WHEN DO LOW-DIMENSIONAL PROBES HELP MULTI-OBJECTIVE OPTIMIZATION?

Anonymous authors
Paper under double-blind review

ABSTRACT

Large-scale multiobjective optimization must explore hundreds or thousands of decision variables under limited evaluation budgets, making broad coverage of the decision space impractical. Low-dimensional probes address this challenge by concentrating evaluations on structured candidate families. Yet terminal performance alone cannot reveal why probing helps. Gains may come from representation, additional queries, or candidate retention, and conflating these effects can misattribute improvements and obscure both the underlying bottleneck and how probing should be improved. We study these effects through a controlled shared-scalar probe embedded in a fixed multiobjective search backbone. Our analysis identifies when a single scalar block is sufficient, separates finite-grid error from a representation-induced floor, and shows that identical nondominated decision sets need not imply identical population states. Under a fixed budget, additional probe queries can improve the evaluated candidate pool but reduce evaluations available for downstream search. We further separate losses from withholding evaluated candidates and finite-population selection, while matched-budget and matched-query comparisons distinguish query value from retention value. Across 64 benchmark configurations, additional-query value varies substantially with problem structure. Targeted controls further support retaining paid alternatives more consistently than the tested adaptive placement strategy.

1 INTRODUCTION

Large-scale multiobjective optimization concerns problems with multiple conflicting objectives and hundreds or thousands of decision variables (Zhang et al., 2018; Tian et al., 2022). In many engineering applications, objective evaluations can be costly, thereby limiting the available evaluation budget (Roussel et al., 2021; Duro et al., 2023; Daulton et al., 2022). Under such budgets, only sparse coverage of the high-dimensional decision space is possible, making broad exploration increasingly impractical (Eriksson & Jankowiak, 2021). This makes it difficult to approach the Pareto front while maintaining diverse trade-off solutions (Wang et al., 2024).

Prior work has explored problem transformation (Zille et al., 2018), low-dimensional reformulation (Qian & Yu, 2017; He et al., 2019), and directed sampling (Qin et al., 2021) to concentrate a limited evaluation budget within lower-dimensional candidate families, thereby enabling denser sampling within such families. More generally, studies of multiobjective archiving show that retaining evaluated solutions can support search efficiency (Li et al., 2024; Bian et al., 2024), while archive reuse can influence subsequent solution generation (Ren et al., 2026). Retaining useful candidates can alter the population passed to subsequent search without requiring additional objective evaluations.

However, low-dimensional probing involves several coupled design choices whose effects need to be distinguished, as shown in Figure 1. Denser probing may remain constrained by the representation of the candidate family, while selective retention can discard complementary solutions that have already been evaluated. Under a fixed evaluation budget, additional probe queries may improve coverage within the candidate family but leave fewer evaluations for subsequent search. These effects call for different design responses. Representation limitations require changing the candidate family, query-allocation trade-offs require weighing additional probe evaluations against their opportunity cost, and retention losses may be reduced by better using already-evaluated candidates without addi-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

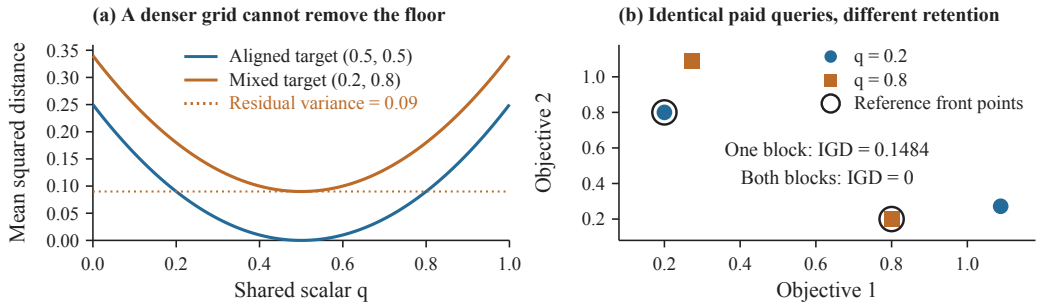

Figure 1: Two illustrative effects in low-dimensional probing. The left panel shows that the mixed target $(0.2, 0.8)$ retains residual variance under denser q sampling, whereas the aligned target $(0.5, 0.5)$ is matched exactly. The right panel shows that, with the same paid queries $q = 0.2$ and $q = 0.8$, one block misses a complementary reference-front point, while both cover both points.

tional objective evaluations. Without separating these effects, gains from low-dimensional probing can be misattributed, obscuring when probing is useful and how it should be improved.

We study a deliberately controlled shared-scalar probe embedded in a fixed multiobjective search backbone. The probe preserves selected coordinates from representative solutions, termed anchors, and assigns a common scalar to the remaining normalized coordinates, so that each queried scalar produces one candidate per anchor. This restricted construction makes representation limits analytically tractable, while matched-query comparisons hold the paid queries and probe cost fixed and vary which evaluated candidates are offered to population selection. We further distinguish paid, offered, and selected candidates to separate losses caused by withholding evaluated candidates from those introduced by finite-population selection. Dominance-safe block selection, equal-cost query-placement controls, and shared-normalization controls are then used to examine whether the observed effects can instead be explained by single-block choice, query placement, or injection normalization. Our contributions are as follows:

1. **Geometry of useful probing.** We characterize when denser probing can and cannot help within a restricted candidate family. A common-minimizer condition identifies when one shared-scalar block suffices, while a weighted-variance decomposition separates reducible finite-grid error from an irreducible representation-induced floor.
2. **Retention and population state.** We identify where value can be lost after candidates have already been evaluated. We decompose retention loss into offering and finite-population selection losses, provide a dominance-safe single-block rule, and show that identical nondominated sets can still yield different subsequent population states.
3. **Controlled empirical evidence.** Matched-budget and matched-query comparisons distinguish query allocation and candidate retention within a search backbone. Additional-query value varies across benchmark structures, while targeted controls support retaining already-paid alternatives more consistently than the tested adaptive placement rule.

