GEOMETRIC REGULARIZATION OF LEARNED BEHAVIOR SPACES IN UNSUPERVISED QUALITY-DIVERSITY OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Discovering increasingly capable behaviors can require preserving solutions whose value only becomes apparent when later discoveries build on them. A central challenge is thus retaining these stepping stones without specifying in advance which behavioral differences will matter for future progress. Quality-diversity algorithms pursue this goal by maintaining a repertoire of distinct policies that the search can revisit and improve. For example, learning a latent representation of behavioral attributes directly from actions of a robot removes the need to specify the behavioral dimensions of any given policy that distinguish meaningfully different solutions from one another. Prior works, like AURORA, train autoencoders on action trajectories to find these important dimensions of variation. But autoencoder reconstruction loss alone does not uniquely determine the latent-space distances between potential solutions. Here, we introduce AURORA-VC, which adds variance and covariance penalties to online trajectory reconstruction. These penalties encourage variation across descriptor dimensions and discourage correlations between behavioral descriptor axes independently of the selected loss or fitness functions used later to assess solution quality. With both PGA-AURORA and AURORA, maximum fitness improves in four of ten tasks at an uncorrected paired $p < 0.05$, with task-dependent effects on repertoire quality. VC is associated with longer survival of high-fitness admissions and reuse of older parents. Survivor-selection interventions support selection as one contributor to the fitness gains. Together, these results show that descriptor learning shapes the search itself, determining which policies compete, survive, and remain available as stepping stones toward higher fitness.

1 INTRODUCTION

Many learning and optimization problems may benefit from access to several useful solutions, whose value depends on the conditions in which they are used. A robot, for example, can capitalize on alternative locomotion strategies when damage makes its usual gait ineffective (Pugh et al., 2016). Alternative behaviors also matter during learning: a behavior with modest current performance may provide a route to better future solutions that would otherwise remain undiscovered (Lehman & Stanley, 2011b). Discarding such behaviors solely because they perform worse in the short-term can therefore eliminate useful possibilities for future exploration and adaptation. These challenges motivate search methods that discover and improve a diverse collection of effective behaviors.

Quality-diversity (QD) algorithms retain an archive of high-performing, behaviorally distinct policies rather than focusing in on a single best policy (Pugh et al., 2015; 2016; Cully & Demiris, 2018). In rugged and deceptive loss landscapes, this repertoire of solutions preserves stepping stones and supports active exploration (Lehman & Stanley, 2011b;a; Mouret & Clune, 2015). A repertoire can also offer alternatives when goals or environmental conditions change. While the assessment of what qualifies a potential solution as high-performing is relatively intuitive given a well-defined loss function, determining what qualifies as a behaviorally distinct policy is less straightforward. The *behavior descriptor* maps a policy's behavior into the space where diversity is measured and strongly influences which solutions a QD algorithm prioritizes. Descriptor distance determines which policies count as behaviorally unique, which policies compete, and thus which parts of the search history

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.

remain available for future refinement and reuse. Hand-designed descriptors, the default in popular algorithms like MAP-Elites (Mouret & Clune, 2015), make these decisions interpretable but often require extensive task-specific knowledge and restrict search to distinctions chosen in advance instead of being able to respond to the specific context of the current policies explored during search.

Instead, AURORA trains an autoencoder on policy trajectories and uses its latent code as the behavior descriptor (Cully, 2019; Grillotti & Cully, 2022a). PGA-AURORA extends this approach to explore new policies found via both genetic algorithm mutations and policy-gradient updates (Nilsson & Cully, 2021; Chalumeau et al., 2023). Our primary experiments use Dominated Novelty Search (DNS) archive selection with both proposal mechanisms (Section 2), while an extension study with PGA-AURORA tests whether our claims extend to an adaptive-radius archive (Appendix H). In both methods, distances between learned descriptors guide archive selection and shape which policies remain available for further exploration. Learning the behavior descriptor removes the need to specify behavioral coordinates, but it does not remove the need to specify behavioral geometry. AURORA trains its latent representation for reconstruction and then uses Euclidean distance in the learned latent space to measure behavioral differences.

Reconstruction alone does not uniquely specify the Euclidean distances used to compare behaviors. In our reconstruction-only PGA-AURORA runs, descriptor variation concentrates in a few latent directions and high-fitness policies have short archive lifetimes, limiting opportunities for reuse (Section 7). This motivates explicitly shaping the geometry of the learned descriptor space.

Ideally, a learned archive representation would spread behavioral variation across its dimensions while limiting redundancy between them. In particular, coordinates should vary across trajectories without merely duplicating one another through linear correlations. VICReg (Variance-Invariance-Covariance Regularization) is a self-supervised representation-learning objective that encourages this kind of balanced, non-redundant representation (Bardes et al., 2022). Specifically, VICReg combines three components: invariance aligns representations of paired views of the same input; variance discourages individual dimensions from collapsing; and covariance discourages different dimensions from encoding redundant information. Motivated by this, we introduce AURORA-VC to condition the learned search metric. AURORA-VC retains trajectory reconstruction and adds VICReg's variance and covariance (V+C) components; it does not use the paired-view invariance term.

Reconstruction alone does not determine which policies are close to one another in the learned descriptor space. V+C regularization shapes this space by encouraging each latent dimension to vary across trajectories and penalizing covariance between dimensions. Although these penalties do not use fitness, they can change which higher-fitness policies are nearby competitors and the distances used to score each policy. This can change which policies enter the archive, how long they remain, and which are available as parents. The retained trajectories also train the next encoder, so changes in archive selection influence subsequent representations. Representation learning and search continually influence one another, making the effect of V+C depend on earlier archive decisions. We examine this through changes in archive selection, policy retention, and parent reuse.

Our contributions are:

- We show that reconstruction loss does not uniquely determine the Euclidean archive metric. A POINTMAZE case study demonstrates that changing descriptor distances can alter archive membership while preserving trajectories, fitness, and reconstruction (Section 3).
- We introduce AURORA-VC, which adds fitness-blind variance and covariance regularization to online trajectory representations, and demonstrate maximum-fitness gains with both AURORA and PGA-AURORA in a ten-task evaluation (Sections 4 and 6).
- We characterize changes in archive competition, turnover, survival, and parent reuse. Continuing to transfer VC-selected survivors yields positive paired median maximum-fitness AUC effects in seven of ten tasks relative to transferring only admission quotas (Section 7).

2 BACKGROUND AND RELATED WORK

Archive descriptors and evaluation. A policy θ produces fitness $f(\theta)$, trajectory τ_θ, and behavior descriptor $b(\theta)$. We call the learned coordinates the *active descriptor*. For evaluation, a fixed *passive archive* accumulates selected active members and retains the best in each cell defined by

a handcrafted descriptor. It reports coverage and QD score but never supplies parents or controls active admission. Maximum fitness is also independent of the passive coordinates.

Fitness-conditioned archive competition. Dominated Novelty Search prioritizes policies by their mean distance to their nearest strictly fitter neighbors, making descriptor geometry central to archive competition (Bahlous-Boldi et al., 2025). Our implementation retains the highest-scoring policies up to a fixed capacity; the formal rule and tie handling appear in Appendix A.2 (Equation 5).

Learning and structuring behavior spaces. Prior work learns behavioral distinctions from observations, maintains multiple learned views, introduces downstream relevance, or quantizes the repertoire (Meyerson et al., 2016; Paolo et al., 2020; Cazenille, 2021; Grillotti & Cully, 2022b; Tsakonas & Chatzilygeroudis, 2026). These methods primarily change which distinctions the representation expresses. We ask what constrains the Euclidean geometry when the learned coordinates themselves define competition.

Autoencoders can reconstruct the same data using different latent coordinate systems, and these coordinates need not align with meaningful behavioral factors (Rippel et al., 2014; Kunin et al., 2019; Locatello et al., 2019). Some methods use the decoder or explicit geometric objectives to constrain local distances (Arvanitidis et al., 2018; Chen et al., 2020; Nazari et al., 2023). VICReg, Barlow Twins, and whitening methods instead shape the overall spread and correlation structure of the learned representation (Bardes et al., 2022; Zbontar et al., 2021; Ermolov et al., 2021). A *fitness-informed* representation objective uses fitness values or comparisons between them to guide representation learning. For example, AURORA-XCON organizes descriptors using triplets selected by fitness similarity; its triplet loss contains no explicit variance or covariance penalty (Coiffard et al., 2025). The V+C penalties use latent variance and covariance without fitness targets (Bardes et al., 2022); their training data still depend indirectly on fitness through archive selection. They impose no explicit requirement that policies with similar fitness should be close in descriptor space. We add variance and covariance penalties to trajectory reconstruction to test whether shaping descriptor geometry can improve search without adding fitness supervision to the encoder.

3 Reconstruction Does Not Define the Archive Metric

In the reconstruction-only autoencoder used by AURORA and PGA-AURORA, the autoencoder learns a behavior descriptor by encoding each policy trajectory τ_i and decoding it into a reconstruction $\widehat{\tau}_i$. The implementation averages squared prediction errors over B normalized trajectories with T sampled steps and p features, using teacher forcing:

$$\mathcal{L}_{\mathrm{rec}}(\phi, \psi) = \frac{1}{B(T-1)p} \sum_{i=1}^{B} \sum_{t=2}^{T} \|o_{i,t} - \widehat{o}_{i,t}\|_2^2. \tag{1}$$

The archive uses the final encoder cell state; the decoder also receives its hidden state and preceding observations (Appendix A). The following functional result isolates why reconstruction alone cannot identify descriptor distances. However, the reconstruction loss does not uniquely determine how encoded trajectories are arranged in the latent space. The archive uses Euclidean distances between outputs of E to determine which policies compete and are retained, so two encoder–decoder pairs can reconstruct trajectories with equal reconstruction errors while still producing different archive decisions.

Proposition 1 (Reconstruction does not identify Euclidean code distance)**.** *Let $E : \mathcal{T} \to \mathbb{R}^d$ and $D : \mathbb{R}^d \to \mathcal{T}$. For every invertible $A \in \mathbb{R}^{d \times d}$, define $E_A(\tau) = AE(\tau)$ and $D_A(z) = D(A^{-1}z)$. Then $D_A(E_A(\tau)) = D(E(\tau))$ for every τ, while*

$$\|E_A(\tau_i) - E_A(\tau_j)\|_2^2 = (z_i - z_j)^{\mathsf{T}} A^{\mathsf{T}} A (z_i - z_j). \tag{2}$$

Proof. Substitution gives $D_A(E_A(\tau)) = D(A^{-1}AE(\tau)) = D(E(\tau))$; expanding the transformed norm gives Equation 2. □

Scaled orthogonal maps $A = cQ$ preserve every neighbor ordering. A general invertible map changes $A^{\mathsf{T}}A$ and can change fitter-neighbor sets, priorities, and the retained repertoire.

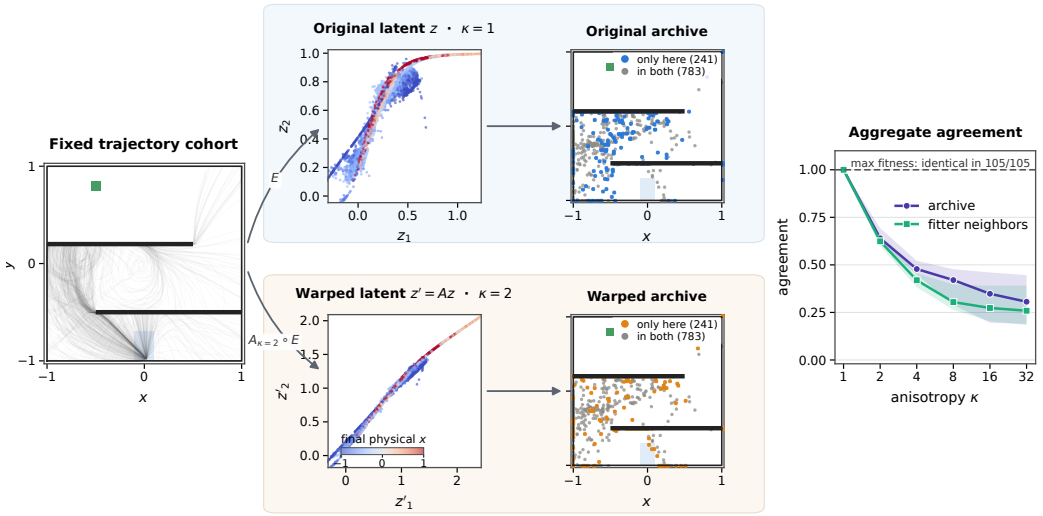

Figure 1: **The same reconstruction can define a different archive.** A fixed pool of 8,192 PointMaze trajectories (*left*) is encoded as learned latent descriptors z (*top*) and warped descriptors $z' = Az$ (*bottom*). Latent axes are not physical positions; colors indicate each trajectory's final physical x coordinate. The maze panels show physical trajectories and retained endpoints. Compensating for the warp in the decoder preserves reconstruction up to numerical error, yet selection changes 241 of the 1,024 retained policies. Stronger warps (larger κ) reduce overlap with the original archive (purple) and agreement between fitter-neighbor sets (green) (*right*). Lines and bands show medians and interquartile ranges across five encoders and four orientations.

Figure 1 demonstrates this phenomenon empirically. Five independently trained two-dimensional autoencoders encode the same policies, trajectories, returns, targets, and candidate order. We apply $A_\kappa = R\,\mathrm{diag}(\kappa, 1/\kappa)R^\mathsf{T}$ and compose the decoder with its analytical inverse. The $\kappa = 1$ control reproduces every decision; reconstruction is preserved up to the declared numerical tolerance, while anisotropic conditions change nearest-fitter identity and the selected repertoire. Maximum fitness remains invariant because the best fixed policy has no fitter competitor. Appendix B provides a detailed analysis of V+C regularization and archive selection in PGA-AURORA-VC, including mathematical proofs and the full POINTMAZE study.

Online AURORA repeats this decision in a feedback loop: retained trajectories train the next encoder, its metric governs competition, and selection determines future parents and training data.

4 AURORA-VC: GEOMETRIC CONDITIONING FOR LEARNED DESCRIPTORS

We introduce AURORA-VC to condition the coordinates used for archive selection. For a trajectory τ_i, let $z_i = E_\phi(\tau_i)$. Given a minibatch of embeddings $Z \in \mathbb{R}^{B\times d}$, define $\widetilde{Z} = Z - \mathbf{1}\bar{z}^\mathsf{T}$ and $\Sigma_B = \widetilde{Z}^\mathsf{T}\widetilde{Z}/\max(B-1, 1)$. AURORA-VC optimizes

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \lambda_V \mathcal{L}_V + \lambda_C \mathcal{L}_C, \tag{3}$$

$$\mathcal{L}_V = \frac{1}{d}\sum_q \max\left(0, \gamma - \sqrt{(\Sigma_B)_{qq} + \epsilon}\right), \qquad \mathcal{L}_C = \frac{1}{d}\sum_{q\neq r}(\Sigma_B)_{qr}^2. \tag{4}$$

Our default setting uses $d = 10$, $\gamma = 1$, $\epsilon = 10^{-4}$, and $\lambda_V = \lambda_C = 1$. As this borrows the variance and covariance components from VICReg (Bardes et al., 2022), we call this VICReg-style conditioning. At initialization and scheduled updates, we train with this loss on retained trajectories, re-encode the archive, and rebuild its priorities (Appendix C).

The $\mathcal{L}_V$ and $\mathcal{L}_C$ penalties address complementary failures. The variance term ($\mathcal{L}_V$) penalizes low marginal scale but does not discourage correlated coordinates; the covariance term ($\mathcal{L}_C$) penalizes off-diagonal linear covariance without imposing marginal scale. Together, they discourage contracted or linearly correlated coordinates, but finite penalties do not algebraically exclude collapse. They constrain second moments without identifying a unique useful metric, guaranteeing local distances, or aligning successive encoders (Appendix B).

5 EXPERIMENTAL DESIGN

We test whether the Variance-Covariance regularization helps learned-descriptor QD discover fitter policies and how it affects the breadth and quality of the discovered repertoire. We study ANTMAZE, ANTOMNI, ANTTRAP, ANTUNI, HALFCHEETAH, HUMANOIDOMNI, HUMANOIDUNI, POINTMAZE, and WALKER from the QDax/Brax task suite (Chalumeau et al., 2024; Freeman et al., 2021), together with KHEPERAX (Grillotti & Cully, 2023). The primary comparison pairs PGA-AURORA and PGA-AURORA-VC at ten matched seeds per task; a second AURORA–AURORA-VC comparison tests transfer without policy-gradient proposals. Both use DNS-style selection; an adaptive-radius control tests transfer beyond that rule. Each canonical run executes $5{,}000$ iterations with 512 proposals per iteration, retains at most $1{,}024$ active policies, and periodically retrains an autoencoder after warm start, using its ten-dimensional latent representations of trajectories as behavioral descriptors for the archive. Appendices A and D specify the environments, optimization, compute, seeds, and cohort validation.

Outcomes and inference. We measure maximum fitness, coverage, and QD score in a fixed 1,024-cell passive CVT archive accumulated from selected active members. A separate tracker records all evaluated candidates where explicitly reported. Coverage is the occupied-cell fraction; QD score sums the best fitness per occupied cell plus a task-specific offset (Appendix D). The reporting descriptors follow Chalumeau et al. (2023); Coiffard et al. (2025), making repertoire outcomes comparable across learned spaces. Since max fitness is the only metric that does not depend on this handcrafted reporting descriptor, it is our primary evaluation metric of interest.

Tables report method medians; two-sided Wilcoxon signed-rank tests use the ten paired seed outcomes. We report raw per-task p values, with paired uncertainty and Holm sensitivity in Appendix E. For KHEPERAX, which reaches the fitness ceiling, optimization comparisons use evaluations to first goal. Appendix E provides full results for both algorithm pairs; Appendix F adds contrastive-fitness and archive-extinction comparisons. Section 6 presents the performance comparisons and sensitivity studies; Section 7 examines how changes in the representation affect archive selection and subsequent search.

6 RESULTS

6.1 VC IMPROVES OPTIMIZATION WITH BOTH PROPOSAL MECHANISMS

Across the 20 task–algorithm settings with and without policy gradients (PGA), VC significantly outperforms reconstruction-only MSE in eight, while MSE outperforms VC in one (Table 1). These comparisons use maximum fitness, except in KHEPERAX, where every run reaches the goal and we compare evaluations to first goal. Figure 2a–b shows the WALKER fitness gains under both proposal mechanisms; Figure 3 illustrates the stronger policies in its passive archive (full archives for all 10 tasks in Appendix G). Appendix E provides the complete tests and learning curves.

Table 1: Terminal medians over ten matched seeds. PA denotes PGA-AURORA; A denotes AURORA. Higher is better, except KHEPERAX evaluations to first goal (lower is better). Within each pair, bold marks the better median for a raw two-sided paired Wilcoxon $p < 0.05$, uncorrected.

	Maximum fitness				**Passive QD score**			
Environment	PA	PA-VC	A	A-VC	PA	PA-VC	A	A-VC
ANTMAZE	-1.41×10^{1}	-7.83×10^{0}	-1.14×10^{1}	-1.23×10^{1}	1.07×10^{4}	1.16×10^{4}	1.23×10^{4}	1.18×10^{4}
ANTOMNI	9.96×10^{2}	$\mathbf{9.97\times10^{2}}$	9.96×10^{2}	$\mathbf{9.96\times10^{2}}$	3.13×10^{6}	2.94×10^{6}	$\mathbf{3.24\times10^{6}}$	3.10×10^{6}
ANTTRAP	4.50×10^{3}	4.39×10^{3}	1.84×10^{3}	1.79×10^{3}	3.40×10^{6}	3.32×10^{6}	3.75×10^{6}	3.67×10^{6}
ANTUNI	6.58×10^{3}	$\mathbf{7.83\times10^{3}}$	1.58×10^{3}	$\mathbf{1.96\times10^{3}}$	2.96×10^{6}	3.01×10^{6}	2.44×10^{6}	$\mathbf{3.01\times10^{6}}$
HALFCHEETAH	7.01×10^{3}	$\mathbf{7.30\times10^{3}}$	4.59×10^{3}	$\mathbf{4.67\times10^{3}}$	$\mathbf{1.12\times10^{7}}$	1.00×10^{7}	$\mathbf{1.10\times10^{7}}$	9.71×10^{6}
HUMANOIDOMNI	1.44×10^{3}	1.39×10^{3}	1.28×10^{3}	1.28×10^{3}	7.84×10^{3}	7.62×10^{3}	7.33×10^{3}	$\mathbf{8.68\times10^{3}}$
HUMANOIDUNI	1.50×10^{3}	1.49×10^{3}	1.38×10^{3}	1.30×10^{3}	5.17×10^{5}	$\mathbf{6.25\times10^{5}}$	5.44×10^{5}	$\mathbf{6.17\times10^{5}}$
KHEPERAX	$\mathbf{1.82\times10^{5}}$	3.44×10^{5}	3.16×10^{5}	4.80×10^{5}	4.24×10^{4}	$\mathbf{5.12\times10^{4}}$	4.12×10^{4}	$\mathbf{5.64\times10^{4}}$
POINTMAZE	-2.28×10^{1}	-2.30×10^{1}	-2.45×10^{1}	-2.38×10^{1}	3.74×10^{5}	3.71×10^{5}	3.65×10^{5}	3.82×10^{5}
WALKER	3.85×10^{3}	$\mathbf{6.51\times10^{3}}$	1.66×10^{3}	$\mathbf{2.23\times10^{3}}$	2.48×10^{6}	$\mathbf{2.65\times10^{6}}$	1.75×10^{6}	$\mathbf{1.89\times10^{6}}$

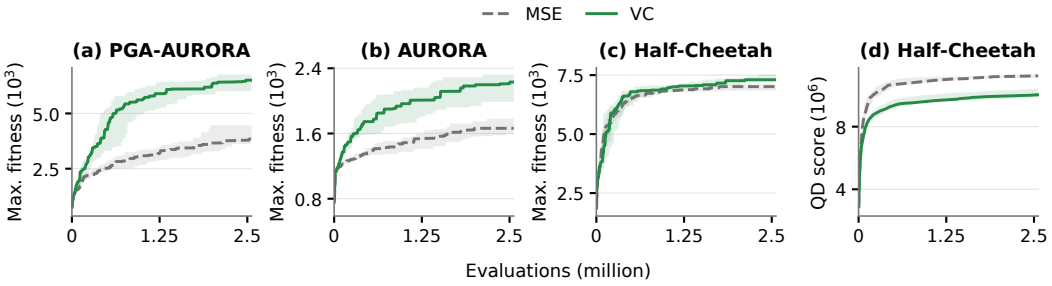

Figure 2: **VC improves peak fitness without uniformly improving repertoire quality.** WALKER maximum fitness improves with both PGA-AURORA (a) and AURORA (b). In HALFCHEETAH, PGA-AURORA finds stronger policies (c) while accumulating a lower passive QD score (d). Lines show ten-seed medians; bands show pointwise 95% bootstrap intervals.

Finding a better policy and building a better repertoire are distinct benefits. Passive QD score also significantly favors VC more often than MSE (eight versus three settings), but these gains need not coincide with better optimization. In HALFCHEETAH, higher maximum fitness accompanies lower QD score (Figure 2c–d). Conversely, the sole significant optimization loss occurs with PGA-AURORA in KHEPERAX, where higher QD score accompanies slower goal discovery. Thus, a geometry that helps discover a strong policy need not distribute quality more broadly across behaviors or reach a goal sooner. The appropriate criterion for VC is the intended search objective, rather than an assumed improvement in every aspect of diversity.

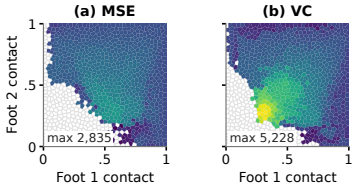

Figure 3: **VC discovers stronger walking policies.** Passive WALKER archives under MSE and VC. Axes show the fraction of time each foot contacts the ground; colors indicate fitness.

Fitness gains also extend beyond DNS. With PGA proposals held fixed, replacing Dominated Novelty Search (DNS) with an adaptive-radius archive still yields a significant overall fitness improvement from VC (Appendix H). The benefit therefore extends to a second archive rule that uses learned distances to decide which policies survive.

6.2 VC BROADENS LATENT VARIATION, WITH DIMINISHING RETURNS

VC spreads variation across more latent directions. To understand the fitness gains, we first ask whether VC spreads variation across more latent directions.

The remaining controls use PGA-AURORA; participation rank measures the effective number of directions contributing to descriptor covariance (Appendix I). At the default width, covariance spectra at encoder updates show more evenly distributed variation under VC throughout training; reconstruction alone concentrates variation in a few directions (Figure 25). VC therefore changes the geometry used for archive competition throughout search, beyond its effect on terminal performance.

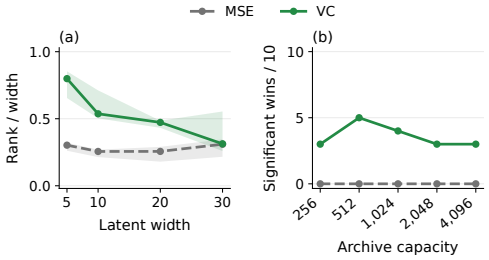

Figure 4: **VC broadens latent variation, with fitness gains across archive sizes.** (a) Rank/width across ten task-level seed medians (median and interquartile range). (b) Significant fitness wins: uncorrected paired Wilcoxon $p < 0.05$, 9–10 pairs. The passive evaluator is fixed.

More coordinates do not guarantee more useful variation. The geometric advantage is strongest in compact representations and diminishes as latent width increases (Figure 4a). Additional coordinates do not produce a proportional increase in effective directions, and fitness gains become less consistent at larger widths. The sweep does not identify an intrinsic task dimension: width also changes model capacity and the relative strength of the $1/d$-normalized covariance penalty. Changing archive capacity also preserves fitness gains across several memory budgets (Figure 4b), with the passive evaluator held fixed. Together, these controls show that VC is useful beyond the default configuration, while higher effective rank alone does not guarantee better search (Appendices J and K).

The two penalties play complementary roles. We next ask whether the benefit requires precise tuning of the variance and covariance weights. V penalizes low coordinate variance, while C discourages redundant coordinates; the weight sweep supports their complementary effect on latent variation and, often, peak fitness. Activating both penalties with non-zero weight matters more consistently than increasing their strength, and no weight pair performs best across all tasks (Appendix L). The useful design principle is to address both weak and redundant variation, while evaluating the resulting geometry by its effect on search.

6.3 Early Geometry Shapes the Opportunities for Later Search

A useful early representation still benefits from adaptation. To test whether continued adaptation improves search beyond VC's early effects, we resume each VC-trained checkpoint in two runs: one that keeps the encoder fixed, while another one that updates it using the variance and covariance penalties without reconstruction loss. Continued adaptation improves the quality accumulated during subsequent search, with more consistent gains in QD-score area under the curve than in peak fitness (Appendix M). An early geometry therefore does not exhaust the value of regularization: continuing to revise archive distances can help search accumulate further quality even without continued reconstruction training.

Late regularization inherits the consequences of early selection. A separate experiment continues or switches between MSE and VC after a shared early history. Continuing VC after a VC history yields more consistent fitness gains than introducing VC after a burn-in period of MSE only (Figure 5; Appendix N). This historical dependence is consistent with the feedback loop introduced earlier: early distances influence which policies survive, and those survivors supply both future parents and encoder training data. Later regularization acts on the opportunities preserved by earlier search; it cannot replay the search under a different early metric. Early geometric regularization can therefore have lasting effects through the policies and training data it leaves available for later search. Section 7 tests how metric changes alter competition and the retained search history.

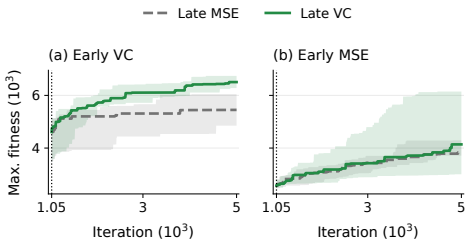

Figure 5: **Early VC leaves a lasting advantage on Walker.** Late MSE and VC updates follow a shared VC history (a) or MSE history (b), starting from identical checkpoints within each pair at iteration 1,050. Lines show ten-seed medians and bands show interquartile ranges on a shared fitness scale. Appendix N reports all tasks.

7 How Geometry Reorganizes Search

The controls support a role for survivor selection in VC's fitness gains. We hypothesize that learned distances change which policies remain available as parents and training data, allowing selection decisions to influence later search. We test whether covariance conditioning alone explains the gains, then trace changes in competition and intervene on the policies retained.

7.1 Covariance Conditioning Alone Does Not Reproduce VC's Gains

Whitening tests whether spreading variance evenly is enough to obtain VC's fitness benefit. It rescales and decorrelates MSE descriptors directly, testing whether covariance conditioning can substitute for training the encoder with VC. If this conditioning were sufficient, periodically whitening MSE descriptors should reproduce VC's gains. Despite nearly equalizing the covariance spectrum, whitening typically yields lower final and time-averaged maximum fitness than VC, while attaining greater passive coverage (Appendix I.4). A well-conditioned covariance therefore does not by itself identify a useful search metric. Together with the continuation control in Section 6, this finding motivates examining how the evolving metric interacts with selection.

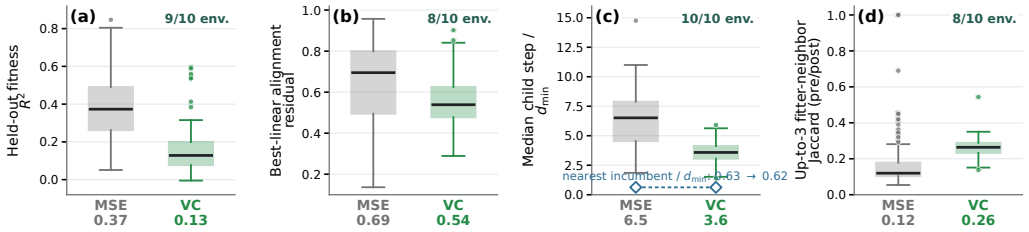

Figure 6: **VC preserves more competitive relationships across retraining.** Distributions of run summaries for MSE and VC: held-out fitness prediction (a), linear alignment error on a fixed trajectory bank (b), parent–child distance relative to the retention cutoff (c), and fitter-neighbor overlap for the same policies before and after encoder retraining (d). Panel counts summarize directions of paired task medians.

7.2 VC Changes Competition Across Encoder Updates

Broader variation does not make fitness easier to predict. We test a possible explanation for the gains by fitting ridge regression from descriptors to fitness, training on half of each sampled archive and evaluating on the held-out half. Prediction becomes less accurate under VC despite its broader latent variation (Figure 6a). The fitness benefit therefore does not require a more linearly accessible fitness signal in the descriptors.

Greater continuity still permits changes in competition. To assess how retraining changes the metric, we encode a fixed bank of trajectories before and after an update and fit a linear map between the embeddings. The map accounts for more of the change under VC (Figure 6b), but it can still stretch or shrink distances. A smaller alignment error alone does not guarantee that policies keep the same fitter neighbors (Appendix Proposition 3).

We therefore measure competition directly while holding policies, trajectories, and fitness fixed. We encode the same incumbents before and after retraining, then recompute their priorities using the new descriptors. This rebuild contains no new candidates and retains every incumbent; membership can change when the next candidate batch arrives. VC preserves more nearest-fitter relationships across retraining, although neighborhoods still change substantially (Figure 6d; Appendix O). Thus, the encoder can reorganize future competition before any policy is removed.

Closer parent–child distances can change local competition. VC reduces parent–child separation relative to the archive's retention cutoff, while nearest-incumbent separation stays similar (Figure 6c). Parent–child distances remain larger than typical nearest-incumbent spacing. This matters because only fitter neighbors contribute to a policy's priority. A sufficiently close parent can lower a worse child's priority; if the child is fitter, it can instead lower the parent's priority (Appendix Proposition 4). The worse member cannot do the same to the fitter member. This identifies a possible local mechanism; these distances do not measure how often parent and child are fitter neighbors. However, archive selection considers the whole pool, so the fitter member is not guaranteed to survive (Appendix Corollary 3).

7.3 Slower Turnover Keeps Earlier Discoveries Available

Under VC, fewer candidates enter the archive, while admitted and removed policies differ more in fitness (Figure 7a–b). High-fitness admissions survive longer, and accepted offspring draw on older parents (Figure 7c–d). Search therefore retains and reuses earlier discoveries while continuing to set fitness records later (Figure 7e).

Longer residence gives a policy more expected opportunities to be sampled as a parent (Appendix Equation 17). Retained trajectories also train subsequent encoders, so archive selection carries earlier discoveries into both future proposals and future distance comparisons. This feedback provides a route through which a change in competition can influence search long after the encoder update that introduced it.

Persistence alone does not establish the cause of the fitness gains: lower admission itself prolongs residence, and later records can reflect slower arrival at a fitness ceiling (Appendix O). Novel passive-cell admissions also remain similar, giving no indication of generally broader behavioral discovery. The diagnostics motivate testing the selection decisions that create this retained history.

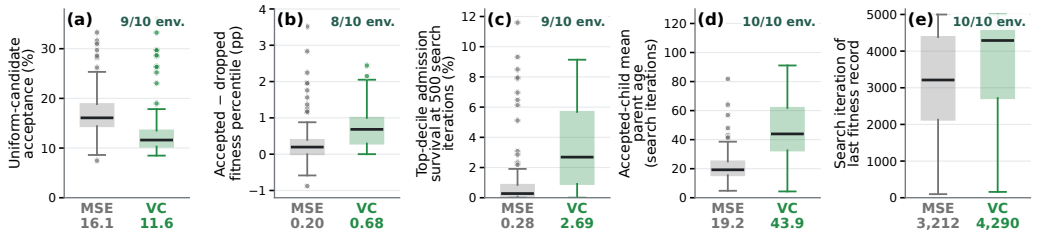

Figure 7: **VC combines slower turnover with reuse of older parents.** Distributions of run summaries for candidate acceptance (a), the fitness-percentile gap between admitted and removed policies (b), survival of top-decile admissions (c), parent age of accepted offspring (d), and timing of the final fitness record (e). Panel counts summarize directions of paired task medians.

7.4 Neighbor Choice and Distance Scoring Act Together

Using VC for both parts of competition gives stronger fitness outcomes than changing either part alone. DNS uses geometry twice: to choose the nearest fitter policies and to measure the separation that determines priority (Equation 5). We separate these operations using MSE and VC encoders in a factorial control with a shared retraining schedule. Mixing the two representations lowers final and time-averaged maximum fitness relative to using MSE for both, whereas using VC for both outperforms either mixture across the nonsaturated task medians (Appendix P.2). The effect of a distance metric therefore depends on which competitors it is used to score.

7.5 Survivor Identity Matters Beyond Admission Counts

Matching an admission count does not preserve VC's advantage. To test whether slower turnover is sufficient, we let the native rule set the number of admissions on each treatment pool, but choose those candidates using distances alone, without fitness. Maximum fitness typically falls under both objectives, and much of VC's advantage disappears (Appendix P.1). Fitness-conditioned choices therefore contribute beyond the rule controlling how many candidates enter.

Transferring selected policies carries more benefit than transferring a quota. In matched MSE archives, we either let MSE choose survivors under a VC-derived admission quota or install the exact VC-selected survivor set. Full identity transfer generally improves maximum fitness averaged over search, while often reducing coverage (Appendix P.4). Choosing different policies changes both the opportunities for later fitness progress and the breadth of the repertoire.

VC selection can also improve fitness when transferred into an archive that continues to store MSE descriptors (Appendix P.3). That transfer changes both counts and identities; the quota comparison above tests the added value of transferring the selected policies. These interventions compare continuing selection rules: after their decisions differ, each rule acts on a population shaped by its own earlier choices. They support survivor selection as one route through which learned geometry influences subsequent search.

8 Conclusion

A learned behavior space is part of the optimizer: its geometry shapes which policies compete, survive, and remain available for future search. Because those policies supply both future parents and encoder training data, early geometric choices can influence search long after the encoder changes.

In this work, we introduce AURORA-VC, which adds fitness-blind variance and covariance penalties to trajectory reconstruction. We observe fitness gains with both PGA-AURORA and AURORA and beyond DNS under an adaptive-radius archive. Our analyses show longer retention of high-fitness policies, and our interventions support survivor selection as a contributor to these gains.

