predicted increase**	**Largest interval width**
Sparse start	0.267	0.280	2.90×10^{-3}	4.13×10^{-4}
Dense start	0.357	0.385	6.47×10^{-3}	6.48×10^{-4}

Proof. Every nonzero source distance has the form $a/R + b\varepsilon$ with integer $0 < a \leq 24 < R/2$ and $b \in \{0, 1, 2\}$. With $M = 2^{32} - 1$, rounding Ma/R gives $a2^{32}/R$ because $a/R < 1/2$. The positive perturbation cannot reduce this value, while the combined perturbation and rounding error remains below one tick after rescaling. The floor is therefore exactly a.

All preferred positions are distinct, legal, and already ordered, so collision resolution leaves them unchanged. Recomputing the median after each insertion preserves the same argument. The discrepancy from the ideal position follows from the two sign perturbations, and the previously established kernel derivative and coarse-energy bounds transfer this coordinate guarantee to the spectrum. The argument applies at every prefix, proving both the base case and the induction step. □

The proposition gives a useful order-level guarantee: repeated admissible insertion does not accumulate allocation drift in this catalogue. Reference motion is absorbed into the correspondence.

The same algebra supports prediction of the next reference state before the next identity is revealed. For any admissible next identity y, the predicted measure includes recentered persistent residents and the updated population weights. With

$$w_j = A_j \mu, \qquad \Delta w_j = A_j(\mu'_y - \mu),$$

the next energy is exactly

$$e'_{j,y} = e_j + 2\langle w_j, \Delta w_j \rangle_F + \|\Delta w_j\|_F^2.$$

This is the same interaction identity established above, now used prospectively. Linearity makes the update equal to a complete rebuild. Normalizing the predicted energies and applying the declared emission yields a candidate next field for every unrevealed identity. The minimum and maximum of any specified observable over those candidates form a valid next-state envelope. A uniform selection model additionally induces a predictive distribution; the envelope itself requires no independence assumption.

For the single-context construction, equal concrete and background calibration populations give field

$$(\psi + 1/N)/2.$$

The observable used here is the field-weighted mean cyclic distance from the center, normalized by the receiving radius. It is linear in the emitted profile, so the spectral bounds transfer directly to this statistic.

The forecast has a particularly direct interpretation: the identity of the next concept need not be known when every admissible next identity moves the chosen statistic in the same direction. The sparse catalogue starts at indices $\{-6, 0, 6\}$ and withholds $\{-4, -2, 2, 4\}$. The dense catalogue starts with all indices from -12 to 12 except $\{-9, -3, 3, 9\}$ and withholds those four. At every state, the forecast is formed over all unrevealed identities before the next one is selected and the reference is rebuilt.

Across both catalogues, all 192 selected transitions match the predicted spectrum and spread. Every forecast interval has its lower endpoint above the current spread, and the repeated paths traverse 32 distinct states. These results confirm the before-reveal envelope for the known catalogues. They do not estimate population accuracy or predict an unknown semantic label.

A.5 Potential Protection and New Contributors

Correspondence of the reference does not imply that an old decision remains locally protected. A new contributor can change the field around an existing location even when the location itself remains represented correctly.

Consider a positive field on a finite graph with a matched anchor, an initial region, and outside targets fixed before evaluation. With

$$U = -\log_2 p,$$

the barrier from anchor a is nonnegative when the initial point is included and infinite for disconnected targets. Equivalently,

$$B = \log_2 \frac{p(a)}{\max_\pi \min_{z\in\pi} p(z)}.$$

The identity follows by exchanging the monotone logarithm with the finite path extrema. A zero barrier does not imply an everywhere-descending path: potentials $(0, -2, -1)$ on a line have zero barrier to the final point despite requiring a local ascent.

Now compare two matched positive fields and define

$$\eta = \log_2(p'/p), \qquad \omega = \max \eta - \min \eta.$$

Every relative potential changes by

$$\eta(a) - \eta(z),$$

whose magnitude is at most ω. Maxima over a fixed path and minima over fixed paths or neighbor sets preserve this uniform bound. Hence, for connected matched comparisons,

$$|G' - G| \le \omega, \qquad |B' - B| \le \omega, \qquad |\ell' - \ell| \le \omega.$$

Thus an initial local protection margin ℓ survives whenever its available slack exceeds the worst relative field tilt. Common positive rescaling of the field adds only a constant to U and therefore leaves these relative quantities unchanged. New graph edges require separate treatment because they can lower a barrier without changing any old vertex value.

A source-derived construction shows how a new semantic contributor can consume this protection. Three initial roots have normalized angular coordinates $-3/5, 0, 3/5$. A new root is placed at $1/16$ in the near case and at $5/16$ in a matched control. With nine global ticks and 201-tick sphere-zero frames, the ordered allocator places the initial roots at $-2, 0, 2$ and the four roots at $-2, 0, 1, 2$ after either birth. The two births therefore have the same structural novelty at the allocated-tick level. Their semantic calibration differs.

The held query reads zero and never enters calibration. Each root generates a lineage with zero orbit, zero planet, and equator occurrence. Singleton latitude, planet, and orbit compilation yields translated copies of one decreasing profile

$$k_0 > \cdots > k_4 > 0.$$

The common signature magnitude cancels from the normalized spectrum. Seventeen-bin half-count calibration gives old likelihoods

$$(1, 3, 1)/19$$

and new likelihood

$$\lambda = \begin{cases} 3/19, & \text{near birth,} \\ 1/19, & \text{control.} \end{cases}$$

Let

$$D = k_0 - k_1, \qquad c = k_1 + k_3 - 2k_2 > 0.$$

The initial edge contrast is

$$\frac{3D - c}{19},$$

and after the new emitter arrives it becomes

$$\frac{3D - c}{19} - \lambda D.$$

Uniform background changes cancel from this contrast, although they remain present in potential ratios.

For the exact source profile,

$$c = \frac{362535413750}{8043590113923},$$

and the critical likelihood is

$$\frac{323367043683689}{2086197675467097}.$$

The near seed crosses this threshold; the farther control does not. Both occupy the same new structural tick. The difference is therefore caused by semantic calibration and not by structural novelty alone.

The construction begins with positive local protection. The near birth creates a downhill edge toward the new context, while the matched control preserves protection. This is a field-level opportunity result: the new contributor changes an existing geometric graph instead of introducing a previously absent graph edge, and the calculation does not label the event itself as a model mistake or a newly true parent.

A.6 Successive Arrivals and Shared-Query Survival

A single arrival can consume local protection. Repeated arrivals turn this into a pathwise question: does protection remain positive at every prefix of the growth history?

The finite-horizon construction begins with the same roots and one supporting child of B at source angle $1/128$. Two catalogues contain forty distinct future child vectors at angles

$$1/128 + n/32768$$

and

$$1/128 + (80 + n)/32768, \qquad n = 1, \ldots, 40.$$

Every arrival is a fresh identity. A B event adds the next supporting B child. The first C event creates the competing root at $1/16$, and subsequent C events add its children. All global child reads are zero although their ground-truth ownership differs. The fine sphere-one capacity is 402, and each state is rebuilt through the declared ordering and allocation branch. The resulting deeper spectra are not additional inputs to the monitored sphere-zero lineage emissions.

After i B events and $j > 0$ C events, let

$$b = 1 + i, \qquad c' = j - 1.$$

The lineage calibration is

$$L(n) = \frac{2n + 3}{2n + 19},$$

while the background likelihood at $t = i + j$ is

$$\frac{2t + 5}{2t + 25}.$$

The protected edge has contrast

$$M_t = [L(b) - L(c')]D - c/19.$$

Before the competing lineage is born, its likelihood is zero. Matching-record increments satisfy

$$L(n + 1) - L(n) = \frac{32}{(2n + 19)(2n + 21)}.$$

Through the forty-step horizon, their lower bound

$$\frac{32}{99 \cdot 101}$$

exceeds $c/(19D)$ for the declared source profile. Hence $b \geq c' + 1$ implies positive contrast, while $b \leq c'$ implies negative contrast. The opposite edge remains protective in the admitted states.

The complete field calculation therefore collapses to one integer balance:

$$W_t = 2 + i - j.$$

Local protection holds exactly when

$$W_t > 0.$$

Each supporting arrival adds one unit of protection; each competing arrival removes one. This is the pathwise quantity hidden by the final taxonomy alone. Two histories can contain the same final numbers of supporting and competing concepts while differing in whether W_t crossed zero at an intermediate prefix.

The forty-step horizon is part of the statement because the likelihood increments decrease as the calibration population grows.

Let τ denote the first loss of protection and let π_t be the probability of a supporting arrival conditional on the revealed history. On a surviving history, failure at the next step can occur only when the current balance is one and the next arrival is competing. The one-step hazard is therefore

$$h_t = (1 - \pi_t)\mathbf{1}\{W_{t-1} = 1\}.$$

Taking conditional expectation over surviving histories gives

$$S_t = \Pr(\tau > t) = \mathbb{E}\left[\mathbf{1}\{\tau > t - 1\}(1 - h_t)\right].$$

The recursion concerns first-loss survival and not unconditional failure probability.

For a constant supporting-arrival probability p, the problem becomes an exact path-counting problem. Among paths with u upward steps, reflection at the first visit to zero leaves

$$\binom{t}{u} - \binom{t}{u+2}$$

surviving paths. Each such path has probability $p^u(1-p)^{t-u}$, giving

$$S_t(p) = \sum_{u=\lfloor t/2 \rfloor}^{t} \left[\binom{t}{u} - \binom{t}{u+2}\right] p^u(1-p)^{t-u}, \qquad 0 \le t \le 40.$$

Out-of-range binomial coefficients are zero.

History-dependent arrival probabilities admit a sharp envelope. If

$$\pi_- \le \pi_t \le \pi_+,$$

couple all processes using the same fresh uniform draw at each step. Their supporting-event indicators, running balances, and every-prefix survival events are ordered. Consequently,

$$S_t(\pi_-) \le S_t \le S_t(\pi_+),$$

and the constant endpoint processes attain the two bounds. The reflection argument is classical; the contribution here is the reduction of field protection to this exact pathwise event.

The same evolving field may affect many initially correct probes simultaneously, so cohort survival need not factor across queries. For uninterrupted-protection indicators I_q,

$$N_t = \sum_q I_q$$

has

$$\mathbb{E}[N_t] = \sum_q S_q$$

and

$$\mathrm{Var}(N_t) = \sum_q S_q(1 - S_q) + 2\sum_{q<r}(S_{qr} - S_q S_r).$$

The covariance term is required because the probes share the same growth history.

In the constructed cohort, twenty probes have source coordinates $-\ell/65536$, twenty have $-3/5 + \ell/65536$, and twenty have $3/5 + \ell/65536$, for $\ell = 1, \dots, 20$. Their persistent reads are respectively zero, minus two, and plus two, and none enters calibration. The zero-read group shares the first-loss event above. For the other two groups, matching likelihood $3/19$ and rival weights at most $1/19$ yield the positive lower bounds

$$\frac{3D - (k_1 - k_4)}{19} > 0$$

and

$$\frac{2D - (k_1 - k_2) - (k_3 - k_4)}{19} > 0.$$

These coefficients come from the same declared source profile. Hence

$$N_t = 40 + 20J_t, \qquad \Pr(J_t = 1) = S_t,$$

with

$$\mathbb{E}[N_t] = 40 + 20S_t$$

and

$$\mathrm{Var}(N_t) = 400S_t(1 - S_t).$$

The cohort result makes the pathwise distinction explicit at population level: many queries can share one vulnerability event, so an evolving taxonomy can produce strongly correlated preservation failures. The law concerns field protection under the declared construction instead of model accuracy.

B Taxonomy Fields

B.1 Source Construction

The `TaxoAtlas` field separates two roles that a single compatibility score can easily conflate. The revealed taxonomy determines *where* structural evidence can be expressed, while semantic calibration determines *how strongly* each contributor matters for the current read. Their combination produces the field supplied to the declared scorer adapter. Ground-truth parent links remain external to this construction and are used only to evaluate the final decision.

Structural arrangement determines emission shape. Fix one finite local coordinate chart X. For contributor h, let μ_h denote the population measure induced by its revealed taxonomy context. At scale m, the declared structural operator A_{hm} extracts the corresponding component of that population, with energy

$$a_{hm} = \|A_{hm}\mu_h\|_F^2. \tag{10}$$

Positive total energy yields normalized scale weights. The registered scale-alignment rules transport these weights to the emission coefficients β_{hm}. If K_{hm} is the normalized kernel associated with scale m, the complete emission of contributor h is

$$\psi_h(x) = \sum_m \beta_{hm} K_{hm}(x). \tag{11}$$

Equation (11) converts structural organization into a spatial evidence profile. Concentrated population structure places more mass in narrow profiles, whereas evidence expressed at broader scales spreads contribution over a larger part of the chart. Coordinate centers, boundary rules, structural operators, and scale alignment are therefore part of the source construction and not free parameters of a later score comparison.

Semantic evidence determines contribution strength. An emission profile describes where a contributor can place evidence. Calibration measures how well a read agrees with the semantic positions of that contributor's visible population. Fix a receiving chart $X = \{x_-, \dots, x_+\}$ of consecutive integer ticks, its reference tick z_0, and contributor h's allocated center z_h. For a query, candidate, or calibration occurrence r, let $\xi_r \in X$ be the tick supplied by the source construction's semantic read rule. This rule uses semantic distances to the frame reference and directional guides without applying resident collision reassignment. Consequently, ξ_r is distinct from an occurrence's

final allocated position. The following definitions operate on these source-specified read ticks and do not infer positions from held-query parent labels.

Let $\mathcal{P}_h$ be contributor h's calibration population of visible occurrence identities. In the controlled lineage fields, it contains the occurrences belonging to that lineage. Each occurrence i enters once with raw weight $w_i = 1/\nu(c_i)$, where c_i is its concept and $\nu(c_i)$ its current visible multiplicity. For a signed integer residual d, define the weighted histogram, total population mass, and matching mass by

$$H_h(d) = \sum_{i \in \mathcal{P}_h} w_i \mathbf{1}\{\xi_i - z_h = d\}, \qquad n_h = \sum_{i \in \mathcal{P}_h} w_i, \qquad c_{rh} = H_h(\xi_r - z_h).$$

The indicator is one when the occurrence and read have the same signed offset from z_h, and zero otherwise. Residual bins contain every integer from $-(x_+ - x_-)$ to $x_+ - x_-$, giving $B = 2(x_+ - x_-) + 1$ bins. These are ordinary signed offsets, even when emission kernels use cyclic distance. Counts may be fractional because they sum occurrence weights. Half-count smoothing gives

$$L_{rh} = \frac{c_{rh} + 1/2}{n_h + B/2}. \tag{12}$$

The background uses the same rule with center z_0 and a population containing all visible occurrences in the chart, each counted once. Its likelihood is L_{r0}, while its spatial profile is uniform in the controlled constructions. Held queries are evaluated through their read ticks without entering any calibration population.

The acquisition construction in Appendix D.3 illustrates why semantic reads and allocated centers are separate. After C appears, $\xi_q = \xi_B = \xi_C = 0$, although $z_B = 0$ and $z_C = 1$. At this first postbirth state, C's population contains only its own unit-mass occurrence. Its residual is therefore -1, matching the query's residual relative to C. On the nine-tick chart there are 17 residual bins, giving $L_{qC} = (1 + 1/2)/(1 + 17/2) = 3/19$. Distinct allocated positions thus need not imply different semantic compatibility with the read.

For an update that keeps the read ticks, centers, residual support, and existing weights fixed, write $c = c_{rh}$ and $n = n_h$. If added occurrences contribute total mass $v \geq 0$, of which $u \in [0, v]$ matches the read's residual, then $L'_{rh} = (c + u + 1/2)/(n + v + B/2)$. Subtracting the old likelihood gives

$$L'_{rh} - L_{rh} = \frac{u - L_{rh} v}{n + B/2 + v}. \tag{13}$$

For $v > 0$, its sign is determined exactly by

$$L'_{rh} \begin{cases} > L_{rh}, & \text{iff } u/v > L_{rh}, \\ = L_{rh}, & \text{iff } u/v = L_{rh}, \\ < L_{rh}, & \text{iff } u/v < L_{rh}. \end{cases} \tag{14}$$

New evidence strengthens a contributor exactly when its matching fraction exceeds the contributor's current calibrated weight. Changes to read ticks, centers, occurrence multiplicities, or bin support require the histogram to be recomputed before applying Equation (12). The append-only interpretation of Equation (13) does not cover such changes without accounting for their effect on the existing counts.

The field combines shape and strength. For read r, the structural profile ψ_h and semantic weight L_{rh} enter multiplicatively. With background profile b and background weight L_{r0},

$$T_r(x) = \sum_h L_{rh} \psi_h(x) + L_{r0} b(x), \qquad Z_r = \sum_h L_{rh} + L_{r0} > 0, \qquad p_r(x) = \frac{T_r(x)}{Z_r}. \tag{15}$$

This is Equation 3 written at the source level. The emission ψ_h determines *where* contributor h places evidence; L_{rh} determines *how much* that contributor matters for read r. Normalization converts their aggregate into relative evidence over the common chart.

Under growth, a read such as *electric bus* can change through two independent routes when *electric vehicle* is introduced. Its arrival can change structural populations and hence emission profiles, and the new semantic comparisons can change calibrated weights through Equation (13). Existing contributors can change through the same mechanisms even when their identities persist. The field can therefore move without any change to the query description itself.

Growth has four routes into the field. Between two taxonomy states, the source field can change through

1. changes in the calibrated weights of persistent contributors;
2. changes in their emission profiles through population, scale-energy, or coordinate changes;
3. births or removals of contributors and changes in the background term; and
4. the resulting change in the global normalizer Z_r.

The fourth term is essential: an unchanged local contribution can lose relative field mass when evidence is added elsewhere. Appendix B.2 bounds these four effects jointly.

Common coordinates make the comparison meaningful. Field comparisons use a common finite chart or a declared mass-preserving alignment between charts. Such an alignment cannot increase total variation. If the alignment itself changes, its contribution must be included in the field-change allowance. Coordinate labels that are not related by the declared alignment are not compared as though they represented the same position.

The source construction therefore assigns two distinct roles: taxonomy structure determines the multiscale emission shape, and semantic comparisons determine the read-specific contribution weight. Their weighted combination forms the normalized evidence field. Appendix B.2 quantifies how far this completed object can move under taxonomy growth.

B.2 Field Response

The source laws in Appendix B.1 identify the routes by which taxonomy growth changes the evidence field: persistent contributors can change their calibration weights or emission profiles, contributors can be born or removed, the background can change, and the total evidence mass can change. The field-response theorem combines these effects into one quantity that can be transferred to each scoring-operation class.

Unnormalized evidence change. Consider a persistent contributor h. Let $\rho_h = \mathrm{TV}(\psi'_h, \psi_h)$ be the total-variation change in its normalized emission. Adding and subtracting $L'_{rh}\psi_h$ gives

$$\begin{aligned}\|L'_{rh}\psi'_h - L_{rh}\psi_h\|_1 &\le \|L'_{rh}(\psi'_h - \psi_h)\|_1 + \|(L'_{rh} - L_{rh})\psi_h\|_1 \\ &= 2L'_{rh}\rho_h + |L'_{rh} - L_{rh}|. \end{aligned} \tag{16}$$

The two terms have different meanings. The first records a change in *where* contributor h places its evidence; the second records a change in *how strongly* it participates for read r. Born and removed normalized emissions contribute their full respective masses, and the background satisfies the same decomposition. Summing these terms gives exactly the allowance A_r in Equation 5, hence

$$\|T'_r - T_r\|_1 \le A_r. \tag{17}$$

Thus A_r accounts for every source-level change before normalization.

Normalization. The normalized field is $p_r = T_r/Z_r$. Even when a local contribution is unchanged, its relative evidence can change because the total mass Z_r changes. Suppose first that $Z'_r \ge Z_r$. Then

$$p'_r - p_r = \frac{T'_r - T_r}{Z'_r} + T_r\left(\frac{1}{Z'_r} - \frac{1}{Z_r}\right). \tag{18}$$

Taking ℓ_1 norms, using $\|T_r\|_1 = Z_r$, and applying Equation (17) gives

$$\begin{aligned}\|p'_r - p_r\|_1 &\le \frac{A_r}{Z'_r} + Z_r\left|\frac{1}{Z'_r} - \frac{1}{Z_r}\right| \\ &= \frac{A_r + |Z'_r - Z_r|}{Z'_r}.\end{aligned}$$

If $Z_r > Z'_r$, exchanging the two states gives the same expression with denominator Z_r. Since total variation is half the ℓ_1 distance between probability distributions and never exceeds one,

$$\mathrm{TV}(p'_r, p_r) \le \min\left\{1, \frac{A_r + |Z'_r - Z_r|}{2\max\{Z_r, Z'_r\}}\right\} = e_r, \tag{19}$$

which proves Theorem 2.