2 Probe Formulation

2.1 Problem Setting and Shared-Scalar Probe

We minimize $f : \mathcal{X} \to \mathbb{R}^M$ on a D-dimensional box $\mathcal{X} = \prod_{j=1}^{D}[\ell_j, u_j]$, with $u_j > \ell_j$. One function evaluation (FE) returns all objectives for one decision. An evaluated initial population I contains N objects, and the total budget is $F_{\max}$. Reference fronts are used for evaluation only.

Normalize decisions by $y_j = (x_j - \ell_j)/(u_j - \ell_j)$. Let J be the coordinates preserved from A anchors $y_1, \ldots, y_A$ selected from the initialization. For anchor a and scalar $q \in [0, 1]$, define

$$y_j^{a,q} = y_{a,j} \ \ (j \in J), \qquad y_j^{a,q} = q \ \ (j \notin J), \qquad q \in [0, 1]. \tag{1}$$

The corresponding physical decision is $x_j^{a,q} = \ell_j + (u_j - \ell_j)y_j^{a,q}$. A *scalar block* $B(q) = (x^{1,q}, \ldots, x^{A,q})$ contains one candidate per anchor, in a fixed order that defines anchorwise block comparisons. The analysis conditions on the chosen J and anchors; it does not assume that J recovers a benchmark's true position coordinates. More scalar values sample the same restricted family more densely without enlarging it.

2.2 Offering, Selection, and Retention

Let $\mathcal{Q} = (q_1, \ldots, q_K)$ be the ordered query sequence, with $q_k \in [0, 1]$. Each query evaluates all A anchors. For distance calculations, let $U = \bigcup_{k=1}^{K} B(q_k)$ be the underlying set of paid probe candidates and $S = I \cup U$ the uncapped paid pool. A return policy offers $O \subseteq U$ to the environmental selector, which selects N objects from $I \cup O$ to form population P_O. At the level of underlying decision sets,

$$O \subseteq U, \qquad P_O \subseteq I \cup O \subseteq I \cup U = S.$$

Offering maps paid probe candidates to O; *selection* maps $I \cup O$ to P_O; *retention* covers both transitions. Single-block and all-block policies offer $O = B(q_k)$ and $O = U$, respectively, with initialization available in both cases. Sets support distance analysis; implementation retains query order and object multiplicities.

3 Geometry, Cost, and Retention Analysis

We analyze representation geometry, query value under a fixed budget, candidate retention, and what front quality reveals about population state.

3.1 Representation Geometry and Common Optima

For anchor a and objective m, consider a separable, nonconstant quadratic distance component on the scanned coordinates $j \notin J$,

$$G_{am}(q) = C_{am} + \sum_{j \notin J} w_{amj}(q - c_{amj})^2 = C_{am} + W_{am}\big[(q - \mu_{am})^2 + \sigma_{am}^2\big], \tag{2}$$

where subsequent sums are over $j \notin J$, $w_{amj} \geq 0$, $W_{am} = \sum_j w_{amj} > 0$, and $\mu_{am} = W_{am}^{-1} \sum_j w_{amj} c_{amj}$. The weighted variance is $\sigma_{am}^2 = W_{am}^{-1} \sum_j w_{amj}(c_{amj} - \mu_{am})^2$. Targets lie in $[0, 1]$; fixed extra distance coordinates contribute to C_{am}. Expanding the square proves the identity. For a finite, nonempty grid $Q \subseteq [0, 1]$,

$$\min_{q \in Q} G_{am}(q) = C_{am} + W_{am}\sigma_{am}^2 + W_{am} \min_{q \in Q}(q - \mu_{am})^2. \tag{3}$$

The term $W_{am}\sigma_{am}^2$ is the *representation-induced floor* on the scanned coordinates, while the last term is the *finite-grid error*. The complete component has continuous minimum $C_{am} + W_{am}\sigma_{am}^2$. Different coordinate targets cannot all be matched by one scalar, so the variance term remains even with continuous scanning. In Figure 1(a), equal weights normalized to sum to one and targets $(0.2, 0.8)$ give $\mu = 0.5$ and $\sigma^2 = 0.09$. Increasing grid density cannot remove this floor.

A common-minimizer principle extends beyond quadratics: if the relevant distance components across all anchors share a minimizer $q_0 \in [0, 1]$, and objectives depend on q only through nondecreasing functions of these components with other factors fixed, the q_0 block weakly dominates the restricted family anchor by anchor.

Proposition 1 (A sufficient condition for one block)**.** *Suppose the preserved coordinates fix each anchor's objective-shape factors, and its objectives are nondecreasing functions of the corresponding distance components. If every nonconstant component in Equation* (2) *has the same mean $\mu_{am} = q_0$, then the q_0 block weakly dominates every other block anchor by anchor. Within this representation, no extra scalar can improve an objective vector beyond that block. The quadratic components have a common continuous minimizer if and only if their means agree.*

Each nonconstant quadratic has its unique minimum at its mean, proving the equivalence; objective monotonicity gives block dominance. To obtain the continuous minimizing block in a paid family,

q_0 must be queried. On a finite grid, an evaluated value that is nearest to every μ_{am} also weakly dominates all other grid blocks, even if the continuous means differ. These statements concern candidate availability. An implemented score can still fail to return a dominating paid block through ties or by excluding it from eligibility.