AURORA-VC conditions global latent statistics without guaranteeing useful local distances, broader repertoires, or faster goal discovery. Our interventions also leave the relative contributions of reusing retained policies and retraining on their trajectories unresolved.

AI USE STATEMENT

Generative AI tools were used to polish the writing and improve coherence throughout the main text and appendices. They also assisted with manuscript organization, section drafting, retrieval and discovery of related work, and code development. Generative AI tools were used to construct the detailed proof in Appendix B. The authors developed the VC method independently of generative AI assistance. The authors take full responsibility for the final work.

REPRODUCIBILITY STATEMENT

The code for our experiments is available anonymously at `https://anonymous.4open.science/r/aurora-vc-iclr`. Appendices C and A provide algorithm pseudocode and implementation details. Appendix D documents the environments, hyperparameters, random seeds, and evaluation budgets. Section 5 and the corresponding experiment-specific appendices describe the evaluation protocols and statistical analyses.

REFERENCES

Georgios Arvanitidis, {Lars Kai} Hansen, and Søren Hauberg. Latent space oddity: On the curvature of deep generative models. In *Proceedings of International Conference on Learning Representations*, January 2018. 6th International Conference on Learning Representations, ICLR 2018 ; Conference date: 30-04-2018 Through 03-05-2018.

Ryan Bahlous-Boldi, Maxence Faldor, Luca Grillotti, Hannah Janmohamed, Lisa Coiffard, Lee Spector, and Antoine Cully. Dominated novelty search: Rethinking local competition in quality-diversity. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '25, pp. 104–112, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714658. doi: 10.1145/3712256.3726310. URL `https://doi.org/10.1145/3712256.3726310`.

Adrien Bardes, Jean Ponce, and Yann LeCun. Vicreg: Variance-invariance-covariance regularization for self-supervised learning. In *ICLR*, 2022.

Leo Cazenille. Ensemble feature extraction for multi-container quality-diversity algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '21, pp. 75–83, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383509. doi: 10.1145/3449639.3459392. URL `https://doi.org/10.1145/3449639.3459392`.

Felix Chalumeau, Raphael Boige, Bryan Lim, Valentin Macé, Maxime Allard, Arthur Flajolet, Antoine Cully, and Thomas PIERROT. Neuroevolution is a competitive alternative to reinforcement learning for skill discovery. In *International Conference on Learning Representations*, 2023. URL `https://openreview.net/forum?id=6BHlZgyPOZY`.

Felix Chalumeau, Bryan Lim, Raphaël Boige, Maxime Allard, Luca Grillotti, Manon Flageat, Valentin Macé, Guillaume Richard, Arthur Flajolet, Thomas Pierrot, and Antoine Cully. Qdax: a library for quality-diversity and population-based algorithms with hardware acceleration. *J. Mach. Learn. Res.*, 25(1), January 2024. ISSN 1532-4435.

Nutan Chen, Alexej Klushyn, Francesco Ferroni, Justin Bayer, and Patrick Van Der Smagt. Learning flat latent manifolds with vaes. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020.

Lisa Coiffard, Paul Templier, and Antoine Cully. Overcoming deceptiveness in fitness optimization with unsupervised quality-diversity. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '25, pp. 122–130, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714658. doi: 10.1145/3712256.3726314. URL `https://doi.org/10.1145/3712256.3726314`.

Antoine Cully. Autonomous skill discovery with quality-diversity and unsupervised descriptors. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '19, pp. 81–89,

New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450361118. doi: 10.1145/3321707.3321804. URL https://doi.org/10.1145/3321707.3321804.

Antoine Cully and Yiannis Demiris. Quality and diversity optimization: A unifying modular framework. *IEEE Transactions on Evolutionary Computation*, 22(2):245–259, 2018. doi: 10.1109/TEVC.2017.2704781.

Aleksandr Ermolov, Aliaksandr Siarohin, Enver Sangineto, and Nicu Sebe. Whitening for self-supervised representation learning. In *International Conference on Machine Learning*, pp. 3015–3024. PMLR, 2021.

Daniel Freeman, Erik Frey, Anton Raichuk, Sertan Girgin, Igor Mordatch, and Olivier Bachem. Brax - a differentiable physics engine for large scale rigid body simulation. In J. Vanschoren and S. Yeung (eds.), *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021. URL https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/d1f491a404d6854880943e5c3cd9ca25-Paper-round1.pdf.

Luca Grillotti and Antoine Cully. Unsupervised behavior discovery with quality-diversity optimization. *IEEE Transactions on Evolutionary Computation*, 26(6):1539–1552, 2022a. doi: 10.1109/TEVC.2022.3159855.

Luca Grillotti and Antoine Cully. Relevance-guided unsupervised discovery of abilities with quality-diversity algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '22, pp. 77–85, New York, NY, USA, 2022b. Association for Computing Machinery. ISBN 9781450392372. doi: 10.1145/3512290.3528837. URL https://doi.org/10.1145/3512290.3528837.

Luca Grillotti and Antoine Cully. Kheperax: a lightweight jax-based robot control environment for benchmarking quality-diversity algorithms. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, GECCO '23 Companion, pp. 2163–2165, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701207. doi: 10.1145/3583133.3596387. URL https://doi.org/10.1145/3583133.3596387.

Daniel Kunin, Jonathan Bloom, Aleksandrina Goeva, and Cotton Seed. Loss landscapes of regularized linear autoencoders. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3560–3569. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/kunin19a.html.

Joel Lehman and Kenneth O. Stanley. Evolving a diversity of virtual creatures through novelty search and local competition. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, GECCO '11, pp. 211–218, New York, NY, USA, 2011a. Association for Computing Machinery. ISBN 9781450305570. doi: 10.1145/2001576.2001606. URL https://doi.org/10.1145/2001576.2001606.

Joel Lehman and Kenneth O. Stanley. Abandoning objectives: Evolution through the search for novelty alone. *Evol. Comput.*, 19(2):189–223, June 2011b. ISSN 1063-6560. doi: 10.1162/EVCO_a_00025. URL https://doi.org/10.1162/EVCO_a_00025.

Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4114–4124. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/locatello19a.html.

Elliot Meyerson, Joel Lehman, and Risto Miikkulainen. Learning behavior characterizations for novelty search. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016*, GECCO '16, pp. 149–156, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342063. doi: 10.1145/2908812.2908929. URL https://doi.org/10.1145/2908812.2908929.

Jean-Baptiste Mouret and Jeff Clune. Illuminating search spaces by mapping elites, 2015. URL https://arxiv.org/abs/1504.04909.

Philipp Nazari, Sebastian Damrich, and Fred A. Hamprecht. Geometric autoencoders: what you see is what you decode. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Olle Nilsson and Antoine Cully. Policy gradient assisted map-elites. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '21, pp. 866–875, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383509. doi: 10.1145/3449639.3459304. URL https://doi.org/10.1145/3449639.3459304.

Giuseppe Paolo, Alban Laflaquière, Alexandre Coninx, and Stephane Doncieux. Unsupervised learning and exploration of reachable outcome space. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2379–2385, 2020. doi: 10.1109/ICRA40945.2020.9196819.

Justin K. Pugh, L. B. Soros, Paul A. Szerlip, and Kenneth O. Stanley. Confronting the challenge of quality diversity. In *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, GECCO '15, pp. 967–974, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450334723. doi: 10.1145/2739480.2754664. URL https://doi.org/10.1145/2739480.2754664.

Justin K. Pugh, Lisa B. Soros, and Kenneth O. Stanley. Quality diversity: A new frontier for evolutionary computation. *Frontiers in Robotics and AI*, Volume 3 - 2016, 2016. ISSN 2296-9144. doi: 10.3389/frobt.2016.00040. URL https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2016.00040.

Oren Rippel, Michael Gelbart, and Ryan Adams. Learning ordered representations with nested dropout. In Eric P. Xing and Tony Jebara (eds.), *Proceedings of the 31st International Conference on Machine Learning*, volume 32-2 of *Proceedings of Machine Learning Research*, pp. 1746–1754, Beijing, China, 22–24 Jun 2014. PMLR. URL https://proceedings.mlr.press/v32/rippel14.html.

Constantinos Tsakonas and Konstantinos Chatzilygeroudis. Vector quantized-elites: Unsupervised and problem-agnostic quality-diversity optimization. *IEEE Transactions on Evolutionary Computation*, 30(4):1669–1683, 2026. doi: 10.1109/TEVC.2025.3631786.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stephane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 12310–12320. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/zbontar21a.html.

APPENDIX

TABLE OF CONTENTS

A Implementation Details **14**

B Formal Analysis of Descriptor Geometry and Archive Selection with POINTMAZE Case Study **16**

C Algorithm Pseudocodes **26**

D Environments and Hyperparameters **28**

E Primary Results for PGA-AURORA and AURORA with VC Regularization **30**

F Comparison with XCON and Archive Extinction **35**

G Passive-Archive Visualizations **39**

H Effect of the Archive Selection Rule on AURORA-VC **50**

I Representation Geometry and Semantic Diagnostics **54**

J Effect of Latent Dimensionality on AURORA-VC **58**

K Effect of Archive Capacity on AURORA-VC **62**

L Effect of Variance and Covariance Weights **65**

M Continued V+C-Only Adaptation versus Freezing **68**

N How Search History Shapes the Effect of VC **70**

O Archive-Dynamics Diagnostics **73**

P Interventions on the Survivor Rule and Mechanism Controls **76**

A Implementation Details

A.1 Trajectory Representation

The learned descriptor is the final cell state of a single-layer LSTM encoder. The decoder is another LSTM initialized with both the final cell state and hidden state of the encoder; its linear output layer predicts trajectory features. Training uses teacher forcing on normalized trajectories: observations $1, \dots, T-1$ predict targets $2, \dots, T$. MSE averages over examples, target steps, and features. The V+C penalties act on the cell-state descriptor, not the full decoder state. Primary tasks use local sensory trajectories except PointMaze, whose two-dimensional task descriptor is the encoder input.

At each scheduled encoder retraining event, features are normalized using feature-wise mean and standard deviation over active members and sampled time steps. Matched MSE and VC runs use the same horizon, sampling pattern, normalization, minibatch procedure, optimizer, and stopping budget. Position descriptors use the last valid pre-action state; contact descriptors average contact indicators over valid trajectory steps. These descriptors never enter active selection, except in the explicitly labeled descriptor-given controls.

At each retraining, encoder and decoder weights continue from the previous update, while Adam's optimizer state is reset. One bootstrap sample of valid archive trajectories is drawn with replacement and reshuffled into minibatches each epoch. The primary sample has $1{,}024$ rows, giving eight 128-example steps per epoch. Early stopping monitors the assigned total training loss, with minimum improvement 5×10^{-4} and patience 10, up to 200 epochs. The MSE and VC share the stopping rule and maximum budget but need not execute the same number of optimizer steps.

A.2 Active Archive and Passive Evaluator

The primary search uses a capacity-bounded archive. Pool incumbents and candidates into P. For member i, let $F_i = \{j \in P : f_j > f_i\}$ and let N_i contain its up-to-three nearest members of F_i. Its priority is

$$s_i = \begin{cases} |N_i|^{-1} \sum_{j \in N_i} \|z_i - z_j\|_2, & |N_i| > 0, \\ +\infty, & |N_i| = 0. \end{cases} \tag{5}$$

The highest-priority $1{,}024$ entries are retained, with deterministic implementation order resolving exact ties. The same rule is reapplied after encoder retraining. Because neighbors must be strictly fitter, changing latent distances changes a fitness-conditioned competition graph rather than a fitness-blind clustering objective.

A separate $1{,}024$-centroid CVT archive receives the selected active members after each update and accumulates the best fitness per handcrafted task cell. This selected-member archive supplies the reported passive coverage and QD score; it provides neither parents nor admission decisions. The instrumented all-candidate tracker includes rejected proposals and is a distinct diagnostic. Under DNS, at least one maximum-fitness policy always survives, so the selected-member maximum equals the best fitness discovered; the adaptive-radius study reports both endpoints explicitly. Plain AURORA uses genetic proposals while PGA-AURORA allocates half of each proposal batch to genetic variation and half to critic-guided policy updates. The stored quantity $d_{\min}$ is the lowest retained DNS priority at the current capacity decision, not a minimum pairwise distance or an admission radius. The capacity controller is inactive for the primary DNS archive.

A.3 VC Encoder Retraining Objective

For centered minibatch embeddings $C = (\max(B-1, 1))^{-1} \sum_i \bar{z}_i \bar{z}_i^{\mathsf{T}}$, the implemented regularizers are

$$L_V = d^{-1} \sum_r \max(0, \gamma - \sqrt{C_{rr} + \epsilon}), \qquad L_C = d^{-1} \sum_{r \neq q} C_{rq}^2. \tag{6}$$

The total loss is $L_{\text{base}} + \lambda_V L_V + \lambda_C L_C$, where L_{base} is reconstruction for AURORA-VC and the triplet objective for the XCON+VC experiments. The baseline VC uses $(\lambda_V, \lambda_C, \gamma, \epsilon) = (1, 1, 1, 10^{-4})$; with ablations and parameter sweeps included in Appendix L. V+C regularization

acts on raw embeddings, after input normalization but without latent whitening. The covariance computation costs $O(Bd^2)$ and requires a $d \times d$ matrix.

B Formal Analysis of Descriptor Geometry and Archive Selection with PointMaze Case Study

Purpose and roadmap. In this section, we present a formal analysis of variance–covariance (VC) regularization to explain how these components shape the descriptor geometry and what they imply for archive selection. In Section B.1, we show that transforming the learned descriptors, while adjusting the decoder to keep the reconstructed trajectories unchanged, can alter the distances used by archive selection. We then identify which descriptor transformations make both VC penalties zero for a fixed embedding covariance, and determine how they change the distances used by archive selection. We explain why a transformation that satisfies these conditions for one minibatch need not satisfy them for other minibatches.

For the archive rule, we derive conditions for preserving competitive neighborhoods, the exact priority change caused by a nearby fitter candidate, and the relationship between how long a policy remains in the archive and how often it is expected to be selected as a parent.

Finally, we return to the controlled PointMaze study and expand on its experimental protocol and results, showing how changes in descriptor geometry can affect competitive neighborhoods and archive membership while preserving reconstruction and holding candidate policies and fitness values fixed.

B.1 Reconstruction Equivalence and Descriptor Distances

Let τ denote a trajectory, let E be an encoder that maps τ to a descriptor $z = E(\tau) \in \mathbb{R}^d$, and let D be a decoder that reconstructs the trajectory from z. Here d is the descriptor dimension, and all vectors are written as columns. Consider an invertible matrix $A \in \mathbb{R}^{d\times d}$, which applies a reversible linear transformation to the descriptors, such as a rotation, scaling, or shear. Let $a \in \mathbb{R}^d$ be a translation vector that shifts every descriptor by the same amount, and define the combined transformation $T_{A,a}(z) = Az + a$. Its inverse is $T^{-1}_{A,a}(z') = A^{-1}(z' - a)$: subtracting a removes the translation, and multiplying by A^{-1} undoes the linear transformation.

We construct a transformed encoder $E' = T_{A,a} \circ E$ and a corresponding decoder $D' = D \circ T^{-1}_{A,a}$. Thus E' transforms the original descriptor, while D' reverses that transformation before applying the original decoder. For every trajectory, the reconstructed trajectory is therefore unchanged:

$$D'(E'(\tau)) = D\big(A^{-1}(AE(\tau) + a - a)\big) = D(E(\tau)).$$

Thus, any reconstruction loss computed from these outputs also remains unchanged.

Archive selection uses Euclidean distances between descriptors. For two trajectories τ_i and τ_j, let the descriptors be $z_i = E(\tau_i)$ and $z_j = E(\tau_j)$. Their transformed descriptor difference is $(Az_i + a) - (Az_j + a) = A(z_i - z_j)$, so the common translation cancels and the squared distance becomes

$$\|E'(\tau_i) - E'(\tau_j)\|_2^2 = (z_i - z_j)^\mathsf{T} A^\mathsf{T} A(z_i - z_j), \tag{7}$$

where $\|\cdot\|_2$ is the Euclidean length, and A^T denotes the transpose of A. The positive-definite matrix $G = A^\mathsf{T}A$ specifies how descriptor differences contribute to the transformed squared distance. For example, multiplying one descriptor coordinate by three multiplies its contribution to the squared distance by nine. Such changes can alter the fitter-neighbor graph, which connects each policy to its nearest policies with strictly higher fitness, even though reconstruction is unchanged. Thus, reconstruction alone does not determine the distances used by the archive selection. As we gradually change A and a, the reconstruction loss remains unchanged, provided that A stays invertible and the decoder is adjusted to reverse each transformation.

Transfer to the LSTM encoder and decoder. In our implementation, the descriptor is the encoder's final cell state z, while the decoder also receives the final hidden state h. The reconstruction identity still holds when we transform z, keep h and the teacher-forcing inputs fixed, and define $D'(z', h) = D(A^{-1}(z' - a), h)$. This construction applies the inverse transformation before decoding; it does not require that the transformation can be implemented solely by changing the LSTM weights. More generally, composing an encoder with any invertible map and its decoder with the inverse preserves reconstruction. The metric characterization below concerns global linear maps; it does

not establish local distance preservation under nonlinear reparameterizations or within the fixed LSTM architecture.

Similarity invariance. The archive-selection results assume fixed finite fitnesses and a fixed deterministic rule for ties that does not depend on embeddings. A linear map preserves every Euclidean distance up to one common positive scale c if and only if $A^\mathsf{T}A = c^2 I_d$, where I_d is the $d \times d$ identity matrix. Equivalently, $A = cQ$ for an orthogonal matrix Q, meaning $Q^\mathsf{T}Q = I_d$; such a matrix rotates or reflects vectors without changing their lengths. Combining this map with a translation gives a similarity transformation, which preserves shape while allowing a common change in scale. Translation cancels from pairwise differences, and cQ multiplies every pairwise distance by c. Using the archive rule in Equation 5, let f_i and s_i denote the fitness and selection priority of policy i, let the positive integer k be the maximum number of strictly fitter neighbors used to compute that priority, and let positive integer K be the archive capacity. Higher priorities are retained, every fitter-neighbor set is unchanged, and every finite priority satisfies $s'_i = cs_i$. Infinite priorities remain infinite, and the common scaling preserves every priority ordering. Thus, the survivor set is unchanged under the same capacity-tie breaking rule.

Let Σ be the covariance of the original descriptors. Its eigenvectors give the principal directions of variation, and its eigenvalues give the variances along these directions; together, the eigenvalues form the covariance spectrum. For non-zero covariance, the transformation to $c^2 Q\Sigma Q^\mathsf{T}$ leaves participation rank unchanged: rotation or reflection preserves the covariance eigenvalues, and the common scale factors cancel from the rank formula. Thus translation, orthogonal transformations, and common positive scaling leave archive selection unchanged, and only changes in relative geometry can affect selection. A general invertible map has no such guarantee, although a non-similarity map can still preserve a particular finite ordering by accident.

Under a general invertible transformation, the covariance becomes $A\Sigma A^\mathsf{T}$. Algebraic rank, the number of linearly independent covariance directions, is unchanged.

B.2 VC Regularization, Descriptor Geometry, and Archive Selection

During encoder training, trajectories are processed in minibatches. For each minibatch, the encoder produces descriptors, and the VC penalties in Equation 4 are computed from their coordinate variance and covariances. These penalties contribute to the training loss alongside reconstruction, and their gradients contribute to the encoder update. We begin by tracing how the VC penalties act on one fixed minibatch to derive their gradients and determine which descriptor geometries fully satisfy the variance and covariance requirements. We then examine how these requirements change across minibatches.

B.2.1 VC penalties and Fixed-Covariance Geometry

Notation and assumptions. For one sampled minibatch $\mathcal{B}$ containing B descriptors, let $Z \in \mathbb{R}^{B\times d}$ store these descriptors, with one descriptor per row and one coordinate per column. To compute the coordinate variances and covariances, we first subtract the mean descriptor $\bar{z}$ from every row of Z:

$$X = Z - \mathbf{1}\bar{z}^\mathsf{T},$$

where $\mathbf{1} \in \mathbb{R}^B$ is a vector of ones that repeats the mean across all rows. Each column of X therefore has mean zero. Let $n = \max(B-1, 1)$ be the covariance normalization, and define the sample covariance $\widehat{\Sigma}_\mathcal{B} = X^\mathsf{T}X/n$. While this minibatch is held fixed, we abbreviate $\Sigma = \widehat{\Sigma}_\mathcal{B}$. Write $x_r = X_{:,r}$ for column r containing the centered values of descriptor coordinate r across the batch. For the regularizers in Equation 4, write $s_r = \sqrt{\Sigma_{rr} + \epsilon}$ for the smoothed coordinate standard deviation; ϵ is the positive smoothing constant and γ is the standard deviation threshold. The coordinate quantity s_r is distinct from the policy priority s_i above; coordinate indices in this calculation are r and q. Let $O = \Sigma - \operatorname{diag}(\Sigma)$, where $\operatorname{diag}(\Sigma)$ retains only the diagonal entries, so O contains the covariances between different coordinates. For the gradient calculations and collapse expansion below, assume $B \geq 2$ and $\gamma > \sqrt{\epsilon} > 0$, as in the implementation.

Gradients of the VC penalties. We write L_V and L_C for the variance and covariance penalties in Equation 4. Away from the variance-hinge boundary $s_r = \gamma$ where the penalty switches between

active and zero, their gradients with respect to the centered descriptors are

$$\nabla_{x_r} L_V = -\frac{\mathbf{1}\{s_r < \gamma\}}{dn\, s_r}\, x_r \qquad\qquad \nabla_X L_C = \frac{4}{dn} XO. \tag{8}$$

Here ∇ denotes a gradient with respect to the indicated vector or matrix, and $\mathbf{1}\{s_r < \gamma\}$ equals one when coordinate r is below a threshold and zero otherwise. For the variance gradient, use $\Sigma_{rr} = \|x_r\|^2/n$ and differentiate the square root to obtain $\nabla_{x_r} s_r = x_r/(ns_r)$. When the variance penalty is active, its derivative adds the minus sign and the factor $1/d$.

For the covariance gradient, write $L_C = \|O\|_F^2/d$, where $\|O\|_F^2$ is the sum of the squared off-diagonal covariances. A small change $\mathrm{d}X$ gives $\mathrm{d}\Sigma = ((\mathrm{d}X)^\mathsf{T} X + X^\mathsf{T}\mathrm{d}X)/n$. Since O is symmetric and has zero diagonal,

$$\mathrm{d}L_C = \frac{2}{d}\,\mathrm{tr}\, O\,\mathrm{d}\Sigma = \frac{4}{dn}\,\mathrm{tr}((XO)^\mathsf{T}\mathrm{d}X),$$

where tr is the sum of a matrix's diagonal entries. The coefficient of $\mathrm{d}X$ gives the covariance gradient. Both gradients have zero column means, so differentiation through mean subtraction gives the same gradients with respect to Z.

If we update the descriptors directly using the variance term alone, a nonzero coordinate below the threshold expands under a gradient step. The covariance step subtracts a weighted combination of correlated coordinates. In the implemented model, the encoder shares parameters across examples and is trained with Adam and reconstruction loss. The actual encoder update therefore need not expand each weak coordinate or decrease both penalties at every step.

The two penalties address different problems. Rescaling coordinates by positive factors leaves their correlations **unchanged**, so the variance penalty alone allows redundant coordinates. The covariance penalty alone allows complete collapse, $X = 0$. For one fixed minibatch, both penalties are zero exactly when

$$x_r^\mathsf{T} x_q = 0 \quad (r \neq q), \qquad \|x_r\|_2^2 \geq n(\gamma^2 - \epsilon). \tag{9}$$

These conditions require the centered coordinate columns to be mutually orthogonal and to have sufficient variance. Since $\gamma^2 > \epsilon$, all columns are nonzero and the covariance is full rank. However, centered columns lie in a space of dimension $B - 1$, so zero loss is impossible when $d > B - 1$. Full rank across a batch does not establish the number of independent directions within a small neighborhood or whether the coordinates represent meaningful behavioral differences. The penalties also place no upper bound on the coordinate variance: a diagonal covariance with very unequal variances can have zero V+C loss.

We can also use the loss values to bound the covariance eigenvalues. Let $\ell_V = L_V$ and $\ell_C = L_C$, and define

$$\alpha = \max\left\{0, [\max(0, \gamma - d\ell_V)]^2 - \epsilon\right\}, \qquad \beta = \max_r \Sigma_{rr}$$

Here α is a lower bound on every coordinate variance, and β is the largest coordinate variance. To obtain the lower bound, note that each coordinate's term $[\gamma - s_r]_+$ is at most the sum of these terms, $d\ell_V$. Thus $s_r \geq \max(0, \gamma - d\ell_V)$, which gives $\Sigma_{rr} \geq \alpha$ after squaring and subtracting ϵ. The off-diagonal part satisfies $\|O\|_2 \leq \|O\|_F = \sqrt{d\ell_C}$, where $\|O\|_2 = \max_{\|u\|_2=1} \|Ou\|_2$ is its largest effect on a unit vector.

For any unit direction u, with $\|u\|_2 = 1$, the variance $u^\mathsf{T}\Sigma u$ is the sum of a diagonal contribution and an off-diagonal contribution. The first lies between α and β, and the second lies between $-\sqrt{d\ell_C}$ and $\sqrt{d\ell_C}$. The smallest and largest directional variances are the eigenvalues $\lambda_{\min}(\Sigma)$ and $\lambda_{\max}(\Sigma)$, so

$$\lambda_{\min}(\Sigma) \geq \alpha - \sqrt{d\ell_C} \qquad \lambda_{\max}(\Sigma) \leq \beta + \sqrt{d\ell_C}. \tag{10}$$

When $\alpha > \sqrt{d\ell_C}$, every direction has positive variance. The covariance condition number, $\kappa_2(\Sigma) = \lambda_{\max}(\Sigma)/\lambda_{\min}(\Sigma)$, then satisfies

$$\kappa_2(\Sigma) \leq \frac{\beta + \sqrt{d\ell_C}}{\alpha - \sqrt{d\ell_C}} \tag{11}$$

This ratio measures how unequal the largest and smallest directional variances are. The bound still depends on β, which VC does not bound from above.

At complete collapse, $X = 0$, all descriptors are identical and both regularizer gradients are zero; the covariance loss is zero, but the variance loss is positive. To examine nearby descriptors, let $U \in \mathbb{R}^{B\times d}$ be a nonzero matrix with zero column means and set $X = tU$ for small t. Expanding around $t = 0$ gives

$$L_V(tU) = \gamma - \sqrt{\epsilon} - \frac{t^2\|U\|_F^2}{2dn\sqrt{\epsilon}} + O(t^4), \qquad L_C(tU) = O(t^4).$$

Here $O(t^p)$ describes terms bounded by a constant times $|t|^p$ near zero; it is unrelated to the off-diagonal covariance matrix O. When the variance penalty has positive weight, the negative quadratic term dominates the fourth-order terms for sufficiently small nonzero t. Thus, complete collapse has zero gradient but is not a local minimum of the positive-weight V+C loss when descriptors can vary freely. This result does not rule out stationary collapse when descriptors must be produced by a neural encoder or when reconstruction is included in the objective.

Local effects of small coordinate changes. We next examine how the penalties respond to a small change in the descriptor coordinates. The transformation in Equation 7 changes the covariance from Σ to $A\Sigma A^\mathsf{T}$. Let $v = \gamma^2 - \epsilon > 0$ be the variance floor, and start from $\Sigma = vI_d$, where every direction has the same variance, exactly at the floor. Write $A = I_d + tH$, where $H \in \mathbb{R}^{d\times d}$ specifies the change and $t > 0$ is small. Then

$$A\Sigma A^\mathsf{T} = vI + vt(H + H^\mathsf{T}) + vt^2HH^\mathsf{T}.$$

The leading off-diagonal term is $vt\,\mathrm{offdiag}(H+H^\mathsf{T})$, where offdiag sets the diagonal entries to zero. The smoothed standard deviation of coordinate r becomes $\gamma + (v/\gamma)tH_{rr} + O(t^2)$. Substituting these expressions into the penalties gives

$$L_C(A) = \frac{v^2t^2}{d}\left\|\mathrm{offdiag}(H + H^\mathsf{T})\right\|_F^2 + O(t^3), \tag{12}$$

$$L_V(A) = \frac{v}{\gamma d}\sum_{r=1}^{d}[-tH_{rr}]_+ + O(t^2), \tag{13}$$

where $L_V(A)$ and $L_C(A)$ are the losses after the transformation and $[u]_+ = \max(0, u)$. The covariance penalty grows quadratically with small symmetric changes that mix coordinates. The variance penalty responds to coordinate contraction, represented by negative diagonal entries H_{rr}.

If H is skew-symmetric, $H^\mathsf{T} = -H$, then $H + H^\mathsf{T} = 0$. Such a change agrees to first order with the rotation $\exp(tH)$, where exp is the matrix exponential. The exact rotation preserves distances; its approximation $I + tH$ need not be orthogonal. Positive diagonal expansion above the variance floor also remains possible.

These local penalties do not select one distance geometry. We call two positive-definite metric matrices equivalent when they differ only by a positive scalar, since that scaling leaves archive selection unchanged. The next proposition shows that, for one fixed covariance, every such metric has a transformation with zero V+C loss.

The fixed-covariance objective. Hold one covariance $\Sigma \succ 0$ fixed, meaning that $x^\mathsf{T}\Sigma x > 0$ for every nonzero direction x. We vary the invertible transformation A and apply its inverse before decoding so reconstruction stays unchanged. Writing $C_A = A\Sigma A^\mathsf{T}$, the objective is

$$\mathcal{J}_\Sigma(A) = \frac{1}{d}\sum_{r=1}^{d}\left[\gamma - \sqrt{(C_A)_{rr} + \epsilon}\right]_+ + \frac{1}{d}\left\|\mathrm{offdiag}(C_A)\right\|_F^2. \tag{14}$$

Giving the two terms any strictly positive weights leaves the zero-loss conditions unchanged. We now find all transformations that meet these conditions and the descriptor distances they produce.

Proposition 2 (Fixed-covariance zero-loss metrics under V+C). *Let $\Sigma \succ 0$ be one fixed descriptor covariance, let $\gamma > 0$ and $\epsilon \geq 0$ satisfy $v = \gamma^2 - \epsilon > 0$, and let A be invertible. The covariance may come from a fixed population, a full fixed archive, or one realized full-rank minibatch. For transformations that preserve reconstruction by applying the inverse before decoding, $\mathcal{J}_\Sigma(A) = 0$ exactly when*

$$A\Sigma A^\mathsf{T} = D, \qquad D \text{ is diagonal}, \qquad D \succeq vI.$$

Here D contains the transformed coordinate variances, each of which must be at least v. For symmetric matrices, $C_1 \succeq C_2$ means that $x^\mathsf{T} C_1 x \geq x^\mathsf{T} C_2 x$ for every x. All zero-loss transformations have the form

$$A = D^{1/2} R \Sigma^{-1/2},$$

where R is orthogonal and D is diagonal with $D \succeq vI$. The square roots and inverse square roots are the symmetric positive-definite ones. This transformation first whitens the covariance to I, then rotates or reflects the coordinates, and finally scales them to the variances in D. The resulting archive metrics are

$$G = A^\mathsf{T} A = \Sigma^{-1/2} R^\mathsf{T} D R \Sigma^{-1/2},$$

and together they give exactly the set

$$\left\{ G = G^\mathsf{T} \succ 0 : G \succeq v\Sigma^{-1} \right\}.$$

Up to a common positive scale, this set includes every positive-definite metric.

Proof. Step 1: identify the zero-loss conditions. Both terms in Equation 14 are nonnegative. The covariance term is zero exactly when $A\Sigma A^\mathsf{T}$ is diagonal, and the variance term is zero exactly when every diagonal entry is at least $v = \gamma^2 - \epsilon$. Thus $A\Sigma A^\mathsf{T} = D$ for diagonal $D \succeq vI$.

Step 2: find all transformations that satisfy them. Given $A\Sigma A^\mathsf{T} = D$, define $R = D^{-1/2} A \Sigma^{1/2}$. Then $RR^\mathsf{T} = I$, so R is orthogonal and $A = D^{1/2} R \Sigma^{-1/2}$. Conversely, substituting any transformation of this form gives $A\Sigma A^\mathsf{T} = D$ and hence zero loss.

Step 3: find the resulting distance matrices. Substituting into $G = A^\mathsf{T} A$ gives $G = \Sigma^{-1/2} R^\mathsf{T} D R \Sigma^{-1/2}$. Every symmetric positive-definite matrix has an orthogonal eigenvector basis. Thus $R^\mathsf{T} D R$ covers exactly the matrices whose eigenvalues are all at least v. Multiplication on both sides by $\Sigma^{-1/2}$ preserves the matrix inequality, and the inverse multiplication by $\Sigma^{1/2}$ reverses the construction. The resulting metrics are therefore exactly those satisfying $G \succeq v\Sigma^{-1}$.

Step 4: allow the common scale that leaves selection unchanged. For any target metric $M \succ 0$, choose a positive scale α satisfying

$$\alpha \geq v \, \lambda_{\max}\left(M^{-1/2} \Sigma^{-1} M^{-1/2} \right).$$

Then $\alpha M \succeq v\Sigma^{-1}$, so Step 3 provides a zero-loss transformation with metric αM. This rescaling multiplies every distance by $\sqrt{\alpha}$ and preserves all fitter-neighbor and priority orderings. Thus every positive-definite metric class has a zero-loss transformation. □

The last step is why every metric class remains possible: any target metric can be scaled enough to meet the variance floor. The construction applies to the specified covariance. It does not show that training finds the transformation or that the same transformation works for other covariances.

Corollary 1 (Zero VC loss does not determine archive distances)**.** *For one fixed full-rank covariance, zero V+C loss excludes collapsed and linearly redundant coordinates, but still allows every positive-definite metric class used by selection.*

Proof. The positive diagonal variance floor gives full rank, and Step 4 of Proposition 2 gives a zero-loss transformation for every positive-definite metric up to scale. □

This is consistent with the local penalty against shear. After whitening, the covariance is isotropic and can be rotated freely. Unequal coordinate expansion then gives a diagonal covariance above the floor, while changing the distance geometry in the original coordinates.

B.2.2 Population and Minibatch Objectives

A transformation that gives zero VC loss for one minibatch need not do so for other minibatches. To examine this difference, draw a minibatch $\mathcal{B}$ from a fixed population of descriptors and choose one invertible transformation A before drawing the batch. The same A must therefore be used for

every sampled covariance $\widehat{\Sigma}_{\mathcal{B}}$. With the decoder adjusted to preserve reconstruction, the expected VC penalty is

$$\mathcal{J}_{\mathrm{mb}}(A) = \mathbb{E}_{\mathcal{B}}\left[\mathcal{J}_{\widehat{\Sigma}_{\mathcal{B}}}(A)\right]. \tag{15}$$

Here $\mathbb{E}_{\mathcal{B}}$ averages the loss over the minibatches that can be sampled. Because the penalties are nonlinear, averaging their values across batches generally differs from computing them once from the population covariance.

Corollary 2 (Expected-minibatch zero requires common batchwise zero)**.** *Assume both VC weights are positive and $v > 0$. Then $\mathcal{J}_{\mathrm{mb}}(A) = 0$ if and only if*

$$A\widehat{\Sigma}_{\mathcal{B}}A^{\mathsf{T}} = D_{\mathcal{B}}, \qquad D_{\mathcal{B}} \succeq vI$$

with probability one under minibatch sampling, where $D_{\mathcal{B}}$ is diagonal. Thus, every sampled batch must have zero coordinate covariances and coordinate variances at least v, apart from events with probability zero. The diagonal covariance $D_{\mathcal{B}}$ can differ between batches, but the transformation A must be shared. Proposition 2 therefore does not show that every positive-definite metric, up to a common scale, can achieve zero expected minibatch loss. Requiring one shared transformation can leave fewer possible metrics, or no transformation with zero expected loss.