The theorem separates local and global change. When *electric vehicle* arrives, the emission attached to *bus* can remain numerically unchanged while its normalized share of the field changes because new evidence changes Z_r. An unchanged local score contribution is therefore not an unchanged relative evidence state.

Emission change from scale redistribution and profile motion. The remaining quantity ρ_h can itself be traced to the `TaxoAtlas` construction. Write

$$\psi_h = \sum_m \beta_{hm} K_{hm}, \qquad \psi'_h = \sum_m \beta'_{hm} K'_{hm},$$

where β_h and β'_h are probability vectors over aligned scales and every K_{hm} is a normalized kernel. Adding and subtracting $\sum_m \beta'_{hm} K_{hm}$ gives

$$\mathrm{TV}(\psi'_h, \psi_h) \le \mathrm{TV}(\beta'_h, \beta_h) + \sum_m \beta'_{hm}\, \mathrm{TV}(K'_{hm}, K_{hm}).$$

The first term measures *how much structural evidence moved between scales*; the second measures *how much the spatial profile at each receiving scale moved*. A contributor can therefore change because its multiscale mixture changes, because its kernels move, or through both mechanisms.

Scale-weight change from structural energy. Let the unnormalized scale energies be

$$a_m = \|A_m \mu\|_F^2, \qquad a'_m = \|A'_m \mu'\|_F^2,$$

with positive totals

$$E = \sum_m a_m, \qquad E' = \sum_m a'_m.$$

Applying the same normalization argument as above, together with $|E' - E| \le \|a' - a\|_1$, gives

$$\mathrm{TV}\left(\frac{a'}{E'}, \frac{a}{E}\right) \le \min\left\{1, \frac{\|a' - a\|_1}{\max\{E, E'\}}\right\}. \tag{20}$$

Thus the scale mixture is stable when the energy vector changes little relative to its total energy.

Each individual energy change satisfies

$$|a'_m - a_m| \le \big(\|A'_m \mu'\|_F + \|A_m \mu\|_F\big)\|A'_m \mu' - A_m \mu\|_F, \tag{21}$$

and the inner difference obeys

$$\|A'_m \mu' - A_m \mu\|_F \le \|A'_m\|_{\mathrm{op}} \|\mu' - \mu\|_F + \|A'_m - A_m\|_{\mathrm{op}} \|\mu\|_F. \tag{22}$$

Equation (21) is the difference-of-squared-norms bound. Equation (22) follows by adding and subtracting $A'_m \mu$. Together they separate two structural causes of spectral change: the population represented by the frame can change, and the structural operator acting on that population can change.

The terms remain coupled through the norm of their sum. New concepts can therefore reinforce or cancel existing scale structure; field movement is not proportional to the number of arrivals. This preserves the interaction already exposed by the energy cross term in Appendix A.1.

Complete source-to-field budget. The preceding bounds trace population and operator changes into energy changes, scale redistribution, emission movement, calibration changes, contributor births or removals, and finally normalization. Together they yield the deterministic field bound $\mathrm{TV}(p'_r, p_r) \le e_r$. The calibration update in Appendix B.1 supplies the weight-change terms, while Equations (20)-(22) supply source-derived emission terms. Appendix C.1 transfers this field budget to complete-score budgets for the eight scoring-operation classes.

B.3 CALIBRATED SEMANTIC ORDER

The `TaxoAtlas` field contains more than pairwise semantic similarity. A candidate receives evidence from its own carrier, from other contributors, from its multiscale emission profile, and from query-conditioned calibration. A larger field value at one candidate location therefore need not, by itself, imply that this candidate is semantically closer to the query. The result below identifies exactly when such an implication is justified.

B.3.1 FROM FIELD ADVANTAGE TO SEMANTIC ADVANTAGE

Let a common cyclic coordinate system contain N ticks. Candidate carrier h occupies tick z_h, has nonnegative calibrated weight L_h, and emits a unit-mass profile ψ_h. With uniform background mass $L_0 \geq 0$,

$$T(z) = \sum_h L_h \psi_h(z) + \frac{L_0}{N}, \qquad F(z) = \frac{T(z)}{Z}, \qquad Z = L_0 + \sum_h L_h > 0. \tag{23}$$

The positive common normalizer preserves candidate order, and the uniform background cancels from pairwise field differences. Candidate comparison can therefore be studied directly through T.

Each aligned emission has the common kernel representation

$$\psi_h(z) = \sum_m \theta_{h,m} K_m(z, z_h),$$

where θ_h is a probability vector over scales. Let θ be a reference scale mixture and let k denote the corresponding centered symmetric profile. For two distinct candidate locations z_i, z_j, define the direct carrier contrast

$$D_{ij} = k(0) - k(\mathrm{dist}_C(z_i, z_j)) > 0. \tag{24}$$

Thus, under the common reference profile, a carrier contributes strictly more evidence at its own location than at the rival location.

The field difference can nevertheless be shifted by two contextual effects. Other carriers may favor one location, and the actual scale mixtures may differ from the reference mixture. Define

$$\begin{aligned} I_{ij} &= \sum_{h \notin \{i,j\}} L_h \left[k(\mathrm{dist}_C(z_i, z_h)) - k(\mathrm{dist}_C(z_j, z_h)) \right], \\ Q_{m,h} &= K_m(z_i, z_h) - K_m(z_j, z_h), \\ E_{ij} &= \sum_h L_h \sum_m (\theta_{h,m} - \theta_m) Q_{m,h}, \\ B_{\mathrm{spec},ij} &= \sum_h L_h \, \mathrm{TV}(\theta_h, \theta) \left(\max_m Q_{m,h} - \min_m Q_{m,h} \right). \end{aligned} \tag{25}$$

Here I_{ij} is the signed contribution of all other carriers under the reference profile. The term E_{ij} is the signed correction induced by the actual multiscale profiles, while $B_{\mathrm{spec},ij}$ bounds its magnitude.

The remaining link is semantic calibration. Let reference unit vectors u_q, u_h and learned unit vectors v_q, v_h define normalized angular distances

$$d_h = \frac{\arccos\langle u_q, u_h \rangle}{\pi}, \qquad \widehat{d}_h = \frac{\arccos\langle v_q, v_h \rangle}{\pi}.$$

Assume

$$|\widehat{d}_h - d_h| \leq \delta_h.$$

For example, normalized angular endpoint errors τ_q, τ_h imply

$$\delta_h = \tau_q + \tau_h$$

by the spherical triangle inequality. If Euclidean unit-vector errors are at most $\rho_x \in [0, 2]$, the chord identity permits

$$\tau_x = \frac{2}{\pi} \arcsin(\rho_x / 2).$$

The declared calibration relation is

$$L_h = b - ad_h + e_h, \qquad a > 0, \qquad |e_h| \le \eta_h. \tag{26}$$

The common slope a expresses the intended monotone relation between semantic closeness and calibrated contribution; e_h measures its residual. For cosine ranking, use

$$r_h = -\widehat{d}_h.$$

Cosine and negative normalized angle induce the same ordering on this domain, although their numerical scales are not affinely identical.

B.3.2 A Certified Order Correspondence

The preceding quantities yield an exact decomposition of field preference.

Theorem 8 (Conditional field-to-semantic correspondence)**.** *Under the preceding cyclic-frame, distinct-carrier, spectral, calibration, and angular-error contract,*

$$T(z_i) - T(z_j) = aD_{ij}(d_j - d_i) + D_{ij}(e_i - e_j) + I_{ij} + E_{ij}. \tag{27}$$

Define

$$\begin{aligned} A_{ij} &= \frac{T(z_i) - T(z_j)}{aD_{ij}}, \\ H_{ij} &= \delta_i + \delta_j + \frac{\eta_i + \eta_j}{a} + \frac{|I_{ij}| + B_{\mathrm{spec},ij}}{aD_{ij}}. \end{aligned} \tag{28}$$

Then every encoder satisfying the stated contract obeys

$$|(r_i - r_j) - A_{ij}| \le H_{ij}. \tag{29}$$

Consequently,

$$A_{ij} > H_{ij} \quad \Longrightarrow \quad r_i > r_j,$$

while

$$A_{ij} < -H_{ij} \quad \Longrightarrow \quad r_i < r_j.$$

Values in the intervening closed interval certify neither order.

Proof. First replace every actual profile by the common reference profile. The carrier at z_i contributes $L_i D_{ij}$ to the field difference, while the carrier at z_j contributes $-L_j D_{ij}$ by symmetry. All remaining carriers contribute I_{ij}. Restoring the actual scale mixtures contributes exactly E_{ij}. The uniform background cancels. Hence

$$T(z_i) - T(z_j) = D_{ij}(L_i - L_j) + I_{ij} + E_{ij}.$$

Substituting Equation (26) yields Equation (27).

It remains to control the spectral correction. For one contributor let

$$w_m = \theta_{h,m} - \theta_m.$$

Because both θ_h and θ are probability vectors,

$$\sum_m w_m = 0,$$

and the positive and negative masses of w both equal $\mathrm{TV}(\theta_h, \theta)$. Let

$$q_{\min} = \min_m Q_{m,h}, \qquad q_{\max} = \max_m Q_{m,h}.$$

The difference between the corresponding positive and negative weighted averages is at most $q_{\max} - q_{\min}$, so

$$\left| \sum_m w_m Q_{m,h} \right| \le \mathrm{TV}(\theta_h, \theta)(q_{\max} - q_{\min}).$$

Multiplying by $L_h \geq 0$ and summing over contributors gives

$$|E_{ij}| \leq B_{\mathrm{spec},ij}.$$

Dividing Equation (27) by $aD_{ij} > 0$ gives

$$A_{ij} = (d_j - d_i) + \frac{e_i - e_j}{a} + \frac{I_{ij} + E_{ij}}{aD_{ij}}.$$

Meanwhile

$$r_i - r_j = \widehat{d}_j - \widehat{d}_i,$$

whose difference from $d_j - d_i$ is at most $\delta_i + \delta_j$. Applying the residual and spectral bounds gives Equation (29).

If $A_{ij} > H_{ij}$, then

$$r_i - r_j \geq A_{ij} - H_{ij} > 0.$$

If $A_{ij} < -H_{ij}$, then

$$r_i - r_j \leq A_{ij} + H_{ij} < 0.$$

Monotonicity of cosine over normalized angles transfers the same strict ordering to cosine similarity. □

Equation (27) separates the observed field advantage into semantic distance, calibration residual, other-carrier context, and spectral mismatch. Encoder distortion contributes the remaining uncertainty in Equation (29). The field certifies semantic preference when its signed advantage exceeds the combined admitted disagreement terms.

For *electric bus*, let i denote *bus* and j a rival candidate. A positive field advantage at *bus* is not sufficient by itself, because neighboring contributors or different scale mixtures may have produced part of that advantage. The condition

$$A_{ij} > H_{ij}$$

states precisely when the advantage is too large to be explained by those effects: every admissible encoder must still rank *bus* above the rival.

For an independently specified unique correct parent y, the conditions

$$A_{yj} > H_{yj}$$

against every current rival certify its semantic ranking. A reversed certificate against any rival certifies failure of that ranking. These conclusions concern the declared angular score. A subsequent structural gate or selector requires its own analysis.

The size of the certificate also has a clear interpretation. A small D_{ij} means that the two carrier locations produce little direct field contrast, so contextual and residual effects are magnified after division by aD_{ij}. This weakens the certificate without establishing that the underlying parent-learning problem is intrinsically difficult. Exact encoder agreement removes the δ terms and leaves calibration residuals, contextual support, and spectral differences unchanged.

B.3.3 What an Angular Record Certifies

The preceding theorem links field order to semantic order under a calibrated representation contract. A separate question arises when the adapter retains the underlying angular record itself. Finite quantization then yields a direct score enclosure.

Proposition 9 (Angular-record score enclosure)**.** *Let $d \in [0, 1]$, let $M \geq 1$ be an integer, and let j be the nearest-integer, half-even rounding of Md. Define*

$$I(j) = \left[\max\left(0, \frac{j - 1/2}{M}\right), \min\left(1, \frac{j + 1/2}{M}\right)\right]. \tag{30}$$

Then

$$d \in I(j), \qquad |d - j/M| \leq \frac{1}{2M}.$$

For the declared score

$$\begin{aligned} s(d) &= \alpha\cos(\pi d), \qquad \alpha > 0, \\ s(d) &\in [\alpha\cos(\pi \sup I(j)), \alpha\cos(\pi \inf I(j))], \\ |s(d) - s(j/M)| &\le \frac{\alpha\pi}{2M}. \end{aligned} \tag{31}$$

If a projected angle $\widehat{d} \in [0,1]$ *additionally satisfies*

$$|\widehat{d} - d| \le \eta,$$

widen $I(j)$ *by* η *and intersect with* $[0,1]$*. The resulting interval encloses* $s(\widehat{d})$*, and*

$$|s(\widehat{d}) - s(j/M)| \le \alpha\pi\left(\eta + \frac{1}{2M}\right).$$

Proof. Nearest rounding changes Md by at most $1/2$. Dividing by M and intersecting with the declared domain gives the distance enclosure. Half-even parity may exclude one endpoint from the exact inverse rounding cell. The displayed closed interval remains a valid outer enclosure.

Cosine is decreasing on $[0,\pi]$, which reverses the endpoint order in the score interval. Its derivative with respect to d has magnitude at most $\alpha\pi$, so the mean value theorem gives

$$|s(d) - s(j/M)| \le \frac{\alpha\pi}{2M}.$$

For the projected angle,

$$|\widehat{d} - j/M| \le |\widehat{d} - d| + |d - j/M| \le \eta + \frac{1}{2M},$$

and the same derivative bound gives the final claim. □

The proposition separates two kinds of uncertainty that should not be conflated. Quantization contributes at most

$$\frac{1}{2M}$$

For the fine angular record used in our evaluation, $M = 2^{32}-1$, so the quantization term is extremely small. For unit-norm inputs in the declared cosine expression, $\alpha = (1+10^{-8})^{-1}$. A different norm product requires its own retained norm contract. Thus, a highly precise retained source distance can yield a correspondingly tight score enclosure. It does not imply that a learned projection has $\eta = 0$. Record precision and representation fidelity remain separate guarantees.

B.4 When Does the Reference Contain Enough Information?

The preceding results bound how an observed reference state changes and how declared scoring operations respond. Those guarantees presuppose that the retained record contains the quantities needed by the score. This requirement is itself characterizable exactly.

Let $\mathcal{W}$ denote the admitted collection of complete worlds. A world contains the source state together with every model parameter, auxiliary variable, support constraint, and decision input permitted by the declared contract. Let

$$Z : \mathcal{W} \to \mathcal{Z}$$

be the record retained for the proposed correspondence, and let

$$s : \mathcal{W} \to \mathbb{R}$$

be the scalar score to be recovered from that record.

For any attained observation $z \in Z(\mathcal{W})$, define its *observation fiber*

$$\mathcal{W}_z = \{w \in \mathcal{W} : Z(w) = z\}.$$

All worlds in $\mathcal{W}_z$ look identical to any downstream rule that receives only Z. The central question is therefore whether those indistinguishable worlds also agree on the quantity the rule is expected to recover.

Exact recovery is equivalent to fiber constancy.

Proposition 10 (Observation sufficiency)**.** *There exists a function*

$$g : Z(\mathcal{W}) \to \mathbb{R}$$

such that

$$s(w) = g(Z(w)) \qquad \textit{for every } w \in \mathcal{W}$$

if and only if s is constant on every observation fiber $\mathcal{W}_z$.

Proof. If $s = g \circ Z$, then any $w, w' \in \mathcal{W}_z$ satisfy

$$s(w) = g(z) = s(w').$$

Thus score constancy on each fiber is necessary.

Conversely, suppose s is constant on every $\mathcal{W}_z$. For each attained z, define $g(z)$ to be this common score value. The definition is independent of the chosen world because all worlds in the fiber agree. Hence

$$s(w) = g(Z(w))$$

for every admitted world. □

The proposition gives an exact information criterion. A retained record is sufficient for a score precisely when it never merges two admitted worlds that require different score values. No downstream computation can repair a violation of this condition because the two worlds present identical inputs to that computation.

This result is independent of computational efficiency and statistical estimation. It asks whether the desired score is a function of the retained information at all.

Information loss has an exact minimax price. Exact factorization may fail while approximate recovery remains possible. The observation fiber determines the best error achievable by *any* adapter that sees only Z.

For nonempty $\mathcal{W}_z$, suppose the possible scores have finite extrema

$$m_z = \inf_{w \in \mathcal{W}_z} s(w), \qquad M_z = \sup_{w \in \mathcal{W}_z} s(w).$$

Then

$$\inf_{g} \sup_{w \in \mathcal{W}_z} |s(w) - g(z)| = \frac{M_z - m_z}{2}. \tag{32}$$

Proof. Fix any output $a = g(z)$. For any two worlds $w, w' \in \mathcal{W}_z$,

$$|s(w) - s(w')| \le |s(w) - a| + |s(w') - a|.$$

Therefore

$$\sup_{\widetilde{w} \in \mathcal{W}_z} |s(\widetilde{w}) - a| \ge \frac{|s(w) - s(w')|}{2}.$$

Taking score values approaching M_z and m_z gives the lower bound

$$\sup_{\widetilde{w} \in \mathcal{W}_z} |s(\widetilde{w}) - a| \ge \frac{M_z - m_z}{2}.$$

Choosing

$$a = \frac{M_z + m_z}{2}$$

places every score in the fiber within $(M_z - m_z)/2$, proving equality. If the score range is unbounded above or below, every finite output has unbounded worst-case error. □

Equation (32) makes the cost of missing information explicit: the irreducible worst-case error is one half of the score variation hidden inside one retained record. A small claimed fidelity residual therefore requires a correspondingly small score diameter on every relevant observation fiber. Initial prediction correctness supplies no such control.

A single uniform finite residual across all observations additionally requires

$$\sup_{z} (M_z - m_z) < \infty.$$

Decision recovery can require less information than score recovery. Exact numerical scores may vary inside an observation fiber while the final decision remains unchanged. Let

$$d : \mathcal{W} \to \mathcal{D}$$

denote the actual deterministic decoded output, including the declared tie convention. An exact observation-based decision rule exists if and only if d is constant on every $\mathcal{W}_z$.

The proof is identical to the scalar case: indistinguishable worlds must require the same output, and fiber constancy permits the decision map to assign that common output.

Score sufficiency implies decision sufficiency, while the converse need not hold. Scores may move substantially without changing their ranking or decoded parent set.

The relevant object is the *actual* decoder. Rank constancy does not replace feasibility constraints, threshold values, sampling probabilities, parsing rules, or support restrictions when those quantities affect the returned output.

What information can the normalized field lose? The normalized `TaxoAtlas` field is intentionally a summary and not a lossless encoding of every source quantity. For a nonzero nonnegative raw measure T with total mass Z, any $c > 0$ satisfies

$$\frac{cT}{cZ} = \frac{T}{Z}.$$

Hence, normalization removes the global raw scale. Two worlds can share the same normalized field while differing in quantities that depend on absolute counts or unnormalized mass.

This identity does not assert that arbitrary rescaling is a lawful taxonomy update. It establishes an information fact: the normalized field alone cannot identify a quantity that varies across such indistinguishable admissible worlds. Count-based, calibration-based, or path-based scorers therefore require the corresponding source records unless a separate theorem recovers those quantities from the retained state.

The same issue appears in several scoring-operation classes.

Hidden parameters can reverse an otherwise identical score record. Suppose Z retains the `TaxoAtlas` reference while the scorer contract permits an unrecorded affine-head weight. Two candidate features $+1$ and -1 receive scores

$$(+1, -1)$$

under head weight $+1$, and

$$(-1, +1)$$

under head weight -1. The observed reference is identical in the two worlds while the ranking reverses.

No function of the retained reference alone can reproduce both decisions. Recording the fixed head, recording its admissible update, or proving a response bound removes this particular ambiguity. Initial correctness does not.

Means alone do not determine a distributional score. Recording only Gaussian centers likewise leaves unresolved variation in the covariance or variance terms. Appendix C.4 gives a covariance-only reversal in which the means remain unchanged while the full divergence ranking changes.

The moment-derived variances of Appendix C.1 provide one sufficient augmented record for the declared diagonal-Gaussian scorer. Another model may use different covariance information and requires its own justified record or response bound.

Thus the relevant principle is not that a distributional scorer is inherently unrecoverable. It is that the observation must retain every score-relevant degree of freedom not otherwise controlled.

A score record does not determine what the decoder can express. The same separation appears at the output interface. A smooth reference field and complete numerical score can remain unchanged while a newly required parent is absent from the decoder's admissible grammar or feasible set. The support obstruction in Appendix C.4 then makes acquisition impossible.

Consequently, output support and feasibility are part of the complete world whenever they can affect the returned parent set. They cannot be inferred from a continuous field without an explicit recovery theorem.