DTLZ2 has a common optimum $q_0 = 0.5$ when all true position coordinates are preserved. Appendix B derives the other DTLZ, WFG and UF cases, including nonquadratic common optima and representation mismatch. Their premises must be checked against the actual routed coordinates.

3.2 Query Value Under a Fixed Budget

Fix the initialization I, the preserved coordinates J and the A anchors. For the ordered query sequence $\mathcal{Q} = (q_1, \ldots, q_K)$ from Section 2, let

$$S(\mathcal{Q}) = I \cup \bigcup_{k=1}^{K} B(q_k) \tag{4}$$

be the uncapped paid candidate pool. For a finite, nonempty objective-space reference set R and a nonempty candidate set P, use the established IGD$^+$ distance (Ishibuchi et al., 2015),

$$I^+(R, P) = \frac{1}{|R|} \sum_{r \in R} \min_{p \in P} \|[f(p) - r]_+\|_2, \tag{5}$$

where $[\cdot]_+$ takes the positive part coordinatewise. All comparisons use the same reference set and objective scaling. IGD$^+$ cannot increase under set inclusion and is unchanged by dominated-point pruning, connecting the metric to block dominance.

Suppose $\mathcal{Q}_2$ extends $\mathcal{Q}_1$ by additional query records, retaining every original query with the same anchors and coordinates. Then

$$S(\mathcal{Q}_1) \subseteq S(\mathcal{Q}_2), \qquad I^+(R, S(\mathcal{Q}_2)) \leq I^+(R, S(\mathcal{Q}_1)).$$

Define the *additional-query discovery gain* as

$$\Delta_{\text{disc}}(\mathcal{Q}_2, \mathcal{Q}_1) = I^+(R, S(\mathcal{Q}_1)) - I^+(R, S(\mathcal{Q}_2)) \geq 0. \tag{6}$$

Set inclusion proves the inequality under the stated nesting conditions; the monotonicity applies to the uncapped paid pool, not necessarily to a capped retrospective archive.

For an applied probe, every query record evaluates all A anchors. Its FE cost and the budget available to subsequent search are therefore

$$C(\mathcal{Q}) = AK, \qquad F_{\text{evo}}(\mathcal{Q}) = F_{\max} - N - AK, \tag{7}$$

with $N + AK \leq F_{\max}$. Here K counts paid query records, including repeated scalars and decisions; repetitions can increase cost without enlarging $S(\mathcal{Q})$.

The geometry sharpens this trade-off. If $\mathcal{Q}_1$ already contains a block that weakly dominates the entire probe family, adding blocks from that family gives $\Delta_{\text{disc}} = 0$. Otherwise, additional queries can improve the paid pool, but their discovery gain is obtained at the expense of later search. Positive discovery gain does not imply better terminal performance: the selected population, remaining FE and FE-dependent schedule all affect subsequent proposals. Zero discovery gain also leaves the selected population undetermined, as the retention analysis below shows. The Midpoint/Scan comparison in Section 4 tests the net value of additional scan queries after returning the cheaper policy's saved FE to evolution; its terminal contrast includes the resulting changes in returned candidates and search schedule.

3.3 Retention from Paid Candidates to Selected Population

With queries fixed, Figure 1(b) shows two paid blocks with complementary front points: either block has positive IGD and IGD$^+$, whereas their union has zero against the same two-point reference. Offering can therefore lose quality before finite-population selection.

Before attributing such a difference to the value of multiple blocks, the single-block rule should respect block dominance. Write $B \prec B'$ when the objective matrix of feasible block B is no larger at every anchor/objective entry than that of B', with at least one strict inequality. This is the strict Pareto order on the concatenated block objectives.

Proposition 2 (Dominance-safe block offering)**.** *For a finite, nonempty family of feasible paid blocks, pruning every block dominated by another leaves a nonempty set, and minimizing any score on that set cannot return a dominated block. If an evaluated block weakly dominates every block anchor by anchor, every surviving block has the same objective matrix as the dominating block; offering any survivor loses no IGD$^+$ quality relative to the union of paid blocks.*

A finite strict partial order has a minimal element. If one block weakly dominates the family, all differing objective matrices are removed; survivors have identical matrices and hence equal reference-point distances. The guarantee concerns offering from the evaluated family and requires no reference front for block choice. Subsequent environmental selection may still lose quality; the rule does not optimize a set-level metric.

Write U for all paid probe candidates and $S = I \cup U$ for the complete paid pool. For an offered subset $O \subseteq U$, the transition is

$$\underbrace{S = I \cup U}_{\text{paid pool}} \longrightarrow \underbrace{I \cup O}_{\text{offered pool}} \longrightarrow \underbrace{P_O}_{\text{selected population}} .$$

The initialization remains available throughout. At the level of underlying decision sets, $P_O \subseteq I \cup O \subseteq S$. With the same reference and metric scaling as in Section 3.2, define the immediate retention gap by

$$\Delta_{\text{ret}}(O) = I^+(R, P_O) - I^+(R, S) \geq 0. \tag{8}$$

The intermediate offered pool separates two nonnegative losses:

$$\begin{aligned} \Delta_{\text{offer}}(O) &= I^+(I \cup O) - I^+(S) \geq 0, \\ \Delta_{\text{sel}}(O) &= I^+(P_O) - I^+(I \cup O) \geq 0, \end{aligned} \tag{9}$$

where R is omitted. Set inclusion gives nonnegativity, and cancellation of the intermediate pool yields

$$\Delta_{\text{ret}}(O) = \Delta_{\text{offer}}(O) + \Delta_{\text{sel}}(O). \tag{10}$$

For a single offered block B, offering loss is precisely discarded-block loss:

$$\Delta_{\text{offer}}(B) = \Delta_{\text{block}}(B) = I^+(I \cup B) - I^+(S). \tag{11}$$

Thus $\Delta_{\text{ret}}(B) = \Delta_{\text{block}}(B) + \Delta_{\text{sel}}(B)$ is a special case of Equation (10). Offering all paid blocks takes $O = U$ and eliminates offering loss, while the finite selector may still lose quality. Writing P_U for that policy's population gives

$$I^+(P_B) - I^+(P_U) = \Delta_{\text{block}}(B) + \Delta_{\text{sel}}(B) - \Delta_{\text{sel}}(U). \tag{12}$$

Removing discarded-block loss therefore does not guarantee a better selected population under this metric. A matched-query comparison holds the paid pool and pre-injection state fixed, and tests how the offering policy and its associated selection context affect the resulting population. The metric comparison uses fixed scaling even when the selector's normalization changes with its offered pool.