Proof. The loss of each minibatch is nonnegative. Its expectation is therefore zero exactly when the loss is zero with probability one. Applying the zero-loss conditions to each sampled covariance gives the result. If a rank-deficient covariance is sampled with positive probability, zero expected loss is impossible: multiplying by invertible A and A^{T} preserves rank, while $D_{\mathcal{B}} \succeq vI_d$ requires full rank. □

Even if every batch covariance is full rank, the same A must make all of them diagonal. After whitening one reference covariance, all other covariances in those whitened coordinates must share one set of orthogonal eigenvectors. In other words, one set of perpendicular directions must diagonalize all of these matrices. For more than two covariance matrices, this condition need not hold. Allowing a separate transformation $A_{\mathcal{B}}$ for each full-rank batch would let us apply Proposition 2 to each batch separately. A shared encoder does not choose a separate transformation for each batch.

Sampling also makes the covariance entries fluctuate around their population values. Let $\Sigma_\star$ be the population covariance and assume the sample covariance is unbiased, so $\mathbb{E}_{\mathcal{B}}[\widehat{\Sigma}_{\mathcal{B}}] = \Sigma_\star$. Assume the descriptor coordinates have finite fourth moments, which ensures that the sample covariance entries have finite variance. Write $C = A\Sigma_\star A^{\mathsf{T}}$ for the transformed population covariance and $\widehat{C}_{\mathcal{B}} = A\widehat{\Sigma}_{\mathcal{B}}A^{\mathsf{T}}$ for the transformed batch covariance. For two different coordinates $q \neq r$,

$$\mathbb{E}[(\widehat{C}_{\mathcal{B}})_{qr}^2] = C_{qr}^2 + \mathrm{Var}[(\widehat{C}_{\mathcal{B}})_{qr}]. \tag{16}$$

The first term is the squared population covariance; the second measures how much the sample covariance varies between batches. A diagonal population covariance removes the first term, but the second can remain positive. Similarly, zero expected variance loss requires $(\widehat{C}_{\mathcal{B}})_{rr} \geq v$ with probability one for every coordinate. A population variance $(C)_{rr} \geq v$ alone does not ensure that each batch meets this floor.

Our implementation uses $B = 128$, $d = 10$, and one shared encoder trained across shuffled minibatches. Training takes a finite sequence of updates on these batches and need not reach zero loss for individual batches or in expectation. Proposition 2 describes the transformations allowed by zero VC loss for one fixed covariance. The experiments show which geometry training reaches and how it affects search.

B.2.3 Archive Stability Across Encoder Retraining

We next ask how much descriptors can change during retraining while preserving archive decisions. The following result compares the change in distances with the gaps that separate included from excluded neighbors and survivors. If those gaps are large enough, retraining leaves the decisions unchanged.

Proposition 3 (Temporal neighbor and priority stability)**.** *Consider the same finite set of policies and fitness immediately before and after one encoder retraining event. Suppose the descriptors satisfy*

$$z_i' = cQz_i + a + e_i \qquad c > 0, \quad Q^\mathsf{T}Q = I, \quad \max_i \|e_i\|_2 \le \delta$$

Here z_i' is the descriptor after retraining, $cQz_i + a$ applies one common similarity transformation of the old descriptors, and e_i is the remaining change for policy i. The bound $\delta \ge 0$ applies to every policy. Then each pair of policies satisfies

$$\left| \|z_i' - z_j'\|_2 - c\|z_i - z_j\|_2 \right| \le 2\delta$$

For a policy with at least $k+1$ strictly fitter alternatives, let $r_{i,k} \le r_{i,k+1}$ be its distances to the k-th and $(k+1)$-th nearest fitter policies before retraining. These are the distances just inside and outside its selected neighborhood. Its k-nearest-fitter set is unchanged whenever

$$c(r_{i,k+1} - r_{i,k}) > 4\delta$$

Every finite priority satisfies $|s_i' - cs_i| \le 2\delta$. Suppose the pool has more than K members and fewer than K members have infinite priority. Order priorities from largest to smallest with $s_{(j)}$ denoting the priority at position j. Let $\Delta_s = s_{(K)} - s_{(K+1)} > 0$ be the finite gap between the lowest retained priority and the highest excluded priority before retraining. The capacity-K survivor set is unchanged whenever $c\Delta_s > 4\delta$. If at least K members have infinite priority, the unchanged fitnesses and fixed tie rule determine the survivors, so their identities are also unchanged.

Proof. Step 1: bound the change in each distance. Subtracting the two descriptor equations cancels the translation and leaves the error $e_i - e_j$, whose length is at most 2δ. The reverse triangle inequality, $|\|u\|_2 - \|v\|_2| \le \|u - v\|_2$, together with $\|Q(z_i - z_j)\|_2 = \|z_i - z_j\|_2$, gives the pairwise bound.

Step 2: preserve the neighbor set. Every previously included neighbor has new distance at most $cr_{i,k} + 2\delta$. Every excluded fitter policy has new distance at least $cr_{i,k+1} - 2\delta$. The assumed gap makes the first bound strictly smaller than the second, so no excluded neighbor can replace an included one. When at most k fitter policies exist, all are included before and after retraining without a gap condition.

Step 3: bound priority change. Let $F_i = \{j : f_j > f_i\}$ contain all strictly fitter policies, and let $m_i = \min(k, |F_i|) > 0$ be the number used to compute the priority. For each subset $S \subseteq F_i$ containing m_i policies, let a_S and a_S' be their mean distances from i before and after retraining. The nearest m_i fitter policies give the smallest mean, so $s_i = \min_S a_S$ and $s_i' = \min_S a_S'$. Each subset mean satisfies $|a_S' - ca_S| \le 2\delta$ by Step 1. Since $c > 0$, taking minima preserves this uniform bound:

$$|s_i' - cs_i| = \left| \min_S a_S' - \min_S ca_S \right| \le \max_S |a_S' - ca_S| \le 2\delta.$$

Step 4: preserve the survivors. When fewer than K priorities are infinite, those members remain selected. After retraining, every previously selected finite priority is at least $cs_{(K)} - 2\delta$, and every previously excluded priority is at most $cs_{(K+1)} + 2\delta$. The assumed gap keeps every selected priority strictly above every excluded priority. When at least K priorities are infinite, only the fixed tie rule determines the survivors. □

Proposition 3 gives two ways to preserve archive decisions: smaller descriptor changes after accounting for a similarity transformation, or larger gaps at the neighbor and survivor cutoffs. VC has no term that aligns the old and new descriptors. The freedom to choose different metrics in Proposition 2 also gives no bound on how much selection-relevant geometry changes between retraining. Whether VC reduces δ or increases the relevant gaps remains an empirical question.

The reported best-linear alignment residual (Figure 6b) measures the error after fitting a linear map between the old and new descriptors of a fixed trajectory bank. This measurement differs from δ in two ways: the fitted map can stretch different directions by different amounts, and the residual combines errors across the bank instead of bounding the error of every policy. For example, $z_i' = Az_i$ gives zero best-linear residual for any invertible A, even though the neighbors can change as in the POINTMAZE study presented in Section 3. Neighbor overlap measures the observed change in neighbor identities directly, but does not verify the sufficient condition in the proposition.

B.2.4 Local Quality Competition and Parent–Child Comparisons

We now hold the encoder fixed and examine what happens when a candidate is added to the pool. Only strictly fitter policies contribute to a policy's priority. A fitter candidate can therefore affect the priority of a worse policy, while that worse policy cannot contribute to the candidate's priority.

Proposition 4 (Local quality asymmetry)**.** *Fix the existing descriptors and use up to k strictly fitter neighbors to compute each priority. Suppose policy i is already in the pool and has at least k fitter alternatives. Let $r_{i,k}$ be the distance to its farthest included neighbor. Add candidate c with $f_c > f_i$ and distance $d_{ic} = \|z_i - z_c\|_2 < r_{i,k}$. Then*

$$s_i^{\text{new}} = s_i^{\text{old}} - \frac{r_{i,k} - d_{ic}}{k}.$$

Candidate c lowers the priority of policy i, while i cannot lower the priority of c because i is not strictly fitter. If $f_c < f_i$, only i can enter the other's strictly-fitter set. If $f_c = f_i$, neither policy enters the other's strictly fitter set.

Proof. The old priority is the mean of the k smallest distances to fitter policies. The new distance $d_{ic} < r_{i,k}$ replaces one occurrence of the largest included distance and leaves the other $k-1$ terms unchanged. Subtracting the two means gives the formula. Since $f_i < f_c$, policy i cannot enter the strictly-fitter set of c. □

The number of existing fitter policies matters. If i has m fitter alternatives with $0 < m < k$, adding c increases the number used in the mean from m to $m+1$, giving

$$s_i^{\text{new}} - s_i^{\text{old}} = \frac{d_{ic} - s_i^{\text{old}}}{m+1}.$$

The priority decreases only when $d_{ic} < s_i^{\text{old}}$. If $m = 0$, adding a fitter candidate changes the priority from infinity to a finite distance.

Corollary 3 (Parent and child)**.** *For a parent and child with unequal fitness, only the fitter member can enter the other's strictly-fitter set. Adding the fitter member to a pool that omits it lowers the worse member's priority under the distance and neighbor-count conditions above. This result does not determine which member has the higher priority or guarantee that the fitter member survives.*

Proof. Apply Proposition 4, or the formula for $m < k$, with the worse member as i and the fitter member as the added policy. □

Other fitter policies can give the better parent or child an even lower priority, and the archive selects survivors from the whole pool. For instance, use one-dimensional descriptors and $k = 3$. Place a parent at 0 with fitness 1, its child at 0.1 with fitness 2, and three policies at $0.11, 0.12, 0.13$, all with fitness 3. The child has priority $(0.01 + 0.02 + 0.03)/3 = 0.02$, while the parent has priority $(0.10 + 0.11 + 0.12)/3 = 0.11$. At capacity four, the parent survives with three highest-fitness policies, which have infinite priority, and the fitter child is rejected. A nearby fitter parent or child can therefore lower the other's priority without determining which one survives. This proposition explains how parent-child proximity affects selection. The experiments measure how VC changes that proximity.

B.3 Archive Persistence and Parent Sampling

A policy that remains in the archive has more opportunities to be selected as a parent. To express this relationship, suppose the archive contains K policies before each parent-sampling step. Let I_{it} equal one when policy i is present at iteration t and zero otherwise. Let $\mathcal{F}_t$ contain the search history before sampling, including the current archive and the number b_t of parent draws to make. Each draw samples uniformly with replacement, so every current member has probability $1/K$ of being chosen and can be chosen more than once. If N_{it} counts the draws of policy i at iteration t and T is a finite number of iterations, then

$$\mathbb{E}[N_{it} \mid \mathcal{F}_t] = \frac{b_t I_{it}}{K}, \qquad \mathbb{E}\left[\sum_{t=1}^{T} N_{it}\right] = \frac{1}{K}\mathbb{E}\left[\sum_{t=1}^{T} b_t I_{it}\right]. \tag{17}$$

The first expectation averages over the parent draws given the search history so far. It follows by adding b_t probabilities I_{it}/K of drawing policy i. The second equality follows by averaging over all possible search histories and adding the expected parent draws across the T iterations. This argument allows parent draws and their descendants to affect future archive membership. When $b_t = b$ is constant, the expected number of draws equals b/K times the expected number of iterations in which policy i is present.

For a simpler fixed-duration case, suppose policy i remains in the archive for L externally fixed iterations and the counts b_t are fixed independently of the draws. There are then $M = \sum_{t=1}^{L} b_t$ independent parent draws. If N_i is the total number of times policy i is drawn, its probability of being drawn at least once is

$$\Pr(N_i \geq 1) = 1 - \left(1 - \frac{1}{K}\right)^M.$$

The probability of missing i on every draw is $(1 - 1/K)^M$, so subtracting it from one gives the formula. If each draw of i independently produces a useful descendant with probability q, the probability of at least one such descendant is $1 - (1 - q/K)^M$. These product formulas require the fixed-duration assumptions above. They cannot be applied by conditioning on the observed lifetime in adaptive search, because that lifetime can depend on the draws themselves. Equation 17 describes expected sampling opportunities; the formulas do not establish the quality of descendants or an effect on final fitness.

B.3.1 Scope of the Causal Interpretation

The formal results explain how descriptor geometry affects archive selection and how archive residence affects expected parent sampling. The changes observed under VC are empirical findings, rather than guarantees of the regularization objective. The interventions to the selection mechanism in Appendix P provide evidence that survivor choices contribute to the search outcomes.

B.4 PointMaze Study and Results

The controlled PointMaze study starts with 720 valid policies from a short search guided by the task's physical descriptors. Deterministic perturbations expand the pool to $8,192$ candidates. Every condition uses the same policy identities, saved trajectories, returns, physical endpoints, and candidate order. We train five independently initialized autoencoders, each producing a two-dimensional descriptor. These provide five learned representations of the same candidate pool.

For each encoder, we apply $A_{\kappa,\theta} = R_\theta \operatorname{diag}(\kappa, 1/\kappa) R_\theta^{\mathsf{T}}$ for $\kappa \in \{1, 2, 4, 8, 16, 32\}$ and four orientations relative to the latent principal axes. Here R_θ rotates the two-dimensional descriptors by angle θ, and $\operatorname{diag}(\kappa, 1/\kappa)$ contains the two scale factors. The transformation stretches one direction by κ and contracts the perpendicular direction by $1/\kappa$. The angle θ controls which directions are stretched and contracted. Applying the inverse transformation before decoding preserves reconstruction exactly in real arithmetic. For the directly decoded orientation-0° audits, the largest absolute decoder-output deviation is 2.03×10^{-3}, below the validation threshold 2.5×10^{-3}; the illustrated $\kappa = 2$ condition differs by at most 5.49×10^{-6}. The identity control case, $\kappa = 1$, leaves the descriptors unchanged. We evaluate selection from the complete pool and sequential archive updates using the same candidate stream in the original order.

We measure reconstruction error, maximum fitness, recall of the up-to-three strictly fitter neighbors, and Jaccard agreement of survivor identities. Neighbor recall is the fraction of the original fitter neighbors that remain neighbors after transformation. Jaccard agreement divides the number of identities shared by the original and transformed survivor sets by the number in their union. For physical coverage and QD score, we use the passive evaluator in Appendix A. It assigns policies to a fixed centroidal Voronoi tessellation (CVT) archive using their physical x–y endpoints. This partition stays fixed while the learned descriptors change, so all conditions use the same geometry for physical evaluation.

Figure 1 summarizes neighbor recall and survivor agreement across all five autoencoders and four orientations. The identity transform reproduces every decision, while anisotropic transformations change fitter-neighbor identities and the selected physical repertoire. The example panels use the generator's fixed demonstration condition: encoder seed 0, $\kappa = 2$, and orientation 0°. Maximum

fitness stays unchanged because the highest-fitness candidates have infinite priority and selection retains at least one of them. Lower neighbor recall or survivor agreement need not consistently increase or decrease passive coverage or QD score. Different survivors can occupy more or fewer physical cells and can have different fitness within these cells. The construction shows that descriptor geometry can change archive selection while preserving reconstruction; it does not show that applying these transformations improves search.

C ALGORITHM PSEUDOCODES

We separate the active search archive $\mathcal{A}$ from the passive evaluator $\mathcal{P}$ that is also updated during the training. In original AURORA, $\mathrm{ActiveUpdate}$ denotes its adaptive unstructured-container insertion rule. In our experiments, this symbol refers to the archive update rule in Equation 5. We combine each archive with the new candidates, score each policy by its distance to up to three strictly fitter neighbors and keep the M policies with highest scores. The passive CVT archive evaluator $\mathcal{P}$ is only used for analysis purposes. It accumulates selected active members under an external descriptor, but never supplies parents or controls admission to $\mathcal{A}$. For VC variants, both initial and subsequent representation updates use Equation 3.

Algorithm 1: Pseudocode for the AURORA algorithm.

Input: initial policies Θ_0; encoder-retraining schedule $\mathcal{U}$; active update rule; genetic batch size N

$(\tau_i, f_i)_{i\in\Theta_0} \leftarrow \texttt{Evaluate}(\Theta_0)$ *// run policies to obtain trajectories τ_i and fitnesses f_i*
Train encoder E_ϕ and decoder D_ψ on $\{\tau_i\}$ with $\mathcal{L}_{\text{rec}}$ *// learn to reconstruct trajectories*
$z_i \leftarrow E_\phi(\tau_i)$ for every $i \in \Theta_0$ *// z_i: learned behavior descriptor of policy i*
$\mathcal{A} \leftarrow \texttt{ActiveUpdate}(\varnothing, \Theta_0; z, f)$ *// initialize the search archive*
$\mathcal{P} \leftarrow \texttt{PassiveUpdate}(\varnothing, \Theta_0)$ *// evaluate diversity using external descriptors*
for $g = 1, 2, \ldots$ **do**
  Sample N parents uniformly from $\mathcal{A}$ *// select policies for reproduction*
  $\mathcal{C} \leftarrow \texttt{GeneticVariation}(\text{parents})$ *// perturb parents to create candidate policies*
  $(\tau_i, f_i)_{i\in\mathcal{C}} \leftarrow \texttt{Evaluate}(\mathcal{C})$ *// run candidates to obtain trajectories and fitnesses*
  $z_i \leftarrow E_\phi(\tau_i)$ for every $i \in \mathcal{C}$ *// compute learned descriptors with the current encoder*
  $\mathcal{A} \leftarrow \texttt{ActiveUpdate}(\mathcal{A}, \mathcal{C}; z, f)$ *// select policies from the archive and candidates*
  $\mathcal{P} \leftarrow \texttt{PassiveUpdate}(\mathcal{P}, \mathcal{A})$ **if** $g \in \mathcal{U}$ **then**
    // Retrain on the behaviors currently retained in the search archive
    Train E_ϕ, D_ψ on trajectories retained in $\mathcal{A}$ with $\mathcal{L}_{\text{rec}}$
    $z_i \leftarrow E_\phi(\tau_i)$ for every $i \in \mathcal{A}$ *// refresh descriptors from stored trajectories*
    $\mathcal{A} \leftarrow \texttt{Rebuild}(\mathcal{A}; z, f)$ *// rebuild the archive in the updated descriptor space*

Output: active search archive $\mathcal{A}$ and passive evaluation archive $\mathcal{P}$

Algorithm 2: Pseudocode for the PGA-AURORA algorithm.

Input: initial policies Θ_0; encoder-retraining schedule $\mathcal{U}$; active update rule; batch size N; replay buffer $\mathcal{B}$; critic Q_ω

$(\tau_i, f_i, \mathcal{T}_i)_{i\in\Theta_0} \leftarrow \texttt{Evaluate}(\Theta_0)$ *// collect trajectories, fitnesses, and transitions $\mathcal{T}_i$*
Insert $\{\mathcal{T}_i\}$ into $\mathcal{B}$ *// store experience for critic training*
Train encoder E_ϕ and decoder D_ψ on $\{\tau_i\}$ with $\mathcal{L}_{\text{rec}}$
$z_i \leftarrow E_\phi(\tau_i)$ for every $i \in \Theta_0$ *// compute learned behavior descriptors*
$\mathcal{A} \leftarrow \texttt{ActiveUpdate}(\varnothing, \Theta_0; z, f)$
$\mathcal{P} \leftarrow \texttt{PassiveUpdate}(\varnothing, \Theta_0)$
for $g = 1, 2, \ldots$ **do**
  Sample parent batches $S_{\text{GA}}, S_{\text{PG}}$ from $\mathcal{A}$
  $\mathcal{C}_{\text{GA}} \leftarrow \texttt{GeneticVariation}(S_{\text{GA}})$ *// perturb parent parameters*
  $\mathcal{C}_{\text{PG}} \leftarrow \texttt{CriticVariation}(S_{\text{PG}}, Q_\omega)$ *// use critic gradients to improve policies*
  $\mathcal{C} \leftarrow \mathcal{C}_{\text{GA}} \cup \mathcal{C}_{\text{PG}} \cup \{\text{greedy actor}\}$, with $|\mathcal{C}_{\text{GA}}| = N/2$ and $|\mathcal{C}_{\text{PG}}| = N/2 - 1$
  $(\tau_i, f_i, \mathcal{T}_i)_{i\in\mathcal{C}} \leftarrow \texttt{Evaluate}(\mathcal{C})$ *// run the combined candidate batch*
  Insert $\{\mathcal{T}_i\}$ into $\mathcal{B}$
  $z_i \leftarrow E_\phi(\tau_i)$ for every $i \in \mathcal{C}$
  $\mathcal{A} \leftarrow \texttt{ActiveUpdate}(\mathcal{A}, \mathcal{C}; z, f)$
  $\mathcal{P} \leftarrow \texttt{PassiveUpdate}(\mathcal{P}, \mathcal{A})$
  Update Q_ω and the critic-guided policy using minibatches from $\mathcal{B}$
  if $g \in \mathcal{U}$ **then**
    Train E_ϕ, D_ψ on trajectories retained in $\mathcal{A}$ with $\mathcal{L}_{\text{rec}}$
    $z_i \leftarrow E_\phi(\tau_i)$ for every $i \in \mathcal{A}$ *// refresh descriptors after retraining*
    $\mathcal{A} \leftarrow \texttt{Rebuild}(\mathcal{A}; z, f)$

Output: $\mathcal{A}$, $\mathcal{P}$, replay buffer $\mathcal{B}$, and critic Q_ω

Algorithm 3: Fitness-relative control based on XCON, as evaluated in Appendix F.

Input: retained trajectories and fitnesses $\{(\tau_i, f_i)\}_{i\in\mathcal{A}}$; encoder E_ϕ; width d; active capacity cutoff $d_{\min}$; epochs K

Sample $|\mathcal{A}|$ valid archive indices with replacement to form I *// bootstrap the training set*

$\alpha \leftarrow d\, d_{\min}$ *// set the adaptive triplet margin*

Initialize a fixed triplet table $T \leftarrow \varnothing$

foreach *sampled position a in I* **do**

  Sample two distinct positions u, v in I such that $f_u \neq f_a$ and $f_v \neq f_a$

  // Choose the positive p by the smaller fitness gap to anchor a; n is the negative

  $(p, n) \leftarrow (u, v)$ if $|f_a - f_u| \leq |f_a - f_v|$, and (v, u) otherwise

  Append (a, p, n) to T

for $k = 1, \ldots, K$ **do**

  Shuffle the fixed table T and split it into minibatches *// reuse the same triplets across epochs*

  foreach *minibatch $B \subset T$* **do**

    $z_j \leftarrow E_\phi(\tau_j)$ for every index j appearing in B *// encode all three members of each triplet*

    $\mathcal{L}_{\mathrm{XCon}} \leftarrow \sum_{(a,p,n)\in B} \left[\|z_a - z_p\|_2^4 - \|z_a - z_n\|_2^4 + \alpha\right]_+$

    Update ϕ by descending $\mathcal{L}_{\mathrm{XCon}}$ *// favor smaller anchor–positive distances*

$z_i \leftarrow E_\phi(\tau_i)$ for every $i \in \mathcal{A}$

Output: updated encoder E_ϕ and incumbent embeddings $\{z_i\}$

Algorithm 4: Pseudocode for the VC representation update. For AURORA-VC, $\mathcal{L}_0 = \mathcal{L}_{\mathrm{rec}}$; for XCON-VC, $\mathcal{L}_0 = \mathcal{L}_{\mathrm{XCon}}$.

Input: retained trajectories; representation parameters Θ containing encoder E_ϕ and, for reconstruction, decoder D_ψ; width d; base loss $\mathcal{L}_0$; weights λ_V, λ_C; target standard deviation γ; stabilizer ϵ; epochs K

Draw one bootstrap training set I from valid retained trajectories (or use the fixed triplet table for XCON)

for $k = 1, \ldots, K$ **do**

  Shuffle I and split it into minibatches

  foreach *minibatch B* **do**

    $Z \leftarrow [E_\phi(\tau_i)]_{i\in B} \in \mathbb{R}^{|B|\times d}$ *// encode trajectories; use only anchors for XCON's VC terms*

    $\bar{z} \leftarrow |B|^{-1} \sum_{i\in B} E_\phi(\tau_i)$

    $\widetilde{Z} \leftarrow Z - \mathbf{1}\bar{z}^{\mathsf{T}}$ *// subtract the batch mean from each coordinate*

    $\Sigma_B \leftarrow \widetilde{Z}^{\mathsf{T}}\widetilde{Z} / \max(|B| - 1, 1)$ *// compute the sample covariance*

    $\mathcal{L}_V \leftarrow \frac{1}{d}\sum_{q=1}^{d} \left[\gamma - \sqrt{(\Sigma_B)_{qq} + \epsilon}\right]_+$ *// penalize coordinates with too little variation*

    $\mathcal{L}_C \leftarrow \frac{1}{d}\sum_{q\neq r} (\Sigma_B)_{qr}^2$ *// penalize covariance between different coordinates*

    Compute the base loss $\mathcal{L}_0$ on the minibatch

    $\mathcal{L} \leftarrow \mathcal{L}_0 + \lambda_V\mathcal{L}_V + \lambda_C\mathcal{L}_C$

    Update the parameters in Θ by descending $\mathcal{L}$

  Stop if the epoch-mean training loss meets the early-stopping rule

$z_i \leftarrow E_\phi(\tau_i)$ for every $i \in \mathcal{A}$

Output: updated representation parameters and incumbent embeddings $\{z_i\}$

After either representation retraining procedure, the enclosing AURORA or PGA-AURORA loop re-encodes incumbents and rebuilds the same active container.

D ENVIRONMENTS AND HYPERPARAMETERS

The hyperparameters used for the AURORA and the PGA-AURORA baselines are given in Table 2.

Table 2: Baseline parameter settings for the experiments.

Parameter	Value
Standard seeds	0–9
Search iterations / candidate batch	5,000 / 512
Evaluated proposals per run	2.56 million
Ordinary metric logging interval	every 5 search iterations
Active archive rule / capacity	DNS-style ranking / 1,024 policies
Inactive threshold-controller field	`target_size=922`
Strictly fitter neighbors	up to 3
Passive CVT cells / initialization samples	1,024 / 50,000
Primary descriptor width	10
Representation minibatch / max. epochs	128 / 200 (XCON minibatch: 512)
Optimizer / learning rate	Adam / 10^{-2}
Early-stopping improvement / patience	5×10^{-4} / 10 epochs
Encoder / decoder	single-layer LSTMs, width 10
Encoder retraining events after warm start	31
VC $(\lambda_V, \lambda_C, \gamma, \epsilon)$	$(1, 1, 1, 10^{-4})$
Iso+LineDD coefficients (default)	$(0.005, 0.05)$
Iso+LineDD coefficients (Kheperax)	$(0.2, 0)$
PGA proposal allocation	256 genetic, 255 PG mutations, 1 greedy actor
Brax / Kheperax policy layers	$[128, 128]$ / $[5]$
PGA critic hidden layers	$[256, 256]$
Replay capacity	10^6
Critic / policy-gradient steps per iteration	3,000 / 150
Transition minibatch	100
Critic / greedy / policy learning rates	$(3 \times 10^{-4}, 3 \times 10^{-4}, 5 \times 10^{-3})$
Discount / target coefficient	0.99 / 0.005
Policy noise / clip / delay	0.2 / 0.5 / 2

Table 3 summarizes the task specifications, encoder inputs, and passive-evaluation settings for the ten environments.

Table 3: Environment contracts. Trajectory shape is sampled steps by features; BD is the external descriptor used only by the passive evaluator.

Environment	Horizon	Policy obs.	Actions	Trajectory shape	Encoder input	BD	QD offset
HALFCHEETAH	1,000	18	6	100×18	sensory	2 contacts	9,231
WALKER	1,000	17	6	100×17	sensory	2 contacts	1,413
ANTOMNI	1,000	87	8	100×25	sensory	2 final-xy	3,000
ANTUNI	1,000	87	8	100×25	sensory	4 contacts	3,240
POINTMAZE	250	2	2	25×2	task BD	2 final-xy	585.775
ANTMAZE	1,000	99	8	100×30	sensory	2 final-xy	57
ANTTRAP	1,000	93	8	100×25	sensory	2 final-xy	3,380
HUMANOIDOMNI	1,000	227	17	100×25	sensory	2 final-xy	0
HUMANOIDUNI	1,000	227	17	100×75	sensory	2 contacts	0
KHEPERAX	200	5	2	50×5	sensory	2 final-xy	$100\sqrt{2}$

Sensory denotes the leading channels of the policy observation vector, with the retained channel count given by the trajectory width. For HALFCHEETAH and WALKER, these are the full 18- and 17-dimensional observations of torso height/orientation, joint angles, and body and joint velocities. For ANTOMNI, ANTUNI, and ANTTRAP, the encoder retains the first 25 observation channels; ANTMAZE retains the first 30 of its 99 policy-observation channels. HUMANOIDOMNI retains the first 25 channels describing the torso height and orientation, joint configuration, and torso linear

velocity. HUMANOIDUNI retains the first 75 channels, additionally including angular and joint velocities and part of the inertia-derived observation block.

For KHEPERAX, sensory input consists of three laser-range and two bumper-contact measurements. These observation trajectories are sampled every 10 simulation steps for the Brax tasks and every four steps for KHEPERAX. Global horizontal position is omitted from policy and encoder inputs in the position-based Brax tasks other than POINTMAZE.

The 2.56-million proposal budget excludes 512 initial policies. Encoder training occurs at initialization and at search iterations $5j(j+1)$ for $j = 1, \ldots, 31$, ending at iteration 4,960. Kheperax time to first goal is measured at the five-iteration logging resolution (2,560 proposals), excluding initialization.

E PRIMARY RESULTS FOR PGA-AURORA AND AURORA WITH VC REGULARIZATION

Variance-Covariance regularization improves the best-policy fitness with both proposal mechanisms, while its effects on passive QD-score and coverage depend on the task. We compare reconstruction-only MSE with VC regularization on ten environments and ten matched seeds for PGA-AURORA v. PGA-AURORA-VC and AURORA v. AURORA-VC. Table 1 gives the final values for the median values of each baseline for each environment, while we describe the detailed results and statistical tests and present the learning curves with endpoint distributions in this section.

E.1 RESULTS FOR PGA-AURORA WITH VC REGULARIZATION

The largest maximum-fitness gains occur in WALKER, where the ten-seed median rises from 3.85×10^3 to 6.51×10^3, and ANTUNI, where it rises from 6.58×10^3 to 7.83×10^3. HALFCHEETAH also improves from 7.01×10^3 to 7.30×10^3. PGA-AURORA-VC wins all ten matched seeds in WALKER and ANTOMNI, nine in HALFCHEETAH, and seven in ANTUNI. The median maximum fitness in ANTMAZE improves from -14.1 to -7.83. ANTTRAP and both Humanoid tasks have mixed seed-level outcomes, and POINTMAZE shows little terminal separation within the given budget.

Passive QD score does not improve in correlation with maximum fitness. In HALFCHEETAH, PGA-AURORA-VC improves maximum fitness in nine of ten seed pairs while reducing passive QD score in all ten. In contrast, passive QD score favors PGA-AURORA-VC in nine of ten HUMANOIDUNI pairs, nine KHEPERAX pairs, and eight WALKER pairs. These outcomes distinguish finding a stronger policy from filling the external descriptor space with strong policies.

Every KHEPERAX run reaches the goal fitness of zero. VC takes more evaluations to reach it for the first time. The median evals to first goal rises from 1.82×10^5 to 3.44×10^5 ($p = 0.0293$). Its final passive QD score rises from 4.24×10^4 to 5.12×10^4.

E.2 RESULTS FOR AURORA WITH VC REGULARIZATION

The fitness improvement observed in PGA-AURORA-VC does transfer to AURORA. For plain AURORA, the marginal median maximum-fitness rises from 1.66×10^3 to 2.23×10^3 in WALKER and from 1.58×10^3 to 1.96×10^3 in ANTUNI. Paired Wilcoxon tests favor VC on maximum fitness in ANTOMNI, ANTUNI, HALFCHEETAH, and WALKER. The remaining tasks do not show a significant maximum-fitness difference at the 0.05 threshold, and KHEPERAX again ties at zero.

At raw $p < 0.05$, passive QD score favors VC in ANTUNI, HUMANOIDOMNI, HUMANOIDUNI, KHEPERAX, and WALKER, but favors MSE in ANTOMNI and HALFCHEETAH; other median differences are not significant at this threshold. In KHEPERAX, the median evaluations to first goal increase from 3.16×10^5 to 4.80×10^5, but this time-to-goal difference is not significant ($p = 0.25$).

E.3 STATISTICAL ANALYSIS OF THE RESULTS

At raw $p < 0.05$, maximum fitness favors VC in ANTOMNI, ANTUNI, HALFCHEETAH, and WALKER for both algorithms. For PGA-AURORA, passive QD score favors VC in HUMANOIDUNI, KHEPERAX, and WALKER, and favors MSE in HALFCHEETAH. For AURORA, the seven significant QD comparisons are listed above in Section E.2. Passive coverage favors VC in KHEPERAX for both algorithms and additionally in ANTUNI and HUMANOIDOMNI for AURORA. Coverage favors MSE in HALFCHEETAH for both algorithms, in POINTMAZE for PGA-AURORA and in HUMANOIDUNI for AURORA. The KHEPERAX entries in the maximum-fitness block report solve-speed, with fewer evaluations indicating faster goal discovery.

Table 4: Statistical tests for PGA-AURORA–PGA-AURORA-VC and AURORA–AURORA-VC over ten matched seeds per environment. PA denotes PGA-AURORA and AU denotes AURORA. Entries are raw two-sided paired Wilcoxon p values, evaluated separately within each environment without multiple-testing correction. Bold marks $p < 0.05$. Arrows show the sign of the median paired VC-minus-baseline difference: $\uparrow$ means an increase, $\downarrow$ a decrease, and $=$ a zero median difference. For KHEPERAX, the maximum-fitness block reports evaluations to first goal, where lower is better; higher is better elsewhere.

	Maximum fitness		Passive QD score		Passive coverage	
Environment	PA–PA-VC	AU–AU-VC	PA–PA-VC	AU–AU-VC	PA–PA-VC	AU–AU-VC
ANTMAZE	0.1602 $\uparrow$	0.6953 $\downarrow$	0.2754 $\uparrow$	0.9219 $\downarrow$	0.5566 $\uparrow$	0.6250 $\uparrow$
ANTOMNI	**0.0020** $\uparrow$	**0.0039** $\uparrow$	0.1309 $\downarrow$	**0.0020** $\downarrow$	0.7500 $=$	1.0000 $=$
ANTTRAP	0.9219 $\uparrow$	0.9219 $\uparrow$	0.6953 $\downarrow$	1.0000 $\uparrow$	0.6953 $\uparrow$	0.9219 $\uparrow$
ANTUNI	**0.0273** $\uparrow$	**0.0039** $\uparrow$	0.7695 $\downarrow$	**0.0020** $\uparrow$	0.1934 $\downarrow$	**0.0098** $\uparrow$
HALFCHEETAH	**0.0039** $\uparrow$	**0.0098** $\uparrow$	**0.0020** $\downarrow$	**0.0020** $\downarrow$	**0.0020** $\downarrow$	**0.0020** $\downarrow$
HUMANOIDOMNI	0.9219 $\uparrow$	0.5566 $\uparrow$	0.7695 $\uparrow$	**0.0020** $\uparrow$	0.7656 $=$	**0.0078** $\uparrow$
HUMANOIDUNI	0.6953 $\uparrow$	0.5566 $\downarrow$	**0.0039** $\uparrow$	**0.0039** $\uparrow$	0.1055 $\downarrow$	**0.0098** $\downarrow$
KHEPERAX	**0.0293** $\uparrow$	0.2500 $\uparrow$	**0.0137** $\uparrow$	**0.0020** $\uparrow$	**0.0195** $\uparrow$	**0.0020** $\uparrow$
POINTMAZE	0.2754 $\downarrow$	0.4922 $\uparrow$	0.5566 $\downarrow$	0.3750 $\uparrow$	**0.0078** $\downarrow$	0.8848 $\uparrow$
WALKER	**0.0020** $\uparrow$	**0.0020** $\uparrow$	**0.0098** $\uparrow$	**0.0098** $\uparrow$	0.3750 $\uparrow$	0.0645 $\uparrow$

Figures 8–10 present the learning curves and endpoint distributions for all ten environments. Figure 8 shows the PGA-AURORA–PGA-AURORA-VC learning curves, Figure 9 shows their terminal seed distributions, and Figure 10 shows the AURORA-AURORA-VC learning curves.