The information hierarchy behind F1-F8. The scoring-operation analysis can now be read as an information hierarchy. The normalized field is sufficient for some explicit readouts, while other families require additional records:

F1	field-derived geometry	for the declared metric/angular readout,
F2	field features + weights	when bilinear/fuzzy parameters matter,
F3	means + variances	for the declared Gaussian readout,
F4	region parameters / boundary records	for overlap or containment,
F5	features + neural parameters or response bounds	for learned heads,
F6	logits + output support + decoder state	for generation,
F7	unary terms + interactions + feasible sets	for joint objectives,
F8	field + structural counts/statistics	for gated structural rules.

The explicit constructions in Appendix C.1 retain sufficient records for their declared formulas and derive response allowances from those records. Equation (48) transfers a reference guarantee to another scorer only when the remaining score variation is controlled by justified fidelity residuals.

Information condition. The retained record partitions admitted worlds into observation fibers. Exact score recovery is possible precisely when the score is constant on each fiber, and the optimal worst-case approximation error is one half of the fiber's score diameter. `TaxoAtlas` need not retain every possible model state. It must retain, or accompany with certified bounds for, the information required by the declared scorer and selector. Recording an arbitrary complete model state would make exact factorization tautological and analytically uninformative. The resulting minimal correspondence contract consists of explicit retained quantities, explicit unresolved variation, and a decision rule whose stability can be certified from them.

C Family Responses and Decisions

C.1 From Field Change to Score Change

The preceding analysis bounds how much the `TaxoAtlas` reference can change under an admitted taxonomy update. This section transfers that reference-level change to scorer outputs. Different scoring rules retain different summaries of the same evolving evidence: a geometric rule may read relative position, a Gaussian rule also reads spread, a neural rule reads learned features, and a structural rule may depend on counts that disappear under normalization. The transfer analysis therefore starts from one common field-change budget and derives the corresponding score-change allowance for each declared readout.

Consider two admissible taxonomy states C, C' on a common finite receiving chart $\mathcal{X}$. Each persistent query or candidate r has field $p_r = T_r/Z_r$, with $Z_r > 0$, and a corresponding updated field p'_r, with $Z'_r > 0$. Coordinates, identities, and scale labels are aligned as required in Appendix B.2. For each persistent concept, let

$$e_r \in [0,1], \qquad \mathrm{TV}(p'_r, p_r) = \frac{1}{2}\sum_{x\in\mathcal{X}} |p'_r(x) - p_r(x)| \le e_r. \tag{33}$$

Appendix B.2 derives e_r from the actual sources of change: likelihood updates, emission changes, births, removals, background changes, and normalization. In particular, if $\|T'_r - T_r\|_1 \le A_r$, then

$$e_r = \min\left\{1, \frac{A_r + |Z'_r - Z_r|}{2\max\{Z_r, Z'_r\}}\right\}$$

is valid.

Equation (33) is the common input to the eight transfers below. It bounds movement of the declared field, not every quantity a model might use. Trainable weights, counts, admissible output strings, feasible parent sets, retrieval support, and other scorer-specific records require their own terms whenever they can change.

Bounded observables and moments. Many scorers compress the full field into a smaller statistic. Total variation converts the field budget directly into a budget for any bounded observable. For a real function f on $\mathcal{X}$,

$$|\mathbb{E}_{p'} f - \mathbb{E}_p f| \le (\max f - \min f)\,\mathrm{TV}(p', p). \tag{34}$$

The signed vector $p' - p$ has total mass zero, and its positive and negative parts each have mass $\mathrm{TV}(p', p)$. Subtracting $\min f$ leaves a function bounded between zero and its range, proving the inequality.

This immediately controls location and spread. Fix d chart observables $x_j \in [-R, R]$, with $R \ge 0$, and a variance floor $\lambda > 0$. Define

$$\mu_{rj} = \mathbb{E}_{p_r} x_j, \qquad v_{rj} = \lambda + \mathrm{Var}_{p_r}(x_j). \tag{35}$$

Then

$$|\mu'_{rj} - \mu_{rj}| \le 2R e_r, \qquad |v'_{rj} - v_{rj}| \le 5R^2 e_r, \qquad \lambda \le v_{rj}, v'_{rj} \le \lambda + R^2. \tag{36}$$

The mean bound follows from the range $2R$. Since x_j^2 has range at most R^2,

$$|(\mu'_{rj})^2 - \mu_{rj}^2| \le 2R|\mu'_{rj} - \mu_{rj}| \le 4R^2 e_r.$$

Subtracting the two variance identities gives the second inequality. Nonnegative variance and $\mathrm{Var}(x_j) \le \mathbb{E}x_j^2 \le R^2$ give the endpoint range. No independence between coordinates is required.

The remaining family bounds repeatedly use this same pattern: identify the quantity a scorer actually reads, then propagate the common field budget through that operation.

F1. Metric and angular scores. For a geometric scorer, the relevant object is relative position. If the query and candidate move together, their comparison can remain almost unchanged even when each moves substantially. A separate endpoint bound is conservative because it ignores this cancellation; the joint refinement appears at the end of this section.

Let

$$u_r(x) = \sqrt{p_r(x)}.$$

Each u_r is a unit vector, including when some probabilities vanish. Define

$$a(e) = \sqrt{2e}, \qquad b(e) = 2\arcsin\sqrt{e/2}.$$

The pointwise inequality $(\sqrt{s} - \sqrt{t})^2 \le |s - t|$ gives $\|u'_r - u_r\|_2 \le a(e_r)$. The chord identity for unit vectors gives $\angle(u'_r, u_r) \le b(e_r)$. Therefore

$$\begin{aligned} &\big|\|u'_q - u'_h\|_2 - \|u_q - u_h\|_2\big| \le a(e_q) + a(e_h), \\ &\big|\langle u'_q, u'_h\rangle - \langle u_q, u_h\rangle\big| \le a(e_q) + a(e_h), \\ &\big|\angle(u'_q, u'_h) - \angle(u_q, u_h)\big| \le b(e_q) + b(e_h). \end{aligned} \tag{37}$$

The first line is the reverse triangle inequality. The second follows after adding and subtracting $\langle u_q, u'_h\rangle$ and using unit norms. The last follows from the angular triangle inequality in both directions. Negating a distance or angle preserves the same absolute change bound.

The square-root dependence is important: it remains valid at zero probability and is not replaced by an unjustified global linear modulus.

The same principle holds for any actual metric d. If

$$d(u, \bar{u}) \le \epsilon_q, \qquad d(v_h, \bar{v}_h) \le \epsilon_h,$$

then the two triangle inequalities give

$$|d(u, v_h) - d(\bar{u}, \bar{v}_h)| \le \epsilon_q + \epsilon_h.$$

Any relation transform belongs inside the input whose movement is bounded. If the scorer instead uses moment vectors, Equation (36) gives

$$\|\mu'_r - \mu_r\|_2 \le 2R\sqrt{d}\, e_r.$$

Cosine and negative angle induce the same ranking on unit vectors, although their numerical thresholds live on different scales.

F2. Linear, bilinear and fuzzy scores. A bilinear scorer depends on both the represented concepts and the weights that compare them. Growth can therefore move the inputs, the weights, or both. A normalized fuzzy score introduces a further dependency through its denominator.

For the explicit features $x_r = p_r$ and a fixed matrix M satisfying $\|M\|_{\mathrm{op}} \le M_0$,

$$|(p'_q)^\top M p'_h - p_q^\top M p_h| \le 2M_0(e_q + e_h). \tag{38}$$

Add and subtract $p_q^\top M p'_h$, use $\|p_r\|_2 \le 1$, and

$$\|p'_r - p_r\|_2 \le \|p'_r - p_r\|_1 \le 2e_r.$$

If the matrix also changes, the term $(p'_q)^\top (M' - M)p'_h$ contributes at most $\|M' - M\|_{\mathrm{op}}$. Such a weight update is not implied by the field bound.

For general inputs, define

$$\delta_q = \|x_q - \bar{x}_q\|_2, \qquad \delta_h = \|x_h - \bar{x}_h\|_2, \qquad \delta_M = \|M - \bar{M}\|_{\mathrm{op}}.$$

Then

$$\begin{aligned} &|x_q^\top M x_h - \bar{x}_q^\top \bar{M} \bar{x}_h| \\ &\quad \le \delta_q \|M\|_{\mathrm{op}} \|x_h\|_2 + \|\bar{x}_q\|_2 \delta_M \|x_h\|_2 + \|\bar{x}_q\|_2 \|\bar{M}\|_{\mathrm{op}} \delta_h. \end{aligned} \tag{39}$$

The decomposition is exact:

$$(x_q - \bar{x}_q)^\top M x_h + \bar{x}_q^\top (M - \bar{M}) x_h + \bar{x}_q^\top \bar{M}(x_h - \bar{x}_h).$$

Applying the induced-norm inequality to these three terms proves the bound. A changing bias contributes its own additive term.

For fixed nonnegative fuzzy weights w, define

$$N_h = \sum_i w_i p_{qi} p_{hi}, \qquad Z_q = \sum_i w_i p_{qi}.$$

Coordinatewise multiplication by a probability vector cannot enlarge the relevant ℓ_1 difference, so

$$|N'_h - N_h| \le 2\|w\|_\infty (e_q + e_h), \qquad |Z'_q - Z_q| \le 2\|w\|_\infty e_q.$$

If $Z_q, Z'_q \ge z_{\min} > 0$, then the complete quotient score satisfies

$$\left| \frac{N'_h}{Z'_q} - \frac{N_h}{Z_q} \right| \le \frac{|N'_h - N_h|}{z_{\min}} + \frac{|N_h|\,|Z'_q - Z_q|}{z_{\min}^2}. \tag{40}$$

This follows after adding and subtracting N_h / Z'_q.

For general memberships in $[0, 1]$ and changing weights, numerator drift is at most

$$\|w' - w\|_1 + \|w\|_1 (\delta_q + \delta_h),$$

where the membership errors use the infinity norm. The denominator has the analogous allowance

$$\|w' - w\|_1 + \|w\|_1 \delta_q.$$

For a ranking contrast

$$\Delta = \sum_i w_i q_i (h_i - r_i),$$

the same expansion yields

$$|\Delta' - \Delta| \le \|w' - w\|_1 + \|w\|_1 (\delta_q + \delta_h + \delta_r).$$

Because $|h_i - r_i| \le 1$, this bound controls the weight-change terms as well. A common positive denominator preserves the sign of this ranking contrast. It does not disappear from a fixed-threshold decision: ranking stability and threshold stability remain distinct.

F3. Distribution scores. A distributional scorer reads more than a center. Two candidates can have the same mean and differ substantially because one is much narrower. The variance floor below is therefore not cosmetic: it prevents small absolute variance changes from producing unbounded divergence changes.

Construct

$$G_r = \mathcal{N}(\mu_r, \mathrm{diag}(v_r))$$

from Equation (35). Both the means and diagonal variances therefore have explicit field-derived meanings. With

$$d_j = \mu_{qj} - \mu_{hj},$$

the query-to-parent KL divergence and Bhattacharyya distance are

$$\begin{aligned} K(q,h) &= \frac{1}{2}\sum_j \left[\log\frac{v_{hj}}{v_{qj}} + \frac{v_{qj}+d_j^2}{v_{hj}} - 1\right], \\ B(q,h) &= \sum_j \left[\frac{d_j^2}{4(v_{qj}+v_{hj})} + \frac{1}{2}\log\frac{v_{qj}+v_{hj}}{2\sqrt{v_{qj}v_{hj}}}\right]. \end{aligned} \tag{41}$$

Assume every endpoint variance lies in $[v_{\min}, v_{\max}]$, with $v_{\min} > 0$, and $|d_j| \le D$. Write δ for absolute coordinate changes. Then

$$\begin{aligned} E_K = \sum_j \Bigg[&\frac{D}{v_{\min}}(\delta\mu_{qj} + \delta\mu_{hj}) + \frac{\delta v_{qj}}{v_{\min}} \\ &+ \left(\frac{1}{2v_{\min}} + \frac{v_{\max}+D^2}{2v_{\min}^2}\right)\delta v_{hj}\Bigg], \\ E_B = \sum_j \left[\frac{D}{4v_{\min}}(\delta\mu_{qj} + \delta\mu_{hj}) + \left(\frac{D^2}{16v_{\min}^2} + \frac{1}{2v_{\min}}\right)(\delta v_{qj} + \delta v_{hj})\right]. \end{aligned} \tag{42}$$

For KL, the mean derivatives have magnitude $|d_j|/v_{hj}$. The variance derivatives are

$$\frac{1}{2}(v_{hj}^{-1} - v_{qj}^{-1})$$

and

$$\frac{1}{2}\left[v_{hj}^{-1} - \frac{v_{qj}+d_j^2}{v_{hj}^2}\right].$$

Their absolute values are bounded by the coefficients in Equation (42). For Bhattacharyya distance, the mean derivative has magnitude

$$\frac{|d_j|}{2(v_{qj}+v_{hj})},$$

and either variance derivative is

$$-\frac{d_j^2}{4(v_{qj}+v_{hj})^2} + \frac{1}{2(v_{qj}+v_{hj})} - \frac{1}{4v_{\cdot j}}.$$

Bounding these terms separately gives the displayed coefficients. The line segment between the two endpoint parameter vectors remains inside the declared mean and variance ranges. Integrating the directional derivative along that segment proves

$$|K' - K| \le E_K, \qquad |B' - B| \le E_B.$$

For the explicit `TaxoAtlas` field readout, Equation (36) permits the substitutions

$$v_{\min} = \lambda, \qquad v_{\max} = \lambda + R^2, \qquad D = 2R,$$
$$\delta\mu_{rj} \le 2Re_r, \qquad \delta v_{rj} \le 5R^2 e_r.$$

The resulting per-coordinate allowances are

$$\begin{aligned} E_{Kj} &= \frac{2R}{\lambda}\, 2R(e_q + e_h) + \frac{5R^2}{\lambda} e_q \\ &\quad + \left(\frac{1}{2\lambda} + \frac{\lambda + 5R^2}{2\lambda^2}\right) 5R^2 e_h, \\ E_{Bj} &= \frac{2R}{4\lambda}\, 2R(e_q + e_h) + \left(\frac{4R^2}{16\lambda^2} + \frac{1}{2\lambda}\right) 5R^2(e_q + e_h). \end{aligned} \tag{43}$$

Summing over coordinates gives the full-score bounds.

Each Bhattacharyya term is nonnegative by the arithmetic-geometric mean inequality. Hence $x \mapsto e^{-x}$ has derivative magnitude at most one on the relevant domain, so the coefficient score e^{-B} changes by at most E_B.

These guarantees concern the declared diagonal-Gaussian readout. They do not identify an arbitrary learned covariance matrix with field spread.

F4. Regions and declared cones. A region scorer asks whether the query lies inside, overlaps with, or remains safely within a candidate region. Here the relevant stability quantity is boundary motion. The minimum permitted query volume matters because the same boundary displacement can have a much larger fractional effect on a very small region.

Associate concept r with the axis-aligned box whose faces are

$$\mu_{rj} \pm a_0 \sqrt{v_{rj}}, \qquad a_0 > 0.$$

Since the derivative of the square root on $[\lambda, \infty)$ is at most $1/(2\sqrt{\lambda})$, every face moves by at most

$$c_r = \left(2R + \frac{5a_0 R^2}{2\sqrt{\lambda}}\right) e_r.$$

Every side length is at most

$$\mathcal{L} = 2a_0 \sqrt{\lambda + R^2},$$

while the query volume is bounded below by

$$V_{\min} = (2a_0\sqrt{\lambda})^d.$$

Suppose corresponding query and candidate faces move by at most ϵ, endpoint side lengths are at most $\mathcal{L}$, and both query volumes are at least $V_{\min} > 0$. An intersection side is the positive part of a minimum upper face minus a maximum lower face. Minimum, maximum, and positive part imply that each such side changes by at most 2ϵ.

Telescoping the product of d nonnegative side lengths, each bounded by $\mathcal{L}$, gives an intersection-volume change at most

$$2d\epsilon\mathcal{L}^{d-1}.$$

The same bound holds for the query-volume change. If I, V denote intersection and query volumes,

$$\frac{I'}{V'} - \frac{I}{V} = \frac{I' - I}{V'} + \frac{I}{V}\,\frac{V - V'}{V'}.$$

Since $0 \le I/V \le 1$,

$$\left|\frac{I'}{V'} - \frac{I}{V}\right| \le \min\left\{1, \frac{4d\epsilon\mathcal{L}^{d-1}}{V_{\min}}\right\}. \tag{44}$$

For the explicit field boxes,

$$\epsilon = \max\{c_q, c_h\}.$$

Directed containment has an even simpler stability quantity. Its slack is the minimum over all lower-face differences $\ell_{qj} - \ell_{hj}$ and upper-face differences $u_{hj} - u_{qj}$. Each difference moves by at most $c_q + c_h$, and the minimum is nonexpansive in the infinity norm. The containment slack therefore moves by at most $c_q + c_h$.

For the declared circular-cone predicate, let the axis be $\sqrt{p_h}$, the query direction be $\sqrt{p_q}$, and the slack be

$$\psi_h - \angle(u_h, u_q).$$

A fixed aperture contributes no drift. If instead

$$\psi_h = \mathbb{E}_{p_h} g$$

for a fixed observable g with a permitted aperture range, its drift is at most

$$\mathrm{range}(g) e_h.$$

Together with the angular bound, the cone slack changes by at most

$$\mathrm{range}(g) e_h + b(e_h) + b(e_q),$$

with the first term omitted for a fixed aperture.

This certificate is specific to the declared circular-cone predicate. Other radial or aperture laws require their own records. Stable containment alone also does not distinguish a direct parent from a more general ancestor.

F5. Neural readouts of field features. A neural head does not remove the reference perturbation; it can amplify it. Its Lipschitz constant measures the largest amplification permitted by the declared fixed network. Parameter updates are a second source of movement and require a separate term.

Let

$$\phi_{rj} = \mathbb{E}_{p_r} f_j, \qquad |f_j| \le B_j,$$

and concatenate the query and candidate feature vectors. Equation (34) gives

$$\|\phi'_{qh} - \phi_{qh}\|_2 \le 2 \sqrt{\sum_j B_j^2} \sqrt{e_q^2 + e_h^2} = \rho_{qh}.$$

For a fixed K-Lipschitz head g,

$$|g(\phi'_{qh}) - g(\phi_{qh})| \le K \rho_{qh}.$$

If layer ℓ has linear operator norm at most B_ℓ and activation Lipschitz constant L_ℓ, repeated application of the layer inequalities gives

$$K \le \prod_\ell B_\ell L_\ell,$$

including the output layer. Fixed biases cancel in differences.

If the head parameters also change and a separate contract gives

$$|g_{\theta'}(\phi') - g_\theta(\phi')| \le \eta_{qh},$$

then adding and subtracting $g_\theta(\phi')$ yields the complete allowance

$$K \rho_{qh} + \eta_{qh}.$$

Retrieval, token truncation, sampled neighbors, and context formatting are covered only when their induced feature changes satisfy the same declared bound.

F6. Token scores and generation. Generative scores expose a distinction that pairwise ranking can hide. A parent can retain the same relative logit preference while its probability decreases because newly admitted outputs consume normalization mass. The transfer therefore separates logit movement from support growth.

For admitted prefix v and token a, let

$$\ell_a(v) = \mathbb{E}_{p_q} f_{v,a} + b_{v,a}, \qquad |f_{v,a}| \le B,$$

with fixed biases. On unchanged support, every logit changes by at most

$$\delta = 2 B e_q.$$

If two logit vectors differ coordinatewise by at most δ, their exponential sums differ by a multiplicative factor in $[e^{-\delta}, e^{\delta}]$. Hence their log-sum-exp values differ by at most δ, and every conditional log probability changes by at most

$$2\delta = 4Be_q.$$

A supported complete sequence of length at most L, including termination, therefore has total log-score drift at most

$$4LBe_q.$$

A candidate-conditioned prompt may use a declared query-and-candidate feature map instead, with the corresponding feature allowance. An average over a fixed set of masked-token logits inherits the largest constituent allowance; a sum inherits the sum. A masked-token logit is not an autoregressive probability unless the rule explicitly defines it as one.

Support growth adds a separate normalization term. At an old prefix, let Z'_{old} be the updated exponential-logit mass of retained old tokens and A'_{new} the mass of newly admitted tokens. For an old token, the new log probability equals its log probability under the updated old support minus

$$\log\left(1 + \frac{A'_{\mathrm{new}}}{Z'_{\mathrm{old}}}\right).$$

This follows by factoring the full normalizer as

$$Z'_{\mathrm{old}}\left(1 + \frac{A'_{\mathrm{new}}}{Z'_{\mathrm{old}}}\right).$$

Thus unchanged old-token ordering does not imply unchanged old-token probability.

A newly required parent string additionally needs a current score bound or a previously declared master support. A fixed vocabulary alone does not guarantee that the permitted identifier grammar can express the new parent. Sequence-level, greedy, and sampling guarantees are given in Appendix C.5.

F7. Joint parent-set scores. A joint scorer can prefer each true parent individually and still reject the correct set. The relevant object is therefore the score of the complete parent configuration, including interactions and set-level penalties.