3.4 Zero Front Loss Does Not Identify Population State

The *population state* comprises the members, multiplicities and ordering passed to subsequent search. Even zero retention gap leaves the selected decisions undetermined. Consider four feasible scalar decisions $a = 0$, $b = 0.25$, $c = 0.5$ and $d = 0.75$ in $[0, 1]$, with objective vectors

$$f(a) = (0, 2), \quad f(b) = (1, 1), \quad f(c) = (2, 0), \quad f(d) = (100, 2),$$

in that input order. Use population size two, one reference vector $(1, 1)$ and injection penalty $\theta = 0$. In this single niche, both slots are filled by lower Pareto rank, then smaller normalized radius $\rho(x) = \|(f(x) - z)/s\|_2$, then input order for residual ties. Here z and s are the pool ideal and range, with componentwise division.

From $\{a, b, c\}$, the ideal is $(0, 0)$ and the range is $(2, 2)$. All three points are nondominated; b has the smallest radius and the tie between a and c favors a, giving $\{a, b\}$. Adding the dominated point d preserves the ideal but changes the range to $(100, 2)$ and the radius order to c, b, a. Selection now gives $\{b, c\}$, although d itself is discarded. A discarded candidate can therefore change normalization and the selected population.

For the finite objective-space reference $R = \{f(b)\} = \{(1, 1)\}$, both selected populations and the full paid pool have $\text{IGD}^+ = 0$. Their retention gaps are equal to zero despite different selected decisions. Here equal metric values coexist with different selected fronts; the stated finite reference is an explicit witness to the non-identification of population state by the gap. Appendix B.2 gives the exact radii and selection outcomes.

An identical nondominated decision set is also insufficient to identify a full population: dominated members, multiplicities and ordering may differ. Those differences can enter subsequent variation even when immediate front quality is unchanged. Other optimizer state also matters for trajectories.

4 Controlled Experiments

We design a set of controlled experiments to examine query value, candidate retention, query placement and population state within a fixed search backbone.

4.1 Experimental Setup

The original experimental panels use $D = 1000$, $N = 100$ and exactly $F_{\max} = 1000$ FE. Policies share the evaluated initialization, coordinate screen, anchors, downstream search, comparator and population selector, with twenty paired runs per configuration. The backbone incorporates competitive-update, density, and reference-vector components (Cheng & Jin, 2015; Li et al., 2014; Cheng et al., 2016). Probes use observed decisions and objectives without access to benchmark identities or known position coordinates. Saved probe FE return to evolution, with selection schedules indexed by elapsed FE. Appendix C provides implementation details, parameter settings, and function-evaluation accounting.

No-scan makes no probe evaluations. Midpoint evaluates and offers the $q = 0.5$ block, costing A FE. Scan evaluates five coarse queries followed by five fine queries around the best coarse scalar, using the prescribed R^2-style score (Brockhoff et al., 2012). It pays $10A$ FE for ten query records, including repetitions, and offers its selected fine block. All-scan makes identical paid queries and computes identical scores, but offers all paid coarse/fine candidates at the same cost.

The main experiment covers 64 configurations: original LSMOP (18) (Cheng et al., 2017), fixed transformations of LSMOP (18), GLSMOP (18) (Gu et al., 2024), and DTLZ/WFG/UF (10) (Deb et al., 2005; Huband et al., 2006; Zhang et al., 2009). These groups cover the structural cases in Section 3, including common scalar optima, endpoint optima and anchor-dependent targets with representation mismatch. Appendix D provides the complete task configurations, transformations, output reconstruction, and broad-study results.

Both co-primary outputs use IGD and the same finite reference front, following established solution-set assessment conventions (Zitzler et al., 2003; Li & Yao, 2020). IGD^+ supports the block-dominance guarantees; we also report metric-sensitivity analyses. Archive ND deduplicates the 1000-query history, caps it at N candidates, then scores its feasible nondominated subset; it is retrospective and never enters the optimizer. Native ND scores the final population's feasible nondominated subset. Archive and Native abbreviate these two outputs in the following figures and tables. Transformed LSMOP uses reference-front ranges for objective normalization; other groups use native units. Mean-IGD ratios are computed within configurations, without pooling raw distances. Two-sided paired conditional exact Wilcoxon signed-rank tests (Wilcoxon, 1945) use Holm correction (Holm, 1979); in the main experiment, each group's family jointly covers all six pairwise contrasts and both outputs. Later controls use their stated families. Throughout the following figures and tables, W/T/L reports reference-policy win/unresolved/loss counts, determined by adjusted $p < 0.05$ and signed-rank direction; T does not establish equivalence. Intervals are pointwise 95% paired bootstrap intervals (Efron, 1979).