E.4 DISCUSSION

Both proposal mechanisms benefit from VC on maximum fitness in several environments, including ANTUNI and WALKER. The transfer to plain AURORA shows that these gains do not require policy-gradient proposals. The QD-scores and coverage results nevertheless depend on the task, and the KHEPERAX time-to-goal comparison shows that faster goal discovery and a higher quality final repertoire need not favor the same objective. The paired tests describe each environment separately; their unadjusted p values do not establish a suite-wide improvement in every outcome. The sensitivity studies in Appendices J–L examine how these conclusions change with representation and archive settings.

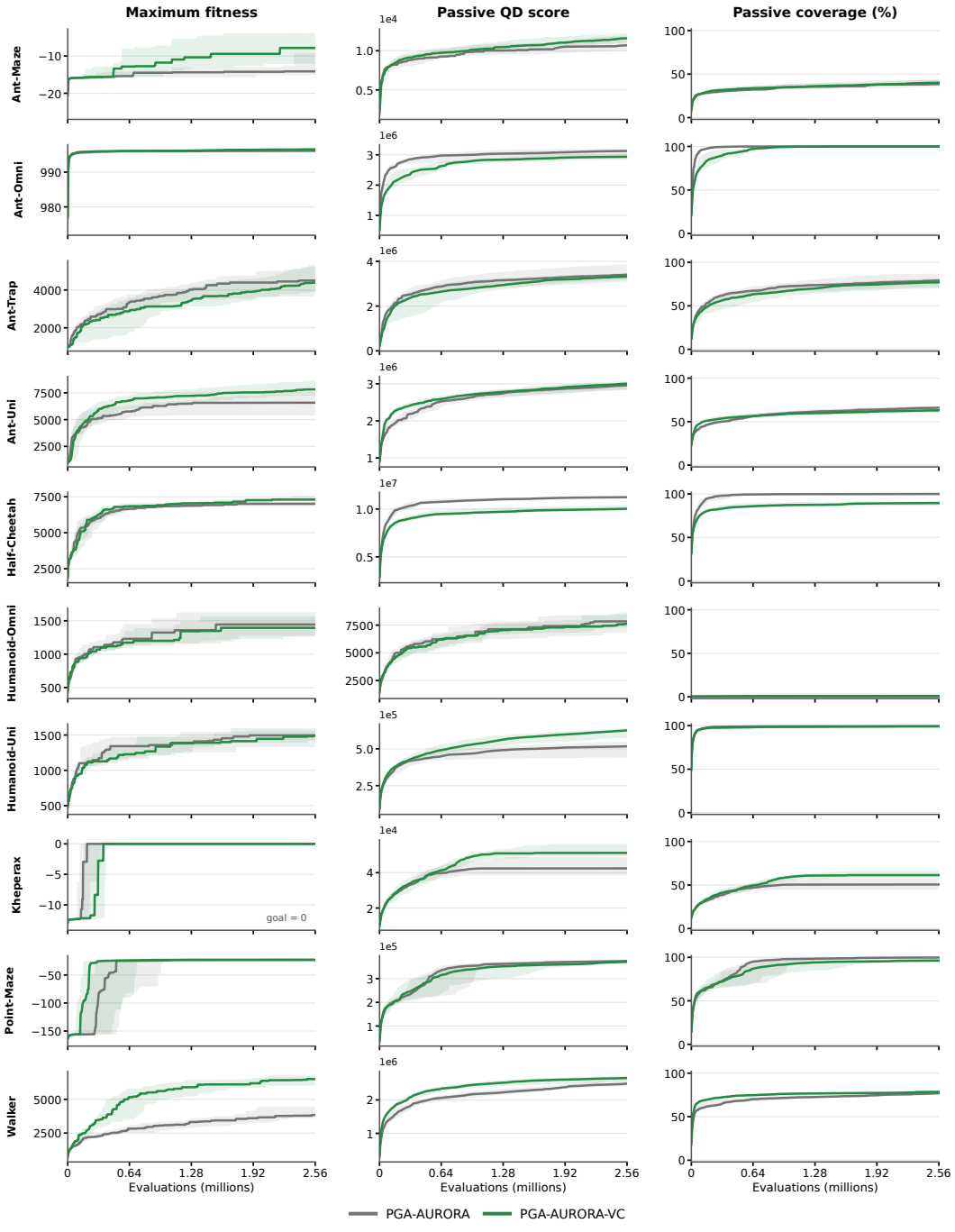

Figure 8: **Primary PGA-AURORA–PGA-AURORA-VC learning curves.** Rows show all 10 training environments. Columns are maximum fitness, passive QD score, and passive coverage. Lines show the median over ten seeds for PGA-AURORA and PGA-AURORA-VC only; bands show pointwise nonparametric 95% seed-bootstrap intervals for the median. Kheperax maximum fitness has goal value zero.

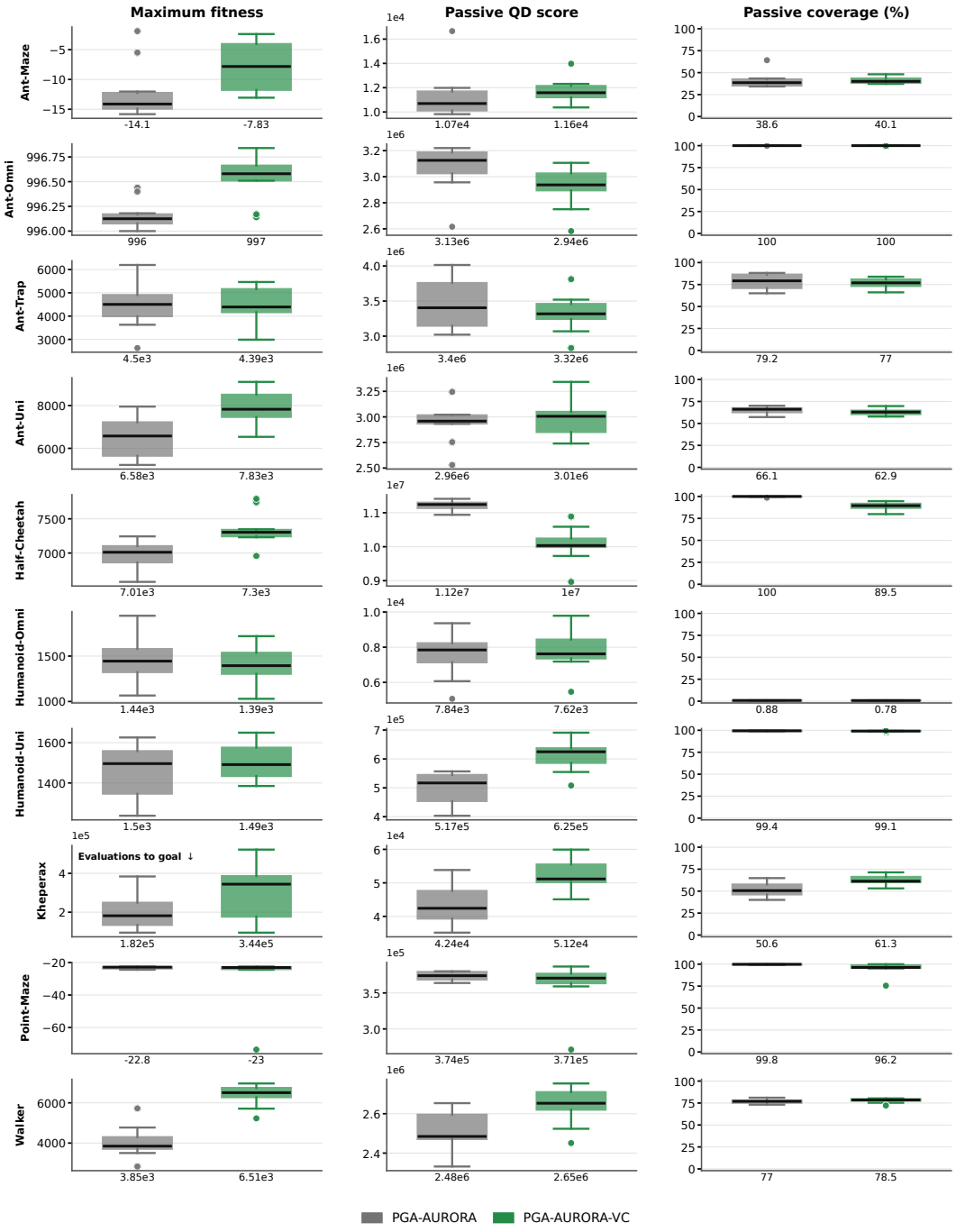

Figure 9: **Primary PGA-AURORA–PGA-AURORA-VC endpoint distributions.** Rows show all 10 training environments. Each box summarizes ten seeds for PGA-AURORA or PGA-AURORA-VC: the box spans the interquartile range, the center line is the median, and whiskers end at the most extreme observations within 1.5 interquartile ranges. Outliers are drawn as points, and the raw median is printed beneath every box. For Kheperax, the maximum-fitness panel reports evaluations to first goal, for which lower is better.

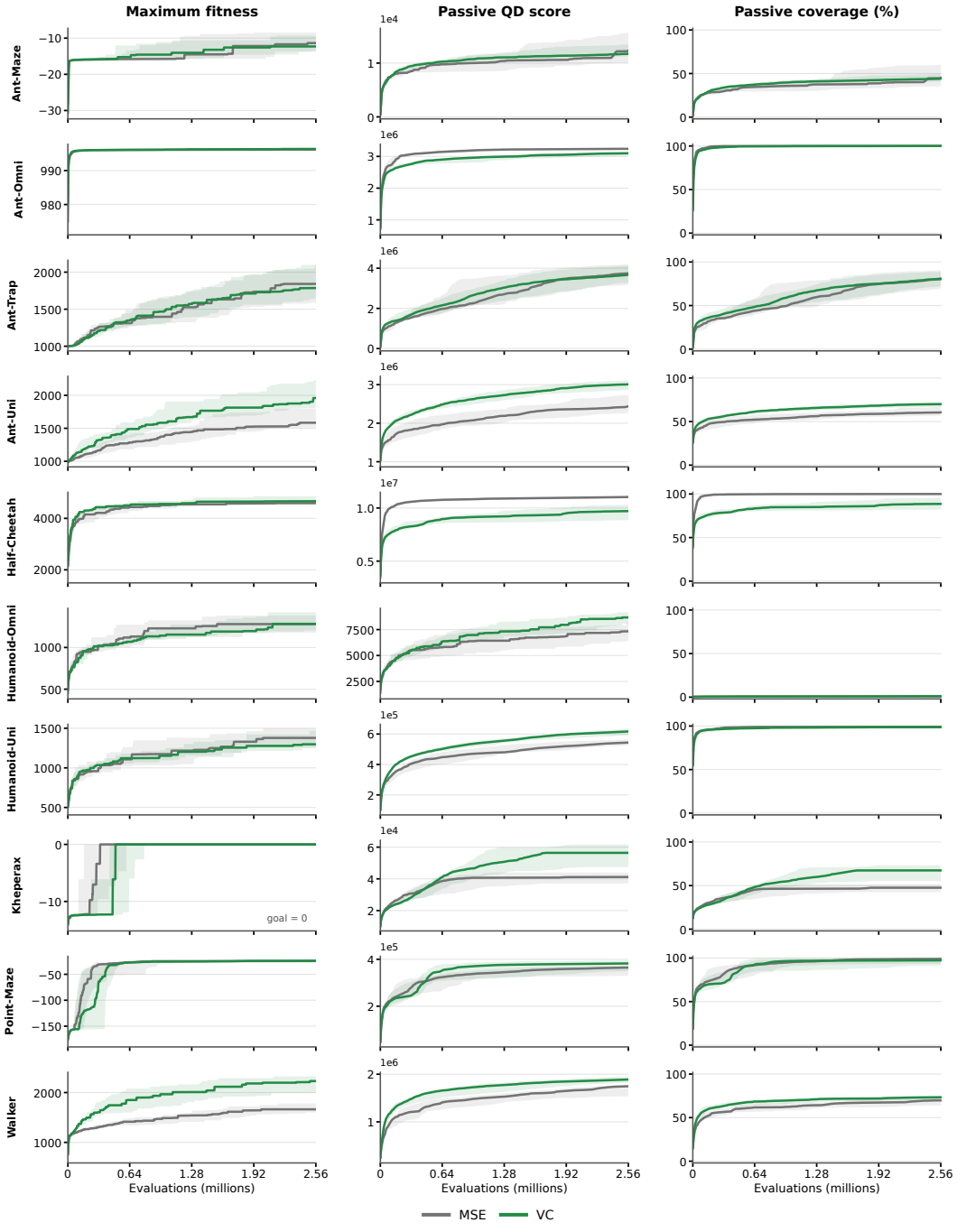

Figure 10: **AURORA–AURORA-VC transfer learning curves.** Lines show the median over ten seeds for AURORA and AURORA-VC; bands show pointwise nonparametric 95% seed-bootstrap intervals for the median. Kheperax maximum fitness has goal value zero.

F COMPARISON WITH XCON AND ARCHIVE EXTINCTION

We compare PGA-AURORA-VC with periodic archive extinction and fitness-relative representation learning to understand how VC relates to other ways of changing exploration. We add VC to each alternative to test whether its contribution persists under a different archive-management rule or representation objective.

The extinction variant, PGA-AURORA-X, retains trajectory reconstruction but periodically reduces the active archive to its champion and a random subset totaling approximately 5% of capacity, then allows it to regrow through ordinary search. PGA-AURORA-XCON combines the same extinction mechanism with a fitness-relative triplet objective based on XCON, which uses fitness differences to determine which trajectories should lie closer in latent space (Coiffard et al., 2025). VC penalizes insufficient coordinate variance and off-diagonal covariance; these terms augment reconstruction in PGA-AURORA-VC and PGA-AURORA-X-VC, and the triplet objective in PGA-AURORA-XCON-VC. The XCON labels in this comparison denote the fitness-relative control defined in Algorithm 3; conclusions apply to that specified objective.

We evaluate six variants in three matched comparisons: PGA-AURORA–PGA-AURORA-VC, PGA-AURORA-X–PGA-AURORA-X-VC, and PGA-AURORA-XCON–PGA-AURORA-XCON-VC. All runs use a 5,000-iteration budget, ten-dimensional descriptors, and a 1,024-policy active-archive capacity. Within each pair, adding VC is the changed treatment; the extinction schedule is held fixed wherever extinction is used. Comparisons across pairs evaluate complete algorithms, since they also change archive retention or representation supervision.

We measure maximum fitness, passive QD score, and coverage using the same fixed 1,024-cell evaluator. Table 5 reports terminal medians and exploratory two-sided paired Wilcoxon tests at an uncorrected $p < 0.05$ threshold. For KHEPERAX, the fitness comparison uses evaluations to first goal because terminal fitness reaches the same ceiling. Figures 11 and 12 show learning curves and endpoint distributions for 500 of these runs, covering all variants except PGA-AURORA-X, which appears in the table.

F.1 RESULTS

VC can improve fitness alongside periodic extinction. The fitness benefit of VC extends to a search procedure that periodically removes most of its active archive. With extinction held fixed, adding VC raises median maximum fitness from 2.72×10^3 to 6.47×10^3 in WALKER and from 6.25×10^3 to 8.20×10^3 in ANTUNI; both paired tests favor VC. The repertoire response is more task dependent: passive QD score also improves in WALKER, while it decreases in ANTOMNI, ANTTRAP, ANTUNI, and HALFCHEETAH at the same threshold. Thus, the clearest benefit in this comparison is improved peak fitness in selected tasks.

Adding VC to XCON has a less consistent fitness effect. The gains seen with reconstruction and extinction do not carry over uniformly to the fitness-relative objective. XCON already reaches high median fitness in WALKER and ANTUNI, and adding VC leaves those medians nearly unchanged. No maximum-fitness comparison between PGA-AURORA-XCON and PGA-AURORA-XCON-VC crosses the testing threshold. The clearest detected change is instead in ANTUNI passive QD score, whose median rises from 2.72×10^6 to 2.84×10^6. The endpoint distributions in Figure 12 show the seed variability underlying these comparisons.

The relative strengths depend on the task and outcome. Direct comparisons place PGA-AURORA-VC above PGA-AURORA-XCON in median maximum fitness on WALKER and HALFCHEETAH, while PGA-AURORA-XCON leads on ANTTRAP and ANTUNI. The learning curves in Figure 11 also show that similar final fitness can accompany different repertoires. In KHEPERAX, for example, every variant reaches the goal, yet the plotted extinction variants attain higher passive QD score and coverage than PGA-AURORA-VC. Together, the curves and endpoint distributions distinguish finding a strong individual policy from accumulating a broad, high-quality repertoire.

F.2 DISCUSSION

We conclude that VC can complement periodic archive extinction, with fitness gains in WALKER and ANTUNI even when the archive is repeatedly reduced and regrown. The comparison also identifies reconstruction with VC as a competitive alternative to fitness-relative representation learning, with each approach showing strengths on different tasks. In PGA-AURORA-VC, reconstruction preserves trajectory information while VC encourages coordinate variation and limits linear redundancy, allowing behavioral differences to inform competition without direct fitness supervision of the encoder. This offers a way to retain distinctions between policies whose current returns are similar but whose trajectories provide different opportunities for subsequent search.

Table 5: Contextual extinction and XCON comparisons: terminal medians over ten matched seeds. PA denotes PGA-AURORA, X denotes periodic archive extinction, and both XCON arms use that same extinction schedule. The KHEPERAX maximum-fitness block reports evaluations to first goal (lower is better); higher is better elsewhere. Bold marks the better marginal median within each displayed pair when the raw two-sided paired Wilcoxon test gives $p < 0.05$, without multiplicity correction.

	PA	PA-VC	PA-X	PA-X-VC	PA-XCON	PA-XCON-VC
Maximum fitness						
ANTMAZE	-1.41×10^{1}	-7.83×10^{0}	-1.27×10^{1}	-1.31×10^{1}	-1.00×10^{1}	-1.26×10^{1}
ANTOMNI	9.96×10^{2}	$\mathbf{9.97\times10^{2}}$	9.96×10^{2}	9.96×10^{2}	9.96×10^{2}	9.96×10^{2}
ANTTRAP	4.50×10^{3}	4.39×10^{3}	4.69×10^{3}	6.17×10^{3}	5.85×10^{3}	3.63×10^{3}
ANTUNI	6.58×10^{3}	$\mathbf{7.83\times10^{3}}$	6.25×10^{3}	$\mathbf{8.20\times10^{3}}$	8.28×10^{3}	8.28×10^{3}
HALFCHEETAH	7.01×10^{3}	$\mathbf{7.30\times10^{3}}$	6.90×10^{3}	7.20×10^{3}	7.03×10^{3}	7.02×10^{3}
HUMANOIDOMNI	1.44×10^{3}	1.39×10^{3}	1.25×10^{3}	1.05×10^{3}	1.16×10^{3}	1.24×10^{3}
HUMANOIDUNI	1.50×10^{3}	1.49×10^{3}	1.26×10^{3}	1.29×10^{3}	1.35×10^{3}	1.30×10^{3}
KHEPERAX	$\mathbf{1.82\times10^{5}}$	3.44×10^{5}	1.63×10^{5}	1.86×10^{5}	2.07×10^{5}	1.51×10^{5}
POINTMAZE	-2.28×10^{1}	-2.30×10^{1}	-2.30×10^{1}	-2.43×10^{1}	-2.29×10^{1}	-2.29×10^{1}
WALKER	3.85×10^{3}	$\mathbf{6.51\times10^{3}}$	2.72×10^{3}	$\mathbf{6.47\times10^{3}}$	6.47×10^{3}	6.46×10^{3}
Passive QD score						
ANTMAZE	1.07×10^{4}	1.16×10^{4}	1.36×10^{4}	1.38×10^{4}	1.31×10^{4}	1.33×10^{4}
ANTOMNI	3.13×10^{6}	2.94×10^{6}	$\mathbf{3.17\times10^{6}}$	2.94×10^{6}	2.88×10^{6}	2.76×10^{6}
ANTTRAP	3.40×10^{6}	3.32×10^{6}	$\mathbf{4.20\times10^{6}}$	3.65×10^{6}	3.61×10^{6}	3.83×10^{6}
ANTUNI	2.96×10^{6}	3.01×10^{6}	$\mathbf{3.17\times10^{6}}$	2.59×10^{6}	2.72×10^{6}	$\mathbf{2.84\times10^{6}}$
HALFCHEETAH	$\mathbf{1.12\times10^{7}}$	1.00×10^{7}	$\mathbf{1.14\times10^{7}}$	9.99×10^{6}	1.07×10^{7}	1.08×10^{7}
HUMANOIDOMNI	7.84×10^{3}	7.62×10^{3}	5.97×10^{3}	5.42×10^{3}	6.64×10^{3}	6.85×10^{3}
HUMANOIDUNI	5.17×10^{5}	$\mathbf{6.25\times10^{5}}$	4.19×10^{5}	5.17×10^{5}	5.46×10^{5}	5.65×10^{5}
KHEPERAX	4.24×10^{4}	$\mathbf{5.12\times10^{4}}$	6.95×10^{4}	6.72×10^{4}	6.94×10^{4}	7.27×10^{4}
POINTMAZE	3.74×10^{5}	3.71×10^{5}	3.75×10^{5}	3.49×10^{5}	3.89×10^{5}	3.91×10^{5}
WALKER	2.48×10^{6}	$\mathbf{2.65\times10^{6}}$	2.27×10^{6}	$\mathbf{2.80\times10^{6}}$	2.82×10^{6}	2.78×10^{6}

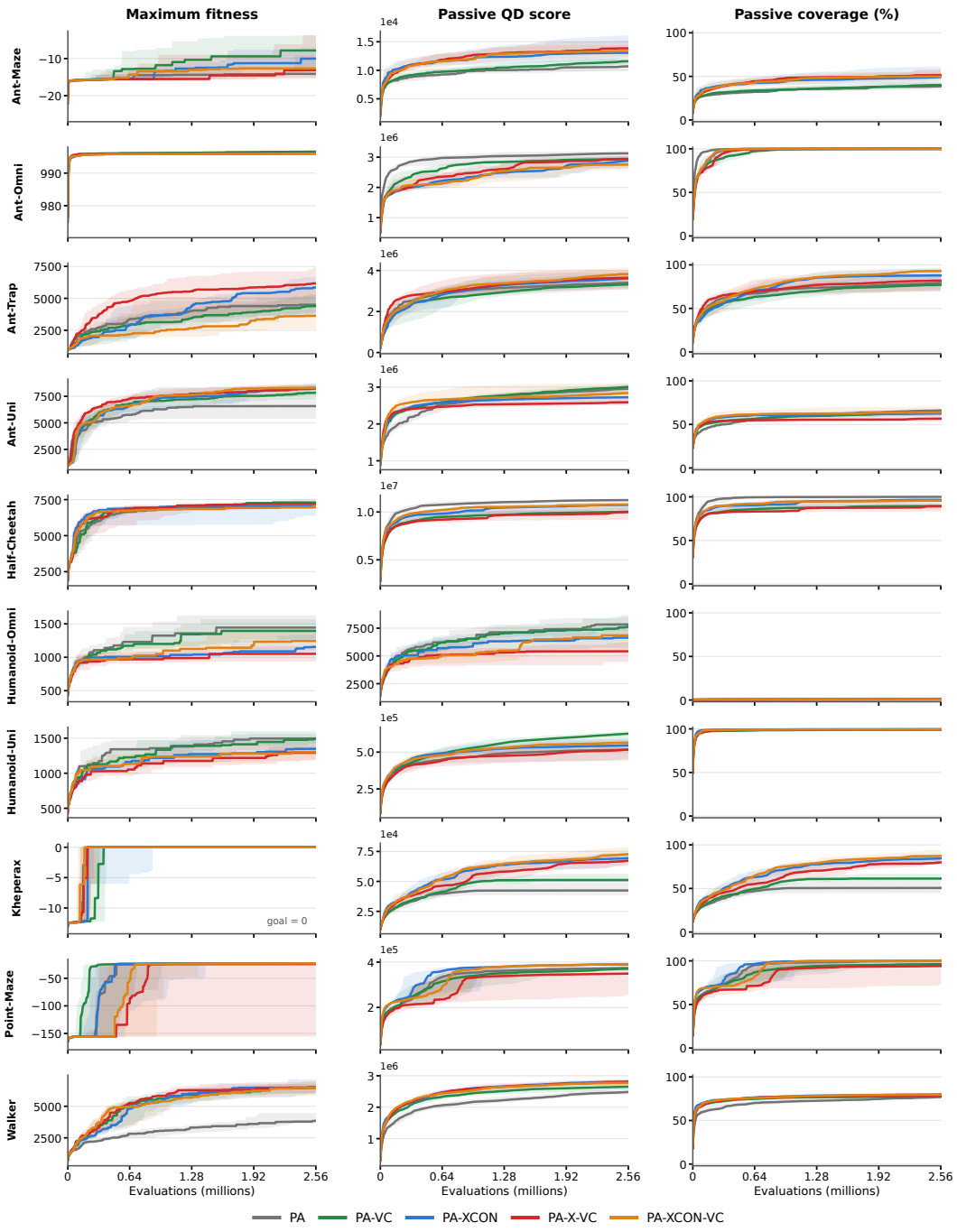

Figure 11: **Learning curves for the contextual XCON and extinction comparison.** Rows are all 10 environments and columns are maximum fitness, passive QD score, and passive coverage. Lines show ten-seed medians for PGA-AURORA, PGA-AURORA-VC, PGA-AURORA-XCON, PGA-AURORA-X-VC, and PGA-AURORA-XCON-VC; bands are exact pointwise 95% seed-bootstrap intervals for the median. PA denotes PGA-AURORA and X denotes extinction; both XCON arms use extinction. Kheperax maximum fitness has goal value zero.

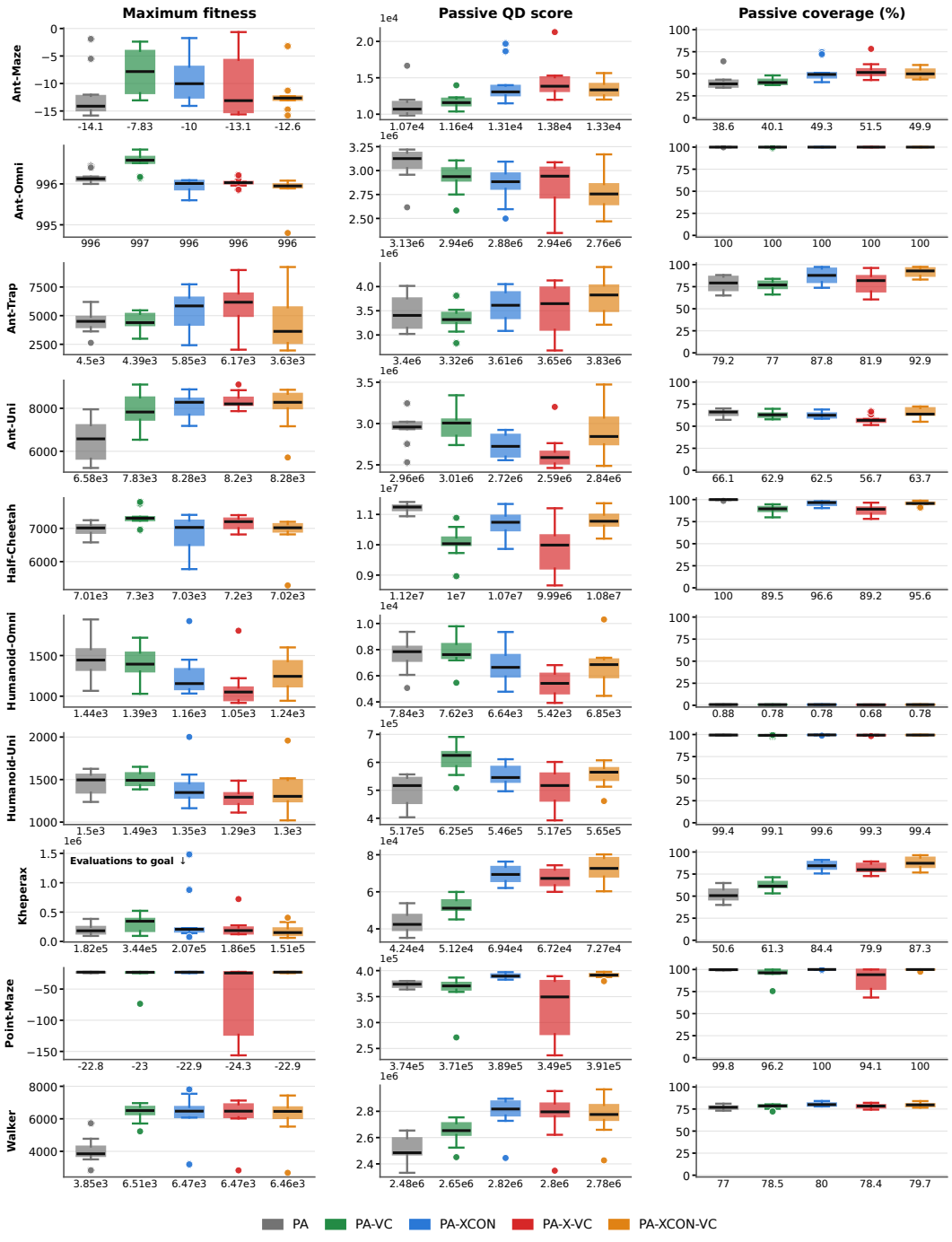

Figure 12: **Endpoint distributions for the contextual XCON and extinction comparison.** Rows and columns follow Figure 11. Each box summarizes ten seeds for PGA-AURORA, PGA-AURORA-VC, PGA-AURORA-XCON, PGA-AURORA-X-VC, or PGA-AURORA-XCON-VC: centers are medians, boxes span the interquartile range, and whiskers end at the most extreme observations within 1.5 interquartile ranges. Outliers are drawn as points, and raw medians are printed below the boxes. For Kheperax, the maximum-fitness panel instead reports evaluations to first goal, for which lower is better.

G PASSIVE-ARCHIVE VISUALIZATIONS

The passive archives show the behaviors discovered by each search process in a common, external descriptor space. Each cell retains the highest-fitness policy among the selected active members encountered in its region. The passive archive is used for evaluation and does not supply parents or affect admission decisions.

Archive structure and colors. This section contains one figure for each of the ten environments. Every figure uses the same layout with rows being different seeds and each column being benchmarks for PGA-AURORA, PGA-AURORA-VC, PGA-AURORA-XCON, PGA-AURORA-X-VC, and PGA-AURORA-XCON-VC.

Position-based descriptors. In ANTMAZE, POINTMAZE, and KHEPERAX, the coordinates are terminal (x, y) positions, normalized for KHEPERAX. ANTTRAP also uses terminal position, showing how the discovered endpoints extend around and beyond the trap. In ANTOMNI and HUMANOIDOMNI, terminal position shows the directions and distances reached during locomotion.

Foot-contact descriptors. For HALFCHEETAH, WALKER, and HUMANOIDUNI, the two coordinates are foot-contact proportions. The four contact proportions of ANTUNI are shown through two-dimensional projections. The occupied regions describe the range of discovered contact patterns, while colors show the fitness of the corresponding gaits.

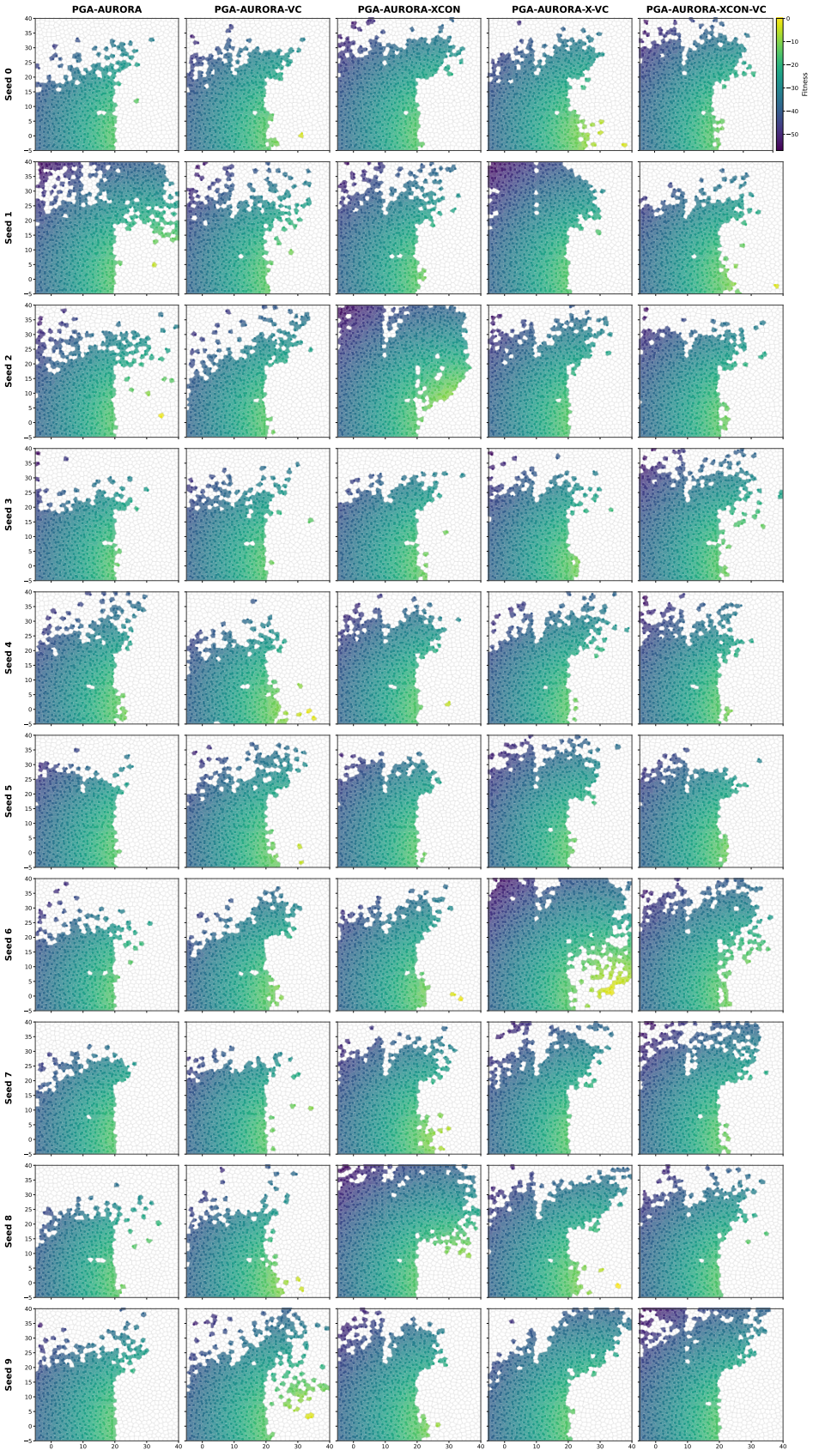

Figure 13: Final passive archive visualizations for ANTMAZE.