Let the unary terms s_{qh} use explicit readouts already covered above. For fixed interaction kernels satisfying

$$|K_{hj}(x, y)| \le B_{hj},$$

define

$$I_{hj} = \sum_{x,y} K_{hj}(x, y)p_h(x)p_j(y).$$

Adding and subtracting $p_h \otimes p'_j$ gives

$$\mathrm{TV}(p'_h \otimes p'_j, p_h \otimes p_j) \le \min\{1, e_h + e_j\}.$$

Taking a product with a fixed probability vector preserves the relevant ℓ_1 difference. Equation (34) therefore yields

$$|I'_{hj} - I_{hj}| \le 2B_{hj} \min\{1, e_h + e_j\}.$$

For

$$S(P) = \sum_{h \in P} s_{qh} + \sum_{h<j \in P} I_{hj} - c(P),$$

a fixed penalty $c(P)$ contributes no numerical drift on a persistent set. Summing the unary and pairwise allowances bounds

$$|S'(P) - S(P)|.$$

A changing kernel or penalty needs its own term.

The comparison must range over the feasible parent sets at the current taxonomy state, including newly admitted candidates and revised constraints. Higher-order interactions require corresponding bounded joint kernels or additional records. Stable individual edges do not by themselves certify that the complete parent set remains optimal.

F8. Structural and statistical scores. A structural scorer can react to growth that occurs outside the query's immediate semantic neighborhood. Counts and structural statistics therefore belong to the source record explicitly; they cannot in general be reconstructed from a normalized field, whose normalization may erase absolute scale.

Let n_h be the retained count for category h, N the declared total, M the number of smoothed categories, and $\alpha > 0$ a fixed smoothing constant. For

$$\mathrm{IC}_h = \log D - \log(n_h + \alpha), \qquad D = N + \alpha M > 0,$$

the derivative $1/x$ on a positive interval gives

$$|\,\mathrm{IC}'_h - \mathrm{IC}_h\,| \le \frac{|D' - D|}{\min\{D, D'\}} + \frac{|n'_h - n_h|}{\min\{n_h, n'_h\} + \alpha}. \tag{45}$$

The same integration gives

$$|\log L' - \log L| \le \frac{|L' - L|}{\min\{L, L'\}}$$

for positive calibration likelihoods. Differences of information-content scores inherit the sum of their respective allowances.

For the declared angular/count rule, the augmented source record supplies n_h, $W > 0$, finite $\gamma \ge 0$, and the actual base score s_h. Its gate is

$$g_h = s_h - 1 + \frac{\gamma}{W^2}\left[\log_2(1 + n_h/2)\right]^2, \qquad S_h = s_h \mathbf{1}\{g_h > 0\}.$$

These quantities permit direct evaluation or interval propagation. The gate-branch proof and zero-count boundary are given in Appendices C.5 and C.4. A different count predicate or a discontinuous path rule requires its own complete structural statistic.

The eighth transfer therefore depends on the augmented structural record as well as the normalized field.

Scope of the transfer bounds. The eight derivations share one structure. Taxonomy growth first produces a bounded change in the declared reference record. Each scoring operation transforms that change into a family-specific complete-score allowance. For persistent outputs, these allowances are deterministic and therefore hold pathwise under stochastic or adversarial growth; no independence assumption is needed.

The guarantee is conditional on the complete declared rule. A changing weight, support, interaction kernel, penalty, feasible set, or gate statistic contributes only when its corresponding change term is available. Membership in the same broad scoring-operation class is insufficient. A different learned implementation additionally requires the fidelity condition of Appendix C.2.

For sets and sequences, component allowances compose over the actual terms used by the declared decoder. Gate signs must be handled through their realized branches. These score-change bounds become decision guarantees only after they are compared with the pre-growth margin, feasibility slack, or selector condition developed in Appendix C.2-C.5.

Joint responses can be substantially sharper. The preceding family bounds are deliberately modular: they remain valid when only separate query and candidate budgets are known. They can be loose when both endpoints respond coherently to the same taxonomy update.

For F1, let

$$d_h = u_q - u_h.$$

The score depends on this relative quantity, so the reverse triangle inequality gives the sharper bound

$$|S'_h - S_h| \le \|(u'_q - u'_h) - (u_q - u_h)\|_2. \tag{46}$$

The right-hand side is itself bounded by the sum of the two endpoint movements and retains their signed cancellation before taking a norm. It is exactly zero whenever the query and candidate have the same relative representation at both endpoints, even if both moved substantially in absolute terms.

The same principle sharpens the F3 Gaussian response. In Equation (42), choose D as the larger endpoint absolute mean difference and choose $v_{\min}, v_{\max}$ as the smallest and largest of the four endpoint variances. The interpolating segment remains within these endpoint ranges. Per coordinate, the resulting KL allowance is

$$\frac{D}{v_{\min}}(\delta\mu_q + \delta\mu_h) + \frac{\delta v_q}{v_{\min}} + \left(\frac{1}{2v_{\min}} + \frac{v_{\max} + D^2}{2v_{\min}^2}\right)\delta v_h,$$

with changes computed from the actual endpoint field moments.

If the query and candidate Gaussians coincide at both endpoints, both KL divergences are exactly zero. The exact zero score-change certificate then supersedes the derivative allowance. This conclusion follows from the readout invariant and does not use the gold parent label.

The distinction is consequential: separate worst-case movement bounds can discard common-mode cancellation, while a joint-response certificate retains it. A retrospective tight bound does not automatically become a prospective guarantee; prospective use requires a valid pre-update bound on the joint response.

C.2 Preservation from Complete Scores

The transfer results in Appendix C.1 bound how far a declared complete score can move after taxonomy growth. Preservation holds when the old decision lead is large enough to absorb that movement.

Let $S_0(a)$ and $S_t(a)$ be complete scores for a persistent output a. The output may be one candidate, an entire parent set, or a complete graph edit. Suppose

$$|S_t(a) - S_0(a)| \le E_a$$

for the realized update. Let y be a still-correct permitted output and r a persistent rival. If

$$m_r = S_0(y) - S_0(r)$$

is the initial decision margin, then

$$S_t(y) - S_t(r) \ge m_r - E_y - E_r. \tag{47}$$

Proof. Write the new contrast as

$$S_t(y) - S_t(r) = m_r + [S_t(y) - S_0(y)] - [S_t(r) - S_0(r)].$$

The two score-change allowances can reduce the original contrast by at most $E_y + E_r$, which gives Equation (47). □

Equation (47) is the generic surviving-margin result. Taxonomy growth can spend at most $E_y + E_r$ of the old lead against rival r. A strictly positive remainder certifies that the old ordering survives. If this holds against every persistent incorrect rival and the current score of y also exceeds every newly admitted rival, y remains the unique maximizer. When comparison is required only against incorrect outputs, every maximizer is correct, allowing another correct output to overtake y. Equality requires the actual tie rule.

A new candidate has no inherited margin m_r. Its comparison must therefore be certified from the current state. This is the first formal separation between *retention* and *acquisition*.

The result applies to the declared decoder when that decoder maximizes these complete scores. Greedy decoding, sampling, thresholding, and constrained set selection require their own decision conditions below.

Uniform conclusions over a scorer class. Suppose a nonempty jointly realizable class of scorers satisfies simultaneous bounds

$$L_a \le S_t(a) \le U_a$$

for every class member. If some permitted correct output y satisfies

$$L_y > \max_{r \text{ incorrect}} U_r,$$

then every maximizer under every admitted scorer is correct. Conversely, if some permitted incorrect output r satisfies

$$L_r > \max_{a \text{ correct}} U_a,$$

then every maximizer is incorrect. Both statements follow by placing the actual scores between their certified endpoints. They require neither independence nor independent attainability of the intervals.

Failure of a sufficient preservation inequality is therefore inconclusive. A uniform failure claim requires a certified reversed comparison over a nonempty jointly admitted class.

Fidelity to an explicit reference readout. The preceding bounds may be derived for an explicit reference score $s(C, a)$ while the deployed scorer uses a complete score $S(C, a)$. Suppose

$$|S(C, a) - s(C, a)| \le \epsilon(C, a), \qquad |S(C', a) - s(C', a)| \le \epsilon(C', a),$$

and the reference analysis gives

$$|s(C', a) - s(C, a)| \le E_a.$$

Adding and subtracting the two reference endpoint scores yields

$$|S(C', a) - S(C, a)| \le E_a + \epsilon(C', a) + \epsilon(C, a). \tag{48}$$

Thus a theoretical guarantee for an explicit readout transfers to another scorer only through a justified fidelity term. The residuals must follow from the actual features, variances, parameters, context, or a valid measurement procedure. Initial correctness alone supplies no such bound. A shared positive affine transformation may be removed from a ranking comparison only when the actual decoder is invariant to it. Thresholds and sampling probabilities remain in their own numerical units.

Along an aligned sequence $C_0, \dots, C_t$, a persistent reference output satisfies

$$|s(C_t, a) - s(C_0, a)| \le \sum_{j=1}^{t} E_{j,a}$$

by telescoping. The corresponding actual-score bound adds only the endpoint fidelity residuals $\epsilon(C_0, a) + \epsilon(C_t, a)$ when these reference increments are used. A new output begins its comparison record at birth.

This endpoint inequality is weaker than uninterrupted trajectory stability. Preservation throughout a growth history requires the relevant margin or feasibility condition to remain valid at every prefix, as the pathwise result in Appendix A.6 makes explicit.

Common movement need not spend ranking margin. Let

$$d_a = s_t(a) - s_0(a)$$

be the reference increment of an old candidate and define its oscillation over old candidates by

$$\omega = \max_a d_a - \min_a d_a.$$

For any old target-rival pair,

$$d_y - d_r \ge -\omega,$$

so an initial reference margin Γ_0 remains at least

$$\Gamma_0 - \omega.$$

If actual current scores lie within ϵ_t of their references, the old-candidate margin remains at least

$$\Gamma_0 - \omega - 2\epsilon_t.$$

Only *relative* movement consumes the ranking margin. A common additive shift has $\omega = 0$ and leaves every old ordering unchanged. Likewise, if canonical pair inputs and parameters remain fixed, every old canonical score remains fixed. Any strictly increasing transformation common to all candidates preserves their order and ties. This includes a positive common denominator and candidate softmax when the decision depends only on ranking.

The conclusion does not extend automatically to numerical thresholds, probabilities, or newly admitted competitors. A fixed network may also change its output when its recomputed context changes.

C.3 Retention, Acquisition and Exact Parent Sets

Preserving an old winner is only one part of taxonomy growth. For *electric bus*, *bus* may remain valid while *electric vehicle* becomes an additional required parent. A different insertion may retire an old direct link while preserving the corresponding ancestor relation.

Let V_t be the finite candidate set at state t, and let the independently defined nonempty direct-parent set be

$$P_t \subseteq V_t.$$

For one transition define

$$R_t = P_{t-1} \cap P_t, \qquad N_t = P_t \setminus P_{t-1}, \qquad D_t = P_{t-1} \setminus P_t.$$

Here R_t contains parents that remain valid, N_t contains newly required parents, and D_t contains formerly direct links that must disappear. A newly required relation may involve a node that already existed before the transition.

The three central requirements are

$$R_t \subseteq \widehat{P}_t \qquad \text{(retention)},$$
$$N_t \subseteq \widehat{P}_t \qquad \text{(acquisition)},$$

and

$$\widehat{P}_t = P_t \qquad \text{(exact recovery)}.$$

Retention and acquisition together provide full recall. Exact recovery additionally removes every retired and otherwise false link. Ratios with empty denominators are left undefined rather than assigned an artificial value.

Access and capacity come before scoring. Let every returned set lie in an accessible candidate set $A_t \subseteq V_t$ and contain at most k_t members. Then every scorer and decoder satisfying these constraints obeys

$$|\widehat{P}_t \cap P_t| \le \min\{k_t, |P_t \cap A_t|\}, \qquad \text{recall}_t \le \frac{\min\{k_t, |P_t \cap A_t|\}}{|P_t|}. \tag{49}$$

Proof. The recovered true-parent set is simultaneously a subset of the output and of $P_t \cap A_t$. Its size is therefore bounded by both k_t and $|P_t \cap A_t|$. Dividing by $|P_t|$ gives the recall ceiling. □

Thus a scoring function cannot recover what the decoder cannot access or express. These ceilings need not be attainable by a restricted scoring class.

If every retained parent is recovered, then necessarily

$$R_t \subseteq A_t, \qquad |R_t| \le k_t,$$

and the remaining acquisition capacity satisfies

$$|\widehat{P}_t \cap N_t| \le \min\{|N_t \cap A_t|,\, k_t - |R_t|\}.$$

Retention already consumes $|R_t|$ output positions. False positives can only reduce the remaining capacity further. If $|R_t| > k_t$, retention itself is impossible.

With one retained parent and one newly required parent, a top-one selector has recall at most $1/2$, independent of score quality. A longer ranked list that is not returned as the parent set is a different output object.

Threshold selection requires two separations. Consider the actual decoder

$$\widehat{P}_t = \{h \in V_t : S_t(h) > \tau_t\},$$

with jointly valid score intervals

$$L_h \le S_t(h) \le U_h.$$

Exact recovery is certified when

$$\min_{h \in P_t} L_h > \tau_t \quad \text{and} \quad \tau_t \ge \max_{r \in V_t \setminus P_t} U_r. \tag{50}$$

The second condition is omitted when no wrong candidate exists.

Proof. Every required parent has actual score at least its lower bound and therefore lies strictly above the threshold. Every other candidate has actual score at most its upper bound and therefore lies at or below the threshold. The strict threshold rule consequently returns exactly P_t. □

The two inequalities play different roles. The first certifies inclusion of every required parent, including R_t and N_t. The second certifies exclusion of retired and false links. Satisfying only the first gives full recall with no false-positive guarantee.

A newly required parent needs a current score bound because it has no inherited target margin. Loose intervals may fail the certificate even when the realized decision is exact. A threshold chosen using the gold parent set is an analytical certificate and not a deployable calibration procedure.

For score top-k, exact recovery follows when

$$k = |P_t|,$$

all required parents are accessible, and the lowest true score is strictly above the highest wrong score. If $k < |P_t|$, Equation (49) forbids exact recovery. If $k > |P_t|$ and the decoder must return exactly k candidates from a sufficiently large candidate set, at least one wrong parent is included. A particular new parent is excluded whenever at least k candidates have strictly larger scores; those competitors need not all be incorrect.

Joint-set recovery is a complete-configuration problem. Let $\mathcal{F}_t$ be the actual finite feasible family and let

$$S_t(P) = \sum_{h \in P} s_h + I(P)$$

be the complete score of configuration P. Suppose each unary term has error at most ϵ_h and the interaction error for P is at most η_P. Then

$$|S_t(P) - \bar{S}_t(P)| \le E_P, \qquad E_P = \sum_{h \in P} \epsilon_h + \eta_P.$$

If $P_t \in \mathcal{F}_t$ and

$$\bar{S}_t(P_t) - E_{P_t} > \max_{P \in \mathcal{F}_t \setminus \{P_t\}} \left[\bar{S}_t(P) + E_P\right],$$

then exact joint maximization returns P_t.

Proof. The lower bound on the true configuration exceeds the upper bound of every feasible alternative. Its realized complete score is therefore strictly largest. □

Feasibility is logically prior to scoring. If $P_t \notin \mathcal{F}_t$, no constrained decoder can return the correct set regardless of the quality of its unary scores. Pairwise guarantees transfer to parent-set recovery only through the complete interaction, feasibility, and selection calculation.

C.4 Family-Specific Obstructions

The preceding results state conditions sufficient for preservation and recovery. The constructions below show why those conditions cannot simply be omitted. Each isolates a different obstruction at the scorer or decoder interface.

Except where stated otherwise, the independently defined target is initially $\{y\}$ and later becomes $\{y, z\}$, where y and z are incomparable true parents and r is incorrect. These are exact score-interface constructions. Their scope is the stated scorer and decoder class.

F1. Geometry can place a wrong candidate between the query and a new parent. On the real line take

$$q = 0, \qquad y = \frac{1}{2}, \qquad r = 1, \qquad z = 2.$$

Initially only y, r are available, and the rule

$$d(q, h) < \frac{3}{4}$$

returns exactly y. After z arrives, the old parent is retained and the new parent is missed. More strongly,

$$d(q, r) < d(q, z),$$

so no distance threshold can include z while excluding r. Every strictly decreasing transformation of distance preserves the same obstruction.

The positive condition is the complementary separation. With current reference distances $\bar{d}_h$ and metric-input allowances E_h,

$$\max_{h \in P_t}(\bar{d}_h + E_h) < R \leq \min_{r \notin P_t}(\bar{d}_r - E_r)$$

certifies exact threshold separation. Conversely,

$$\bar{d}_z - E_z > \bar{d}_r + E_r$$

certifies the incompatible ordering throughout the admitted uncertainty set. A transformed relation vector and any numerical conversion between cosine and angle belong in these premises.

F2. A shared affine head cannot separate a convex collision. Suppose

$$\phi_r = \xi\phi_y + (1 - \xi)\phi_z, \qquad 0 < \xi < 1.$$

For every shared affine score

$$S_h = w^\top \phi_h + b,$$

linearity gives

$$S_r = \xi S_y + (1 - \xi)S_z.$$

Hence

$$S_y > \tau \quad \text{and} \quad S_z > \tau$$

imply

$$S_r > \tau.$$

No affine head with a common threshold can retain both true parents while excluding r on these fixed features. A common positive denominator preserves the same affine relation.

A concrete fuzzy instance is

$$q = (1, 1), \qquad y = (1, 0), \qquad z = (0, 1), \qquad r = (1/2, 1/2).$$

Initially $w = (1, 0)$, the query denominator is one, and $\tau = 3/4$ selects only y. The new z scores zero. Under any reweighting with a positive common denominator, the score of r remains the mean of the scores of y and z, so exact threshold separation remains impossible within this class. Candidate-specific biases, nonlinear features, or a joint term change the analyzed scorer class.

Preservation additionally requires weight and norm records. With fixed unit memberships

$$q = (3/5, 4/5), \qquad y = (4/5, 3/5), \qquad r = (3/5, 4/5),$$

weights $(3, 1)$ give normalized fuzzy scores

$$48/65, \quad 43/65,$$

whereas weights $(1, 3)$ give

$$16/25, \quad 19/25,$$

reversing the order while every membership stays fixed.

Likewise,

$$\frac{\langle x_q, x_h \rangle}{\|x_q\|} = \|x_h\| \cos(q, h)$$

for nonzero vectors. With

$$q = (1, 0), \qquad y = (4/5, 3/5), \qquad r = (6/5, 8/5),$$

the cosine scores are $4/5, 3/5$, while the dot scores are $4/5, 6/5$. Candidate norms therefore matter unless they are equal or explicitly bounded.

F3. A correct center need not imply compatible distributions. Take one-dimensional query and old-parent Gaussians with mean zero and variance one. Let the wrong candidate have mean $1/2$, variance one, and the new true parent have mean zero, variance four. Substitution into Equation (41) gives

$$\begin{aligned} K_y &= 0, \qquad K_r = \frac{1}{8}, \qquad K_z = \frac{\log 4 - 3/4}{2} > \frac{1}{8}, \\ B_y &= 0, \qquad B_r = \frac{1}{32}, \qquad B_z = \frac{1}{2}\log(5/4) > \frac{1}{32}. \end{aligned} \tag{51}$$

The first strict inequality follows from $\log 2 > 1/2$, obtained by integrating $1/x > 1/2$ on $1 \le x < 2$. The second follows from $\log(5/4) \ge 1/5$, using $1/x \ge 4/5$ on $[1, 5/4]$.

Initially,

$$K < 1/16 \qquad \text{or} \qquad B < 1/64$$

selects only y. After z arrives, either rule misses it, and no threshold on the chosen divergence can include z while rejecting r. A common monotone transformation cannot reverse this order.

The positive condition again has the opposite form. Under the diagonal-Gaussian domains of Appendix C.1, exact acquisition is certified whenever

$$\max_{\text{true}}(\bar{d} + E) < T \le \min_{\text{wrong}}(\bar{d} - E)$$

for the chosen divergence. Both means and variances are required to justify E.

The variance floor in that result is essential. Set all means to zero and initial variances to

$$(v, v, 2v)$$

for q, y, r, then change only y's variance to $4v$. Initially the true-parent divergence is zero. Afterward the KL distances of r, y are

$$\frac{\log 2 - 1/2}{2}, \qquad \frac{\log 4 - 3/4}{2},$$

and the Bhattacharyya distances are

$$\frac{1}{2}\log\frac{3}{2\sqrt{2}}, \qquad \frac{1}{2}\log\frac{5}{4}.$$

Both rankings reverse independently of $v > 0$, while the absolute variance change $3v$ tends to zero. Hence no global absolute-variance response constant exists over all strictly positive variances.