4.2 Query Value and Retention

Three contrasts distinguish the tested query and offering policies. Scan versus Midpoint assesses additional queries under single-block offering; All-scan versus Midpoint assesses them when all paid blocks are offered. Both use the same total FE budget, return Midpoint's saved FE to evolution, and include changes in offered candidates, remaining search FE and the FE-indexed schedule; neither

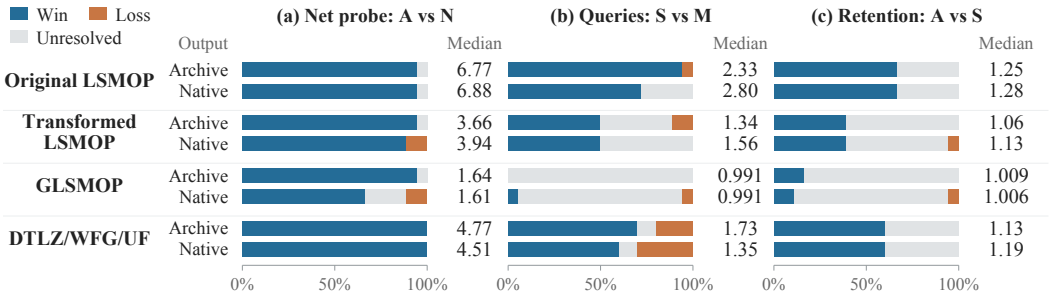

Figure 2: Bars show W/T/L shares under the Holm families in Table 1. Values are median comparator/reference mean-IGD ratios; > 1 favors the first policy.

Table 1: 5,120 runs over 64 configurations. Cells show W/T/L (median comparator/reference mean-IGD ratio), for Archive ND; Native ND. Ratios > 1 favor the reference. Holm correction covers all six contrasts and both outputs within each group.

Reference	Comparator	Original LSMOP (18)	Transformed LSMOP (18)	GLSMOP (18)	DTLZ/WFG/UF (10)
All-scan	No-scan	17/1/0 (6.77); 17/1/0 (6.88)	17/1/0 (3.66); 16/0/2 (3.94)	17/1/0 (1.64); 12/4/2 (1.61)	10/0/0 (4.77); 10/0/0 (4.51)
Scan	Midpoint	17/0/1 (2.33); 13/5/0 (2.80)	9/7/2 (1.34); 9/9/0 (1.56)	0/18/0 (0.991); 1/16/1 (0.991)	7/1/2 (1.73); 6/1/3 (1.35)
All-scan	Scan	12/6/0 (1.25); 12/6/0 (1.28)	7/11/0 (1.06); 7/10/1 (1.13)	3/15/0 (1.01); 2/15/1 (1.01)	6/4/0 (1.13); 6/4/0 (1.19)
Scan	No-scan	17/0/1 (5.49); 16/1/1 (4.86)	17/0/1 (3.17); 15/1/2 (3.27)	18/0/0 (1.62); 12/6/0 (1.53)	10/0/0 (3.66); 10/0/0 (2.78)
Midpoint	No-scan	17/1/0 (1.84); 12/5/1 (1.56)	17/1/0 (1.61); 15/1/2 (1.62)	16/2/0 (1.66); 12/5/1 (1.71)	6/4/0 (1.31); 6/4/0 (1.28)
All-scan	Midpoint	17/0/1 (3.13); 17/1/0 (3.89)	10/8/0 (1.62); 10/7/1 (2.32)	0/18/0 (1.00); 0/18/0 (0.998)	9/1/0 (1.77); 9/1/0 (1.73)

isolates discovery gain alone. All-scan versus Scan holds paid queries, probe FE and pre-injection state fixed while changing offering and its selection context. No-scan comparisons assess the net benefit of the probe package. Figure 2 and Table 1 show the results. In Figure 2 and the following figures and tables, N, M, S and A denote No-scan, Midpoint, Scan and All-scan, respectively.

On original LSMOP, Scan versus Midpoint gives 17/0/1 on Archive ND and 13/5/0 on Native ND, with median Midpoint/Scan mean-IGD ratios of 2.33 and 2.80 across configurations. All-scan versus Midpoint instead gives 17/0/1 and 17/1/0, with median Midpoint/All-scan ratios of 3.13 and 3.89. Additional queries thus repay their cost on most original-LSMOP cases under either offering rule, with broader native support when all paid alternatives are offered. Transformed LSMOP and DTLZ/WFG/UF show less uniform outcomes, including losses. On GLSMOP, Scan versus Midpoint is largely unresolved despite frequent gains over No-scan from all three probing policies. Query value thus depends on the construction: additional queries can justify their cost in some structural regimes, while benefit from probing alone does not establish a benefit from denser adaptive querying.

With identical paid queries, All-scan versus Scan gives 12/6/0 on both original-LSMOP outputs. Median Scan/All-scan mean-IGD ratios across these configurations are 1.25 for Archive ND and 1.28 for Native ND. The descriptive lost-quality fraction $\overline{\Delta_{\mathrm{ret}}(O)}/\overline{I^+(R, P_O)}$, computed from run averages within each configuration, has medians of 0.271 for Scan and zero for All-scan across original LSMOP. It measures distance excess, not the fraction of discarded points. Effects are less uniform on transformed LSMOP, including a native loss, and largely unresolved on GLSMOP; six of ten DTLZ/WFG/UF configurations favor All-scan on both outputs. The query histories underlying both Archive ND readouts share the same paid scan prefix, so terminal archive differences arise from post-injection queries. Retaining paid alternatives is therefore supported across several constructions, with case-dependent effects.