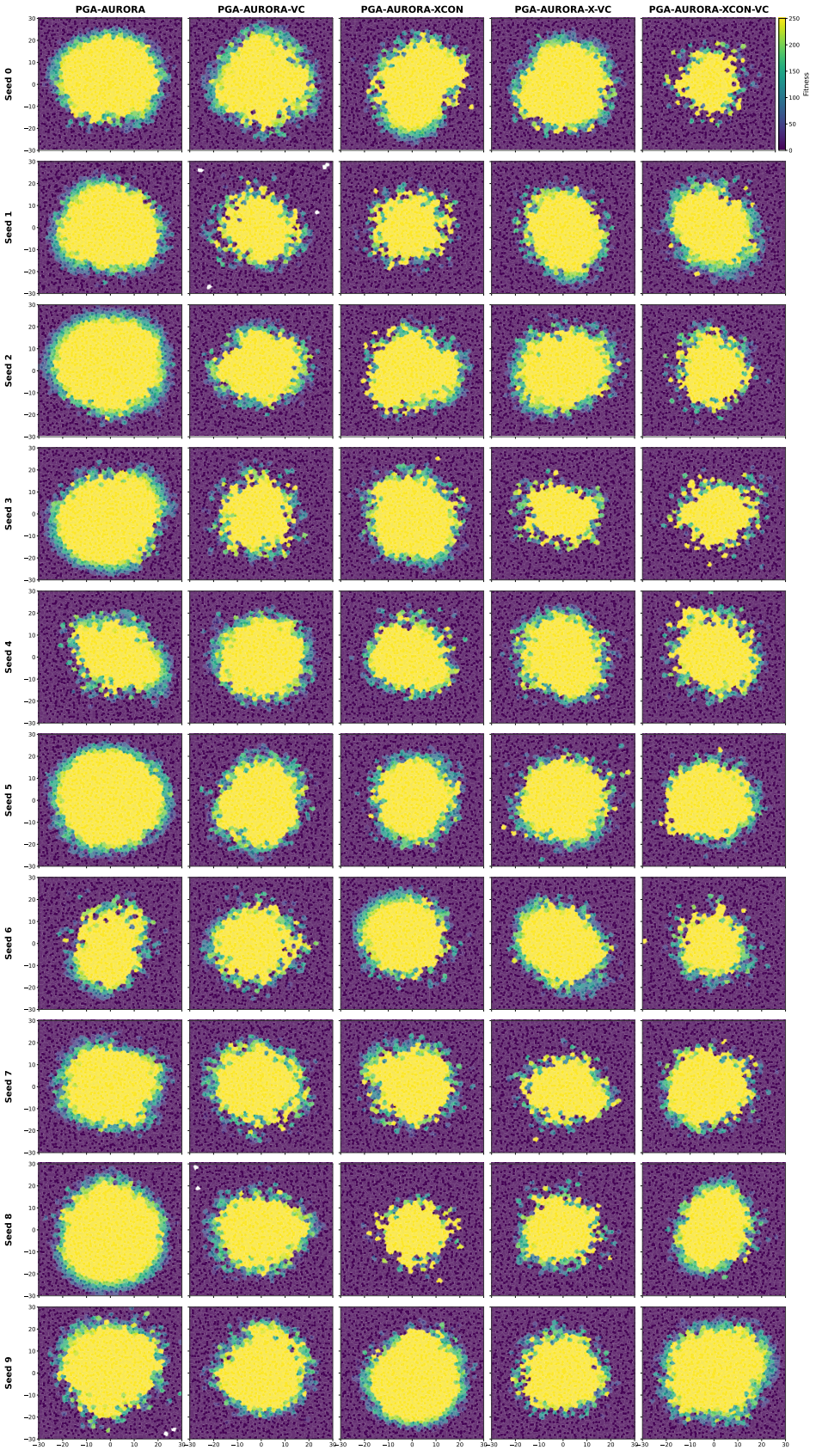

Figure 14: Final passive archive visualizations for ANTOMNI.

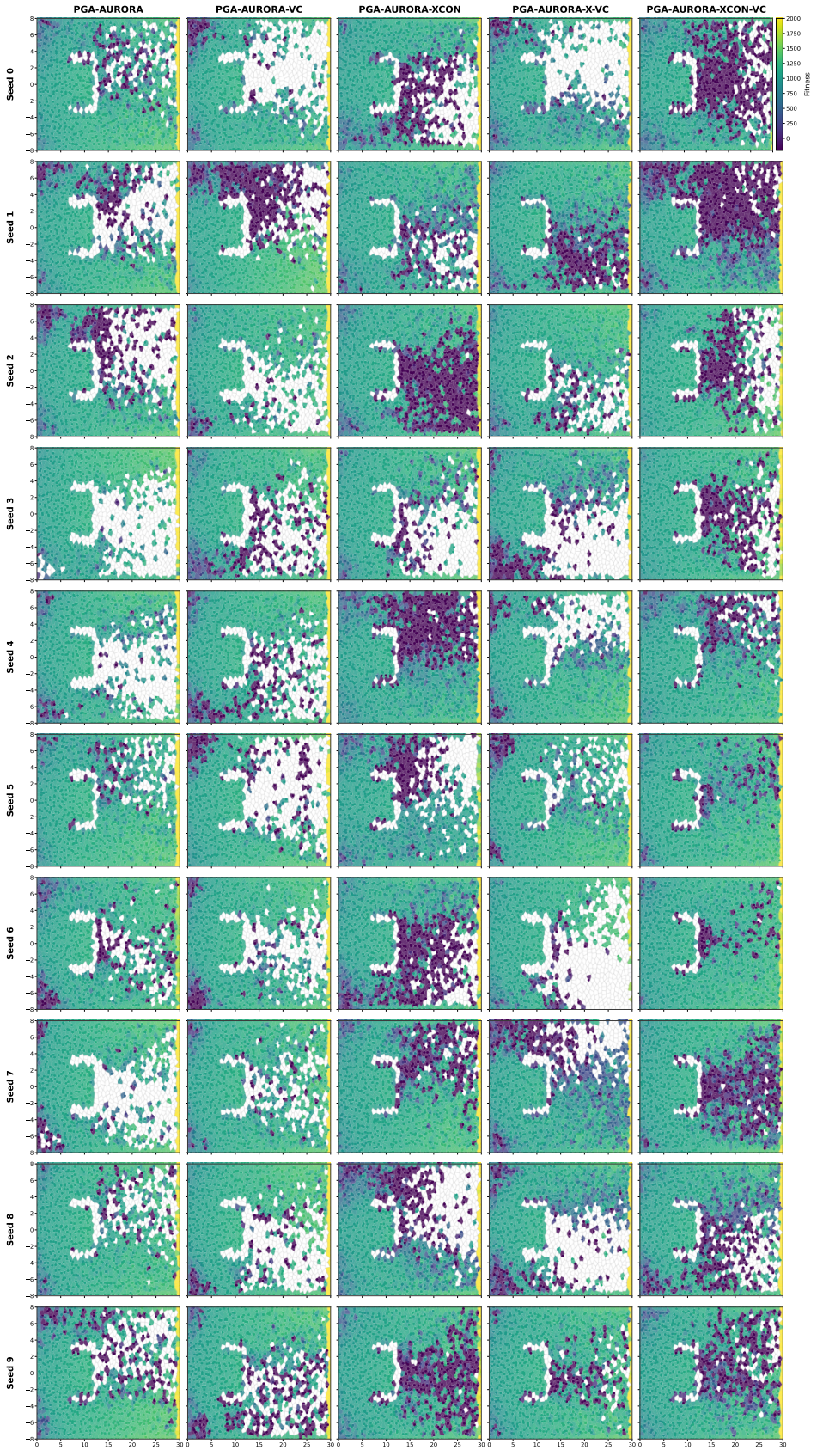

Figure 15: Final passive archive visualizations for ANTTRAP.

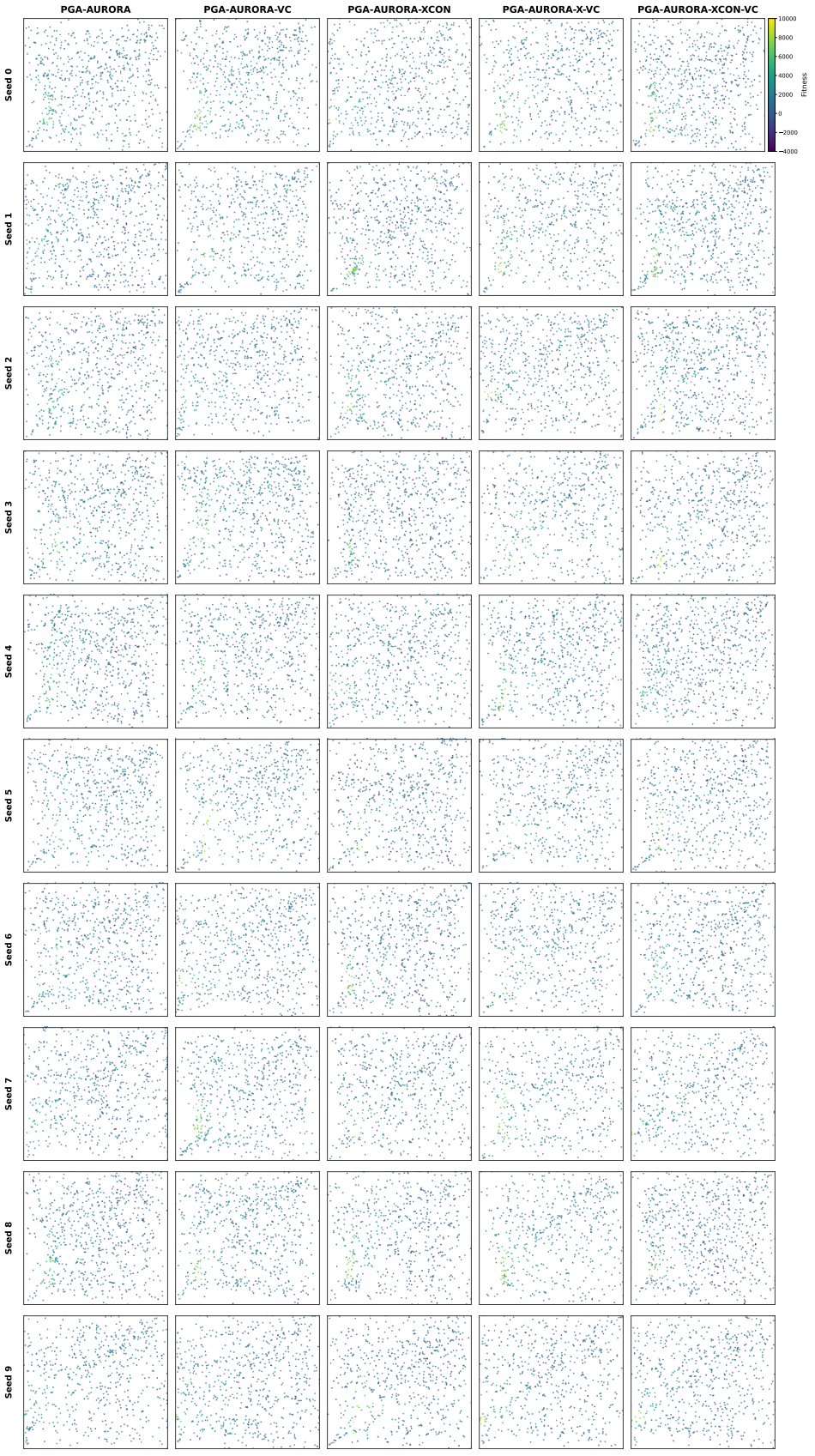

Figure 16: Final passive archive visualizations for ANTUNI.

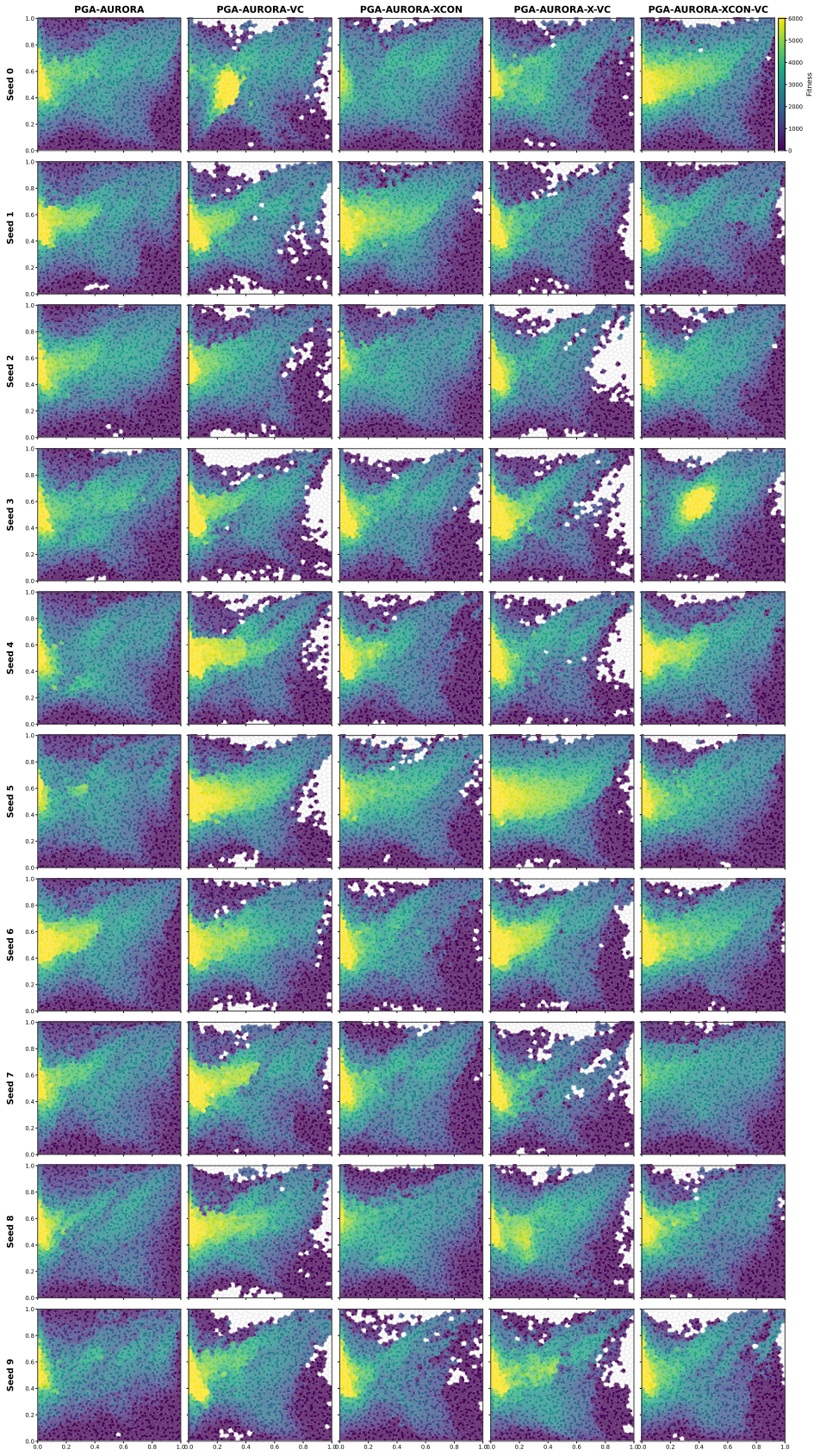

Figure 17: Final passive archive visualizations for HALFCHEETAH.

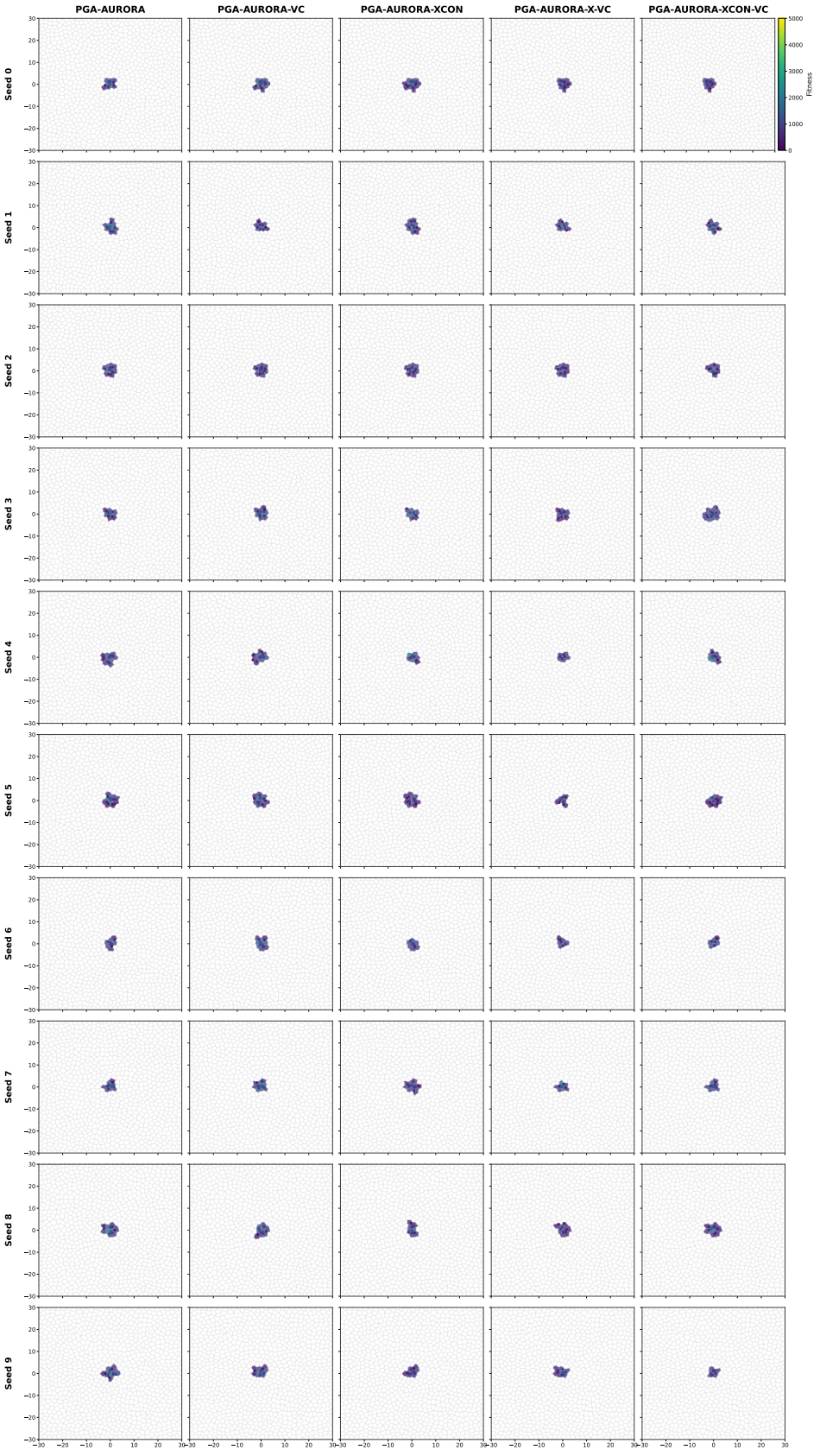

Figure 18: Final passive archive visualizations for HUMANOIDOMNI.

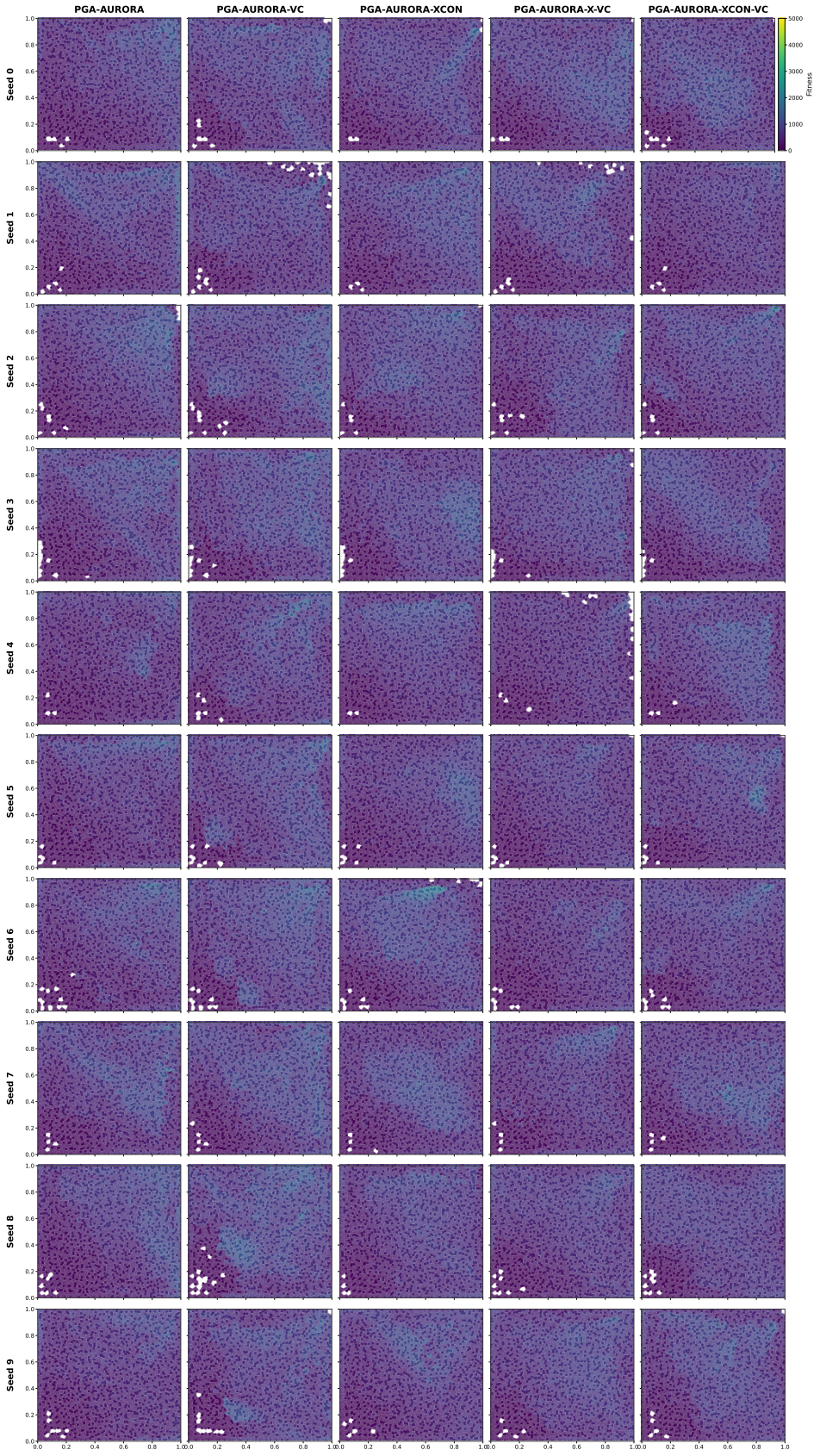

Figure 19: Final passive archive visualizations for HUMANOIDUNI.

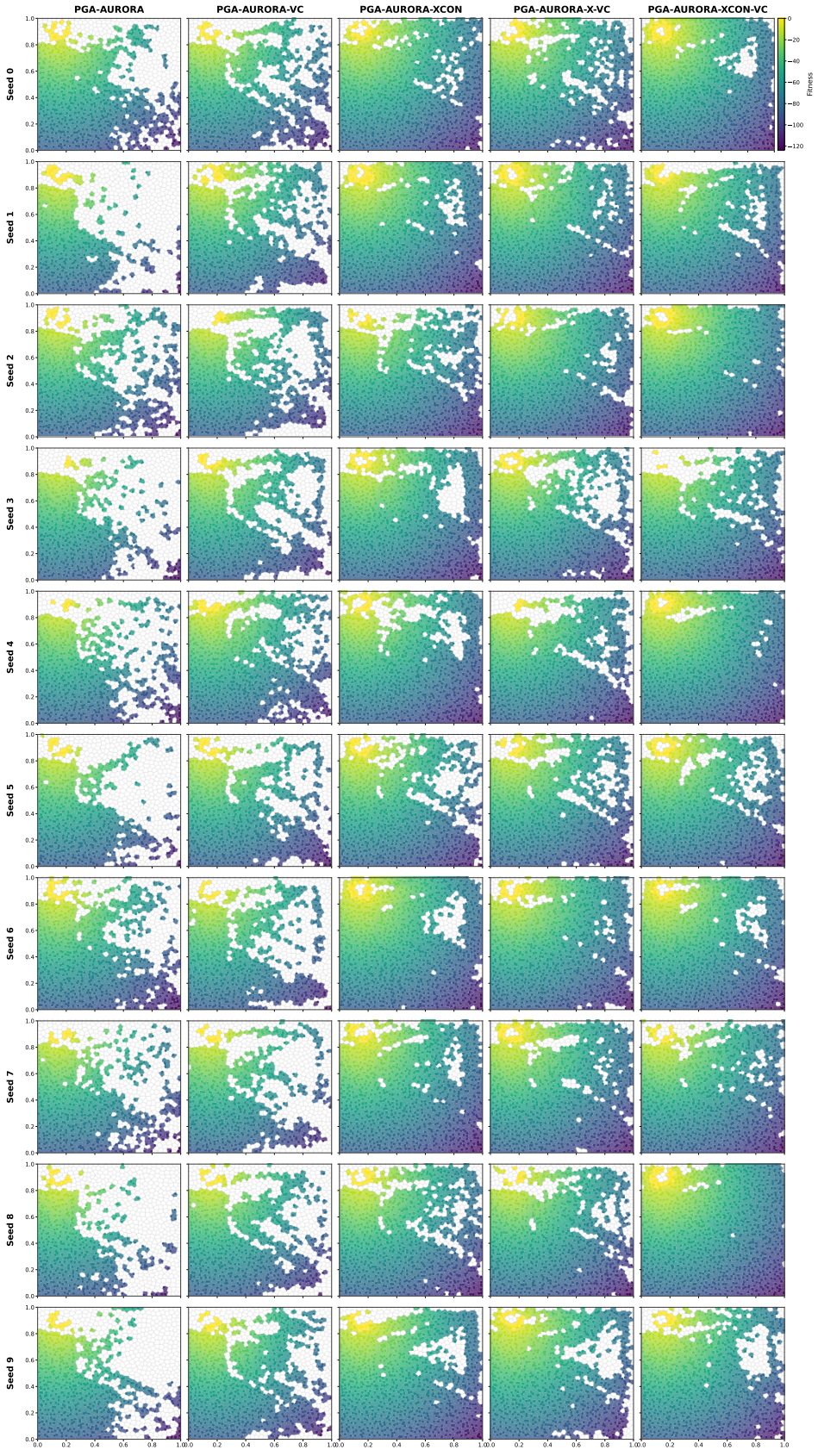

Figure 20: Final passive archive visualizations for KHEPERAX.

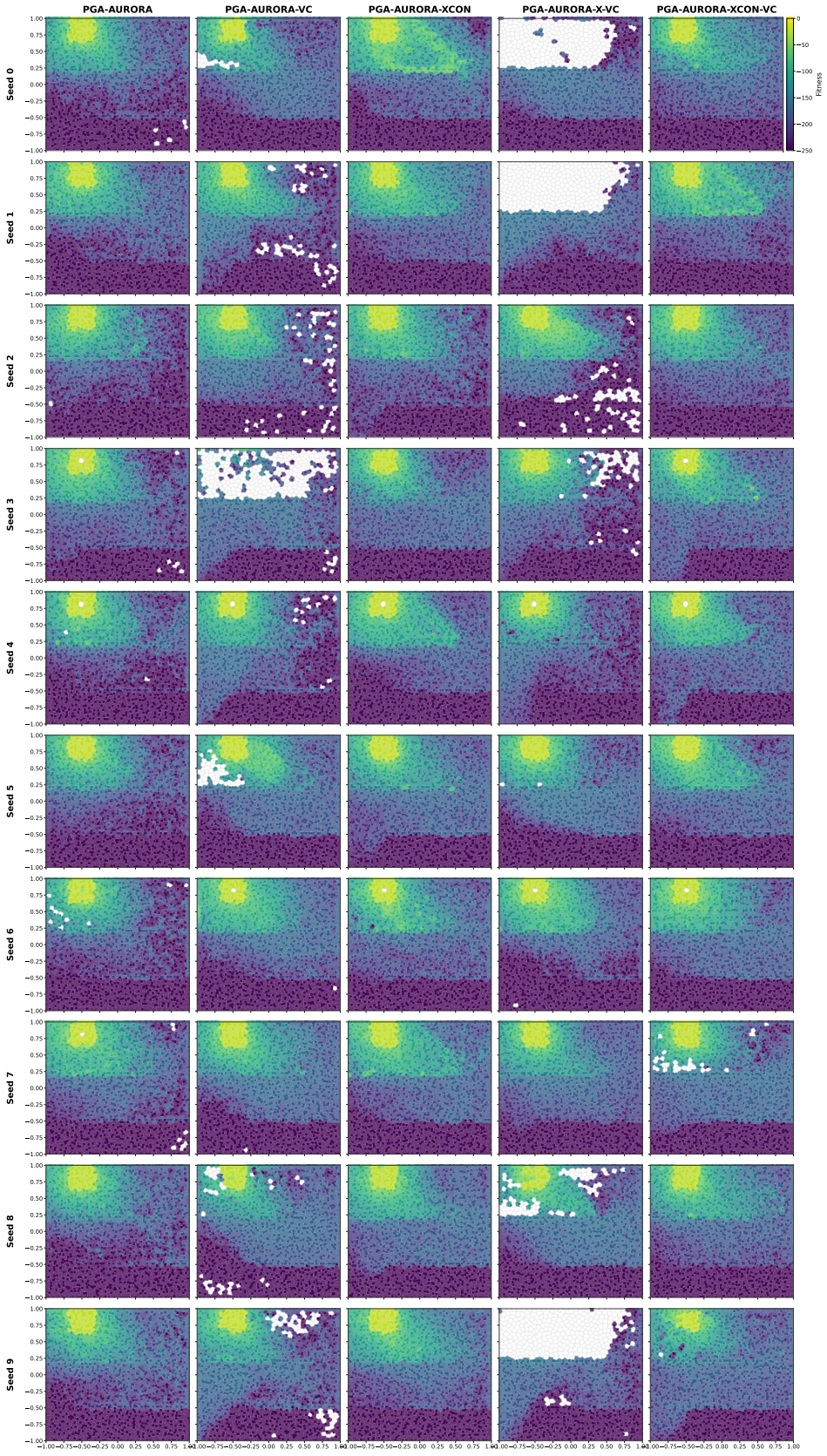

Figure 21: Final passive archive visualizations for POINTMAZE.

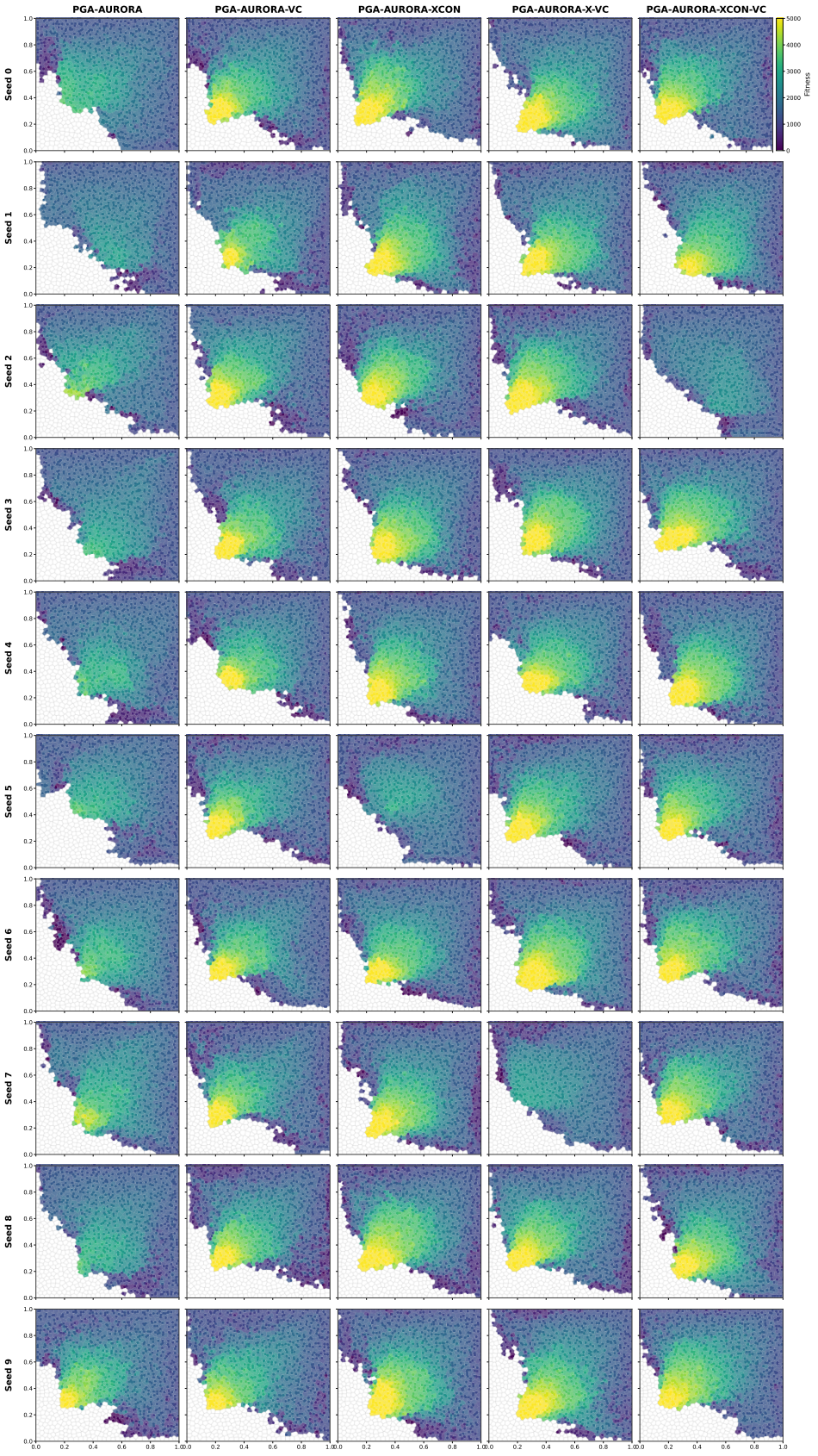

Figure 22: Final passive archive visualizations for WALKER.

H EFFECT OF THE ARCHIVE SELECTION RULE ON AURORA-VC

We investigate whether the effect of VC extends beyond the Dominated Novelty Search (DNS) archive selection rule used in the main experiments (Bahlous-Boldi et al., 2025). DNS ranks policies by their mean Euclidean distance to up to three nearest strictly fitter neighbors in learned descriptor space and retains those with the highest scores (Equation 5). We replace DNS with a custom adaptive-radius variant of the unstructured archive used in the original AURORA paper (Cully, 2019), with archive-size control inspired by AURORA-CSC (Grillotti & Cully, 2022a). We keep PGA proposals and compare reconstruction-only MSE with reconstruction plus VC. The aim is to test whether the fitness and archive-retention effects also appear when learned distances govern selection through local radius conflicts.

Under the new rule, a candidate is eligible if no incumbent lies within radius r, or if exactly one incumbent lies within r and the candidate is strictly fitter. Candidates with two or more nearby incumbents are rejected. Eligible candidates and incumbents then undergo a greedy rebuild in descending fitness order, with stable tie-breaking: a policy survives only if it lies farther than r from every policy already retained. The radius adapts after each batch to keep occupancy near 1,024 policies; initialization uses the tenth percentile of nearest-neighbor distances, followed by proportional control and a bounded eight-round radius search. After encoder retraining, the radius is rescaled by the change in median nearest-neighbor distance before rebuilding, so a representation update can immediately remove incumbents.

We evaluate MSE and MSE+VC across all ten standard environments, using ten matched random seeds per environment. Each run uses 5,000 updates with 512 proposals per update, ten-dimensional descriptors, and a target archive size of 1,024 policies. The DNS results from the main experiments provide context for comparing the effect of VC across archive selection rules.

The primary outcome is terminal maximum fitness in a fixed 1,024-cell reporting archive accumulated from selected active-archive members. A separate tracker receives every evaluated policy, including rejected candidates; Table 6 reports its maximum fitness separately. Neither tracker influences search. Task effects are medians of within-seed VC-minus-MSE differences, with cell-wise 95% percentile bootstrap intervals from 50,000 resamples. The registered global test excludes KHEPERAX, scales each task by its pooled tenth-to-ninetieth-percentile range across the 40 DNS and radius-rule endpoints, and averages the nine scaled task medians. Its one-sided paired-label randomization test uses one million draws over the 90 radius-rule seed pairs.

H.1 RESULTS

Fitness gains extend to the adaptive-radius rule. VC improves the registered paired median fitness in six of the nine tasks outside KHEPERAX, with the clearest gains in WALKER and HALFCHEETAH (Table 6). Their median gains are approximately $2{,}992$ and 527, respectively, and all ten seed pairs favor VC in both tasks. The normalized global effect is $+0.274$, with one-sided $p = 0.000383$. Effect directions agree with the earlier DNS comparison in seven of nine tasks, while ANTUNI and HUMANOIDOMNI change direction. The separately reported all-evaluated endpoint favors VC in eight task medians.