F4. Containment alone does not identify direct parenthood. Take

$$Q = [1/4, 3/4], \qquad Y = [-1, 1], \qquad Z = [0, 2], \qquad R = [-2, 3].$$

The taxonomy may place R above incomparable direct parents Y, Z, with the query below both. Initially only Y is available. Once Z, R are exposed, the correct direct-parent set is

$$\{Y, Z\},$$

while all three query-overlap ratios equal one. A common overlap threshold or containment predicate therefore cannot distinguish the two direct parents from the broader ancestor. Ancestry filtering, volume priority, or a joint decision introduces additional information.

Positive containment certificates require true reference slacks larger than their face-change allowances and wrong slacks smaller than the negatives of those allowances. A redundant ancestor need not satisfy the latter condition. Likewise, identical axes and apertures yield identical slack for the explicit circular-cone predicate of Appendix C.1; a threshold cannot distinguish a true parent from a wrong candidate carrying the same complete record.

The volume floor in the response theorem is also necessary. In one dimension let

$$Q = Y = [0, v], \qquad R = [0, v/2].$$

Their overlap ratios are one and $1/2$. Move Y to

$$[3v/4, 7v/4].$$

Its overlap becomes $1/4$, so R wins. The face displacement $3v/4$ tends to zero with v, while the score reversal remains fixed.

F5. A shared deterministic head cannot resolve an input collision. If

$$\phi_z = \phi_r,$$

then

$$g_\theta(\phi_z) = g_\theta(\phi_r)$$

for every deterministic shared head, regardless of its expressivity. A common threshold cannot include only z. For example,

$$\phi_y = 1, \qquad \phi_r = \phi_z = 0, \qquad g(x) = x, \qquad \tau = 1/2$$

gives initial exactness and later misses the new parent. Refitting the head alone cannot resolve an input collision. Candidate identity or additional context changes the input premise.

Conversely, without a response bound, an arbitrarily small context change can create an arbitrarily large decision change. Let the old correct score be

$$1 - ac$$

and the wrong score be zero. At $c = 0$, the correct output wins for every $a > 0$. For any $\delta > 0$, choosing

$$a = \frac{2}{\delta}$$

makes it lose at $c = \delta$. Initial perfection therefore supplies no uniform positive perturbation radius over an unrestricted neural family. A bounded Lipschitz constant together with a positive margin gives the conditional protection proved in Appendix C.1.

F6. Expressibility and decoding are separate from sequence score. Let Ω_t be the permitted complete strings and π_t the parser from strings to parent sets. If no

$$a \in \Omega_t$$

parses to a set containing newly required parent z, then no decoder constrained to that support can acquire z. A constrained sampler assigns acquisition probability zero. A closed identifier grammar can satisfy this condition; an open-text generator need not.

Even when the correct output is expressible and globally optimal, greedy decoding can fail. At the later state, let one string b parse exactly to

$$\{y, z\}$$

with probability $2/5$. Let strings a_1, a_2, parsing to

$$\{y\} \quad \text{and} \quad \{y, r\},$$

each have probability $3/10$. A prefix tree first chooses branch a with total probability $3/5$ or branch b with probability $2/5$, and then splits the a branch equally. The unique globally most probable complete string is b, yet greedy decoding chooses the a branch and returns an incorrect set regardless of the next tie decision.

Initially all mass may be assigned to a_1, which correctly expresses $\{y\}$. The support and probabilities are declared parts of the update. Any beam that prunes the prefix of b exhibits the same failure for that beam configuration.

F7. A set penalty can make multi-parent recovery impossible. Consider

$$S(P) = \sum_{h \in P} s_h - \lambda \frac{|P|(|P| - 1)}{2}, \qquad 0 \le s_h \le 1, \qquad \lambda > 1.$$

Assume the nonempty feasible family is downward-closed: every subset of a feasible set is feasible. For any feasible P with at least two members and any $h \in P$,

$$S(P \setminus \{h\}) - S(P) = -s_h + \lambda(|P| - 1) > 0.$$

Deleting h is permitted and strictly improves the score. Hence no maximizer can contain two or more parents, uniformly over all permitted unary scores.

With every subset feasible and initial target

$$H = \{y\},$$

setting $s_y = 1$ makes $\{y\}$ uniquely optimal. After z becomes required, even

$$s_y = s_z = 1$$

cannot make

$$\{y, z\}$$

optimal. At $\lambda = 2$, the pair scores zero while each singleton scores one.

The obstruction is therefore architectural and not a failure of unary evidence: the objective itself forbids the required multi-parent solution. Downward closure is essential to the negative result. If the only feasible set were $\{y, z\}$, that set would maximize regardless of the penalty. The positive complete-set certificate in Appendix C.3 does not require downward closure.

F8. Missing structural evidence can make a semantically strong parent ineligible. A count predicate

$$n_h \geq m, \qquad m \geq 1,$$

rejects a desired candidate while its observed count remains below m. Likewise, any shared structural-only score assigns equal values to candidates with equal complete structural records. A common threshold cannot distinguish a true z from a wrong r under such a collision. Two leaf siblings with the same depth and count give one example for that summary.

For the strict angular/count gate, define

$$a(n) = \log_2(1 + n/2), \qquad \beta = \gamma/W^2, \qquad W > 0, \qquad \gamma \geq 0,$$

with base semantic score

$$s \in [-1, 1].$$

The gate and complete score are

$$g = s - 1 + \beta a(n)^2, \qquad S = s\mathbf{1}\{g > 0\}.$$

At $n = 0$,

$$a(0) = 0, \qquad g = s - 1 \leq 0.$$

Thus the gate rejects a zero-count candidate for every permitted semantic score, including $s = 1$. A newly valid parent can therefore be semantically maximally compatible and still remain ineligible.

For $n \geq 0$, $s \leq 1$, and $\beta > 0$, the exact eligibility boundary is

$$n > 2\left[2^{\sqrt{(1-s)/\beta}} - 1\right]. \tag{52}$$

Proof. Acceptance requires

$$\beta a(n)^2 > 1 - s.$$

Both sides are nonnegative. Taking square roots, exponentiating, and rearranging preserves the strict inequality and gives Equation (52). □

The least eligible integer is the floor of the right-hand side plus one. Even $s = 1$ requires $n \geq 1$. When $\beta = 0$, no permitted s passes. If $s \in [\ell, u]$, then

$$\ell - 1 + \beta a(n)^2 > 0$$

certifies acceptance for every admissible score, whereas

$$u - 1 + \beta a(n)^2 \leq 0$$

certifies rejection.

No universal finite discovery delay follows from unrelated taxonomy growth. For any finite horizon, arrivals may occur entirely on other branches while the candidate's observed descendant count remains zero. A probabilistic discovery-time result therefore requires an arrival law that links future arrivals to the missing structural evidence.

A three-state gate and capacity construction. The following three-state construction isolates semantic strength, structural eligibility, and output capacity. Let the independent target assign incomparable parents bus b and electric vehicle z to held-out query q. The first visible graph contains b, its child d, and an unrelated root a. The second adds root z. The third adds child w of z. The query remains outside the visible structural counts. The available-taxonomy targets are

$$\{b\}, \qquad \{b, z\}, \qquad \{b, z\}.$$

Choose fixed unit vectors

$$q = (1, 0), \qquad b = (4/5, 3/5), \qquad z = (12/13, 5/13),$$
$$d = (0, 1), \qquad a = (-1, 0), \qquad w = (0, -1).$$

Their cosine scores for b, z, d, a, w are

$$4/5, \qquad 12/13, \qquad 0, \qquad -1, \qquad 0.$$

The same calculation permits the common norm-offset factor

$$\alpha = (1 + 10^{-8})^{-1}.$$

For structural statistic

$$\mathrm{depth}_h + \log_2(1 + n_h),$$

every graph has minimum zero and maximum one, hence $W = 1$. Set $\gamma = 1$.

At count one,

$$\log_2(3/2) > 1/2$$

because

$$(3/2)^2 > 2.$$

Thus

$$a(1)^2 > 1/4.$$

Since

$$\alpha > 15/16,$$

the gate slack for b exceeds

$$(4/5)\alpha - 3/4 > 0.$$

At the middle state the newborn z has count zero and is rejected, despite

$$12/13 > 4/5.$$

After z acquires one child, its gate slack exceeds

$$(12/13)\alpha - 3/4 > 0,$$

and its accepted score becomes strictly larger than that of b. Every remaining complete score is zero.

The two output rules then separate different failure mechanisms. Top-one returns

$$b, \quad b, \quad z,$$

whereas threshold $S > 1/2$ returns

$$\{b\}, \quad \{b\}, \quad \{b, z\}.$$

At the middle state, both rules retain the old parent and miss the new one, giving parent recall $1/2$ and precision one. The miss occurs even though the new parent's base semantic score already exceeds the old parent's. At the final state, the threshold decoder recovers both parents, while top-one loses retention because it can return only one.

The construction therefore separates three quantities that an end-to-end accuracy number would conflate: semantic preference, structural eligibility, and decoder capacity.

C.5 Composition and the Actual Decoder

The family-specific score bounds are not yet decision guarantees. The final answer is produced by the complete inference rule, which may normalize scores, combine components, apply gates, restrict access, or decode a structured output. Stability must therefore be carried through the actual decoder.

Common transformations preserve rankings, not necessarily decisions. A strictly increasing transformation g common to all candidates preserves every pairwise order:

$$x > y \quad \Longleftrightarrow \quad g(x) > g(y).$$

A correspondingly transformed threshold $g(\tau)$ also preserves the strict threshold set. Keeping the old numerical threshold need not.

For example, two equal old logits give softmax probabilities

$$1/2, \quad 1/2.$$

Adding a third equal-logit candidate leaves every old pairwise ordering unchanged and changes the old probabilities to

$$1/3, \quad 1/3.$$

Any fixed threshold in

$$(1/3, 1/2)$$

therefore changes the returned set even though no old logit comparison changed. Candidate-set growth is part of the complete normalization record.

Sums, mixtures and products propagate component uncertainty differently. For fixed weights and component contrasts satisfying

$$|\Delta_i - \bar{\Delta}_i| \le \epsilon_i,$$

the triangle inequality gives

$$\left| \sum_i w_i \Delta_i - \sum_i w_i \bar{\Delta}_i \right| \le \sum_i |w_i| \epsilon_i.$$

If every component has the form

$$\Delta_i = a_i D + \varepsilon_i$$

relative to a common contrast D, then the aggregate reference coefficient is

$$\sum_i w_i a_i.$$

A ranking interpretation in the same direction requires this coefficient to be positive. Changing weights contribute an additional error term.

For probability weights w, w' on a common indexed support, suppose

$$|\Delta_i'| \le B, \qquad |\Delta_i' - \Delta_i| \le e_i.$$

Then

$$\left| \sum_i w_i' \Delta_i' - \sum_i w_i \Delta_i \right| \le \sum_i w_i e_i + B \|w' - w\|_1. \tag{53}$$

Proof. Expand the difference as

$$\sum_i w_i (\Delta_i' - \Delta_i) + \sum_i (w_i' - w_i) \Delta_i'.$$

Nonnegative probability weights and the triangle inequality give the result. □

A newly introduced component may have zero old weight. Its direct perturbation term then vanishes, while its new weight still appears through $\|w' - w\|_1$.

For multiplicative components $a_i, a_i' \in [0, 1]$, the telescoping product identity gives

$$\left| \prod_{i=1}^k a_i' - \prod_{i=1}^k a_i \right| \le \sum_{i=1}^k |a_i' - a_i|.$$

A single zero factor can therefore erase a favorable signal carried by every other component.

Mixture reweighting can also reverse an ordering with completely unchanged components. Contrasts $+1, -1$ have mixture value $+1/2$ under weights

$$(3/4, 1/4)$$

and value $-1/2$ under

$$(1/4, 3/4).$$

The scorer components did not change; the complete decision did.

A gate introduces a separate feasibility state. For

$$S = s\mathbf{1}\{g > 0\},$$

suppose jointly valid marginal intervals are

$$s \in [\ell, u], \qquad g \in [a, b].$$

A safe interval for the complete gated score is

$$\begin{cases} [\ell, u], & a > 0, \\ [0, 0], & b \le 0, \\ [\min\{0, \ell\}, \max\{0, u\}], & a \le 0 < b. \end{cases} \tag{54}$$

The first case certifies acceptance and the second certifies rejection. In the third case the gate branch itself is unresolved, so the score may lie either in the base-score interval or at zero. Dependence between s and g can tighten this enclosure.

Hard pruning and score zeroing are different operations. Pruning changes accessibility. Zeroing leaves a numerical candidate that can still be selected through ties, negative rival scores, or sufficiently large top-k. Priority rules and lexicographic tie-breaking require their own strict comparisons.

Retrieval imposes a pointwise ceiling. Suppose the final decoder must choose from accessible set A_t, while G_t denotes the set of correct permitted answers. Its pointwise success indicator is bounded by

$$\mathbf{1}\{A_t \cap G_t \neq \varnothing\}.$$

When

$$A_t \cap G_t = \varnothing,$$

every accessible answer is wrong regardless of scorer quality. Expectations under any declared law inherit this coverage ceiling without an independence assumption.

The statement concerns decoders constrained to A_t. A generator capable of naming parents outside retrieved context follows a different access model. Parent-set capacity is handled more precisely by Equation (49).

Global sequence maximization reduces to a complete-sequence margin. Fix a finite common support Ω, a fixed parser π, and sequences of length at most L, including termination. Let

$$\ell(a) = \sum_i \log p(a_i \mid a_{<i})$$

be the complete log score, with reference score $\bar{\ell}(a)$. Suppose every used conditional log probability is finite and differs from its reference by at most ϵ on every shared prefix. Summation gives

$$|\ell(a) - \bar{\ell}(a)| \le L\epsilon$$

for every

$$a \in \Omega.$$

If a valid sequence a^* with

$$\pi(a^*) = P_t$$

has a lower complete-score bound exceeding the upper bound of every invalid sequence, every global maximizer parses to P_t. Other valid strings need not score below a^*. On unchanged support and target, an initial global sequence margin greater than

$$2L\epsilon$$

is the corresponding preservation condition.

Zero conditional probabilities and support changes lie outside this finite log-error premise and require their own treatment.

Greedy decoding requires prefix margins, not only a global margin. Suppose the reference greedy path is valid and, at each visited prefix, the chosen token leads every alternative by more than

$$2\epsilon$$

in conditional log score. At the first prefix, both actual token errors are at most ϵ, so the reference choice remains uniquely optimal. Conditional on that unchanged choice, the next actual prefix equals the reference prefix. Induction through the terminating token preserves the entire path and its parsed output.

A large global sequence margin alone does not imply these prefix margins. The F6 construction in Appendix C.4 gives the counterexample. This is a greedy-decoding certificate rather than a beam-search certificate.

Sampling requires event-level probability control. For current target parent z, define

$$E_z = \{a \in \Omega : z \in \pi(a)\}.$$

The complete-sequence log bound implies

$$p(a) \geq e^{-L\epsilon}\bar{p}(a).$$

Summing over the same event gives

$$\Pr(E_z) \geq e^{-L\epsilon}\,\overline{\Pr}(E_z).$$

The comparison concerns the same current-state event on shared finite support. It cannot replace a newly required acquisition event by an old target event that did not require z.

A second route uses conditional total variation. Suppose the next-token distributions at step i differ by at most

$$\delta_i$$

uniformly over common prefixes through finite horizon L. Couple the two next-token laws maximally whenever the prefixes still agree. The conditional probability of a first disagreement at step i is at most δ_i. A union bound gives complete-sequence disagreement probability at most

$$\min\left\{1, \sum_i \delta_i\right\}.$$

Identical complete strings have identical parsed outputs, so the same quantity bounds both sequence total variation and the absolute change in any fixed parsed-event probability. Termination may be represented by an absorbing token. No independence between prefixes is required.

Decision stability is a property of the complete rule. A reference-level change produces family-specific score allowances. An old decision survives when its margin exceeds the relevant score movement. A new parent requires its own current-state acquisition condition. Exact parent-set recovery additionally requires sufficient access, output capacity, feasibility, and false-link separation. Normalization, mixture weights, gates, retrieval, and structured decoding can still alter the final answer. The guarantees therefore concern the declared reference, complete score, feasible output set, and selection rule together. They do not imply a learned-model generalization rate, an acquisition-delay law without an arrival model, or an efficiency advantage over direct score evaluation.

D CONSTRUCTED COMPARISONS

D.1 A FIXED TARGET UNDER REMOTE STRUCTURAL GROWTH

The previous results show that taxonomy growth can perturb both semantic and structural evidence. The present construction removes that ambiguity. The correct parent is fixed, every semantic vector is fixed, and several semantic scorers keep the same winner throughout. Only the revealed taxonomy grows. This isolates whether structural growth alone can overturn a previously correct gated decision.

Write

$$z(\vartheta) = (\cos\vartheta, \sin\vartheta).$$

The held query has angle $-\pi/8$ and independently specified sole parent B throughout the entire history. Initially the candidate taxonomy contains roots

$$A,\ B,\ D$$

at angles

$$-3\pi/5, \qquad 0, \qquad 3\pi/5,$$

together with one child $\mathrm{B000}$ of B at angle $\pi/128$.

Arrival one introduces a competing root C at angle $\pi/16$. Arrival $j \in \{2, \dots, 40\}$ introduces child C_{j-1} of C at

$$\vartheta_{\mathrm{C}_m} = \pi\left(\frac{1}{128} + \frac{80+m}{32768}\right), \qquad m = 1, \dots, 39.$$

All present concepts remain candidates, with rows aligned to their identities. The held query is excluded from the structural and field populations. No semantic vector changes at any stage. Descendant statistics alone are refreshed from the revealed taxonomy.

Proposition 11 (Fixed semantics, changing structural decision)**.** *Along this forty-arrival history:*

1. *pure cosine selects B uniquely at every state;*
2. *the declared positive fuzzy realization selects B uniquely at every state;*
3. *the identity-covariance Gaussian KL and Bhattacharyya realizations select B uniquely at every state;*
4. *the complete gated score with $\gamma = 5$ selects B through arrival 26 and the incorrect parent C from arrival 27 through arrival 40; and*
5. *the compensated gate with $W_0 = 1$ and*

$$\gamma_j = 5W_j^2$$

retains B throughout.

Proof. The proof separates what never changes from what does.

The semantic order never changes. The normalized angular distances from the held query to B, C, A, D are respectively

$$\frac{1}{8}, \qquad \frac{3}{16}, \qquad \frac{19}{40}, \qquad \frac{29}{40}.$$

Every arriving child of C has angle strictly between 0 and $\pi/4$. Its angular distance from the query is therefore strictly between

$$\frac{\pi}{8} \quad \text{and} \quad \frac{3\pi}{8}.$$

All relevant angular distances lie below π, where cosine is strictly decreasing. Hence B is the unique cosine winner at every state, including against every newly introduced child.

Thus the taxonomy can grow for forty arrivals without changing the semantic cosine ordering of the correct parent.

The fuzzy realization preserves the same order. Map every two-dimensional unit source vector z to

$$\begin{aligned}\phi(z) &= \left(\frac{1+z_1/2}{2}, \frac{1+z_2/2}{2}, \frac{1-z_1/2}{2}, \frac{1-z_2/2}{2}\right), \\ u(z) &= \frac{\phi(z)}{\sqrt{9/8}}, \qquad w = (1,1,1,1).\end{aligned} \tag{55}$$

Every coordinate of $\phi(z)$ lies in $[1/4, 3/4]$, so all memberships are strictly positive. Moreover,

$$\|\phi(z)\|_2^2 = 1 + \frac{\|z\|_2^2}{8} = \frac{9}{8}, \qquad \sum_i \phi_i(z) = 2.$$

Therefore $u(z)$ has unit Euclidean norm and the query denominator is positive.

Direct expansion gives

$$u(z_q)^\top u(z_h) = \frac{8 + z_q^\top z_h}{9}, \qquad \sum_i u_i(z_q) = \frac{2}{\sqrt{9/8}} > 0.$$

Hence the complete weighted-overlap score

$$\frac{\sum_i w_i u_i(z_q) u_i(z_h)}{\sum_i w_i u_i(z_q)}$$

is a common positive affine function of cosine. It therefore has exactly the same unique winner B at every state.

The Gaussian realizations preserve the same order. For each candidate and the query, set

$$\mu_h = z_h, \qquad \Sigma_h = I_2.$$

Equal covariance matrices cancel the determinant and variance-ratio terms. Consequently,

$$D_{\mathrm{KL}}(q\|h) = \frac{\|z_q - z_h\|_2^2}{2} = 1 - z_q^\top z_h,$$

and

$$D_{\mathrm{B}}(q,h) = \frac{\|z_q - z_h\|_2^2}{8} = \frac{1 - z_q^\top z_h}{4}.$$

Thus both

$$-D_{\mathrm{KL}} \qquad \text{and} \qquad \exp(-D_{\mathrm{B}})$$

preserve cosine order. The correct parent B therefore remains the unique winner throughout the forty arrivals.

This reduction relies on the stated common covariance and does not extend to arbitrary learned covariance matrices.