DTLZ2 provides a boundary case linking these results to Section 3. With true position coordinates preserved in the checked runs, its midpoint is a common optimum and additional-query IGD^+ discovery gain is zero. For $M = 3$, All-scan nevertheless outperforms Scan at termination although both immediate retention gaps are zero. Terminal benefit therefore need not reflect better immedi-

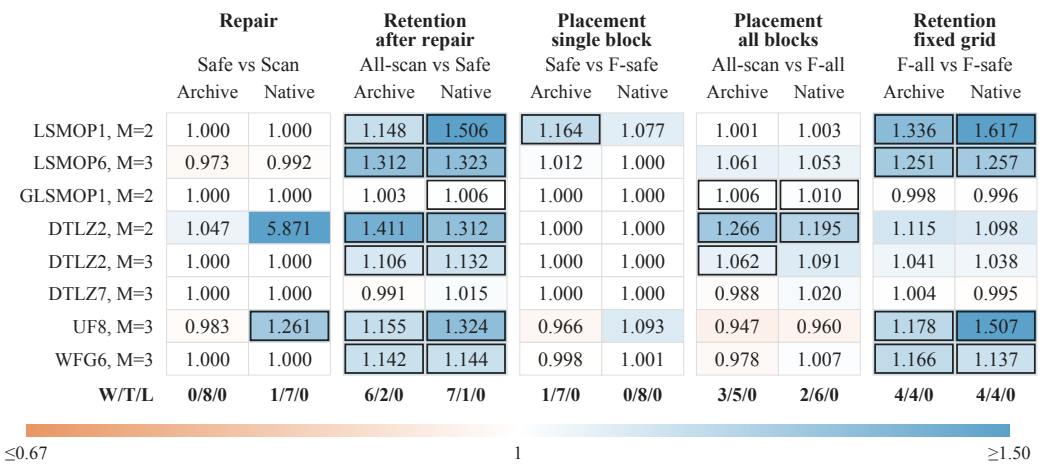

Figure 3: Comparator/reference mean IGD; ratios > 1 favor the first policy. Dark outlines mark joint 160-test Holm significance; unoutlined cells are unresolved. Full results including pointwise 95% intervals are provided in Appendix E.

ate discovery or front quality. Section 4.3 tests retention after single-block repair, compares query placement at equal cost, and examines injection normalization and full-population differences.

4.3 Targeted Controls and Population-State Analysis

Two targeted controls test retention after single-block repair and equal-cost placement changes, then examine injection normalization and population state on DTLZ2. Appendix E provides the complete protocols and results, including separate population- and optimizer-state interventions.

4.3.1 Single-Block Repair and Equal-Cost Placement

We test whether improving the single-block rule removes the retention advantage and whether adaptive query placement adds value at equal cost. Five policies are compared on eight targeted configurations with twenty paired runs distinct from the broad study. Safe extends eligibility from fine blocks to all ten paid blocks, removes anchorwise-dominated choices, and applies the unchanged clipped score to the survivors, offering one block without additional FE. Fixed-safe and Fixed-all replace adaptive fine placement with a fixed grid while matching the respective single- and all-block retention rules. All five policies spend $10A$ probe FE and share the remaining protocol. The ten pairwise contrasts and both outputs form one 160-test Holm family. Fixed-safe and Fixed-all are abbreviated as F-safe and F-all in the following figures and tables.

Figure 3 shows retention gains. All-scan/Safe gives 6/2/0 on Archive ND and 7/1/0 on Native ND (median Safe/All-scan mean-IGD ratios 1.15 and 1.23). Fixed-all/Fixed-safe gives 4/4/0 on both outputs. Safe/Scan has one Native ND win on UF8; the record audit finds identical offered, post-injection and final decisions in 119/160 pairs. All-scan/Fixed-all gives 3/5/0 and 2/6/0; single-block placement contrasts are mostly unresolved. Across these eight cases, retaining paid alternatives has more consistent support than adaptive placement; unresolved contrasts do not establish equivalence.

4.3.2 Normalization and Population State

To examine why a terminal retention advantage can coexist with zero immediate front loss, we test whether pool-specific injection scaling explains the effect and inspect the resulting population state. On DTLZ2 ($M = 3$), Safe and All-scan use a shared ideal and range from the common initial-plus-paid pool at injection only; later normalization remains population dependent, and other selection and search rules are unchanged. We compare immediate nondominated decision sets and complete populations using the same twenty paired runs, reusing default records from Section 4.3.1 without adding independent runs.

Table 2: Injection normalization on DTLZ2 ($M = 3$). Terminal Safe/All-scan mean-IGD ratios above one favor All-scan.

Norm.	Terminal effect [95% CI]		Population state at injection				Intervention: shared/pool	
	Archive ND	Native ND	Zero gap (populations)	Same ND (20 pairs)	Diff. pop. (median)	Dominated g (Safe; All)	Safe selection changed	All trajectory identical
Pool-based	1.11 [1.06,1.16]*	1.13 [1.07,1.19]*	40/40	20/20	44/100	82.80; 42.22	–	–
Shared	1.12 [1.08,1.16]*	1.16 [1.11,1.20]*	40/40	20/20	44/100	82.78; 42.22	16/20	20/20
Shared/pool interaction (All-scan/Safe)			Archive ND			Native ND		Inference
Exp. mean paired log-effect change			0.987 [0.949,1.022]			0.979 [0.945,1.011]		Both unresolved

Same ND: identical nondominated decision sets; Diff. pop.: differing members; g: dominated-member means. State entries are descriptive; $g = \sum_{j=3}^{1000}(x_j - 0.5)^2$ (lower is better). All-scan is the negative control. CIs are pointwise 95%; *: 14-test Holm significance.

Table 2 shows that All-scan remains favored on both outputs under shared normalization, so pool-specific injection scaling is insufficient to explain the terminal benefit. All eighty immediate populations have zero retention gap. Under each normalization, all twenty Safe/All-scan pairs have identical nondominated decision sets but differ by a median 44 of 100 population members. The source matching identifies 84 initial and 16 probe members in every Safe population, versus means of 40.55 and 59.45 for All-scan. Shared scaling changes Safe's selection in 16/20 runs; All-scan is an unchanged negative control. Both log-effect interactions measure Safe's normalization sensitivity and remain unresolved. These descriptive checks do not isolate causal contributions of composition, quality or order. Separate component controls and selected-case external comparisons are reported in Appendix F.