Peak fitness and repertoire breadth respond differently. The paired-effect panels in Figure 23 show that fitness gains do not coincide with broader coverage. Under the radius rule, passive QD score increases in six task medians, while coverage decreases in seven (Table 7). WALKER improves both peak fitness and QD score, whereas HALFCHEETAH gains fitness while losing about twelve percentage points of coverage. Figure 24 shows the corresponding learning trajectories and seed variation.

The new rule also changes archive turnover. VC reduces the fraction of candidates retained after their own update in every task and increases incumbent retention through retraining in eight tasks. These changes occur while typical archive occupancy stays close to the shared target under both objectives. The completed runs recorded no full-capacity rejections, controller fallbacks, or radius-clipping events. Thus, the retention differences accompany the changed learned geometry while the adaptive controller maintains a comparable archive budget.

H.2 DISCUSSION

We conclude that VC's task-dependent fitness benefits extend to a second archive selector that uses learned distances and fitness. The accompanying turnover changes indicate that its influence on retained search history also appears under adaptive-radius competition.

Table 6: Paired VC-minus-MSE maximum-fitness effects under adaptive-radius selection. Each task has ten matched seeds. The registered outcome uses the selected-member reporting archive; intervals are cell-wise 95% paired-bootstrap intervals from 50,000 resamples. DNS effects use earlier controls from this repository for descriptive context. The last column separately reports the all-evaluated discovery endpoint under adaptive-radius selection. W/L/T counts positive, negative, and zero seed differences for the registered endpoint. All KHEPERAX runs reach the goal in the all-evaluated tracker; one VC run's selected-member endpoint does not.

Environment	DNS Δ	Adaptive radius Δ [95% CI]	W/L/T	All-evaluated Δ
ANTMAZE	$+6.85$	$+4.12$ [$+0.1603$, $+9.95$]	9/1/0	$+7.45$
ANTOMNI	$+0.3983$	$+0.3242$ [$+0.1757$, $+0.5267$]	10/0/0	$+0.3479$
ANTTRAP	$+141.4$	$+423.8$ [$-1{,}413.2$, $+1{,}124.7$]	6/4/0	$+377.0$
ANTUNI	$+1{,}453.6$	-14.17 [$-1{,}832.3$, $+2{,}000.6$]	5/5/0	$+8.59$
HALFCHEETAH	$+266.1$	$+527.5$ [$+301.0$, $+656.8$]	10/0/0	$+385.6$
HUMANOIDOMNI	$+1.04$	-31.35 [-374.1, $+120.1$]	5/5/0	-85.80
HUMANOIDUNI	$+6.42$	$+15.75$ [-72.53, $+45.53$]	7/3/0	$+15.75$
KHEPERAX	0	0 [0, 0]	0/1/9	0
POINTMAZE	-0.0224	-0.0038 [-0.0262, $+24.10$]	5/5/0	$+0.0032$
WALKER	$+2{,}269.3$	$+2{,}992.2$ [$+2{,}406.5$, $+3{,}970.0$]	10/0/0	$+2{,}976.1$

Table 7: Repertoire and retention effects under adaptive-radius selection. Entries are medians of ten within-seed VC-minus-MSE differences; pp denotes percentage points. QD score and coverage use the selected-member reporting archive. Candidate retention is the fraction of all generated candidates retained after their own update and controller pruning. Retraining retention is the surviving fraction of incumbents, first reduced to a median over the 31 scheduled events within each run.

Environment	QD-score Δ	Coverage Δ (pp)	Candidate retention (pp)	Retraining retention (pp)
ANTMAZE	-469.1	-2.54	-2.81	$+1.97$
ANTOMNI	-5.51×10^{5}	-0.6348	-3.20	$+2.21$
ANTTRAP	-6.68×10^{5}	-10.99	-2.78	$+2.47$
ANTUNI	$+14{,}436.1$	-1.17	-3.43	$+3.54$
HALFCHEETAH	-1.48×10^{6}	-12.06	-2.98	$+1.82$
HUMANOIDOMNI	$+1{,}128.1$	$+0.1953$	-1.89	$+0.3821$
HUMANOIDUNI	$+68{,}521.9$	-0.3906	-1.81	$+1.66$
KHEPERAX	$+5{,}861.6$	$+6.79$	-0.0042	-1.50
POINTMAZE	$+13{,}222.1$	-1.07	-2.17	-4.54
WALKER	$+3.22\times10^{5}$	$+1.32$	-2.72	$+1.39$

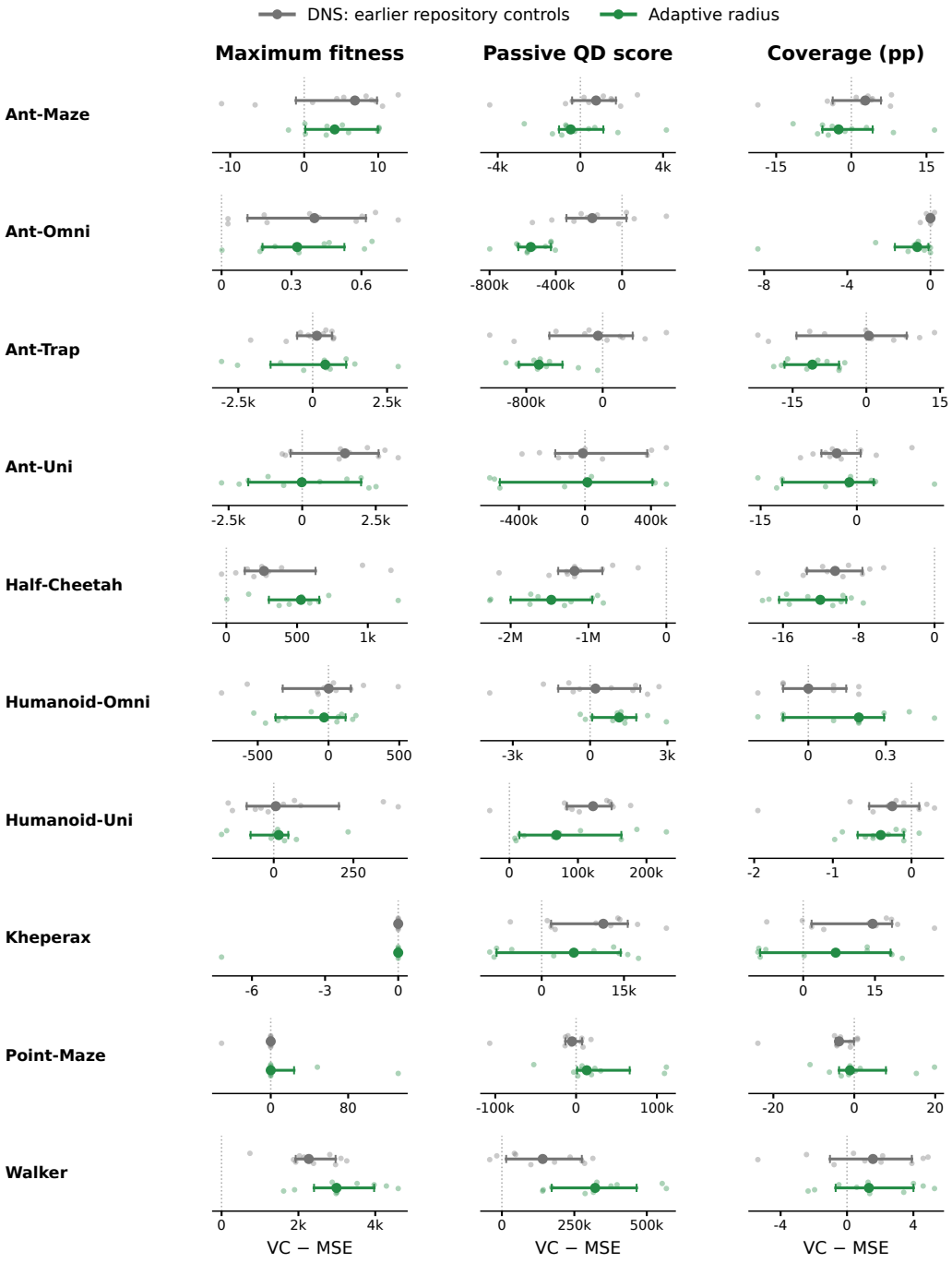

Figure 23: **Paired effects of VC under two archive selection rules.** Rows show the ten environments; columns show terminal maximum fitness, QD score, and coverage in the selected-member reporting archive. Grey marks the earlier DNS controls and green the new adaptive-radius runs, each with ten MSE–VC seed pairs per task. Small points are individual seed differences; large points and bars show paired medians and cell-wise 95% bootstrap intervals from 50,000 resamples. Positive differences favor VC. Fitness and QD score retain task-specific units, with k and M denoting thousands and millions; coverage differences are percentage points. Intervals are descriptive and not adjusted for multiple comparisons, and the two cohorts do not constitute a randomized selector comparison. KHEPERAX is shown descriptively and excluded from the registered global test.

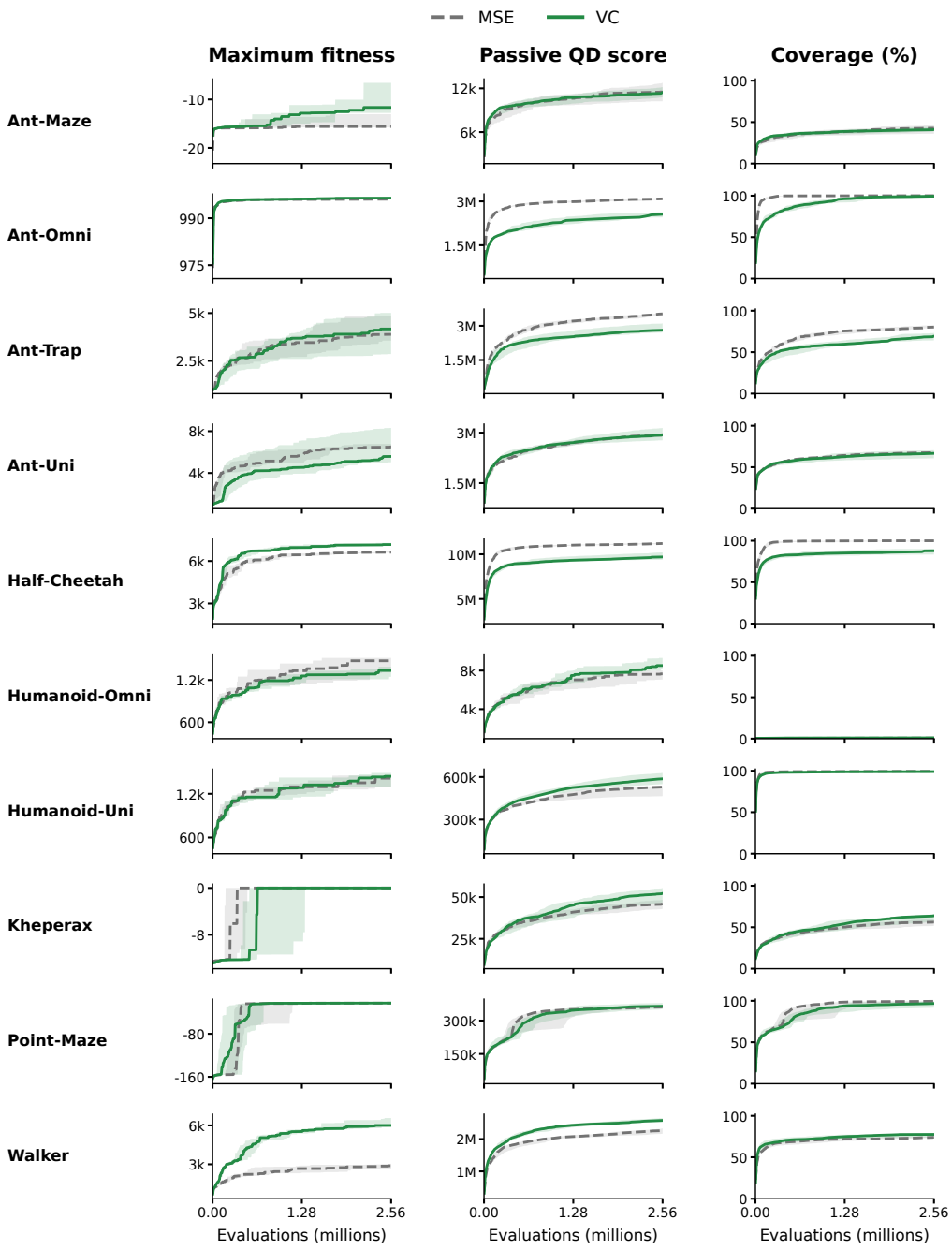

Figure 24: **Learning curves under adaptive-radius archive selection.** Rows and outcome columns follow Figure 23, using only the 200 new radius-rule runs. Grey dashed and green solid lines show ten-seed MSE and VC medians; shaded bands show interquartile ranges. Curves contain all 1,000 checkpoints over 5,000 updates, and the horizontal axis counts generated proposals, excluding initialization. All outcomes use the selected-member reporting archive; the all-evaluated discovery endpoint is reported separately in Table 6. Outcome scales are task specific, except coverage, which uses a common percentage scale.

I Representation Geometry and Semantic Diagnostics

I.1 Experiment Design

We examine whether VC spreads descriptor variation across more latent directions, how this geometry develops during search, and whether the resulting representations make behavioral quantities more accessible to linear prediction.

Measuring the effective number of directions. For the valid members of an active archive, let C be the sample covariance of their centered embeddings and let $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$ be its eigenvalues. When the total variance is nonzero, define the normalized variance shares $p_r = \lambda_r / \sum_q \lambda_q$ and the participation rank

$$r_{\mathrm{PR}} = \frac{(\sum_r \lambda_r)^2}{\sum_r \lambda_r^2} = \frac{1}{\sum_r p_r^2}.$$

If k directions carry equal shares of all the variance, then $r_{\mathrm{PR}} = k$; it approaches one when a single direction dominates and reaches d for a flat nonzero spectrum. Thus, ten coordinates can all vary while having participation rank one if they are scaled copies of the same signal. Ten uncorrelated coordinates with equal variance instead have participation rank ten. The statistic is invariant to a common rescaling and to orthogonal changes of basis, so merely enlarging or rotating the descriptor cloud cannot increase it. Here, "effective rank" refers specifically to this participation ratio, rather than numerical matrix rank or entropy-based effective rank. When the configured dimension varies, r_{PR}/d measures the fraction of available covariance directions used (Appendix J).

Why the two penalties can increase participation rank. The variance term encourages coordinates with small marginal variance to reach a common standard-deviation floor; the covariance term penalizes linear redundancy between coordinates. Their complementary roles can be seen directly from

$$r_{\mathrm{PR}} = \frac{(\operatorname{tr} C)^2}{\|C\|_F^2} = \frac{\left(\sum_r C_{rr}\right)^2}{\sum_r C_{rr}^2 + \sum_{r \neq q} C_{rq}^2}. \tag{18}$$

Holding the diagonal fixed, reducing the off-diagonal covariance energy decreases the denominator and increases participation rank. The variance penalty encourages weak coordinates to contribute, while the covariance penalty discourages those contributions from repeating the same direction. Together, these pressures can spread variation across more directions; when the diagonal variances are similar and off-diagonal terms are small, r_{PR} approaches d. This is a conditional geometric explanation, not a monotonicity guarantee for training: the variance hinge does not equalize variances above its floor, and reconstruction training changes both the diagonal and off-diagonal entries. Appendix B.2.1 gives the corresponding gradient and fixed-covariance analysis.

Geometry throughout training. The training grids use all 200 primary MSE/VC runs: ten environments, two objectives, and ten matched seeds, with descriptor width $d = 10$. We recompute each spectrum from the occupied active-archive embeddings immediately after encoder retraining and archive rebuilding. There are 32 snapshots per run, from initialization through the final scheduled encoder update at search iteration 4,960, yielding 6,400 snapshots. Eigenvalues are sorted and normalized separately within each run and checkpoint before taking the median across seeds. Consequently, a heatmap row denotes an ordered covariance direction, not a persistent coordinate or the same eigenvector across time. The adjacent curves show the median of the per-run participation ranks with seed interquartile ranges; rank is not computed from the median spectrum. Separate diamond markers show the terminal active-archive rank at iteration 5,000. The terminal component comparison additionally uses V-only and C-only runs and evaluates the active embeddings at the end of the 5,000-iteration search budget.

Behavioral prediction on fixed trajectories. The semantic diagnostics compare terminal MSE and VC encoders on shared frozen trajectory banks. We count covariance-active coordinates, count coordinates whose held-out trajectory-statistic ridge prediction exceeds $R^2 = 0.3$, and evaluate joint linear prediction from all coordinates of both the passive descriptor and the broader trajectory-statistic battery. Comparisons on a common bank hold the evaluated trajectories fixed while changing the encoder. These diagnostics measure second-moment geometry and the accessibility of the specified targets to linear prediction; they do not assign intrinsic meanings to latent axes.

I.2 RESULTS

The two penalties have complementary geometric effects. V+C has the highest marginal ten-seed terminal participation-rank median in nine environments; C-only is highest in HALFCHEETAH. The complete weight factorial in Appendix L examines this component comparison over a two-dimensional range of penalty strengths.

Full spectra reveal how the available dimensions are used. Figure 25 shows the complete training grid for all ten environments. Concentration in the first heatmap rows indicates that a few directions explain most of the descriptor variance; a more even distribution across rows indicates broader use of the latent space. The participation-rank curves summarize this distribution at each checkpoint and expose variation across seeds and training stages. VC has a higher marginal ten-seed median participation rank than MSE at all 32 recorded post-rebuild checkpoints in every environment. At the last encoder update, the task-specific VC medians range from 7.60 to 9.21 effective directions, compared with 1.70 to 3.52 for MSE, out of ten available directions. These checkpoint comparisons are descriptive and do not imply that every individual VC seed has higher rank. They also concern the immediate post-rebuild archive: subsequent search can concentrate the retained embeddings again before the next retraining event. For example, the HALFCHEETAH VC median falls from 8.31 immediately after the last update at iteration 4,960 to 3.80 in the final archive at iteration 5,000. The terminal markers make this distinction visible rather than treating post-retraining rank as constant between updates.

Greater covariance activity does not imply more linearly readable semantics. On the frozen probe banks, VC raises the median number of covariance-active coordinates from five to nine of ten configured coordinates; its marginal ten-seed median is higher in nine environments. However, fewer coordinates pass the held-out trajectory-statistic threshold: the median falls from five to one, with a lower VC value in all ten tasks. Joint linear decodability of the passive descriptor decreases in nine tasks, and joint linear decodability of the broader statistic battery decreases in all ten. Thus, the additional covariance-active coordinates do not correspond to additional linearly recoverable axes for the measured targets.

I.3 DISCUSSION

The per-coordinate semantic count depends on the coordinate basis, whereas a joint linear probe over all coordinates is invariant to orthogonal changes of basis. The decrease in joint decodability therefore cannot be explained solely by rotating the coordinates. It also does not imply that behavioral information is absent or that nonlinear probes could not recover it.

VC acts on empirical global second moments of sampled archive embeddings. Neither penalty guarantees non-collapse or independence: the exactly collapsed point has zero variance gradient, and zero linear covariance does not imply nonlinear independence. The objective also does not guarantee exact whitening, local isometry, semantic alignment, or temporal alignment. Participation rank can change because the encoder changes, the retained population changes, or both. The fixed-pool POINTMAZE construction and same-incumbent retraining probes separate these sources of variation, but neither makes participation rank sufficient to predict search quality.

I.4 STATIC-GEOMETRY AND ADAPTATION CONTROLS

These controls investigate what is gained by learning with VC throughout search. We compare VC with whitening MSE descriptors, then test continued adaptation against freezing after the same VC training history.

The whitening control periodically fits a regularized ZCA transform to MSE embeddings and uses the transformed distances for competition. It uses the same teacher-forced LSTM architecture and updated decoder-state carry as the primary runs (Appendix A), with 100 runs across ten environments and ten seeds. All ZCA runs complete 5,000 iterations and record all 32 scheduled transforms. We compare marginal ten-seed outcomes and AUC, defined on the common grid of updates $5, 10, \ldots, 5{,}000$.

The continuation control uses 100 paired VC checkpoints across the same environments and seeds. Each pair restores identical complete search state at iteration $1{,}050$ and runs two branches to iteration

5,000: one freezes the encoder, and the other continues training using only V+C. The paired design and task-level estimates are detailed in Appendix M.

I.4.1 RESULTS

Whitening achieves the intended covariance conditioning without reproducing VC's fitness outcomes. At the final whitening event, pooled median participation rank rises from 2.218 to 10.000, and the covariance condition number falls from 25,412.2 to 1.002. Nevertheless, ZCA has lower maximum-fitness AUC and final fitness than VC in eight task medians, while attaining higher coverage AUC in eight. These directions describe full-run algorithm comparisons.

Continuing adaptation after an identical VC history gives a broader pattern of favorable outcomes than freezing. The paired median effect is positive for QD-score AUC in all ten tasks, coverage AUC in nine with one tie, and maximum-fitness AUC in eight. Table 8 places the fitness-AUC directions from the two controls alongside one another.

Table 8: Static-geometry and adaptation controls, reported as environment-level directions for passive maximum-fitness AUC. Both comparisons use ten environments and ten seeds; the continuation contrast pairs branches from identical VC checkpoints.

Contrast	Environment direction count	Supporting breadth result
Ordinary ZCA minus VC	Higher in $2/10$, lower in $8/10$	Coverage AUC higher in $8/10$
Continued V+C-only adaptation minus freezing	Higher in $8/10$, lower in $1/10$, tied in $1/10$	QD-score AUC higher in $10/10$

I.4.2 DISCUSSION

We conclude that covariance conditioning alone captures only part of what changes when VC is used during search. ZCA produces an almost isotropic representation with different fitness and coverage outcomes, while the exact-checkpoint comparison supports continued V+C-only adaptation after VC has shaped early search. Together, these controls motivate evaluating the evolving search process alongside snapshot geometry. This distinction is consistent with the remaining metric ambiguity under zero V+C loss in Proposition 2.

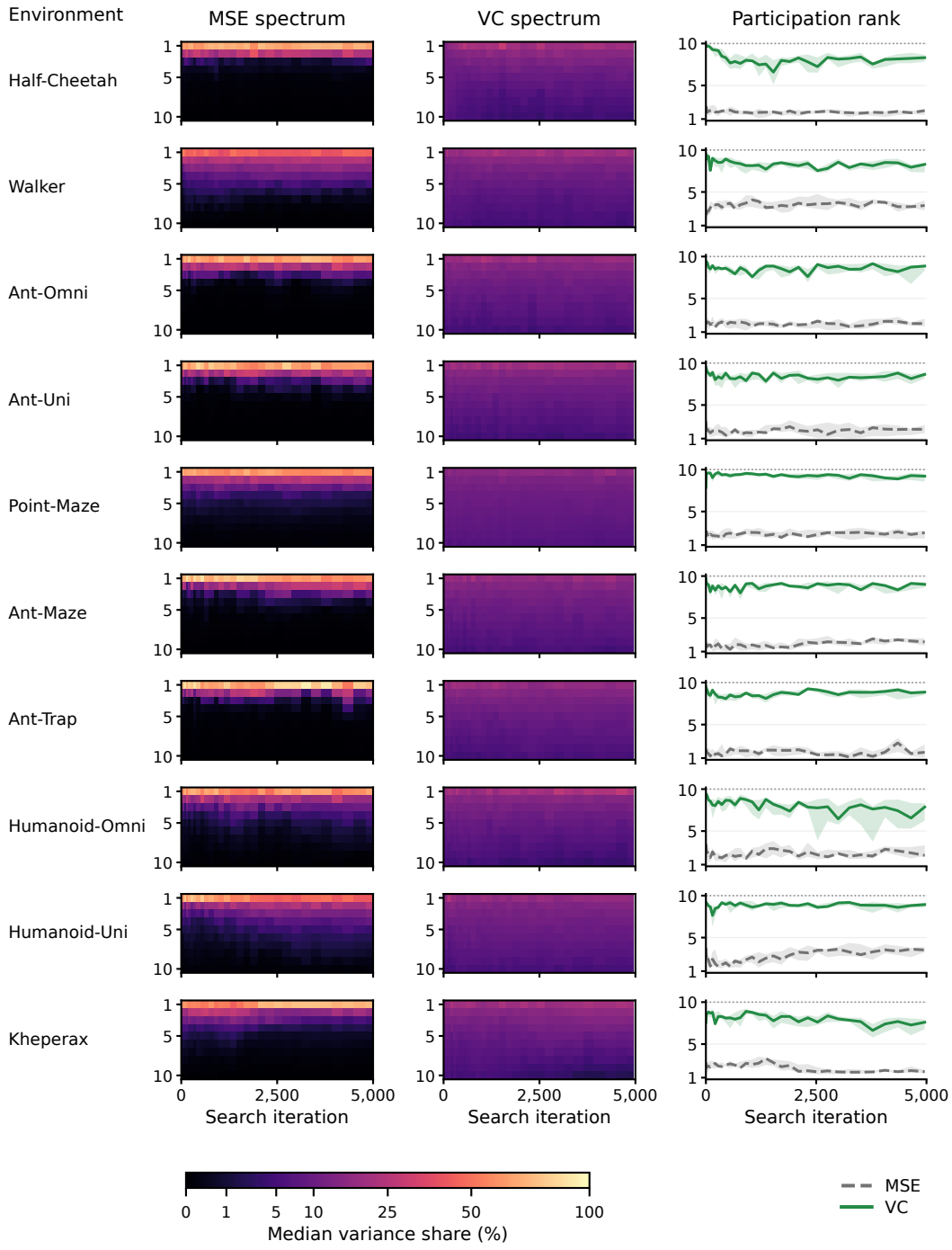

Figure 25: **Latent dimension use throughout training across all ten environments.** Each row shows MSE and VC covariance spectra and participation rank for ten matched seeds. Heatmaps report median percentages of total variance in sorted covariance directions at all 32 post-rebuild encoder updates through iteration 4,960, using a shared square-root color scale; directions are ranked separately at each checkpoint. Curves show median per-run ranks and interquartile ranges (MSE grey dashed, VC green solid); the dotted line marks $d = 10$. Diamonds with interquartile bars mark final archives at iteration 5,000. All 6,400 snapshots from 200 runs are included; these descriptive trajectories reflect both encoder and archive changes.

J Effect of Latent Dimensionality on AURORA-VC

This experiment investigates whether the effect of VC depends on the number of coordinates available in the learned representation. We vary latent dimensionality to test whether the geometric changes and search benefits observed at the default dimension persist in smaller and larger representations.

We compare MSE and VC at latent dimensions $d \in \{5, 10, 20, 30\}$ in all ten environments, using ten matched seeds per environment and dimension. All conditions use the same 5,000-iteration search budget, 1,024-policy active archive, representation-retraining schedule, and proposal mechanism. A separate 1,024-cell passive evaluator remains fixed, so changing latent dimension does not change the resolution at which search outcomes are measured.

We measure terminal active-archive participation rank r_{PR} and normalized rank r_{PR}/d. These distinguish the effective number of covariance directions from the fraction of the configured dimension used. We measure search outcomes using the passive evaluator's maximum fitness, QD score, and coverage, as defined in Appendix E. Figure 26 places participation rank alongside these outcomes, and Figure 27 shows normalized rank; both display individual seeds, medians, and interquartile ranges. Figure 28 shows the within-seed VC-minus-MSE differences, their medians, and cell-wise 95% percentile bootstrap intervals from 50,000 paired resamples.

J.1 Results

The geometric advantage weakens in the largest representation. VC increases participation rank consistently at the smaller dimensions, but its advantage becomes less uniform as more coordinates are available. The paired median rank difference favors VC in every task at $d = 5$ and $d = 10$, and in most tasks at $d = 20$. At $d = 30$, however, only four tasks favor VC, and the across-environment medians are nearly equal (9.39 for VC and 9.29 for MSE). The normalized-rank panels further show that adding coordinates does not generally produce a proportional increase in the effective dimensions used.

Fitness gains are most consistent at smaller dimensions. Increasing latent dimensionality does not make the search benefit of VC more consistent. The paired median maximum-fitness effect is positive in eight tasks at each of $d = 5$ and $d = 10$, compared with five at each of $d = 20$ and $d = 30$. Thus, the strongest cross-task pattern of fitness gains occurs in the smaller representations, while larger ones produce more mixed effects. Figure 28 shows how these effects vary across tasks and seeds.

Repertoire quality does not follow rank or peak fitness. Passive QD score and coverage show no comparable broad advantage for VC. At $d = 5$, coverage decreases in most tasks despite the consistently higher participation rank and frequent peak-fitness gains. Using more latent directions therefore does not produce a broader repertoire in the external behavior space. In KHEPERAX, both objectives reach the goal ceiling in every condition, so terminal maximum fitness cannot distinguish them; QD score and coverage still reveal differences between the resulting repertoires.

At fixed average squared off-diagonal covariance, L_C scales as $d - 1$ because it divides $d(d - 1)$ terms by d. Thus fixed weights do not hold the strength of covariance regularization constant across widths; this sweep jointly changes model capacity and regularization scaling.

J.2 Discussion

We conclude that the benefits of VC extend beyond the default ten-dimensional setting, with more consistent fitness gains in the smaller representations tested here. Increasing the configured dimension provides additional coordinates without guaranteeing that they are effectively used or improve search. Latent dimensionality should be evaluated through both descriptor geometry and search outcomes. The sweep does not identify a dimension that is best for every task or establish participation rank as the cause of the fitness gains.

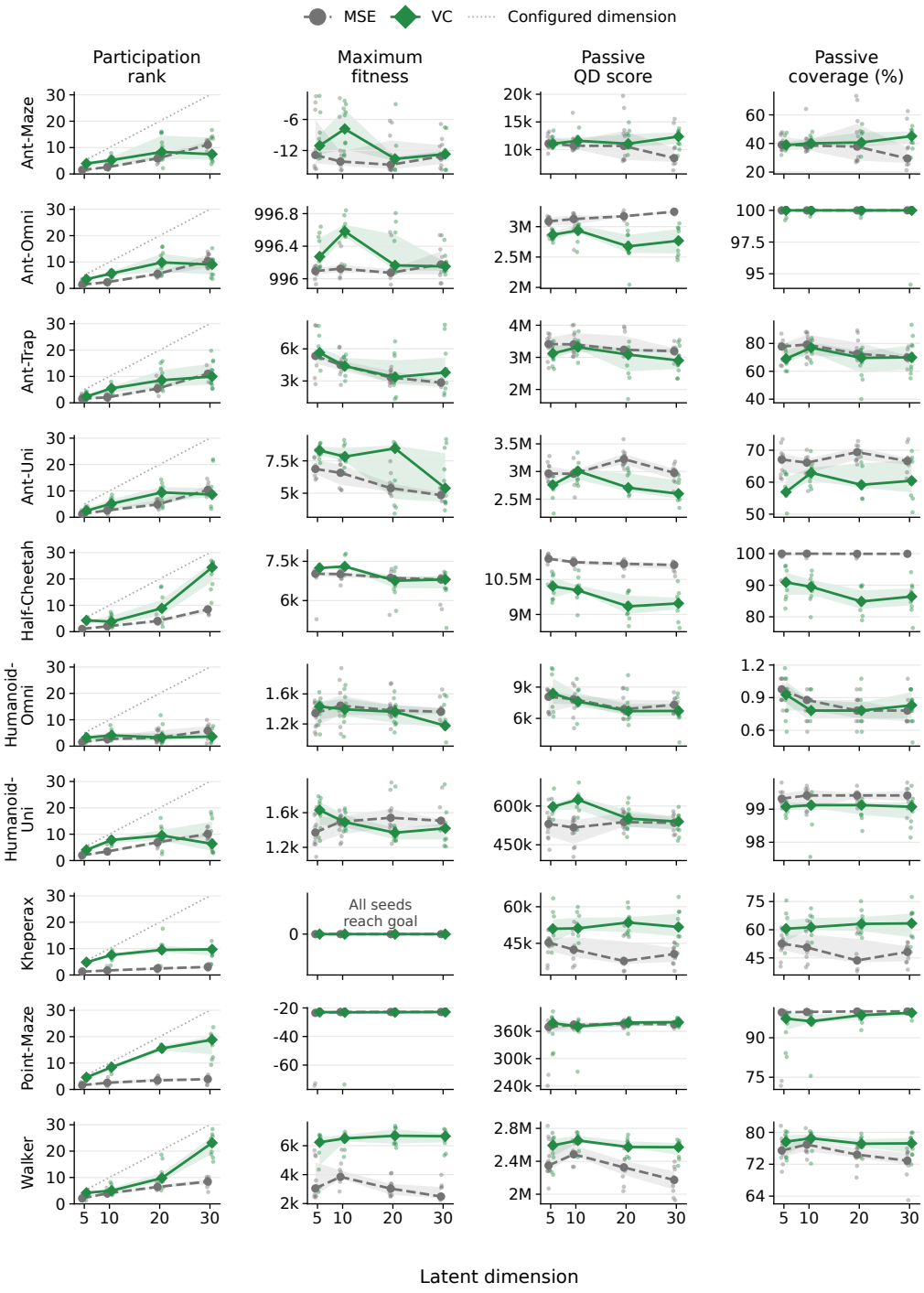

Figure 26: **Descriptor geometry and search outcomes across latent dimensions.** Rows show the ten environments; columns show terminal active-archive participation rank, passive maximum fitness, passive QD score, and passive coverage at $d = 5, 10, 20, 30$. Grey dashed and green solid lines show MSE and VC medians, bands show interquartile ranges, and points show all ten matched seeds in each condition. Rank panels share a scale, with the dotted line marking the configured dimension, the maximum attainable rank. Outcome scales are specific to each task; k and M denote thousands and millions. All 800 runs use the same 1,024-cell passive evaluator. Kheperax terminal fitness is tied at its goal value of zero.

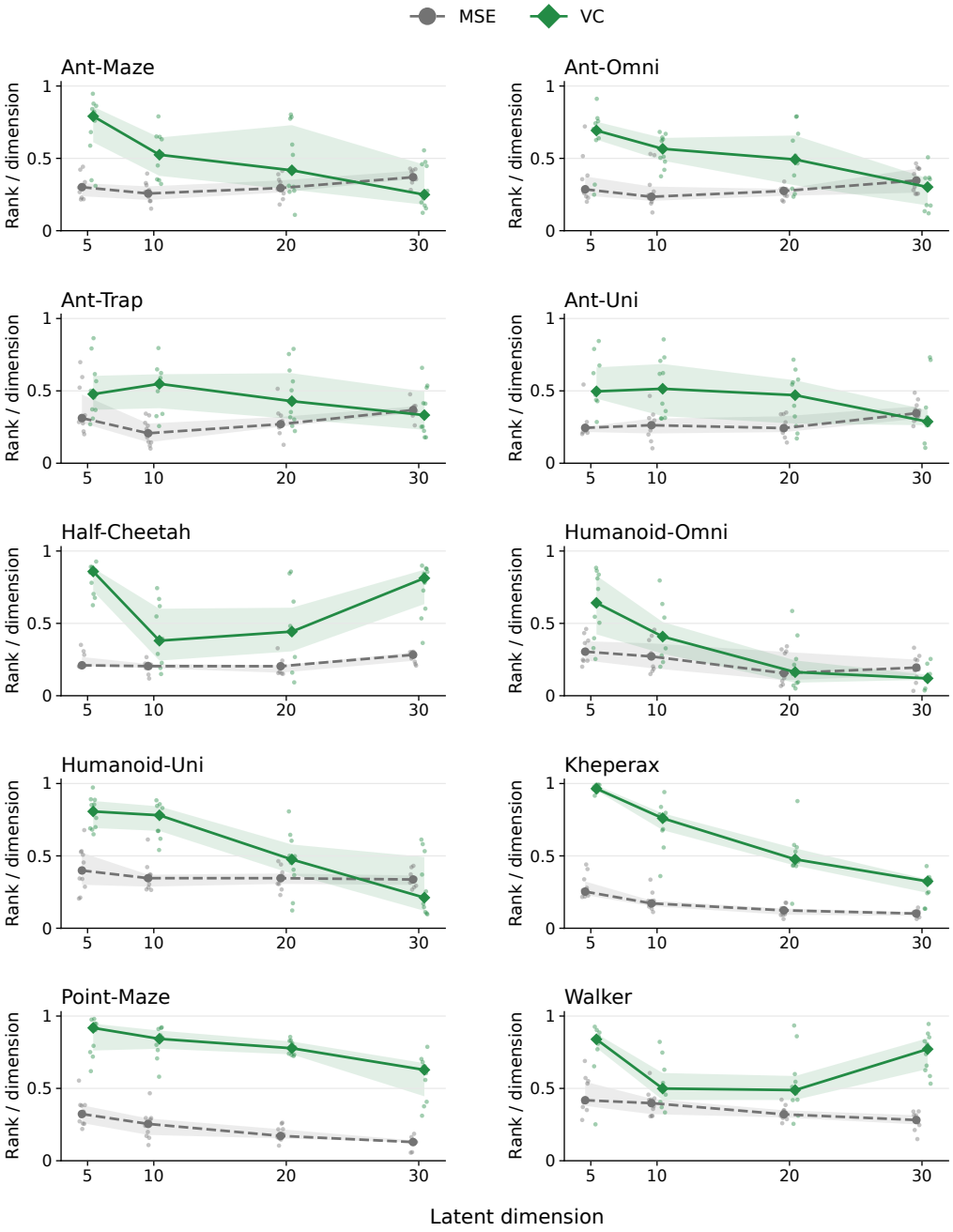

Figure 27: **Fraction of the configured latent dimension used.** Panels normalize the participation rank in the first column of Figure 26 by the configured dimension d, on a shared $[0, 1]$ scale. Lines are ten-seed medians, bands are interquartile ranges, and points show individual seeds. A value of one denotes a flat nonzero covariance spectrum. Normalization permits comparison of dimensional utilization across d; it does not change the sign of an MSE–VC contrast at any fixed d.