Only the structural gate develops a moving boundary. For the gated realization, all semantic vectors have unit norm. Let

$$\alpha = (1 + 10^{-8})^{-1}.$$

The semantic scores of the fixed correct parent and competing root are

$$s_B = \alpha \cos(\pi/8), \qquad s_C = \alpha \cos(3\pi/16),$$

so

$$0 < s_C < s_B < 1$$

at every state.

The target B has exactly one descendant throughout the history. After $j \geq 1$ arrivals, C has

$$n_C = j - 1$$

descendants. Every other candidate is a leaf and therefore receives complete score zero under Equation (58).

Using the raw structural radius

$$d_h + \log_2(1 + n_h)$$

with root depth one, roots A, D attain the minimum radius one and a child attains radius two initially. After subtracting the minimum from the maximum, the global structural range is

$$W_0 = 1, \qquad W_j = \max\{1, \log_2 j\}, \qquad j \geq 1.$$

For B, the semantic score and descendant count remain fixed. Its gate nevertheless weakens because W_j grows. The strict acceptance condition reduces exactly to

$$W_j < W_B^*, \qquad W_B^* = \frac{\sqrt{5}\,\log_2(3/2)}{\sqrt{1 - \alpha\cos(\pi/8)}}. \tag{56}$$

Equality rejects because the gate is strict.

This equation isolates the mechanism: the local evidence for B is unchanged, while remote growth increases the global denominator W_j until the fixed semantic and structural support of B is no longer sufficient.

The transition occurs exactly between arrivals 26 and 27. The boundary satisfies

$$\log_2 26 < W_B^* < \log_2 27.$$

For example,

$$0.9238795 < s_B < 0.9238796,$$
$$0.5849625 < \log_2(3/2) < 0.5849626,$$
$$4.7004 < \log_2 26 < 4.7005,$$
$$4.7548 < \log_2 27 < 4.7549.$$

Substitution into

$$g_B(j) = s_B - 1 + 5\left[\frac{\log_2(3/2)}{W_j}\right]^2$$

gives

$$g_B(26) > 0, \qquad g_B(27) < 0.$$

Since W_j is nondecreasing, B is accepted through arrival 26 and rejected from arrival 27 onward.

The numerical inequalities can be made fully finite. Using

$$\cos(\pi/8) = \frac{\sqrt{2 + \sqrt{2}}}{2},$$

rational radical bounds follow by squaring rational endpoints. For $x > 1$, the logarithm expansion

$$P_K(x) = 2\sum_{k=0}^{K-1} \frac{t^{2k+1}}{2k+1}, \qquad t = \frac{x-1}{x+1},$$

satisfies

$$0 < \log x - P_K(x) \leq \frac{2t^{2K+1}}{(2K+1)(1-t^2)}.$$

Taking $K = 160$ for

$$x \in \{3/2, 2, 26, 27\}$$

and dividing the resulting logarithm intervals by the positive interval for $\log 2$ yields the displayed rational bounds.

Thus, the phase transition is certified analytically rather than located by numerical search.

The weaker semantic rival becomes structurally eligible. The competing root C is accepted whenever it has at least one descendant. For $n \geq 1$,

$$(1 + n/2)^2 \geq 1 + n,$$

hence

$$\frac{\log_2(1 + n/2)}{\log_2(1 + n)} \geq \frac{1}{2}.$$

Its structural term is therefore at least

$$\frac{5}{4},$$

and

$$g_C \geq s_C - 1 + \frac{5}{4} > 0.$$

At birth, C is a leaf and is rejected.

While B remains eligible,

$$s_B > s_C > 0,$$

so B beats the accepted C and every rejected leaf. Once the global range pushes B across its gate boundary, its complete score becomes zero while C retains positive complete score. The unique winner therefore changes to the semantically weaker and incorrect parent C.

The semantic ranking never reversed. The structural eligibility did.

Compensation removes exactly this source of drift. Set

$$\gamma_j = 5W_j^2.$$

Then

$$\frac{\gamma_j}{W_j^2} = 5$$

at every state. The gate therefore retains the target's initial positive structural slack despite the increasing global range. Since all semantic vectors remain fixed and

$$s_B > s_C$$

throughout, Proposition 16 implies that the compensated rule retains B as the unique winner at every arrival. □

Isolated mechanism. The forty-arrival history separates three mechanisms that would otherwise be conflated in an end-to-end accuracy change.

First, semantic evidence is fixed. Cosine, the positive fuzzy realization, Gaussian KL, and Bhattacharyya all retain the correct parent B throughout.

Second, the target relation is fixed. B remains the independently specified sole parent at every state.

Third, the gated rule changes only because taxonomy growth changes its structural context. The exact boundary

$$\log_2 26 < W_B^* < \log_2 27$$

predicts the decision flip before evaluating arrival 27.

The construction holds the semantics and correct parent fixed while remote structure grows. The resulting increase in the global range W_j reduces the target gate slack until B is rejected after arrival 26.

Compensation restores the missing invariance by holding

$$\gamma_j / W_j^2$$

fixed. It does not improve the semantic representation because none was needed: the semantic ordering was already correct.

This is an existential finite comparison under the declared structural refresh. The same source vectors support valid order-preserving realizations for the other displayed scorers, so the construction does not imply a universal failure of semantic scoring classes. The sole-parent truth is also essential. In the acquisition history below, C is itself a correct parent, so returning C alone becomes an incomplete parent set.

Table 5: Role of each empirical protocol. A state or transition is part of a declared history.

Protocol	Controlled question	Unit and supported conclusion
Catalogue forecast	Can the next reference state be bounded before the arriving identity is revealed?	Candidate arrival / prefix; checks the source update identity and forecast envelope.
Stochastic protection	Can the old protected state survive every prefix of a growth trajectory?	Path of at most 40 arrivals; first-loss and uninterrupted-survival analysis.
Preservation	If the correct parent never changes, can its old decision be certified after each update?	200 transitions across 201 states; certificate coverage on realized-correct outputs.
Acquisition	If a second parent becomes required, can the rule retain the old parent and acquire the new one?	40 postbirth states; retention, acquisition, exact-set recovery and false links.
Environment formula replay	Do candidate competition and structural filtering create the same kinds of failures under observed taxonomy structure?	Query / snapshot / declared order; initially-correct losses and endpoint totals.
Six-source acquisition	Does preserving old links imply recognition of newly available parents?	Old and newly available reference links; retention, acquisition and false-link control.
Full-source profiles	Are the declared field quantities computable over larger observed taxonomies?	Spectrum / coordinate probe; descriptive construction coverage instead of parent accuracy.

D.2 Experimental Evidence Map

The empirical protocols test different links in the theoretical chain and therefore have different units of analysis. They are not pooled into one accuracy estimate. In particular, final-state correctness, uninterrupted preservation, acquisition of a newly required parent, and exact parent-set recovery are distinct outcomes.

The fixed-target construction and the acquisition construction use closely related growth mechanisms with different ground-truth targets. In the fixed-target history, the competing root remains incorrect. In the acquisition history, the same kind of arrival becomes an additional required parent. This distinction is deliberate: the same structural event can test either preservation of an old decision or acquisition of a new one, depending on the independently specified target.

Named-formula realizations likewise retain their own candidate sets, thresholds, gates, and output rules. The protocols, therefore, diagnose components of the reference-score-selector chain rather than compare eight trained systems.

D.3 Shared Acquisition History

The acquisition construction makes the change in the required answer explicit. The complete ground-truth graph contains incomparable roots A, B, C, D, one child B000 of B, children C001 through C039 of C, and a held query whose direct parents are B and C. At time zero, C and its descendants are hidden, so the only available required parent is B. At time one, C is revealed and becomes newly required. Subsequent steps reveal one child of C at a time, while the required parent set remains {B,C}. The held query never enters the calibration populations. Its target links, therefore, cannot influence the reference through calibration. The semantic source is the unit circle. Root angles divided by π are:

$$-\frac{3}{5}, \qquad 0, \qquad \frac{1}{16}, \qquad \frac{3}{5},$$

for A, B, C, D, while the query angle is $-1/8$. The inherited source-order construction uses B as a reference. Quantization places the preferred semantic coordinate of C at zero; the order-preserving collision rule assigns it the distinct root coordinate one. A and D occupy -2 and $+2$. The receiving chart has nine cyclic ticks. At every snapshot, the declared population, spectrum, emission,

Table 6: Eight declared rules on the common acquisition history. Larger complete scores are preferred. These are controlled scorer constructions, not eight trained models.

Construction	Complete rule
Distance	$-\|\sqrt{p_q} - \sqrt{p_h}\|_2$.
Bilinear	$-\|p_q - p_h\|_2^2$, represented by an augmented bilinear form.
Moment Gaussian	Negative directed KL between Gaussians with mean $m(p)$ and variance $v(p)$.
Box overlap	Intersection length divided by query-interval length for $[m(p) - \sqrt{v(p)}/4, m(p) + \sqrt{v(p)}/4]$.
Fixed neural score	$-\|p_q - p_h\|_1$, realizable by a fixed ReLU readout.
Normalized token score	Log-softmax of the preceding L1 scores over the current candidate set.
Penalized joint set	Individual matching rewards plus pair interactions and a set-size penalty, defined below.
Structural gate	Declared gated angular arithmetic on square-root fields with current structural counts and normalization.

calibration, and residual-count laws are rebuilt from the revealed taxonomy. The construction is intentionally favorable to both true parents. Once C is present,

$$p_q = p_B = p_C.$$

Thus, both required parents receive ideal supplied field inputs. A failure to return $\{B, C\}$ cannot be attributed to a worse supplied representation of one true parent. For a profile p, define

$$m(p) = \mathbb{E}_p[x], \qquad v(p) = 1 + \mathrm{Var}_p(x) > 0.$$

The eight complete scoring rules are listed in Table 6. Their numerical scales differ; the shared history does not make them comparable as scores. For every scalar rule, the selection threshold is fixed at

$$\frac{S_0(B) + \max_{h \neq B} S_0(h)}{2}.$$

It is therefore calibrated once, between the initially correct parent and the strongest initial rival, and is never retuned using later labels. Every rule returns exactly B at the initial state. The joint-set rule instead evaluates every feasible subset and resolves equal optima by a fixed lexicographic rule. Numerical outputs evaluate the declared formulas, while the obstruction arguments below are symbolic.

D.4 Output Obstructions and Controls

The construction isolates three downstream mechanisms because the supplied representation of the true parents is already favorable.

Normalization can change admission without changing either true logit. For the normalized-token rule,

$$\ell_h = -\|p_q - p_h\|_1.$$

A fixed probability threshold a is equivalent to the log-score cutoff $\log a$. After C is born,

$$\ell_B = \ell_C = 0.$$

Nevertheless, C is admitted only when

$$e^{\ell_C} > a \sum_j e^{\ell_j}.$$

The candidate-set normalizer can therefore alter admission while both true logits remain unchanged. The unnormalized L1-score control uses the corresponding initial score threshold and remains exact throughout the construction.

A set objective can reject a perfectly represented new parent. For parent set P, define

$$J(P) = \sum_{h \in P} \left(\frac{1}{2} - \|p_q - p_h\|_1 \right) - 2\binom{|P|}{2} + \frac{1}{10} \sum_{h<j \in P} \langle p_h, p_j \rangle.$$

For any $|P| = m \geq 2$, the marginal contribution of the last element is at most

$$\frac{1}{2} - 1.9(m-1) < 0.$$

Its individual reward is at most $1/2$, whereas each new pair costs at least $2 - 1/10 = 1.9$. Because all subsets are feasible, deleting that element strictly improves the objective. No optimum can therefore contain two parents.

B and C tie as the best singletons, and the fixed tie rule returns B. Thus the newly required C is missed even though both true parents receive the maximum unary match. The obstruction lies in the complete set objective, not in their individual representations.

A structural gate can make a new true parent temporarily ineligible. Under the declared gated score, a candidate with zero descendants would need an angular factor strictly larger than one to pass. The permitted factor is at most one. A newly revealed root can therefore receive complete score zero despite ideal semantic matching.

This is an eligibility obstruction, not a universal statement about structural scorers. A top-k rule may still return a zero-scored candidate through its tie convention, whereas the positive-cutoff rule used here excludes it. The resulting delay is specific to the declared gate and selector.

Together these controls isolate three distinct failure sources under exact supplied evidence: normalization, set preference, and structural feasibility. They explain why exact true-parent inputs do not imply exact parent-set recovery.

D.5 Preservation Certificates: Correctness versus Certifiability

The preservation experiment separates two questions that are easy to conflate: *did the decision remain correct?* and *could the available bound certify that correctness?* A sufficient certificate may answer the second question conservatively even when the first answer is unambiguously positive.

The construction contains 201 taxonomy states and 200 transitions. The query has the same unique correct parent throughout, and all eight declared output rules remain correct at every state. The experiment therefore contains no realized prediction failure. It instead asks how much information a preservation bound needs in order to certify these already-correct decisions.

The sharpest contrast occurs for the distance and moment-Gaussian scores:

Rule	Realized correct transitions	Separate worst-case certificate	Refined certificate
Distance	200/200	0/200	200/200
Moment Gaussian	200/200	0/200	195/200

Thus the original separate bounds are valid yet maximally uninformative for these two rules on this construction. The refinements recover nearly all of the missing certifiability without changing a single prediction.

Why the separate distance bound can be loose. A distance score depends on the *relative* query-candidate representation. Bounding query movement and candidate movement independently discards their direction before the two changes are compared. The resulting triangle-inequality allowance is safe. A common movement is nevertheless charged twice even when it has little effect on their separation.

Retaining the joint response instead gives

$$\left| \|u'_q - u'_h\|_2 - \|u_q - u_h\|_2 \right| \leq \left\| (u'_q - u'_h) - (u_q - u_h) \right\|_2, \tag{57}$$

the same relative-motion principle used in Equation 8 and Appendix C.1. The right-hand side preserves cancellation before taking the norm. If query and parent move together, their individual movements can be large while their relative displacement is small or even zero.

On the 200 preservation transitions, the joint-response certificate succeeds on 200/200 cases, compared with 0/200 under the separate endpoint allowance. The improvement comes entirely from retaining information that the conservative decomposition had discarded.

Why the Gaussian bound can be loose. The same phenomenon appears in a richer form for the moment-Gaussian score. A global derivative bound must remain valid over a broad admissible range of means and variances. The realized transition occupies a much smaller segment of that domain.

For each transition, the refined certificate applies the derivative argument of Appendix C.1 to the convex segment joining the actual endpoint parameters. Its variance floor is the minimum of the four endpoint variances, its variance ceiling is their maximum, and its mean-gap allowance is the larger endpoint mean difference. These quantities remain rigorous bounds on the complete segment while avoiding ranges that the realized transition never visits.

The refinement certifies 195/200 Gaussian transitions. The remaining five decisions are still correct under direct score evaluation. The available sufficient bound does not separate the true parent from its rival. They are therefore *inconclusive*, not counterexamples to the theorem.

An exact special case makes the role of joint information particularly clear. If the query and candidate Gaussians coincide at both endpoints, both directed divergences are exactly zero. The correct score-change bound is then zero, regardless of how loose a separately constructed derivative allowance might be.

Certificate failure is not prediction failure. For a true parent h and rival r, the generic preservation statement has the form

$$\text{new margin} \geq \text{old margin} - \text{certified relative score movement}.$$

A positive lower bound certifies preservation. A nonpositive lower bound says only that the supplied information is insufficient to prove the sign of the new margin. It does not imply that the sign changed.

The preservation construction contains correct-and-certified transitions and correct-but-uncertified transitions, with no certified-but-incorrect case.

Interpretation. The experiment does not improve model accuracy: the decisions were already correct. It instead identifies which information makes the theoretical guarantee informative. Separate worst-case allowances discard common-mode response and local endpoint structure; joint-response and endpoint-adapted bounds retain them.

Validity, tightness, and realized correctness are therefore distinct properties of a preservation analysis. A useful certificate must be valid first. Its practical coverage then depends on how much decision-relevant structure it preserves.

The refinements above are retrospective checks because they use realized endpoint information. They become prospective guarantees only when the corresponding joint response or endpoint range can itself be bounded before the update. Their counts are therefore kept separate from the forty-state acquisition experiment, which evaluates a different question.

E GATED SCORER COMPATIBILITY

A gated scorer combines two logically distinct requirements. The semantic score asks whether a candidate is compatible with the query; the structural gate asks whether that candidate is eligible to contribute its semantic score at all. Taxonomy growth can therefore change the final decision even when semantic compatibility is unchanged. A newly introduced *electric vehicle* can be the strongest semantic match for *electric bus* and still be suppressed when the structural evidence required by the gate has not yet appeared.

Rejection alone, however, is not a decision error. The declared rule maps a rejected candidate to score zero, so a rejected correct parent can still tie or beat negative rivals. A strict top-one failure requires a positive accepted incorrect rival. The results below retain this complete-score distinction and characterize when the gate certifies a decision, when it forces an error across an entire fidelity

class, how far a representation must move to cross the gate, and when one global gate strength can satisfy all desired acceptance and rejection constraints.

E.1 Semantic Compatibility Meets Structural Eligibility

For a query and candidate parent h, let $s_h \in [-1, 1]$ be its semantic score, $n_h \geq 0$ its distinct strict-descendant count, $W > 0$ the declared structural range, and $\gamma \geq 0$ the gate strength. Define

$$\begin{aligned} a(n) &= \log_2(1 + n/2), \qquad b_h = \gamma \left[\frac{a(n_h)}{W}\right]^2, \\ g_h &= s_h - 1 + b_h, \qquad S_h = \begin{cases} s_h, & g_h > 0, \\ 0, & g_h \leq 0. \end{cases} \end{aligned} \tag{58}$$

The gate is strict, and equality maps to zero. Equivalently, candidate h is eligible exactly when

$$s_h > 1 - b_h.$$

The structural term therefore moves the semantic acceptance threshold. More structural support lowers the threshold; a larger global range W weakens the same local support. At $n_h = 0$, $a(0) = 0$, so $g_h = s_h - 1 \leq 0$ for every permitted semantic score. A zero-descendant candidate is rejected even at the maximal score $s_h = 1$.

This boundary separates semantic preference from decision feasibility. A candidate can be the strongest semantic match and still contribute complete score zero.

Proposition 12 (Complete-score interval propagation)**.** *For an admitted semantic-score enclosure $s_h \in [\ell_h, u_h]$, a valid enclosure for the complete score is*

$$\mathcal{I}_h = \begin{cases} [\ell_h, u_h], & \ell_h - 1 + b_h > 0, \\ [0, 0], & u_h - 1 + b_h \leq 0, \\ [\min(0, \ell_h), \max(0, u_h)], & \textit{otherwise}. \end{cases} \tag{59}$$

A candidate is certified as the unique maximizer when its lower complete-score endpoint exceeds every other candidate's upper endpoint.

Proof. If $\ell_h - 1 + b_h > 0$, every admitted semantic score passes the gate, so the complete-score image is $[\ell_h, u_h]$. If $u_h - 1 + b_h \leq 0$, every admitted score fails, including equality, and the image is $\{0\}$. In the remaining case both branches are possible, so the image lies in $[\ell_h, u_h] \cup \{0\}$ and hence inside the displayed convex-hull enclosure. Strict separation of complete-score intervals yields the same strict separation for every jointly admitted score vector. □

The mixed branch is an enclosure rather than an attainability claim. Overlap is inconclusive. In particular, semantic ordering alone does not determine the final ranking: an accepted negative score can lose to a rejected zero.

E.2 When a Gate Forces Error Across a Scorer Class

The interval result becomes substantially stronger when the correct parent is uniformly rejected while one incorrect rival is uniformly accepted with positive score. The gate then forces a top-one error throughout the entire admitted fidelity class.

Theorem 13 (Uniform gated-class error and panel ceiling)**.** *Let $\mathcal{M}$ be a nonempty class of jointly realizable scorer models or nonanticipating adaptation policies with perfect initial predictions on a declared finite probe panel. Fix an admitted taxonomy history and a time t. Each probe has a finite nonempty candidate set containing its independently specified unique correct parent. Assume that every member of $\mathcal{M}$ satisfies the declared current score intervals under this same history and the complete rule in Equation* (58)*.*

For one probe with correct parent y and incorrect rival r, suppose

$$u_y - 1 + b_y \leq 0, \qquad \ell_r - 1 + b_r > 0, \qquad \ell_r > 0. \tag{60}$$

Then every member of $\mathcal{M}$ has $S_y = 0 < S_r$ and therefore makes a top-one error, independently of other candidates and tie-breaking.

For a panel of $K \geq 1$ probes, let J_t contain the probes with such uniform certificates. Every model in the class obeys

$$\sup_{M \in \mathcal{M}} \mathrm{Acc}_M(t) \leq 1 - \frac{|J_t|}{K}. \tag{61}$$

If one common model-independent law on admissible histories satisfies the same class-wide contracts, then

$$\sup_{M \in \mathcal{M}} \mathbb{E}[\mathrm{Acc}_M(t)] \leq 1 - \mathbb{E}\left[\frac{|J_t|}{K}\right]. \tag{62}$$

The certificate set is defined from the common history and uniform contracts rather than selected separately after observing each model's errors.