5 Conclusion and Limitations

5.1 Conclusion

This study separates representation, query allocation, and candidate retention in low-dimensional multiobjective probing. The analysis distinguishes reducible finite-grid error from a representation-induced floor and decomposes retention loss into offering and finite-population selection losses. Controlled experiments show that the value of additional probe queries depends on problem structure. After dominance-safe single-block repair, retaining already-paid alternatives is favored on most targeted configurations and receives more consistent support than the tested adaptive placement rule. The DTLZ2 controls further show that zero retention gap and identical nondominated decision sets do not identify the full population passed to subsequent search. These findings support checking inexpensive probe choices, accounting for query costs, and examining the loss of evaluated candidates when designing low-dimensional probes.

5.2 Limitations

The broad evidence uses one backbone, $D = 1000$, $N = 100$ and a 1000-FE budget; the larger-budget follow-up covers only five selected cases. The benchmarks examine selected structural conditions rather than a representative distribution of applications, and paired runs quantify within-case search randomness only. The geometric claims require their stated routed-coordinate premises; reference fronts and known position sets support diagnosis rather than guide the optimizer. Shared normalization tests one injection boundary, while the population-state diagnostics remain descriptive and do not isolate the causal contributions of dominated-member quality, composition, or ordering. Equal function-evaluation budgets also do not imply equal wall-clock cost. Further studies across search backbones, evaluation budgets, and application settings are needed to assess how broadly these findings generalize. The supplementary experiments in Appendices D–F examine larger budgets, nonrepeated fixed grids, population- and optimizer-state interventions, frozen routing, and external-method comparisons. These checks extend the evidence within selected settings, but do not establish broad generality across search backbones, problem dimensions, or applications.

AI USE STATEMENT

In this work, we used generative AI tools to assist with translation and data analysis. We have not used generative AI tools to develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, design research methodology, implement methods, and interpret results. And generating synthetic datasets, cleaning and reformatting datasets are not applicable to this work. Additionally, we used generative AI tools to edit a research paper to improve readability, identify relevant literature, and format references. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Appendix B presents the assumptions and derivations underlying the analytical results. Appendix C specifies the shared search procedure, implementation details, parameter settings, and function-evaluation accounting. Appendix D documents the benchmark configurations, transformations, output reconstruction, evaluation metrics, and paired-run design. Appendices E and F provide the protocols for the targeted controls, component studies, and external comparisons, explicitly distinguishing their interventions, output definitions, and budget boundaries.

REFERENCES

Chao Bian, Shengjie Ren, Miqing Li, and Chao Qian. An archive can bring provable speed-ups in multi-objective evolutionary algorithms. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 6905–6913. International Joint Conferences on Artificial Intelligence Organization, 2024.

Dimo Brockhoff, Tobias Wagner, and Heike Trautmann. On the properties of the R2 indicator. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*, pp. 465–472. Association for Computing Machinery, 2012.

Ran Cheng and Yaochu Jin. A competitive swarm optimizer for large scale optimization. *IEEE Transactions on Cybernetics*, 45(2):191–204, 2015.

Ran Cheng, Yaochu Jin, Markus Olhofer, and Bernhard Sendhoff. A reference vector guided evolutionary algorithm for many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 20(5):773–791, 2016.

Ran Cheng, Yaochu Jin, Markus Olhofer, and Bernhard Sendhoff. Test problems for large-scale multiobjective and many-objective optimization. *IEEE Transactions on Cybernetics*, 47(12):4108–4121, 2017.

Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Multi-objective Bayesian optimization over high-dimensional search spaces. In James Cussens and Kun Zhang (eds.), *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pp. 507–517. PMLR, 2022.

Kalyanmoy Deb and Himanshu Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601, 2014.

Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

Kalyanmoy Deb, Lothar Thiele, Marco Laumanns, and Eckart Zitzler. Scalable test problems for evolutionary multiobjective optimization. In Ajith Abraham, Lakhmi Jain, and Robert Goldberg (eds.), *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*, Advanced Information and Knowledge Processing, chapter 6, pp. 105–145. Springer, London, 2005.

João A. Duro, Umud Esat Ozturk, Daniel C. Oara, Shaul Salomon, Robert J. Lygoe, Richard Burke, and Robin C. Purshouse. Methods for constrained optimization of expensive mixed-integer multi-objective problems, with application to an internal combustion engine design problem. *European Journal of Operational Research*, 307(1):421–446, 2023. doi: 10.1016/j.ejor.2022.08.032.

Bradley Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1): 1–26, 1979.

David Eriksson and Martin Jankowiak. High-dimensional Bayesian optimization with sparse axis-aligned subspaces. In Cassio de Campos and Marloes H. Maathuis (eds.), *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pp. 493–503. PMLR, 2021.

David Eriksson, Michael Pearce, Jacob R. Gardner, Ryan D. Turner, and Matthias Poloczek. Scalable global optimization via local Bayesian optimization. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32, pp. 5496–5507. Curran Associates, Inc., 2019.

Haoran Gu, Handing Wang, and Jingjing Ma. Test suites and performance of algorithms in large-scale multiobjective evolutionary optimization. In *2024 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8. IEEE, 2024.

Cheng He, Lianghao Li, Ye Tian, Xingyi Zhang, Ran Cheng, Yaochu Jin, and Xin Yao. Accelerating large-scale multiobjective optimization via problem reformulation. *IEEE Transactions on Evolutionary Computation*, 23(6):949–961, 2019.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

Simon Huband, Philip Hingston, Luigi Barone, and Lyndon While. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation*, 10(5):477–506, 2006.