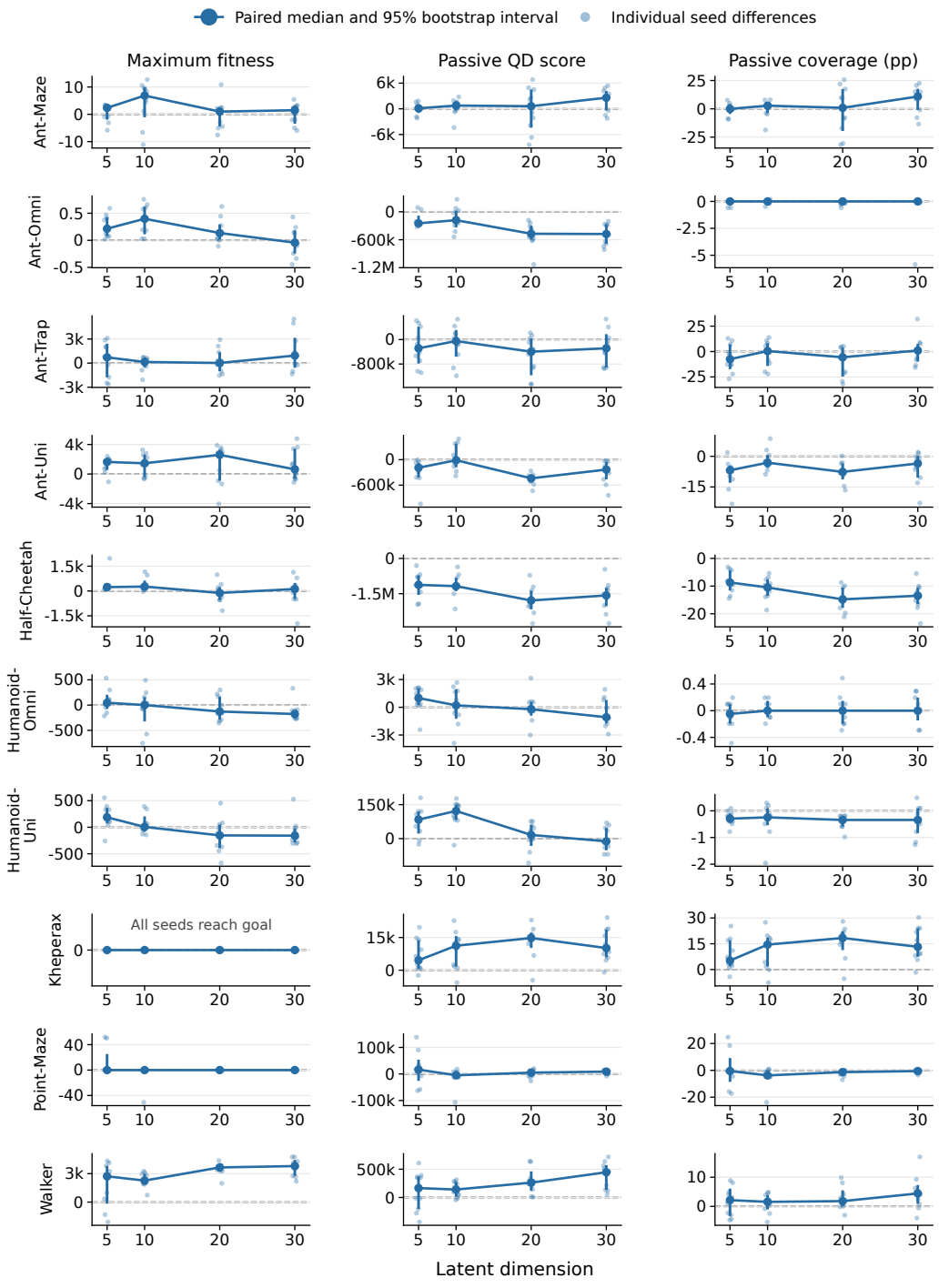

Figure 28: **Paired VC-minus-MSE search effects across latent dimensions.** Rows follow Figure 26; columns show maximum fitness, passive QD score, and passive coverage. Small points show individual seed differences; connected points show their medians and vertical intervals show cell-wise 95% paired bootstrap intervals from 50,000 resamples. The dashed zero line separates effects favoring VC from effects favoring MSE. Fitness and QD differences retain task-specific units, and coverage differences are percentage points. Intervals are not simultaneous or multiplicity-adjusted.

K Effect of Archive Capacity on AURORA-VC

This experiment investigates whether the effect of VC depends on how many policies the active archive can store. We vary archive capacity to test whether the geometric changes and search benefits observed at the default capacity persist under smaller and larger archive capacities.

We compare MSE and VC at active archive capacities of 256, 512, $1{,}024$, $2{,}048$, and $4{,}096$ policies in all ten environments. Latent dimension remains ten, and the 5,000-iteration search budget, proposal mechanism, and representation-retraining schedule are held fixed. Every condition uses a separate 1,024-cell passive evaluator, so changing active capacity does not change the resolution at which search outcomes are measured. The capacity-1,024 condition reuses the primary runs.

We measure terminal active-archive participation rank and the passive evaluator's maximum fitness, QD score, and coverage. Figure 29 places geometry and search outcomes alongside one another, showing individual seeds, medians, and interquartile ranges. Figure 30 shows the within-seed VC-minus-MSE differences, their medians, and cell-wise 95% percentile bootstrap intervals from 50,000 paired resamples.

K.1 Results

The geometric contrast persists across the capacity range. VC generally spreads descriptor variation across more effective dimensions under both small and large archive budgets. Its marginal median participation rank exceeds MSE in at least eight of the ten environments at every tested capacity, and the across-environment median is consistently higher. The first column of Figure 29 shows that this geometric difference persists as the number of retained policies changes. The effect on participation rank is not limited to the default archive size.

Fitness gains are most frequent at intermediate capacities. The persistent rank increase is accompanied by a less uniform fitness benefit. Positive paired median fitness effects are most frequent at capacities 512 and 1,024, where eight tasks favor VC, compared with six at capacity 256 and five at 4,096. Thus, neither increasing the memory budget nor making it more restrictive consistently strengthens the advantage of VC. The paired differences in Figure 30 also show substantial variation across tasks and seeds.

Archive-wide outcomes do not follow maximum fitness. Neither passive QD score nor coverage shows a uniform VC advantage across capacities. Even at the smallest archive size, where VC improves the paired median QD score in most tasks, coverage decreases more often than it increases. Several task–capacity combinations favor MSE on these repertoire measures despite higher participation rank under VC.

K.2 Discussion

We conclude that VC's geometric effect persists across the tested memory budgets, while its search benefit depends on archive capacity and task. Intermediate capacities produce the broadest pattern of positive fitness effects, but this does not identify an archive size that is best for every task or outcome. Because the passive evaluator remains fixed, the outcome differences reflect the policies found rather than a change in evaluation resolution. The sweep does not establish participation rank as the cause of the fitness gains.

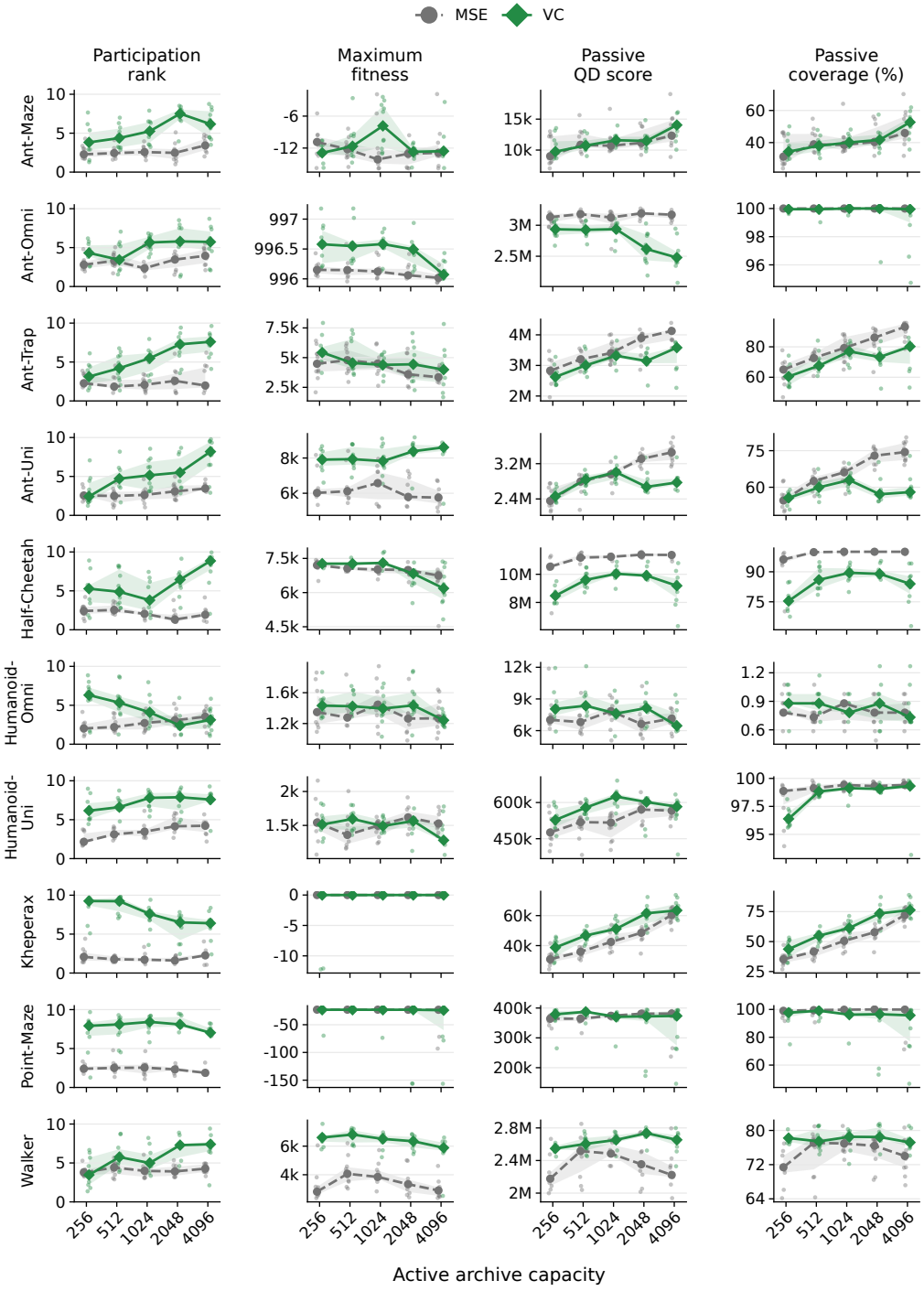

Figure 29: **Descriptor geometry and search outcomes across archive capacities.** Rows show the ten environments; columns show terminal active-archive participation rank, passive maximum fitness, passive QD score, and passive coverage. Capacities are 256, 512, 1,024, 2,048, and 4,096; each horizontal step doubles capacity. Grey dashed and green solid lines show MSE and VC medians, bands show interquartile ranges, and points show all matched seeds. Rank panels share a scale, with a maximum attainable rank of ten; outcome scales are specific to each task, with k and M denoting thousands and millions. Only active capacity varies: all analyzed runs use the same 1,024-cell passive evaluator.

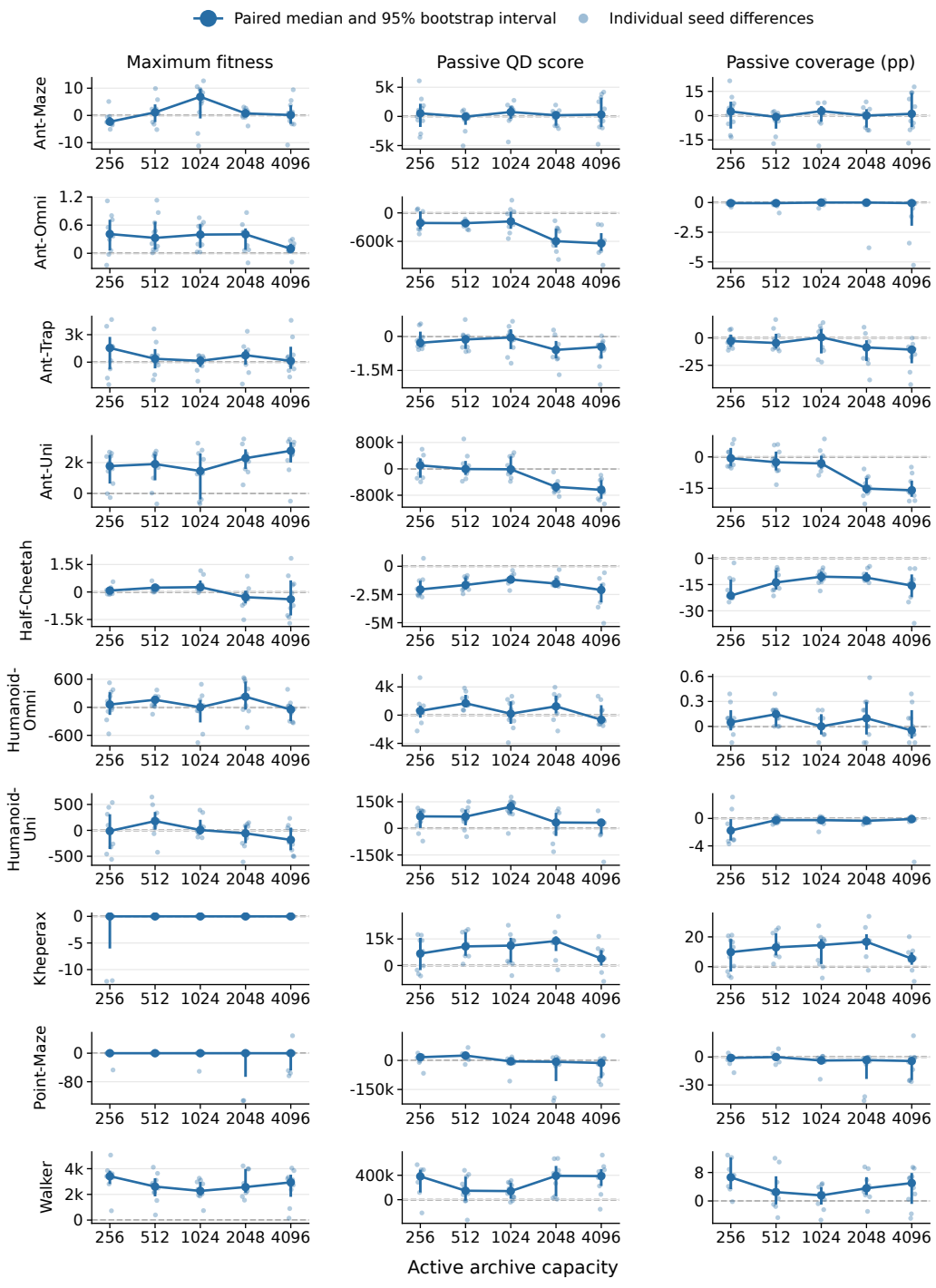

Figure 30: **Paired VC-minus-MSE search effects across archive capacities.** Rows follow Figure 29; columns show maximum fitness, passive QD score, and passive coverage. Small points show individual seed differences; connected points show their medians, with cell-wise 95% paired bootstrap intervals from 50,000 resamples. Positive values favor VC, and the dashed line marks zero. Fitness and QD effects retain task-specific units; coverage effects are percentage points.

L Effect of Variance and Covariance Weights

We investigate how the variance and covariance weights affect descriptor geometry and search outcomes. The primary experiments set both weights to one. We vary them separately and together to examine their individual effects, their interaction, and whether larger weights of both components consistently improve search.

We test all 25 combinations of variance weight λ_V and covariance weight λ_C, with each weight taking values in $\{0, 0.25, 0.5, 0.75, 1\}$. All 25 combinations include seeds 0–9 in every environment. Every condition uses the same experimental protocol and has complete terminal active and passive archives.

Controlled settings. Within each environment–seed block, the search budget, 1,024-member active archive, 1,024-cell passive evaluator, latent width, representation schedule, optimizer, and all non-VC hyperparameters are held fixed. Only (λ_V, λ_C) changes. The $(0, 0)$ cell is reconstruction only; $(\lambda_V, 0)$ and $(0, \lambda_C)$ isolate the variance and covariance axes, while both positive values test the interaction of both components. Geometry is computed from the occupied descriptors in the terminal active search archive. Passive coverage, QD score, and maximum fitness are computed from the terminal fixed external evaluator.

Measurements and statistical analysis. For covariance eigenvalues $\{\eta_j\}_{j=1}^{d}$, the primary geometry endpoint is the participation-ratio effective rank

$$r_{\mathrm{PR}} = \frac{(\sum_j \eta_j)^2}{\sum_j \eta_j^2}.$$

The two component-specific diagnostics are weakest-axis balance, $\min_j \mathrm{sd}(z_j)/\operatorname{mean}_j \mathrm{sd}(z_j)$, and relative off-diagonal covariance, defined as the root-mean-square off-diagonal covariance divided by mean diagonal variance. Every grid cell is first summarized across its ten matched seeds. For any endpoint Y, the seed-paired interior interaction is

$$I_Y(v, c) = Y(v, c) - Y(v, 0) - Y(0, c) + Y(0, 0), \qquad v, c > 0.$$

We compute each interaction within a matched seed and then take the median over the ten seeds. The raw grid medians and interaction summaries below are recomputed from the complete ten-environment, ten-seed factorial. Figure 31 shows the raw ten-seed median, with fixed rank and coverage color scales and environment-specific color scaling for maximum fitness and QD score.

L.1 Results

V and C have complementary geometric effects. The two penalties work together to spread descriptor variation across more effective dimensions: V penalizes low marginal scale, while C discourages redundant coordinates. Their combined effect on participation rank usually exceeds the sum of their separate effects, with positive interactions in 144 of the 160 environment–weight contrasts. The main change occurs when the penalties become active; increasing their weights further does not consistently raise rank. Indeed, the largest weight pair, $(1, 1)$, gives the highest median rank in none of the ten environments. This pattern suggests that activating both penalties matters more consistently than maximizing their strength.

Geometry and performance separate. Peak fitness often shows the same complementarity between V and C, while repertoire-wide outcomes are less consistent. For example, the interaction is positive in 126 of the 160 contrasts for maximum fitness, but in only 34 for coverage; QD-score interactions are split almost evenly between positive and negative. Thus, the tendency for the two penalties to reinforce each other is much clearer for the best policy than for behavioral breadth or accumulated repertoire quality. Additional effective latent dimensions do not guarantee coverage of more behaviors in the external evaluation space, so geometric conditioning and search outcomes must be assessed separately. Terminal maximum fitness saturates across the KHEPERAX grid; this endpoint does not resolve differences in time to reach the goal.

L.2 Discussion

Within each factorial, assigning a weight pair changes the complete online search trajectory, so the surfaces estimate algorithm-level treatment effects of those objectives under the ten-seed protocol.

They do not isolate participation rank as a mediator, identify a universally optimal weight, or show that active covariance directions align with the passive descriptor. Ten seeds improve the stability of the surfaces but do not turn a 25-cell sensitivity study into a selection procedure. The complete ten-environment analysis supports a zero-to-positive conditioning threshold and complementary V/C roles; fine ordering among positive weights and any corresponding performance ranking remain descriptive.

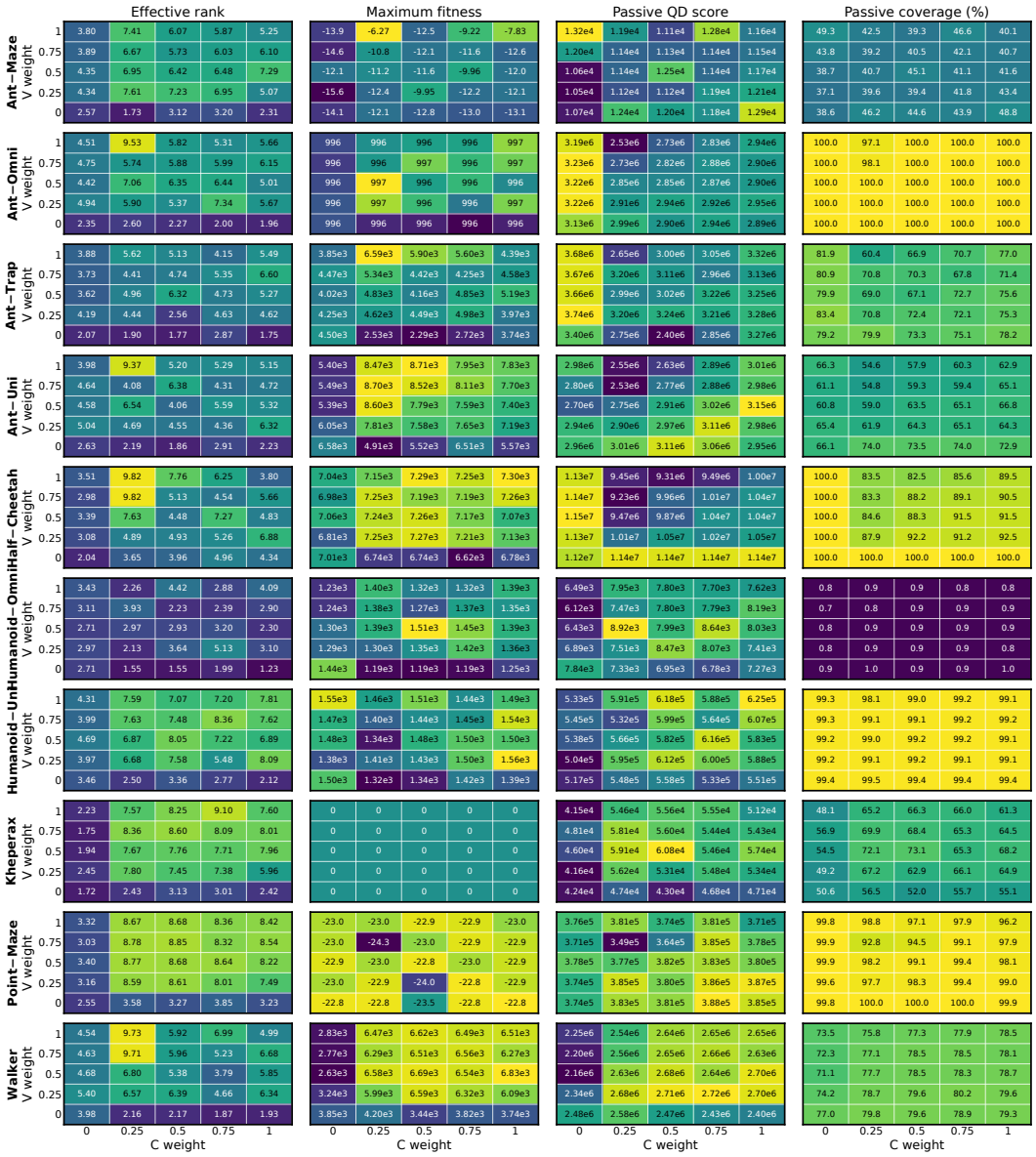

Figure 31: **Ten-seed variance–covariance weight sensitivity across 2,500 validated runs.** Rows are the ten tasks; columns are terminal active-archive participation rank, passive maximum fitness, passive QD score, and passive coverage. Each panel shows the 25 weight combinations, with λ_C on the horizontal axis and λ_V on the vertical axis. Cell text is the raw median across ten matched seeds. Rank and coverage use fixed color ranges; maximum-fitness and QD colors are scaled within task only, so the printed values carry the quantitative scale. Kheperax maximum fitness is frequently saturated, so its QD and coverage panels remain informative. The sweep is not used to select a post-hoc treatment.

M CONTINUED V+C-ONLY ADAPTATION VERSUS FREEZING

This experiment investigates the effect of continuing encoder training with only the variance and covariance penalties compared with freezing the encoder after a shared VC training history.

We evaluate both procedures in ten environments with ten seeds per environment. For each environment and seed, both branches start from the same VC checkpoint after search iteration 1,050. The frozen branch stops encoder training and keeps the checkpoint weights fixed. The V+C-only branch continues scheduled encoder training with $L_V + L_C$, without reconstruction loss, and uses the updated descriptors for archive competition. Policy search continues in both branches through search iteration 5,000.

M.1 RESULTS

Trajectory-integrated effects mostly favor continued adaptation. The paired seed median is positive for QD-score AUC in all ten tasks, for coverage AUC in nine tasks with a tie in Humanoid-Omni, and for maximum-fitness AUC in eight tasks with a negative effect in Humanoid-Omni and a tie at the Kheperax goal. At raw $p < 0.05$, QD-score AUC favors continuation in ANTMAZE, ANTTRAP, ANTUNI, HUMANOIDUNI, KHEPERAX, POINTMAZE, and WALKER. Coverage AUC favors continuation in ANTTRAP, ANTUNI, HUMANOIDUNI, and KHEPERAX; maximum-fitness AUC does so in ANTOMNI, ANTUNI, and HALFCHEETAH. The environment direction counts describe the signs of the observed paired seed medians and should not be read as ten separate discoveries.

Final peak-fitness effects remain heterogeneous. At search iteration 5,000, the paired seed median for maximum fitness is positive in eight tasks and tied in Humanoid-Omni and Kheperax. It is $+778.1$ in ANTUNI (10/10 positive seed pairs; raw $p = 0.0020$), $+224.1$ in HALFCHEETAH (9/10; $p = 0.0098$), $+181.7$ in ANTTRAP (6/10; $p = 0.7520$), and $+56.0$ in WALKER (6/10; $p = 0.6680$). The numerically small $+0.135$ effect in ANTOMNI is consistent across all ten pairs and significant at raw $p = 0.0020$. Point-Maze changes by only $+0.010$, with a deterministic paired bootstrap interval spanning zero.

Archive-wide outcomes are not reducible to peak fitness. The paired seed median for final passive QD favors continuation in nine tasks and is negative only in Ant-Omni. The raw paired seed medians include $+369{,}863$ in Ant-Trap, $+159{,}307$ in Ant-Uni, $+111{,}150$ in Walker, and $+110{,}827$ in Half-Cheetah. The paired seed median for final coverage is positive in eight tasks and tied in Ant-Omni and Humanoid-Omni. At raw $p < 0.05$, final QD score favors continuation in ANTMAZE, ANTTRAP, ANTUNI, HUMANOIDUNI, KHEPERAX, and WALKER. Final coverage favors continuation in ANTMAZE, ANTTRAP, ANTUNI, HUMANOIDUNI, and KHEPERAX. The contrast can therefore change archive-wide accumulation even when its peak-fitness effect is small or uncertain.

M.2 DISCUSSION

Because both arms restore the same complete state, their post-handoff difference is a causal effect of the executed late procedure in these 100 checkpoint pairs. The experiment supports continued V+C-only adaptation over freezing most clearly for passive QD accumulation, with less uniform evidence for peak fitness. It does not isolate the variance and covariance terms from one another, compare V+C-only updates with the ordinary reconstruction-plus-VC objective, or show that a change in participation rank, admission, or lineage mediates an outcome. Mechanism instrumentation was disabled for these branches, so no such intermediate quantity is inferred from the performance curves.

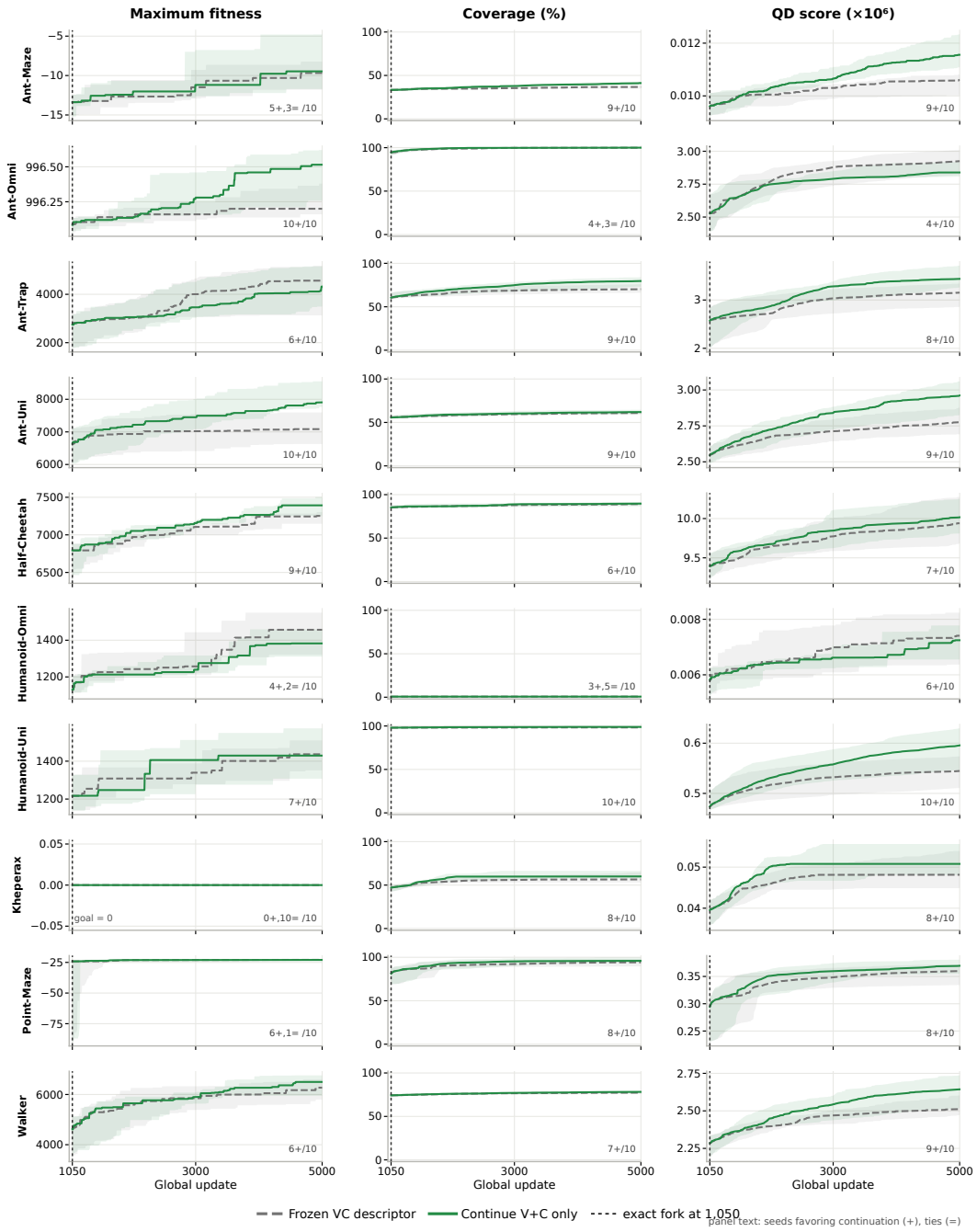

Figure 32: **Continuing V+C-only adaptation versus freezing from 100 exact VC checkpoints.** Rows are the ten tasks and columns are passive maximum fitness, coverage, and QD score. Both curves in each panel share the search-iteration-1,050 source state; lines are ten-seed medians and bands are interquartile ranges through search iteration 5,000. The small annotation gives the number of paired seeds whose final value favors continuation; an equals count is shown when ties occur. Because each pair shares its complete pre-handoff state and randomness, post-handoff differences isolate the executed late procedure.

N How Search History Shapes the Effect of VC

This experiment investigates how earlier representation learning shapes the effect of VC later in search. We evaluate continuing or switching the training objective after histories produced by MSE or VC, with the goal of testing whether the benefit of late VC depends on the inherited search state.

We evaluate all four schedules in ten environments with ten seeds per environment. For each environment and seed, search first runs with either MSE or VC through iteration $1{,}050$, which is approximately one-fifth of the way through the $5{,}000$-iteration training budget. This allows the initial objective to shape the representation and archive while leaving most of training to evaluate the effect of continuing or switching objectives. Each saved state then provides two continuations: one trains with MSE and the other with VC, while policy search continues through iteration $5{,}000$. The four schedules are MSE→MSE , MSE→VC, VC→MSE, VC→VC. Here, MSE uses reconstruction loss, while VC retains reconstruction and adds the variance and covariance penalties. The experiment contains 400 completed search histories: 200 continuations under the original objective and 200 branches that switch objectives.

Within each environment, seed, and early objective, both continuations start from the same policies, archive fitnesses, and emitter and replay state. Each branch trains under its assigned loss, re-encodes the archive, and continues the retraining.

We compare the two late objectives within each early history. We measure final maximum fitness, passive coverage, and passive QD score, together with each outcome's time-normalized area under the curve (AUC) from iteration $1{,}050$ through $5{,}000$. Coverage differences are reported in percentage points.

N.1 Results

Continuing VC after a VC history usually improves peak fitness. Compared with switching to MSE at iteration $1{,}050$, continuing VC has a positive paired median effect on final maximum fitness in eight tasks, a negative effect in PointMaze, and a tie at the Kheperax goal. The largest gains are $+936.5$ in Walker, $+662.5$ in AntUni, and $+272.9$ in HalfCheetah. At raw $p < 0.05$, final maximum fitness favors late VC in AntOmni, AntUni, HalfCheetah, and Walker, and favors late MSE in PointMaze. The paired median effect on maximum-fitness AUC also favors late VC in eight tasks. At raw $p < 0.05$, maximum-fitness AUC favors late VC in AntOmni, AntUni, HalfCheetah, HumanoidUni, and Walker.

Adding VC after an MSE history gives less consistent fitness gains. Compared with continuing MSE, switching to VC has a positive paired median effect on final maximum fitness in six tasks, negative effects in AntOmni, HumanoidOmni, and PointMaze, and a tie in Kheperax. The paired median gains include $+1{,}439.2$ in AntUni, $+931.3$ in AntTrap, $+306.5$ in Walker, and $+65.6$ in HalfCheetah. Only AntUni has a significant final-fitness difference at raw $p < 0.05$ ($p = 0.0156$). The paired median maximum-fitness-AUC effect is positive in six tasks, with only AntUni crossing the raw 0.05 threshold.

Peak-fitness gains need not improve the full repertoire. After a VC history, continuing VC reduces final coverage in six task-level paired medians, increases it in three, and ties in AntOmni. Its final QD-score effect is negative in six tasks and positive in four. In HalfCheetah, the $+272.9$ final-fitness gain accompanies a coverage reduction of 7.96 percentage points and a QD-score reduction of 1.06×10^6. Coverage AUC favors late VC in only two tasks and QD-score AUC in four; both AUC measures favor late MSE in PointMaze at raw $p < 0.05$.