Proof. For every admitted target score, $s_y - 1 + b_y \leq u_y - 1 + b_y \leq 0$, so the strict gate returns $S_y = 0$. Every admitted rival satisfies $s_r - 1 + b_r \geq \ell_r - 1 + b_r > 0$ and $s_r \geq \ell_r > 0$, so $S_r = s_r > 0$. Thus the correct parent cannot maximize the complete score for any jointly realizable member of the class.

Each certified probe contributes zero to the number of correct predictions, while every remaining probe contributes at most one. Hence $\mathrm{Acc}_M(t) \leq 1 - |J_t|/K$ for every M, proving Equation (61). Under the common history law the same inequality holds pointwise for each admitted policy; taking expectations and then the supremum gives Equation (62). No independence between query errors or arrivals is required. □

A local gate incompatibility can therefore induce a panel-level accuracy ceiling without assuming independent probes. The conclusion applies to adaptation that remains inside the stated reference-fidelity class. It does not cover arbitrary representations that merely share an architecture name. Uniqueness of the correct parent is material: suppressing one valid parent need not force a top-one error when another valid parent remains above every rival. Multi-parent recovery instead follows the set-level conditions in Appendix C.3. Enlarging the admitted intervals can remove a certificate without certifying success.

E.3 How Far Must a Representation Move to Become Eligible?

A forced rejection can also be stated geometrically. If a score-faithful adapter maps semantic scores to a retained scalar coordinate, the gate induces a boundary in that coordinate. The distance from the canonical reference to this boundary is a necessary representation change for acceptance.

Proposition 14 (Reference departure and boundary attainability)**.** *Let A be a fixed strictly decreasing score-faithful adapter on the declared scalar score domain J. Let the strict gate threshold $\tau_y = 1 - b_y$ lie inside that domain and put $c_y = A(\tau_y)$. Let ϱ_y be the exact canonical reference coordinate, with $\varrho_y \geq c_y$. Define*

$$X_{\mathrm{acc}} = \{A(s_y(z)) \mid z \text{ is jointly admissible and } s_y(z) > \tau_y\}.$$

Every $x \in X_{\mathrm{acc}}$ satisfies

$$x < c_y, \qquad |x - \varrho_y| > \varrho_y - c_y. \tag{63}$$

With the convention $\inf \varnothing = +\infty$,

$$d_* := \inf_{x \in X_{\mathrm{acc}}} |x - \varrho_y| \geq \varrho_y - c_y. \tag{64}$$

For nonempty X_{acc}, equality holds if and only if $\sup X_{\mathrm{acc}} = c_y$. In that case the infimum is not attained. Independently of approachability, a fidelity requirement $|x - \varrho_y| \leq e_y$ with $\varrho_y - e_y \geq c_y$ excludes every accepted input.

Proof. Strict decrease of A maps $s_y > \tau_y$ to $x < c_y$. Since $\varrho_y \geq c_y$, every accepted coordinate satisfies

$$|x - \varrho_y| = \varrho_y - x > \varrho_y - c_y,$$

proving the lower bound, including the empty-set convention.

For nonempty X_{acc}, its supremum is at most c_y, and

$$\inf_{x \in X_{\mathrm{acc}}} (\varrho_y - x) = \varrho_y - \sup X_{\mathrm{acc}}.$$

Equality in Equation (64) is therefore equivalent to $\sup X_{\mathrm{acc}} = c_y$. No accepted point equals the strict boundary c_y, so equality of the infimum does not imply attainment. Finally, $|x - \varrho_y| \le e_y$ and $\varrho_y - e_y \ge c_y$ imply $x \ge c_y$, contradicting acceptance. □

The proposition turns a gate into a fidelity requirement. If the canonical reference lies on the rejected side, acceptance needs at least the boundary distance

$$\varrho_y - c_y$$

of departure. A fidelity ball that cannot cross this boundary certifies rejection throughout the admitted class. Continuity of a scalar adapter can establish approachability in the scalar domain; joint attainability in a constrained representation remains a separate premise. If no admitted score passes the gate, the required departure is infinite.

E.4 Binding the Gate to the Constructed Field

The preceding result becomes source-level when the gate statistic can be recovered from the declared `TaxoAtlas` construction. The fixed-target growth construction provides such a binding: growth under a new root changes the structural range while the old parent's semantic match remains fixed, and the field contrast algebraically determines the competing-lineage count.

For the postbirth states in Appendix D.1, let G be the unnormalized field contrast and let $D > 0$ and $\chi > 0$ be the construction's fixed kernel quantities. The established identity is

$$G = \left(\frac{5}{21} - L_C\right) D - \frac{\chi}{19}, \qquad L_C = \frac{2n_C + 3}{2n_C + 19}. \tag{65}$$

Solving gives

$$\begin{aligned} L_C &= \frac{5}{21} - \frac{G + \chi/19}{D}, \\ n_C &= \frac{19L_C - 3}{2(1 - L_C)}, \qquad W = \max\{1, \log_2(1 + n_C)\}. \end{aligned} \tag{66}$$

The first identity uses $D > 0$. For the second, $L_C(2n_C + 19) = 2n_C + 3$ gives $2(1 - L_C)n_C = 19L_C - 3$. The construction has $L_C < 1$, so division is valid. The recovered count is a legal integer because G comes from a state of this construction; an arbitrary field contrast is not granted that property. A normalized field additionally requires its common normalizer to recover G.

For the unit-norm angular instantiation,

$$s(d) = \alpha \cos(\pi d), \qquad \alpha = (1 + 10^{-8})^{-1}.$$

The target B has one descendant, reference distance $1/8$, and gate strength $\gamma = 5$. When its gate threshold lies strictly inside $(-\alpha, \alpha)$, the decreasing angular adapter maps that threshold to

$$c_B(G) = \frac{1}{\pi} \arccos\left(\frac{1 - 5[\log_2(3/2)/W(G)]^2}{\alpha}\right). \tag{67}$$

Whenever $1/8 \ge c_B(G)$, Proposition 14 yields the necessary departure

$$d_* \ge \frac{1}{8} - c_B(G).$$

Equality requires joint boundary approachability. The implication is specific to this field construction and its source-angle fidelity contract; agreement with a coarse integer tick alone does not identify the required angular coordinate.

The construction therefore maps the field contrast G to the structural count n_C, then to the range W, the gate boundary $c_B(G)$, and the necessary fidelity departure. A field change can thus be translated into an exact statement about how far a score-faithful representation must move before the gated decision can change.

E.5 Removing Remote-Range Drift Without Removing the Gate

The gate strength enters through

$$\beta = \frac{\gamma}{W^2}.$$

Remote taxonomy growth can increase W without changing a candidate's descendants or semantic score, thereby weakening the candidate's structural support. Holding β fixed removes this particular source of drift. It does not determine which candidates should pass or which accepted candidate should win.

Proposition 15 (Acceptance-set invariance)**.** *Fix $n > 0$ and a nondegenerate admitted score interval $J = [\ell, u]$. For $\beta_i \geq 0$, let*

$$\tau_i = 1 - \beta_i a(n)^2$$

and assume $\ell < \tau_i < u$ for $i \in \{0, 1\}$. The acceptance sets

$$\{s \in J \mid s > \tau_i\}$$

are equal if and only if $\beta_1 = \beta_0$. The new acceptance set contains the old one if and only if $\beta_1 \geq \beta_0$.

Proof. Since $n > 0$, $a(n)^2 > 0$, so τ_i is strictly decreasing in β_i. Equal strengths yield equal thresholds and hence equal acceptance sets. If the thresholds differ, their midpoint lies in J and in exactly one of the two upper intervals, so the acceptance sets differ. Likewise, $\beta_1 \geq \beta_0$ is equivalent to $\tau_1 \leq \tau_0$, which is exactly the condition for the new upper interval to contain the old one. □

The interior-threshold assumption matters. Saturated acceptance sets can coincide for unequal strengths. Preserving decisions on finitely many observed scores is weaker than preserving the full acceptance set over J. At $n = 0$, $a(n) = 0$ and no score at most one passes regardless of β.

Proposition 16 (Compensated acceptance and ranking)**.** *Fix $W_0 > 0$ and $\gamma_0 \geq 0$. For each $W_t > 0$, set*

$$\beta_0 = \frac{\gamma_0}{W_0^2}, \qquad \gamma_t = \gamma_0 \left(\frac{W_t}{W_0}\right)^2. \tag{68}$$

If a candidate's semantic score and descendant count are unchanged, its gate and complete score are unchanged. With unchanged semantic score and $n_h(t) \geq n_h(0)$, an initially accepted candidate stays accepted.

Suppose h remains the correct available parent, was initially accepted, has $n_h(t) \geq n_h(0)$, and satisfies $s_h(t) = s_h(0) > 0$. If $s_h(t) > s_r(t)$ for every current rival r, including newly available rivals, then h is the unique maximizer of the complete score.

Proof. Equation (68) gives

$$\frac{\gamma_t}{W_t^2} = \beta_0.$$

Hence the gate slack becomes

$$s_h(t) - 1 + \beta_0 a(n_h(t))^2.$$

It is unchanged when both semantic score and descendant count are unchanged. Since $a(n)$ is nonnegative and increasing, nondecreasing count cannot reduce the slack when the semantic score is fixed. Initial strict acceptance is therefore preserved. Under the ranking premises, $S_h = s_h > 0$. Every rival either passes and retains semantic score $S_r = s_r < s_h$, or fails and receives $S_r = 0 < s_h$. Thus h is the unique maximizer. □

Compensation therefore removes one specific dependence: remote growth through the global range W. It does not remove semantic drift, new competitors, or the need for selectivity.

Semantic drift still spends gate and ranking slack. Suppose the semantic score also moves after compensation.

Proposition 17 (Preservation with bounded semantic drift)**.** *Under Equation* (68)*, suppose the correct available parent h has initial slack $g_h(0) > 0$, nondecreasing descendant count, and*

$$s_h(t) \geq s_h(0) - e_h(t).$$

Let $s_r(t) \leq u_r(t)$ for every current rival. The conditions

$$e_h(t) < g_h(0), \qquad s_h(0) - e_h(t) > \max\{0, \max_{r \neq h} u_r(t)\} \tag{69}$$

are sufficient for h to be the unique correct maximizer. An empty rival maximum is interpreted as $-\infty$.

Proof. Nondecreasing count gives

$$g_h(t) \geq g_h(0) - e_h(t) > 0,$$

so the target remains accepted. Its complete score is at least $s_h(0) - e_h(t)$, which is positive and exceeds every admitted rival semantic score. A passing rival cannot reach it, and a rejected rival receives zero. Hence the complete-score separation is strict. □

For an old rival with $|s_r(t) - s_r(0)| \leq e_r(t)$, an initial semantic lead larger than $e_h(t) + e_r(t)$ supplies the required current comparison. Newly available rivals require their own current bounds. The condition is pathwise: compensation preserves the gate law, while semantic stability must still be certified along the realized growth history.

E.6 Can One Global Gate Be Selective Enough?

A single global strength faces two opposing requirements. Weakly supported true parents need sufficiently large β to pass. Strongly supported false parents require β to remain sufficiently small. The next result gives the exact interval in which both demands can be satisfied.

Theorem 18 (Exact global selectivity feasibility)**.** *Let $\mathcal{T}$ be a finite nonempty set of desired query-parent pairs and $\mathcal{R}$ a finite set of undesired pairs. All these pairs have positive descendant counts. Their semantic-score constraints $[\ell_h, u_h] \subseteq [-1, 1]$ are fixed while the global strength is varied. For necessity, assume these are exact interval constraints, or that each relevant desired lower endpoint and undesired upper endpoint is attainable in the admitted score class. Set*

$$L = \max_{h \in \mathcal{T}} \frac{1 - \ell_h}{a(n_h)^2}, \qquad U = \min_{h \in \mathcal{R}} \frac{1 - u_h}{a(n_h)^2}, \tag{70}$$

where $U = +\infty$ when $\mathcal{R}$ is empty. A finite nonnegative strength β accepts every desired pair for every admitted score and rejects every undesired pair precisely when

$$\beta > L, \qquad \beta \leq U. \tag{71}$$

Such a strength exists if and only if $L < U$. For merely outer score enclosures, these inequalities remain sufficient, whereas $L \geq U$ need not establish infeasibility in a smaller realizable class.

Proof. For a desired pair, even the smallest admitted score must pass:

$$\ell_h - 1 + \beta a(n_h)^2 > 0.$$

Because $a(n_h)^2 > 0$, this is equivalent to

$$\beta > \frac{1 - \ell_h}{a(n_h)^2}.$$

Satisfying every desired pair therefore requires $\beta > L$. For an undesired pair, even the largest admitted score must fail:

$$u_h - 1 + \beta a(n_h)^2 \leq 0,$$

which is equivalent to

$$\beta \le \frac{1 - u_h}{a(n_h)^2}.$$

Satisfying every undesired pair therefore requires $\beta \le U$.

Monotonicity of gate slack in the semantic score makes these endpoint conditions sufficient over each interval. Exact interval constraints or endpoint attainability make them necessary. Their intersection is nonempty exactly when $L < U$. If $U < +\infty$, any $\beta \in (L, U]$ is feasible; if $U = +\infty$, any finite $\beta > L$ suffices. With outer enclosures only, the same conditions remain sufficient, while an unattainable endpoint cannot support the converse. □

Global gate design, therefore, reduces to the interval condition $L < \beta \le U$, where the lower bound admits every desired pair and the upper bound rejects every undesired pair. If $L \ge U$, the desired acceptance and rejection requirements conflict within the declared class. Rescaling $\gamma = \beta W^2$ cannot repair this incompatibility because it does not change β.

A desired zero-descendant pair is sharper still. With semantic score at most one, no finite gate strength can make it pass. Zero-descendant undesired pairs are already rejected and impose no upper-strength constraint. The theorem concerns eligibility and not ranking. An undesired candidate may pass without becoming the winner, and an eligible parent set may still exceed the capacity of a top-k decoder.

E.7 The Limit of Compensation

Compensation removes remote-range drift by holding β fixed. It can also expose a semantically stronger incorrect rival that the original gate happened to suppress. Consider valid scalar score inputs with

$$W = 10, \qquad \gamma_0 = 5, \qquad W_0 = 1,$$

a correct parent with $(s_h, n_h) = (0.95, 4)$, and an incorrect rival with $(s_r, n_r) = (0.97, 1)$. The original strength is $\beta = 0.05$. Since $\log_2 3 > 1$. The target slack is positive. Also, $\log_2(3/2) < 0.6$, so the rival slack is less than $-0.03 + 0.05(0.6)^2 < 0$. The original rule, therefore, selects the target. Compensation gives $\beta_0 = 5$, accepts both candidates, and then the larger semantic score $0.97 > 0.95$ makes the rival win. These scores are realizable as unit-vector similarities. This is a scorer-interface counterexample and not an additional compiled taxonomy history. The counterexample shows that eligibility, selectivity, and correct ranking are distinct design objectives. Within the separate fixed-target growth construction, compensation can preserve the correct ranking even when no global strength accepts the target and rejects the specified rival, because the target's larger semantic score resolves the decision after both pass. A gate can therefore stabilize eligibility without solving the complete recommendation problem.

F Specific Scorer Formula Correspondence

The eight scoring-operation classes describe mathematical operations. They are not a complete set of published systems. To connect the analysis to concrete recent methods, this section instantiates three scoring formulas whose required inputs and output rules can be stated exactly. The correspondence is formula-level: `TaxoAtlas` supplies declared inputs to the corresponding operation, and the preceding theory characterizes that operation under those inputs. It does not identify the resulting calculation with the output of an unobserved trained representation, checkpoint, retrieval stage, or end-to-end decoder.

I. FUSE (Conditional overlap as a bilinear readout). FUSE (Xu et al., 2024) uses a conditional-overlap score over membership vectors in $[0, 1]^d$, with positive query measure and shared measure weights. For a `TaxoAtlas` field p_r, choose

$$u_r = \sqrt{p_r},$$

coordinatewise, and use uniform positive weights. Since p_r is a probability vector,

$$u_r \in [0, 1]^d, \qquad \|u_r\|_2 = 1.$$

The resulting supplied-input score is

$$s(q,h) = \frac{\langle u_q, u_h \rangle}{\sum_x u_q(x)}. \tag{72}$$

For a fixed query, the denominator is common to every candidate. Candidate ranking is therefore determined by

$$\langle u_q, u_h \rangle = 1 - \frac{1}{2}\|u_q - u_h\|_2^2.$$

The declared FUSE readout consequently falls under the bilinear/geometric transfer analysis: a bound on square-root-field movement yields a bound on its numerator, while numerical threshold decisions must additionally retain the positive query denominator. This is an exact identity for the supplied membership vectors. It does not claim that the `TaxoAtlas` square-root field reproduces FUSE's learned fuzzy-map representation or the output of its complete trained system. Such a transfer would require the separate fidelity condition of Equation (48).

II. TaxoBell (Moment fields as Gaussian parameters). TaxoBell (Mishra et al., 2026a) evaluates separate Gaussian ranking variants using a box center as the Gaussian mean and squared offset as variance. The `TaxoAtlas` moment readout in Appendix C.1 supplies quantities of the same mathematical type:

$$\mu_r = \mathbb{E}_{p_r}[x], \qquad v_r = \lambda + \mathrm{Var}_{p_r}(x), \qquad \lambda > 0.$$

For one coordinate, the directed KL score is determined by

$$D_{\mathrm{KL}}(q\|h) = \frac{1}{2}\left[\log\frac{v_h}{v_q} + \frac{v_q + (\mu_q - \mu_h)^2}{v_h} - 1\right], \tag{73}$$

while the Bhattacharyya distance is

$$D_{\mathrm{B}}(q,h) = \frac{(\mu_q - \mu_h)^2}{4(v_q + v_h)} + \frac{1}{2}\log\frac{v_q + v_h}{2\sqrt{v_q v_h}}. \tag{74}$$

The corresponding coefficient is

$$\exp[-D_{\mathrm{B}}(q,h)].$$

These are distinct ranking variants, not terms that should be summed into one score. Both depend on center and spread. The variance floor and moment-response bounds in Appendix C.1 therefore supply exactly the additional quantities required by the declared diagonal-Gaussian analysis. The correspondence is again at the formula interface. It does not identify `TaxoAtlas` moments with TaxoBell's learned box parameters or claim a theorem for arbitrary learned covariance structure.

III. Polaris: semantic score multiplied by structural eligibility. For Polaris (Mishra et al., 2026b), the published scoring expression uses an angular factor given by the dot product divided by the product of vector norms plus 10^{-8}. With unit `TaxoAtlas` square-root-field directions,

$$a_h = \frac{\langle u_q, u_h \rangle}{1 + 10^{-8}}. \tag{75}$$

The scoring rule combines this semantic term with a count/range-dependent indicator:

$$S_h = a_h \, \mathbf{1}\left\{ a_h - 1 + \gamma\left[\frac{\log_2(1 + n_h/2)}{W}\right]^2 > 0 \right\}. \tag{76}$$

Here n_h is the current strict-descendant count. In the two controlled numerical realizations, $\gamma = 5$, and

$$W = \max_j \left[d_j + \log_2(1 + n_j)\right] - \min_j \left[d_j + \log_2(1 + n_j)\right],$$

where the range is taken over all currently present graph nodes j, including nodes outside the scored root-candidate set. Roots have depth one, and a zero range is replaced by one.

Equation (76) exposes the two dependencies analyzed in F8. The semantic score a_h depends on query-candidate compatibility, while eligibility additionally depends on the candidate's descendant count and on the global structural range W. Taxonomy growth can therefore change the complete

score even when the semantic vectors remain fixed. Appendix E characterizes this gate directly, and Appendix D.1 gives a finite history in which the semantic order remains unchanged while growth in W flips the selected parent at an analytically predicted arrival. In our controlled realization, structural statistics are recomputed after each admitted insertion and the declared positive set cutoff is applied. These choices define our controlled formula realization and are not claims about a deployed Polaris system or trained checkpoint.

The shared-field realization scores root candidates through `TaxoAtlas`-derived angular evidence and the refreshed structural gate. The separate source-angle realization retains the declared source angles for all nodes in the constructed graph and applies the same initial cutoff contract. These are two controlled realizations of the same inspected scoring operation and not two evaluations of a trained Polaris model.

In the multi-parent acquisition construction, the source-angle realization returns

$$B \quad \text{at postbirth step } 1,$$
$$\{B, C\} \quad \text{at steps 2-26,}$$

and

$$C \quad \text{at steps 27-40.}$$

The independently defined target after birth is {B,C}. Hence, the final phase loses the still-required B link; returning C alone yields an incomplete parent set rather than selecting an entirely incorrect parent. The separate fixed-target construction in Appendix D.1 keeps B as the unique parent throughout and therefore studies a different question. The two histories are not pooled.

Scope of the correspondences. The three mappings have a deliberately narrow interpretation: a published mathematical operation becomes analyzable only after its required `TaxoAtlas` inputs and output rule are declared. For FUSE, the exact supplied-input identity exposes a bilinear/geometric ranking operation. For TaxoBell, moment fields instantiate the mean-and-variance arguments of the declared Gaussian variants. For Polaris, the declared expression exposes the interaction between semantic compatibility and structural eligibility. The conclusions therefore concern the displayed operations under the stated representation maps, update rules, structural records, and decoders. They do not imply that every implementation bearing the same method name satisfies the corresponding `TaxoAtlas` contract, nor that the controlled calculations reproduce a trained checkpoint. The broader method census in Appendix H identifies related operations in additional systems without promoting those stage-level similarities to formula-level correspondence.

G Dataset Protocols and Evidence

G.1 Theory-to-Evidence Coverage

The empirical evidence has three roles. Formal results establish what can and cannot be guaranteed. Controlled histories show that the characterized limiting cases are attainable. Observed-taxonomy replays test whether the corresponding kinds of change occur when real taxonomy structure and supplied semantic vectors are used. No empirical protocol substitutes for a theorem, and the protocols are not pooled into one benchmark score.

G.2 Matched Conditional Ranking Protocol

The matched ranking study isolates candidate-set growth from retrieval failure. The same held queries are evaluated in a fixed taxonomy and in its evolved state. At each state, the parent-complete neighborhood contains every reference parent and is then filled with the closest F1 candidates until the nominal neighborhood size is reached. Ties at the boundary are retained, so the realized candidate set may exceed the nominal size. This protocol grants *access* to every correct parent; it does not identify which candidates are correct. A degradation after growth therefore cannot be attributed to a correct parent being absent from the neighborhood. F1 to F6 and F8 use the declared formulas in Appendices C. F7 is excluded because it scores complete parent sets instead of individual candidates. Optimistic Hit@1 counts a top-score tie as correct if it contains any correct parent. Tie-safe Hit@1 applies the declared deterministic tie rule. Lost and gained counts are paired query transitions. The F8 result gives the clearest paired open-world effect. On Environment, optimistic Hit@1

Table 7: Theory-to-evidence coverage. Each empirical protocol tests a distinct consequence of the characterization.

Question	Mathematical support	Controlled evidence	Observed grounding
Does the reference preserve the distinctions needed by a scorer?	Conditional semantic correspondence and adapter enclosures	Shared source-to-taxonomy construction	Environment, Science and Food field profiles
Can an old valid parent be lost after growth?	Margin survival and family transfer bounds	Normalization and gate obstructions	Matched F8 losses; Environment rival arrivals
Can a newly required parent be missed?	Acquisition as a separate obligation; family-specific feasibility conditions	Joint-set rule misses 40/40; gate misses the new root while structurally unsupported	Six-source acquisition replay
Can correct ranking still produce false links?	Threshold and selector conditions	Gaussian cutoff admits false links in 39/40 states	False-link counts in the six-source replay

Table 8: All ranked families in the nominal five-choice replay. Values are optimistic Hit@1 percentages before and after taxonomy growth. F8 begins from the strongest fixed-world operating point; F1-F6 are already near the floor and therefore offer little room for interpretable degradation.

	F1	F2	F3	F4	F5	F6	F8
Environment fixed	2.0	0.0	5.9	5.9	2.0	2.0	49.0
Environment evolved	5.9	5.9	3.9	3.9	7.8	7.8	21.6
Science fixed	1.9	0.0	0.0	0.0	1.9	1.9	37.0
Science evolved	1.9	1.9	1.9	1.9	0.0	0.0	31.5

falls from 49.0% to 21.6%. On Science it falls from 37.0% to 31.5%. These are not unpaired differences: Environment contains 20 lost and 6 gained queries, while Science contains 13 lost and 10 gained queries at the nominal five-choice setting. The sweep contains 56 dataset-size-family settings, 2 taxonomies $\times$ 4 neighborhood sizes $\times$ 7 ranked families, and 2,940 query-level comparisons. The reported MRR, rank changes, tie-safe outcomes, and paired transitions are computed from these comparisons.

G.3 Environment Formula Replay

The Environment replay asks a more local question: can an initially correct decision be disrupted when the query representation and old correct-parent score remain fixed? The visible Environment/EuroVoc bundle contains 221 concepts and 222 occurrences with fixed supplied semantic vectors. A deterministic partition gives 82 held single-parent leaf queries, 98 baseline concepts and 41 distinct arriving leaves. Every held query's correct parent remains available throughout. The target itself never changes. This diagnostic partition is separate from the published benchmark split. Three supplied rules are evaluated: angular similarity, the declared complete structural gate, and a compensated version that keeps the gate's normalized coefficient fixed as the global range changes. Graph statistics are refreshed after every arrival while the semantic vectors remain fixed. The angular replay gives the cleanest candidate-competition example. Four initially correct queries are lost when an arriving candidate outranks the unchanged correct parent. For *plant resources*, the correct parent *natural resources* remains at approximately 0.731, while arriving *animal resources* overtakes it by only 2.80×10^{-4}. The other lost queries are *national park*, *erosion*, and *Norwegian Sea*. The old score did not need to deteriorate; the competitive set changed around it.

The gated replay exposes a different trajectory effect. For *metal pollution*, the correct parent *pollution* continues to pass the gate, while rival *pollutant* becomes eligible and has the larger score during steps 19-36. A later normalization change rejects the rival again, so the initially correct decision is lost and subsequently recovered. Compensation keeps the rival eligible and makes the error persist. Final-state correctness and uninterrupted preservation are therefore distinct. The compensated rule

Table 9: F8 as the nominal candidate neighborhood grows. Each dataset uses the same paired queries at every size. A sufficiently difficult fixed-world candidate set can create a floor that hides additional open-world degradation.

Taxonomy	Choices	Fixed Hit@1	Evolved Hit@1	Lost	Gained
Environment	5	49.0%	21.6%	20	6
Environment	10	27.5%	7.8%	12	2
Environment	20	5.9%	5.9%	2	2
Environment	40	5.9%	5.9%	2	2
Science	5	37.0%	31.5%	13	10
Science	10	7.4%	7.4%	4	4
Science	20	3.7%	1.9%	2	1
Science	40	3.7%	1.9%	2	1

Table 10: Environment supplied-vector replay on 82 held queries. Each retention denominator is the rule's own initially correct cohort. These are formula diagnostics instead of trained-method results.

Rule	Initial correct	Final correct	Initial cohort correct at end	Initial cohort never lost
Angular	24	20	20/24	20/24
Gated angular	8	13	8/8	7/8
Compensated gate	8	16	7/8	7/8

improves the total number of final correct queries while losing one member of its initially correct cohort.

Arrival order changes the trajectory without changing the endpoint scores. The order controls use the primary declared order, its reverse, and 32 random orders. Their final scores are identical. Angular initial-cohort losses remain four and compensated-gate losses remain one. Gated losses vary between zero and one. These controls mirror the pathwise theory of Appendix A.6: the same endpoint can be reached through growth histories with different intermediate preservation behavior. The 34 orders are declared as controls rather than samples from a claimed real-world arrival distribution.

Structure and semantics are separately necessary for these supplied rules. Two controls isolate their roles. With graph structure alone and unrestricted admissible semantic scores, 35 correct parents have no visible descendants and receive a gated score of zero. Other zero-scored candidates prevent a strict unique-winner certificate. Across all 82 probes, the broader structure-only interval analysis certifies no strict correct winner. This is an inconclusive certificate and not a claim that every structure-based method fails. Conversely, keeping the graph fixed while permuting vectors across identities destroys much of the original semantic alignment. Across 32 fixed permutations, initial angular correctness ranges from 0 to 3, with median 0, compared with 24 under the original assignment. Gated correctness ranges from 0 to 6, with median 2, compared with 8. For these supplied scoring formulas, the concept-to-vector assignment matters. It does not establish that a `TaxoAtlas` field statistic predicts trained-model performance. A fine angular adapter with $2^{32} - 1$ intervals encloses all 8,036 baseline pair scores and certifies the source-formula classifications. This is a sufficiency result for the retained fine angular record, not a general claim about coarse coordinates or learned representations. This replay evaluates the supplied-vector scoring formulas rather than `TaxoAtlas` field inputs.

G.4 Full-Source Descriptive Profiles

The full-source compilation asks whether the declared reference quantities can be constructed at larger observed-taxonomy scale. It is descriptive and not a parent-prediction experiment. The normalized Environment, Science and Food sources contain the coverage summarized below. The complete Environment input has 261 concepts and is distinct from the 221-concept visible bundle used in the preceding formula replay. Concepts may have multiple structural occurrences. Coordinate probes inspect declared field positions within local coordinate systems; they are not held-out query placements. Environment and Science use exact field and concentration calculations where

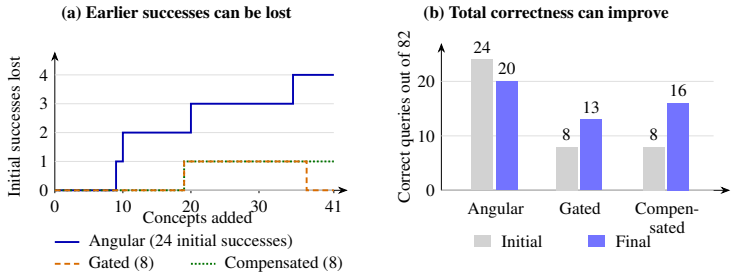

Figure 3: Environment replay with 82 queries and 41 arrivals. (a) Losses among initially correct outputs. (b) Endpoint totals; compensation gains three final successes while losing one earlier success. Supplied-vector formulas.

Table 11: Full-source descriptive field coverage. Counts establish construction coverage rather than scorer accuracy. Food uses the approximate field readout described below.

Source	Concepts	Occurrences	Spectra	Coordinate systems	Probes
Environment	261	262	94	33	162
Science	429	453	180	62	307
Food	1486	1577	423	120	597

available. Food uses the completed full-lattice readout from exact spectra. Across 469 completed Environment/Science reference probes, the floating readout differs from the exact computation by at most 5.33×10^{-15} bits in entropy and 1.82×10^{-12} in support-scaled concentration. These discrepancies are descriptive numerical checks, not certified error bounds for Food or for threshold decisions. Source sizes, supports, and construction stages differ, so the profiles are not used to claim that one taxonomy is intrinsically harder than another.

G.5 Parent Acquisition on Six Source Taxonomies

The six-source replay tests the empirical distinction between retaining an old parent and acquiring a newly available one. The protocol uses six normalized sources, 64 held queries and 32 separate calibration concepts per source. All 96 concepts are excluded from the candidate taxonomy. Three arrival orders and five snapshots use fixed supplied semantic vectors. A query's reference parents at a snapshot are exactly its asserted source parents that are available at that time. An unavailable parent does not induce a replacement shortcut edge. The replay therefore measures acquisition through parent availability, not retirement of an old direct parent after refinement. Three supplied rules are evaluated: cosine similarity, the declared structural gate, and a control retaining the initial normalization range. For each rule, one threshold is selected on the initial calibration panel only, maximizing micro-F1 over $-1, -0.95, \ldots, 1$, with ties selecting the highest threshold. No later query label retunes it. This thresholding procedure is an experimental diagnostic and not a deployed decoder of a named method. The aggregate contrast is sharp. Cosine retains 72/72 initially correct links while acquiring only 17/89 newly available required parents. The gated rule retains 36/37 initially correct links and acquires 13/89. Near-perfect retention therefore does not imply strong acquisition. The source-level results show the same distinction. On Environment, cosine retains 20/20 initial links and acquires 6/13 new links while returning 124 false links. The structural gate retains 7/7, acquires 2/13, and returns 22 false links. On WordNet Noun, the gate retains 5/5 initial links and acquires none of the 17 newly available parents. These counts separately measure retention, acquisition, and false-link control. They do not attribute every missed link to taxonomy growth because supplied vectors and an initially calibrated threshold may already be imperfect. The replay covers 384 distinct evaluation concepts. Repeated arrival orders are treated as dependent replays rather than independent queries. The reported retention, acquisition, missed-link, and false-link counts use the reference parent sets determined by the source partitions and currently available candidates. WordNet Verb uses a frozen repaired source with 13,767 concepts and 13,238 edges. One cycle-closing edge is excluded by the declared deterministic traversal before partitioning; the raw source remains

Table 12: Retention and acquisition on six source taxonomies. Final counts use the primary declared order. Each retained fraction uses that rule's initially correct-link denominator; each acquired fraction uses all newly available reference links. False-link counts are not normalized across candidate universes of different sizes.

	Cosine			Structural gate		
Source	**Retained**	**Acquired**	**False**	**Retained**	**Acquired**	**False**
Environment	20/20	6/13	124	7/7	2/13	22
Science	14/14	1/11	94	5/5	1/11	20
Food	16/16	4/17	359	8/8	7/17	96
MeSH	3/3	1/17	44	5/5	1/17	32
WordNet Noun	10/10	2/17	691	5/5	0/17	765
WordNet Verb	9/9	3/14	879	6/7	2/14	183

Table 13: Closest conceptual precedents and the additional object characterized here.

Precedent	**Studied object**	**Present analysis**
TaxoOrder (Song et al., 2021)	Insertion order and dependencies within a supplied expansion batch	Given growth histories; retention, acquisition, and exact parent-set consequences of complete scores and selectors
Temporal construction and TaxoAdapt (Tuan et al., 2016; Kargupta et al., 2025)	Changing evidence or corpus-driven taxonomy maintenance	Declared taxonomy updates followed through reference, score, and parent-set decision
ICON (Shi et al., 2024)	Implicit concepts, parent/child insertion, and structural update	Direct-parent recovery; child selection and unrestricted joint graph repair remain outside the guarantee
Ranking regions (Asudeh et al., 2018; Chen & Jagadish, 2026)	Ranking sensitivity to changes in scoring weights over specified items	Taxonomy-induced representation changes, candidate births, and complete parent-set selection
Margin certification (Tsuzuku et al., 2018)	Decision margin under a supplied perturbation bound	Taxonomy-induced perturbation derived first, then transferred to the complete decision rule
Graphon transfer (Ruiz et al., 2020)	Transfer across graph sizes under common graphon, signal, and filter assumptions	Explicit evolving-taxonomy reference and selector conditions; no generic GNN-transfer claim

unchanged. The reported results therefore concern this repaired source instead of asserting that the original graph is acyclic.

H Related Methods and Scoring Operations

Closest Conceptual Comparisons The closest precedents address different parts of the problem studied here: taxonomy evolution, ranking stability, perturbation margins, and transfer across changing graph structure. Table 13 compares their stated analysis objects with the object characterized in this paper. The comparison concerns assumptions and mathematical scope and not performance or overall method capability. The elementary ingredients-metric stability, margin survival, Lipschitz transfer, and normalization bounds-are standard tools. The present analysis composes them along a different chain:

$$\text{taxonomy growth} \;\to\; \text{reference change} \;\to\; \text{complete-score change} \;\to\; \text{parent-set decision}.$$

Coverage of the Eight Scoring-Operation Classes Recent and earlier taxonomy systems use a broad range of scoring mechanisms. Prior taxonomy systems include angular and metric comparisons, bilinear or overlap scores, distributional comparisons, geometric regions, learned neural

readouts, token-generation scores, joint configuration objectives, and explicit structural or statistical rules. These correspond to the eight operation classes analyzed in this paper. Representative examples make the connection concrete. FUSE uses fuzzy-membership overlap (Xu et al., 2024); TaxoBell studies Gaussian distributional scores (Mishra et al., 2026a); Polaris combines angular compatibility with structural statistics (Mishra et al., 2026b); and recent neural and generative approaches use contextual readouts or token-level decisions. Appendix F gives exact formula-level correspondences for selected methods where the required mathematical contracts can be specified. The correspondence is deliberately operation-level. The presence of a distance, Gaussian, neural, or gated component in a published method does not imply that the complete system satisfies the assumptions of the corresponding theorem. Transfer additionally requires the representation map, update law, relevant parameters, feasible output set, and selector to satisfy the declared contract.

This distinction is important because different stages of an end-to-end system may play different roles. A training regularizer need not be the inference score. Retrieval used to construct a prompt need not restrict which outputs a generator may legally produce. A candidate ranking does not determine how many parents the final system returns. Stage-level similarity therefore cannot be promoted automatically to an end-to-end guarantee.

Scope of the Literature Mapping. The resulting literature claim is intentionally bounded: operations represented by F1 to F8 recur in existing taxonomy systems. It does not claim that every system belongs uniquely to one family, that every implementation satisfies the required `TaxoAtlas` correspondence, or that the mapping is an exhaustive novelty survey. The stronger statements in this paper are therefore made only where the required contract is explicit. Appendix F instantiates selected published formulas, while the controlled constructions specify the complete mathematical operation and selector directly. This keeps the relationship to existing methods informative without treating a broad architectural resemblance as theorem coverage.

I Reproducibility and Claim Scope

The evidence in this paper has several reproducibility levels, reflecting how each result was obtained. Controlled constructions are reconstructible directly from their mathematical specification. The matched fixed/open-world study is reconstructible from query-level outcomes. Observed-taxonomy formula replays additionally depend on frozen taxonomy and semantic artifacts. Full-source `TaxoAtlas` profiles provide conformance checks for saved reference objects rather than independent recompilation of every atlas state.

Controlled Constructions The controlled constructions fix the taxonomy histories, source geometry, reference construction, scoring formulas, selectors, and tie rules before evaluation. The acquisition history contains one initial state followed by 40 postbirth states; the preservation construction contains 201 states and 200 transitions. Reconstruction reproduces the principal controlled findings, including 200/200 joint-response distance certificates, 195/200 joint-response Gaussian certificates, and 192/192 before-reveal source-law checks across the two declared catalogues. These quantities test different claims. The preservation construction measures certificate coverage while the realized decision remains correct. The acquisition construction changes the required parent set and measures retention, acquisition, false-link control, and exact recovery. The before-reveal checks validate prospective reference-state envelopes over the declared candidate arrivals. These are deterministic calculations on finite specified histories. Their reproducibility does not depend on model training, random initialization, or external inference services.

Matched Fixed/Open-World Study. The conditional-ranking study is retained at query level. Its aggregate result contains 56 rows, 2 taxonomies $\times$ 4 neighborhood sizes $\times$ 7 rankable scoring-operation classes, and 2,940 query-level records. The reported optimistic Hit@1, tie-safe Hit@1, MRR, and paired loss/gain statistics are reconstructible from these records. Every correct parent is inserted before the remainder of the candidate neighborhood is filled. A ranking loss in this protocol therefore cannot be attributed to omission of the correct parent from retrieval.

Observed-Taxonomy Replays The Environment sequential diagnostic and the six-source acquisition replay use observed taxonomy structure together with frozen supplied semantic vectors. Reproducing these studies therefore requires both the declared scoring logic and the corresponding

frozen source artifacts. For Environment, the replay reconstructs the graph snapshots, scorer outputs, arrival-order controls, and semantic-permutation controls reported in Appendix G.3. These are supplied-formula diagnostics rather than executions of a trained end-to-end taxonomy-expansion system. The six-source replay covers Environment, Science, Food, MeSH, WordNet Noun, and the declared repaired WordNet Verb source. It uses frozen source partitions, candidate-availability schedules, supplied semantic vectors, initial-only calibration, and fixed selectors. Its principal aggregate results-72/72 retained versus 17/89 acquired links for cosine, and 36/37 retained versus 13/89 acquired for the structural gate-are reconstructible from the replay records. False-link counts are reconstructed separately and are not pooled across sources with different candidate universes. A complete rerun additionally requires the corresponding official taxonomy files and frozen semantic artifacts. The released records support verification of the formulas and reported aggregation at the scope claimed here.

Full-Source `TaxoAtlas` Conformance The Environment, Science, and Food full-source artifacts verify that the declared `TaxoAtlas` quantities can be materialized at larger scale. Saved spectra, fields, coordinate records, and descriptive summaries are checked for the corresponding normalization and consistency properties. These checks establish conformance of the saved reference calculations. They are not equivalent to independently recompiling every `TaxoAtlas` state from the original taxonomy and semantic sources. The reported Food result uses the completed declared readout.

Named Formula Correspondence The correspondences in Appendix F are reproducible at the mathematical-operation level. For selected methods, the released materials evaluate the displayed formula on the declared supplied representation.

$$\text{declared inputs} + \text{declared operation} \longrightarrow \text{declared score}.$$

It does not turn the controlled realization into an evaluation of a trained checkpoint or imply that every implementation associated with the same method name satisfies the same correspondence.

Claim Boundary The reproducibility boundary follows the scientific boundary. The released materials support the declared chain from taxonomy state through reference, scoring operation, selector, and diagnostic outcome at the scope stated in the paper. They do not establish child selection, unrestricted joint graph repair, arbitrary learned-representation correspondence, or a universal acquisition-delay distribution without an arrival law. The pathwise survival result applies to its declared arrival process, and the field-response and opportunity results retain their stated construction assumptions. The completed six-source evidence concerns the declared supplied-vector cosine and structural-gate formulas. It is not a native eight-family-by-six-source evaluation.