After an MSE history, adding VC increases final coverage in four task-level paired medians, decreases it in five, and ties in one. Final QD score increases in five tasks and decreases in five. At raw $p < 0.05$, final QD score favors late VC in HumanoidUni, with a paired median gain of $37{,}856$, and favors late MSE in AntTrap and AntUni. The paired median effects on coverage AUC and QD-score AUC are positive in six tasks each. Only the HumanoidUni QD-score-AUC gain crosses the raw 0.05 threshold; no coverage-AUC test does. These task counts describe the directions of paired median effects and should not be read as counts of significant differences.

N.2 Discussion

The effect of late VC depends on the search state inherited from earlier training. Continuing VC after a VC history gives more consistent peak-fitness gains than adding VC after an MSE history, while effects on coverage and QD score remain mixed. The late objective acts on policies, archives, replay data, and representations already shaped by the early objective. Because the two late branches within each checkpoint start from the same complete state, their differences isolate the effect of the assigned late training procedure and its subsequent search updates. Continued VC regularization can improve peak fitness even after VC has shaped the early search state. The mixed coverage and QD-score effects mean that this benefit should not be interpreted as a general improvement in repertoire quality.

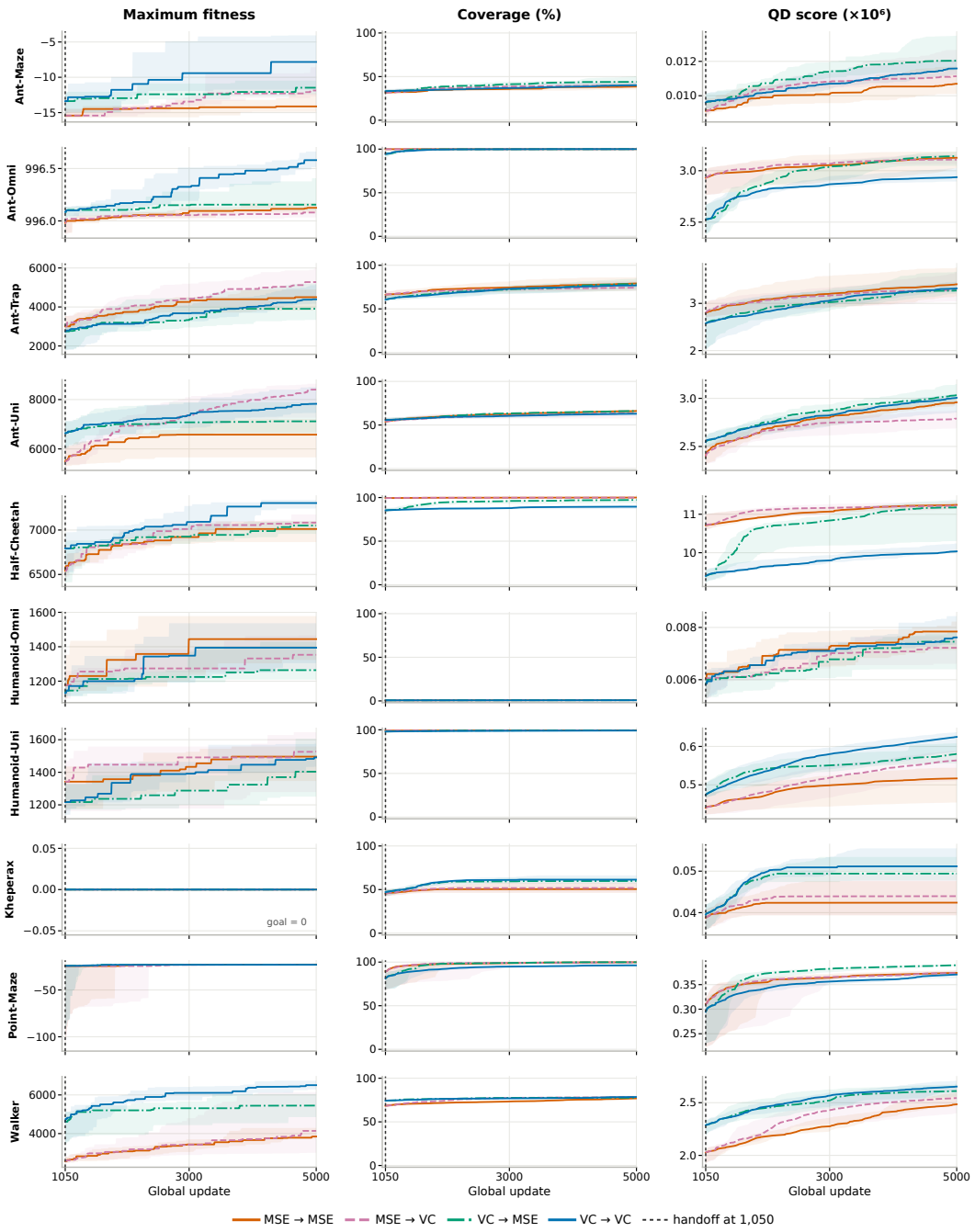

Figure 33: **The late representation objective changes a common inherited search state in a task-dependent way.** Rows are the 10 tasks outlined in the paper; columns are passive maximum fitness, passive coverage, and passive QD score. Each panel combines MSE→MSE and VC→VC source histories with the exact-checkpoint MSE→VC and VC→MSE continuations. Lines are ten-seed medians, shaded regions are interquartile ranges, and the dotted line marks the search-iteration-1,050 handoff. All 400 completed search histories continue through search iteration 5,000.

O ARCHIVE-DYNAMICS DIAGNOSTICS

This experiment investigates how VC changes the evolving archive during search. We track representation updates, candidate admissions, policy lifetimes, and parent reuse to examine whether the fitness gains are accompanied by more persistent competitive relationships and greater reuse of retained policies.

We compare the 200 instrumented PGA-AURORA runs with MSE and VC in ten environments and ten matched seeds per environment. Each run uses ten-dimensional embeddings, a 1,024-policy active archive, and a 5,000-iteration search budget. Unless stated otherwise, arm summaries are pooled medians of 100 run-level values, and environment direction counts use the sign of the paired seed median within each task.

Fitness prediction and representation change. At five prespecified retraining events, $\{4, 8, 15, 23, 31\}$, we split the current incumbents in half and fit ridge regression from all embedding coordinates to fitness, standardizing coordinates using only the training half. Held-out R^2, summarized by its within-run median across events, measures how easily fitness can be predicted from the representation. At the same events, we encode a frozen bank of 4,096 trajectories before and after training, center the embeddings, and fit an unconstrained linear map A between them. We measure the relative alignment residual

$$r_{\text{lin}} = \frac{\min_A \|Z^- A - Z^+\|_F}{\|Z^+\|_F}. \tag{19}$$

Because A can stretch or shear distances, a lower residual does not establish the distance-preservation bound in Proposition 3.

To measure changes in competition directly, we compare neighbor sets and priorities at all 31 post-warm-start retraining events. Stage A encodes the incumbents before retraining, stage B re-encodes the same ordered trajectories afterward, and stage C rebuilds priorities without changing embeddings or membership. This holds the evaluated policies fixed while the representation changes. Across all 6,200 events, the maximum absolute stage-B/stage-C participation-rank difference is below 6×10^{-14}, confirming that rebuilding alone does not change the measured geometry.

Candidate displacement and population histories. For uniformly sampled valid candidates with a known primary parent, we measure child–parent and nearest-incumbent distances in units of $d_{\min}$, the lowest priority retained at that candidate batch's capacity decision. These distances describe movement under the active metric, not parameter-space mutation size. Admissions define births, the initial population starts at iteration zero, and removals during candidate insertion define deaths. The top-decile survival analysis uses post-initial admissions, with fitness classified relative to the archive at the most recent retraining event. We also track membership survival between retrainings and the age distribution of the current archive. Scheduled rebuilding contains only the incumbents and removes no members.

For accepted children and new maximum-fitness records, we average age over the resolved parents of each child, then take the within-run median across children. Parent-age summaries use resolved ages; 701 of 8,904 post-warm-start record events lack resolved age or fitness. The older-than-100 share retains unresolved records in its denominator without counting them as older. Accepted and displaced fitness percentiles are compared as separate within-run distributions, rather than as matched candidate–victim pairs.

O.1 RESULTS

Closer linear alignment does not imply easier fitness prediction. A global linear map accounts for more of the representation change under VC: the fixed-bank residual falls from 0.695 to 0.539. However, held-out fitness prediction becomes less accurate, with R^2 falling from 0.373 to 0.128. The changed representation does not simply make fitness easier to recover with this linear probe. After accounting for task and objective, post-hoc run-level associations between fitness-readout accuracy and either neighbor preservation or the alignment residual are small, with no detectable seed-level relationship.

Retraining preserves more competitive relationships, but still reorganizes them. Fitter-neighbor Jaccard rises from 0.120 to 0.264 when evaluated on identical incumbents before and after retraining

(Figure 34b). Ordinary neighbor overlap and priority agreement also increase; all three measures favor VC in eight task-level paired medians. These changes indicate greater continuity in competition across encoder updates. However, the nearest fitter neighbor still changes for roughly three-quarters of incumbents under VC, so the competitive structure remains substantially reconfigured at each update.

Candidate motion changes relative to the active metric. Children are closer to their parents in survival-cutoff units under VC, with a lower paired median distance in every task. The pooled median falls from 6.52 to 3.58, while distance to the nearest incumbent remains nearly unchanged at about 0.63. Thus, parent–child movement decreases but remains much larger than ordinary candidate-to-incumbent spacing; it does not show that children stay in their parents' competitive neighborhoods. The fitness-percentile gap between accepted and displaced populations also increases, from 0.195 to 0.684 percentage points.

Persistence increases while admissions decrease. VC archives turn over more slowly: fewer candidates enter, members survive longer, and older policies remain available to search. Uniform-candidate acceptance falls from 16.1% to 11.6%, while top-decile survival through 500 iterations rises from 0.276% to 2.69% (Figure 34a). Membership survival between retrainings is higher in nine task-level paired medians, and the archive-age distributions show older retained members (Figure 34c). Lower inflow itself prolongs residence, so these observations do not establish stronger protection of elite policies at each insertion.

Records occur later and use older retained parents. The longer-lived archive is accompanied by later record setting and reuse of older parents. The final maximum-fitness record shifts from a pooled median iteration of $3{,}212$ under MSE to $4{,}290$ under VC (Figure 34d), and the median per-record mean resolved-parent age rises from 9.3 to 16.9 iterations. These observations show that retained policies contribute to later search, but record timing can also reflect slower arrival at a fitness ceiling.

The diagnostics do not show uniformly broader behavioral exploration. The latent graph becomes less locally correlated in fitness, has more mutual-kNN components, and has more unequal competitive in-degrees. However, its largest component still contains more than nine in ten members under VC, so the additional components do not indicate strongly isolated niches. Latent elite dispersion also falls in WALKER and ANTUNI despite their favorable fitness results. In the fixed passive space, the median across task-specific run medians of novel-cell admissions remains near 726 per run in both arms, with flat novel-proposal rates and flat to slightly lower capture ratios under VC. The measured changes therefore concern how search retains and reuses policies without showing a uniform expansion of behavioral discovery.

O.2 DISCUSSION

We conclude that VC changes both the persistence of competitive relationships and the archive's retention and reuse of policies. The combination of lower inflow, longer survival, and later records from older parents is consistent with retained search history contributing to continued progress. The increased persistence suggests that encoder adaptation can retain continuity in the archive's competitive structure. Longer residence gives retained policies more opportunities to be sampled as parents, allowing search to build on earlier discoveries. Their trajectories also contribute to later encoder training, so the archive carries accumulated experience into both candidate generation and the representation that guides future competition.

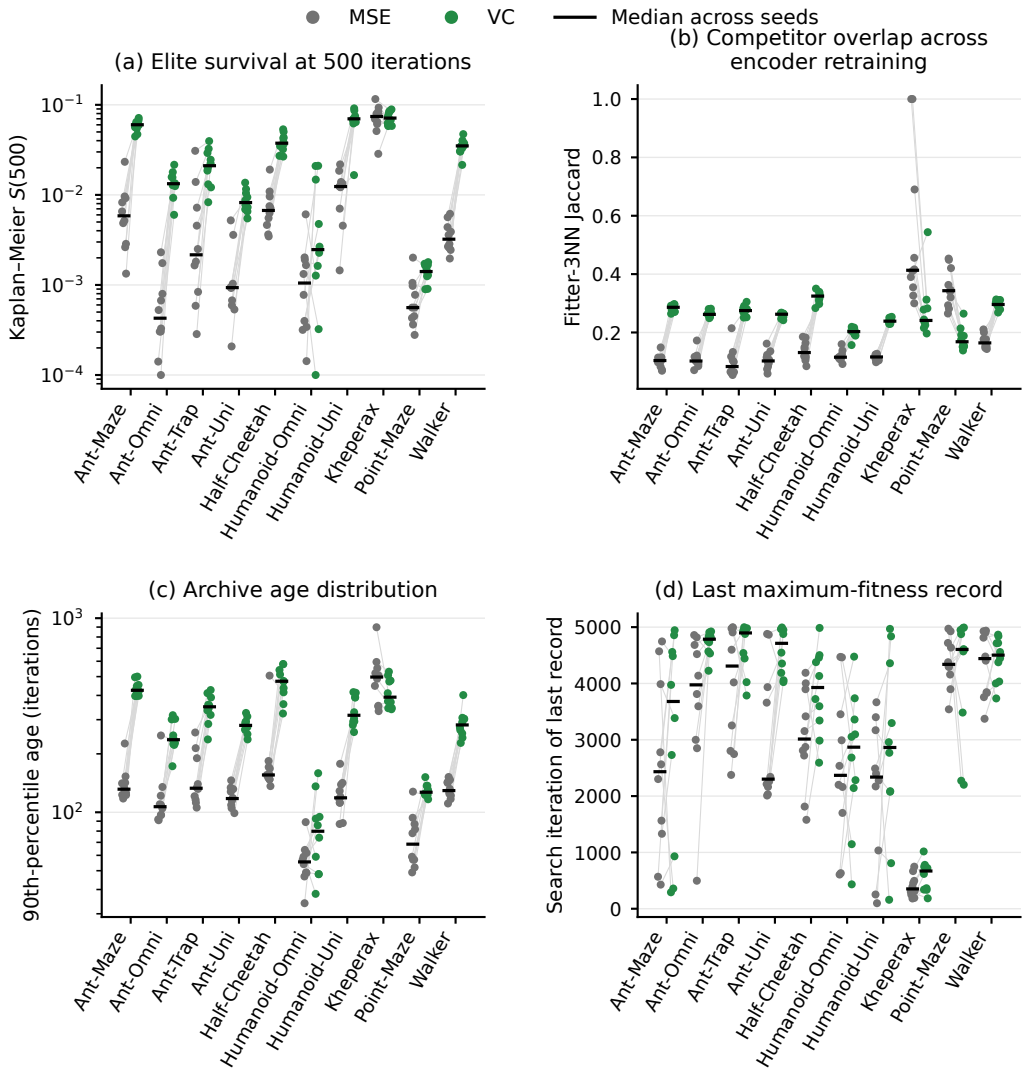

Figure 34: **Population dynamics across 200 instrumented runs.** Grey and green points are MSE and VC seeds; lines connect matched seed pairs, and black bars show marginal ten-seed task medians. Panels report (a) Kaplan–Meier survival through 500 search iterations for post-initial top-decile admissions; (b) pre/post-retraining fitter-neighbor Jaccard for identical incumbents; (c) the within-run across-retraining archive-age p_{90}; and (d) the search iteration of the final new maximum-fitness record. Panels (a) and (c) use logarithmic axes; the two survival estimates below 10^{-4} are displayed at 10^{-4}.

P INTERVENTIONS ON THE SURVIVOR RULE AND MECHANISM CONTROLS

This appendix investigates whether VC's task-dependent fitness gains arise partly from changes in competition and survivor selection. Our hypothesis is that conditioning learned distances changes which fitter policies compete, which policies survive, and which retained policies supply future proposals and encoder-training data.

We combine matched-seed search experiments and checkpoint continuations across ten environments. We choose candidates with a fitness-blind geometric score while matching the native admission count on each treatment pool, separate the metrics used for neighbor ordering and distance scoring, and transfer VC admission quotas or exact survivor sets into archives that retain MSE descriptors. These interventions examine whether the identities of retained policies contribute beyond the number admitted and whether VC's selection decisions affect search when the archive retains MSE coordinates.

The final experiment compares amplification of fitness-predictive and random latent directions to test whether emphasizing fitness-related variation produces a more useful selection metric. Across these studies, we assess maximum fitness alongside QD score and coverage to distinguish gains in peak fitness from changes in behavioral breadth.

P.1 MATCHED-COUNT, FITNESS-BLIND SURVIVOR SELECTION

We investigate whether VC's late-search advantage depends on which candidates survive beyond the rule that determines how many are admitted. At each update, we compute the native strictly-fitter decision on the treatment's current pool, record its admission count K_t, and replace its selected candidates with exactly K_t candidates chosen by a fitness-blind geometric score. The score is mean Euclidean distance to up to three nearest other valid pool members, regardless of fitness. We retain the highest-scoring K_t candidates and fill the remaining capacity with the highest-scoring incumbents, using pool order to break ties.

We run 200 treatment continuations across ten environments, ten seeds, and the MSE and VC objectives. Both objective branches resume from 100 VC checkpoints at update $1{,}050$ and continue to update $5{,}000$. The native controls are the matched VC-to-MSE continuations and the source VC run suffixes; checkpoint hashes verify the shared starting states, including random-number state. We compare final outcomes and AUC, defined as the trapezoidal time average over the common post-checkpoint grid sampled every five updates. Effects are paired seed medians with cell-wise 95% bootstrap intervals from $50{,}000$ resamples; exact sign-flip tests separately target paired means. We also measure how blindness changes the VC advantage using $(\text{blind VC} - \text{blind MSE}) - (\text{native VC} - \text{native MSE})$.

P.1.1 RESULTS

Matching the admission count still changes which policies enter and remain in the archive. All 790,000 checks confirmed exact agreement between the treatment and its native-shadow count on the same pool, while candidate and survivor identities differed (Table 9). The blind selector admitted lower-fitness candidates in all 20 environment–objective comparisons.

Table 9: Run-reduced manipulation checks for the matched-count blind selector, reported as medians over 100 runs per objective.

Branch	Admissions/update	Candidate Jaccard	Survivor Jaccard	Score-rank correlation	Admitted-fitness difference
MSE with blind selection	143.142	0.499	0.798	0.743	-41.053
VC with blind selection	108.198	0.448	0.840	0.800	-83.860

These changes redirect search toward broader coverage but lower peak fitness. Blind selection reduces both final maximum fitness and maximum-fitness AUC in nine task medians under each objective (Table 10). The final-fitness VC advantage is positive in eight tasks under native selection but only two under blind selection, and the difference-in-differences is negative in nine tasks. VC therefore loses much of its advantage when fitness-conditioned identities are removed, although

WALKER retains a positive effect. The table distinguishes these task directions from cases whose cell-wise confidence interval excludes zero.

Table 10: Environment-level direction summary for the matched-count blind selector, where every contrast is blind minus native except the stated difference-in-differences.

Outcome	Contrast	Directions $+/-/=$	CI excludes zero $+/-$	Median of environment medians
Final maximum fitness	MSE blind − native	$0/9/1$	$0/8$	-120.723
Final maximum fitness	VC blind − native	$0/9/1$	$0/9$	-263.485
Final maximum fitness	Difference-in-differences	$0/9/1$	$0/4$	-154.920
Maximum-fitness AUC	MSE blind − native	$0/9/1$	$0/8$	-59.248
Maximum-fitness AUC	VC blind − native	$0/9/1$	$0/9$	-158.993
Maximum-fitness AUC	Difference-in-differences	$0/9/1$	$0/4$	-99.802
Final coverage	MSE blind − native	$9/0/1$	$3/0$	$+0.805$
Final coverage	VC blind − native	$8/1/1$	$6/0$	$+3.518$
Final coverage	Difference-in-differences	$8/1/1$	$2/0$	$+1.073$
Coverage AUC	MSE blind − native	$10/0/0$	$4/0$	$+0.596$
Coverage AUC	VC blind − native	$9/1/0$	$8/0$	$+3.009$
Coverage AUC	Difference-in-differences	$8/2/0$	$3/0$	$+0.529$

P.1.2 DISCUSSION

We conclude that fitness-conditioned survivor selection contributes to the late-search advantage of VC beyond applying the native admission-count rule. Which policies are retained affects the balance between finding high-fitness individuals and expanding coverage. The count constraint applies to each treatment pool's shadow decision; treatment and native-control runs subsequently develop different pools and admission sequences, so the result concerns the continuing selection policy.

P.2 COMPETITION-GEOMETRY FACTORIAL

We investigate how the metric used to identify fitter neighbors interacts with the metric used to score their distances. We test whether either part of VC competition transfers independently, or whether changing both produces the stronger effect.

We run a two-by-two factorial with 400 fresh runs across ten environments, ten matched seeds, and four arms. MM uses MSE for both neighbor ordering and distance scoring; VM uses VC ordering with MSE scoring; MV uses MSE ordering with VC scoring; VV uses VC for both. The surrounding search algorithm and dual-encoder implementation are shared, including a retraining schedule tied to MSE epochs, so VV differs from the primary VC implementation. We measure final passive outcomes at update $5{,}000$ and their trapezoidal time averages over updates 1–$5{,}000$. Paired seed medians have cell-wise 95% bootstrap intervals from $20{,}000$ resamples. All 400 runs completed, and checks confirmed the assigned neighbor, distance, and rebuild metrics and agreement between logged endpoints and saved passive archives.

P.2.1 RESULTS

Changing only one part of competition consistently weakens fitness in this factorial. Both hybrids, VM and MV, have negative paired median effects relative to MM in all nine nonsaturated tasks, for both final maximum fitness and maximum-fitness AUC. Using VC for both operations reverses that pattern: VV exceeds both hybrids in all nine tasks, with a positive factorial interaction in every task. VV also improves on MM in six of the nine task medians for both fitness outcomes.

Tables 11 and 12 show the arm medians, while Table 13 reports the paired effects and their uncertainty. For example, the paired VV-minus-MM fitness-AUC gain in WALKER is $1{,}266.8$, with interval $[613.4, 2{,}637.3]$. The paired effects account for seed matching and are not differences between marginal medians. KHEPERAX remains in the descriptive tables but is excluded from the nine-task summary because all arms reach its final-fitness ceiling.

Table 11: Marginal ten-seed medians of passive maximum-fitness AUC in the competition-geometry factorial.

Environment	MM	VM	MV	VV
AntMaze	−13.95	−15.73	−15.70	−13.88
AntOmni	995.9	995.5	995.6	996.1
AntTrap	4,077.4	3,081.6	2,755.5	3,549.6
AntUni	5,965.5	4,450.1	4,664.2	7,090.2
HalfCheetah	6,490.4	5,237.1	5,835.5	6,881.9
HumanoidOmni	1,177.3	963.4	880.4	1,224.9
HumanoidUni	1,379.0	1,069.0	1,119.3	1,283.7
Kheperax	−1.041	−0.9787	−0.8414	−1.283
PointMaze	−33.39	−59.03	−93.89	−37.71
Walker	3,457.0	2,390.8	2,107.8	5,158.1

Table 12: Marginal ten-seed medians of final passive maximum fitness in the competition-geometry factorial.

Environment	MM	VM	MV	VV
AntMaze	−12.61	−15.61	−15.67	−12.61
AntOmni	996.1	995.8	995.8	996.5
AntTrap	5,305.4	3,457.4	3,121.1	4,883.2
AntUni	6,551.1	4,621.1	4,897.3	7,730.1
HalfCheetah	6,967.1	6,306.5	6,298.0	7,246.6
HumanoidOmni	1,277.1	1,049.5	988.8	1,388.0
HumanoidUni	1,537.6	1,277.5	1,240.8	1,434.4
Kheperax	0	0	0	0
PointMaze	−22.86	−23.43	−47.77	−22.98
Walker	4,879.2	2,486.4	2,291.2	6,363.6

Table 13: Paired VV-minus-MM effects in the competition-geometry factorial, shown as paired seed medians with cell-wise 95% deterministic paired bootstrap intervals.

Environment	Fitness AUC $VV - MM$	Seeds $+/=/-$	Final fitness $VV - MM$	Seeds $+/=/-$
AntMaze	1.152 [−1.730, 3.468]	7/0/3	0.4139 [−4.051, 4.717]	6/0/4
AntOmni	0.1874 [0.0787, 0.3563]	9/0/1	0.2609 [0.0387, 0.6176]	9/0/1
AntTrap	−191.1 [−1,609.3, 1,004.4]	5/0/5	−582.7 [−1,493.1, 928.8]	4/0/6
AntUni	1,027.9 [426.4, 1,911.7]	8/0/2	1,429.0 [964.5, 1,911.5]	9/0/1
HalfCheetah	318.3 [24.04, 491.9]	8/0/2	195.8 [166.8, 313.6]	9/0/1
HumanoidOmni	0.4198 [−130.7, 203.9]	5/0/5	37.51 [−156.9, 234.0]	5/0/5
HumanoidUni	−86.39 [−200.5, 29.58]	3/0/7	−135.2 [−272.2, 6.612]	3/0/7
PointMaze	−0.1894 [−34.77, 9.862]	5/0/5	−0.00951 [−1.456, 0.0308]	5/0/5
Walker	1,266.8 [613.4, 2,637.3]	9/0/1	1,542.0 [426.8, 3,324.3]	10/0/0

P.2.2 Discussion

We conclude that neighbor ordering and distance scoring should be considered together when evaluating how VC changes competition. The benefit of VC in this factorial is strongest when the same learned metric determines both which policies compete and how their separation is scored. This finding concerns the tested dual-encoder procedure and its shared retraining schedule; it supports a coordinated change in competition within that setting.

P.3 Survivor-policy transplant

In this experiment, we investigate whether VC can improve search through its survivor rule while the active archive continues to store and use MSE descriptors. We transfer the selection decision into an otherwise MSE-based archive to test the contribution of this part of the search loop.

We add 100 transplant runs, T, across ten environments and ten matched seeds, using the 200 existing MM and VV runs from the competition-geometry factorial as controls. T maintains a VC shadow

that selects survivors from the live MSE incumbent-plus-candidate pool. It installs those exact survivors in MSE order while retaining MSE descriptors, cutoff, emitter inputs, and rebuilds. Thus, $T - MM$ transfers both the admission count and survivor identities; $VV - T$ compares the resulting search with a procedure that also uses VC archive geometry. VV retains the factorial's retraining schedule tied to MSE epochs. All runs last 5,000 iterations, and effects are paired seed medians with cell-wise 95% bootstrap intervals. GPU assignment changed exact initial parameter hashes in some matched cells; hardware-restricted analyses serve as sensitivity checks.

P.3.1 RESULTS

The transplant changes search even though the archive continues to use MSE coordinates. All 500,000 treatment updates installed the VC-selected count and identities while passing the MSE descriptor, installation-order, cutoff, and routing checks. The admitted sets differed substantially from native MSE decisions, confirming that the intervention changed selection.

These selection changes improve maximum-fitness AUC in seven of ten task medians and final fitness in six of nine nonsaturated tasks (Table 14). The clearest favorable effects occur in ANTUNI and WALKER, while HUMANOIDUNI and POINTMAZE favor MM on fitness AUC. The remaining comparison with VV is more mixed: fitness-AUC directions split evenly, and every $VV - T$ fitness interval crosses zero. T also has higher coverage AUC than VV in nine tasks, although the corresponding intervals all cross zero.

Table 14: Survivor-policy transplant effects, reported as paired seed medians with cell-wise 95% deterministic paired bootstrap intervals.

Environment	$T - MM$ AUC	$VV - T$ AUC	$T - MM$ final	$VV - T$ final
ANTMAZE	2.682 [−0.1196, 6.516]	−0.5570 [−7.034, 1.662]	0.4206 [−2.526, 9.177]	−0.3694 [−8.451, 2.460]
ANTOMNI	0.2484 [0.1176, 0.3776]	−0.0556 [−0.1291, 0.0668]	0.2741 [0.1282, 0.5160]	0.0741 [−0.1479, 0.1476]
ANTTRAP	119.5 [−1,251.5, 645.4]	−180.6 [−1,090.4, 527.5]	−417.0 [−1,107.0, 652.1]	−140.9 [−869.0, 818.3]
ANTUNI	1,041.2 [660.8, 1,928.9]	76.99 [−1,075.6, 500.0]	1,702.4 [958.4, 2,693.2]	−327.8 [−1,349.4, 680.6]
HALFCHEETAH	137.1 [−738.8, 327.0]	137.0 [−37.52, 893.0]	172.3 [−754.5, 235.0]	120.8 [−51.36, 1,004.6]
HUMANOIDOMNI	−0.7704 [−139.8, 216.1]	69.18 [−120.7, 130.9]	135.4 [−0.0443, 292.3]	74.55 [−290.6, 235.6]
HUMANOIDUNI	−64.88 [−176.2, −5.600]	−9.978 [−88.40, 129.7]	−92.72 [−317.0, 50.04]	−24.46 [−188.6, 111.8]
KHEPERAX	0.0631 [−0.0707, 0.9027]	−0.3316 [−0.8060, 0.0241]	0 [0, 0]	0 [0, 0]
POINTMAZE	−8.074 [−83.10, −0.6059]	11.03 [−34.95, 35.87]	−0.0365 [−24.57, 0.00755]	0.0668 [−1.472, 24.58]
WALKER	1,755.7 [431.3, 2,537.5]	26.58 [−502.8, 495.2]	1,785.8 [−44.78, 3,463.2]	55.51 [−475.9, 431.8]

P.3.2 DISCUSSION

We conclude that the VC survivor rule can produce fitness benefits while the active archive retains MSE descriptors. Selection therefore provides a route through which the learned geometry influences later search by changing the policies available for future proposals. The transplant transfers both counts and identities, while $VV - T$ compares complete procedures that develop different search histories. The next experiment compares this full transfer with transferring only the VC admission quota.

P.4 COUNT VERSUS SURVIVOR IDENTITY

In this experiment, we investigate whether the benefit of transferring the VC survivor rule depends on the identities retained beyond importing its admission quota. We compare two ways of using a live VC shadow: K takes its scalar quota and lets MSE select the survivors, while T installs the exact VC-selected survivor set. Both arms retain MSE coordinates and archive state.

We run 100 fresh pairs, or 200 search trajectories, across ten environments and ten seeds for 5,000 updates. Each pair uses the same GPU and matches all seven recorded model, optimizer, normalization, and logical-initialization hashes. Execution order is balanced within every environment. K applied its assigned quota and T installed its assigned identities at every update; first-pool admission counts matched across all 100 pairs, although the logs do not contain a complete identity hash for that pool. We report paired $T - K$ effects, with cell-wise 95% bootstrap intervals from $20{,}000$ resamples for maximum-fitness AUC.

P.4.1 RESULTS

Transferring the selected identities usually improves peak fitness beyond transferring the quota alone. Maximum-fitness AUC favors T in seven of ten task medians and 62 of 100 pairs (Table 15). Final fitness favors T in six tasks, favors K in three, and ties in one. The largest favorable AUC effects with intervals above zero occur in ANTUNI and WALKER, while POINTMAZE favors K with an interval below zero.

The repertoire measures reveal a different response. T improves QD-score AUC in only two task medians and lowers coverage AUC in nine, showing that its peak-fitness advantage is accompanied by narrower coverage in most tasks. The table reports task-level fitness-AUC estimates and intervals; the available final-fitness summary provides task directions only.

Table 15: Maximum-fitness AUC effect of exact VC survivor identity beyond the VC admission count, reported as the paired seed median of $T - K$ with a cell-wise 95% deterministic paired bootstrap interval.

Environment	Median $T - K$ [95% CI]	Seeds $+/=/-$
ANTMAZE	1.387 [-4.377, 4.150]	$6/0/4$
ANTOMNI	0.149 [0.028, 0.361]	$9/0/1$
ANTTRAP	12.41 [-778.02, $1{,}525.3$]	$5/0/5$
ANTUNI	846.86 [371.76, $1{,}759.2$]	$9/0/1$
HALFCHEETAH	387.27 [-145.70, 548.73]	$7/0/3$
HUMANOIDOMNI	91.77 [19.61, 151.04]	$8/0/2$
HUMANOIDUNI	-76.43 [-172.75, 163.52]	$4/0/6$
KHEPERAX	-0.145 [-0.641, 0.312]	$4/0/6$
POINTMAZE	-16.98 [-64.31, -0.707]	$1/0/9$
WALKER	$1{,}771.6$ [$1{,}233.0$, $2{,}661.3$]	$9/0/1$

P.4.2 DISCUSSION

We conclude that the identities retained by VC carry useful information for subsequent fitness progress beyond its scalar admission quota in several tasks. The contrast also shows how selecting different policies changes the balance between peak fitness and repertoire breadth. Each arm computes its quota on its own evolving pool, so this result compares continuing identity transfer with continuing quota transfer, including their effects on later candidate pools and counts.

P.5 FITNESS-DIRECTION METRIC SENSITIVITY

We investigate whether emphasizing a latent direction that predicts fitness changes competition and improves subsequent search more than emphasizing a comparable random direction. It tests a possible link between fitness readability in the representation and the usefulness of its distances for selection.

At each VC checkpoint, we double distances along a fitted fitness-predictive direction and compare this with doubling along a random direction intended to have matched projected variance. The descriptive MSE control instead halves distances along the respective directions. The accepted cohort contains 100 VC pairs and 99 MSE pairs; the MSE branches for KHEPERAX seed 3 have no accepted completion receipts and are excluded. Pairs share the environment, seed, source checkpoint, objective, and amplification strength. We measure endpoints at update $5{,}000$ and AUC as the trapezoidal time average over updates $1{,}051$–$5{,}000$. Task effects are paired seed medians with 95% bootstrap intervals; aggregate intervals use $50{,}000$ resamples of the ten environment medians.

P.5.1 RESULTS

The variance check supports the intended comparison for VC: 96 of 100 pairs match within 10%. Only 16 of 99 MSE pairs meet that tolerance, so we interpret the VC comparison and retain the MSE outcomes as descriptive results. The assigned checkpoints, fitting samples, held-out samples, and direction parameters were verified for every analyzed pair.

Within VC, the fitness and random perturbations produce similar changes in competitive relationships. Their raw-to-treated fitter-neighbor Jaccards are 0.784 and 0.788, and the interval for the

paired graph-change contrast includes zero. Neither final nor trajectory-average maximum fitness shows a consistent advantage for the fitness direction (Table 16). The clearer response is in repertoire quality: final passive QD score decreases in eight task medians, with an aggregate interval below zero. QD-score AUC also decreases in eight tasks, although its aggregate interval crosses zero.

Table 16: Aggregate checkpoint metric-sensitivity effects, where each entry is the median of environment medians with its 95% interval and the environment directions for fitness-direction minus random-direction perturbation. VC doubles the chosen direction; the descriptive MSE control halves it and has nine pairs in KHEPERAX and ten elsewhere.

Source	Outcome	Effect [95% CI]	Directions $+/-/=$
MSE	Final passive maximum fitness	0.0023 $[-71.225,\ 18.122]$	$5/4/1$
MSE	Maximum-fitness AUC	-0.888 $[-94.917,\ 0.012]$	$2/7/1$
VC	Final passive maximum fitness	-0.0078 $[-69.655,\ 62.949]$	$3/5/2$
VC	Maximum-fitness AUC	-0.0046 $[-21.718,\ 29.911]$	$2/6/2$
VC	Final passive QD score	$-2{,}155$ $[-15{,}429,\ -44.5]$	$2/8/0$
VC	Passive QD-score AUC	-766 $[-7{,}878,\ 3{,}253]$	$2/8/0$

P.5.2 DISCUSSION

We conclude that amplifying a fitness-predictive direction provides no consistent fitness advantage over the variance-matched random direction in the valid VC comparison. Its clearer effect on passive QD score distinguishes fitness readability from the geometry needed to support a high-quality repertoire. These continuations measure neighborhood and outcome changes, but their logs do not identify downstream survivor retention or parent reuse as mediators.