Hisao Ishibuchi, Hiroyuki Masuda, Yuki Tanigaki, and Yusuke Nojima. Modified distance calculation in generational distance and inverted generational distance. In António Gaspar-Cunha, Carlos Henggeler Antunes, and Carlos Coello Coello (eds.), *Evolutionary Multi-Criterion Optimization, Part II*, volume 9019 of *Lecture Notes in Computer Science*, pp. 110–125, Cham, 2015. Springer.

Miqing Li and Xin Yao. Quality evaluation of solution sets in multiobjective optimisation: A survey. *ACM Computing Surveys*, 52(2):26:1–26:38, 2020.

Miqing Li, Shengxiang Yang, and Xiaohui Liu. Shift-based density estimation for Pareto-based algorithms in many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 18 (3):348–365, 2014.

Miqing Li, Manuel López-Ibáñez, and Xin Yao. Multi-objective archiving. *IEEE Transactions on Evolutionary Computation*, 28(3):696–717, 2024.

Hong Qian and Yang Yu. Solving high-dimensional multi-objective optimization problems with low effective dimensions. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 875–881. AAAI Press, 2017.

Shufen Qin, Chaoli Sun, Yaochu Jin, Ying Tan, and Jonathan Fieldsend. Large-scale evolutionary multiobjective optimization assisted by directed sampling. *IEEE Transactions on Evolutionary Computation*, 25(4):724–738, 2021.

Shengjie Ren, Zimin Liang, Miqing Li, and Chao Qian. Not just for archiving: Provable benefits of reusing the archive in evolutionary multi-objective optimization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(43):37152–37160, 2026.

Ryan Roussel, Adi Hanuka, and Auralee Edelen. Multiobjective Bayesian optimization for online accelerator tuning. *Physical Review Accelerators and Beams*, 24(6):062801, 2021. doi: 10.1103/PhysRevAccelBeams.24.062801.

Ye Tian, Ran Cheng, Xingyi Zhang, and Yaochu Jin. PlatEMO: A MATLAB platform for evolutionary multi-objective optimization. *IEEE Computational Intelligence Magazine*, 12(4):73–87, 2017.

Ye Tian, Langchun Si, Xingyi Zhang, Ran Cheng, Cheng He, Kay Chen Tan, and Yaochu Jin. Evolutionary large-scale multi-objective optimization: A survey. *ACM Computing Surveys*, 54 (8):174:1–174:34, 2022.

Haofan Wang, Li Chen, Xingxing Hao, Rong Qu, Wei Zhou, Dekui Wang, and Wei Liu. Learning-guided cross-sampling for large-scale evolutionary multi-objective optimization. *Swarm and Evolutionary Computation*, 91:101763, 2024.

Ziyu Wang, Masrour Zoghi, Frank Hutter, David Matheson, and Nando de Freitas. Bayesian optimization in high dimensions via random embeddings. In Francesca Rossi (ed.), *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pp. 1778–1784. AAAI Press, 2013.

Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

Qingfu Zhang and Hui Li. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, 2007.

Qingfu Zhang, Aimin Zhou, Shizheng Zhao, Ponnuthurai Nagaratnam Suganthan, Wudong Liu, and Santosh Tiwari. Multiobjective optimization test instances for the CEC 2009 special session and competition. Technical Report CES-487, School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK, 2009.

Xingyi Zhang, Ye Tian, Ran Cheng, and Yaochu Jin. A decision variable clustering-based evolutionary algorithm for large-scale many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 22(1):97–112, 2018.

Heiner Zille, Hisao Ishibuchi, Sanaz Mostaghim, and Yusuke Nojima. A framework for large-scale multiobjective optimization based on problem transformation. *IEEE Transactions on Evolutionary Computation*, 22(2):260–275, 2018.

Eckart Zitzler, Marco Laumanns, and Lothar Thiele. SPEA2: Improving the strength Pareto evolutionary algorithm. TIK-Report 103, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Zurich, Switzerland, 2001.

Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M. Fonseca, and Viviane Grunert da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.

APPENDIX: WHEN DO LOW-DIMENSIONAL PROBES HELP MULTIOBJECTIVE OPTIMIZATION?

A Related Works
B Analytical Details and Benchmark-Specific Checks
C Shared Search Backbone and Implementation
D Broad Study and Budget Sensitivity
E Targeted Controls and Population-State Analysis
F Component Controls and External Comparisons

A RELATED WORKS

Structured search methods include WOF/LSMOF reformulations and LMOEA-DS/learning-guided directed sampling (Zille et al., 2018; He et al., 2019). These end-to-end methods motivate our controlled study of candidate families (Qin et al., 2021; Wang et al., 2024). An end-to-end comparison combines the effects of the search representation, query allocation and subsequent population dynamics, so it does not identify which part makes a probe useful. Restricting the candidate family can concentrate a limited budget, but denser sampling within that family cannot recover decisions that the representation excludes. High-dimensional Bayesian optimization instead uses random embeddings, local trust regions or sparse surrogate priors (Wang et al., 2013; Eriksson et al., 2019; Daulton et al., 2022; Eriksson & Jankowiak, 2021); our fixed shared-scalar probe uses no surrogate model.

SPEA2 archive truncation and NSGA-III/RVEA reference-based selection compress evaluated candidates into finite populations (Zitzler et al., 2001; Deb & Jain, 2014; Cheng et al., 2016). A probe may evaluate several blocks but pass only a subset to environmental selection. A candidate can thus be paid for without ever entering the selector's input, which separates query coverage from the value of retaining evaluated points. Our shared-backbone controls distinguish this interface effect from the geometric limitations of the probe family.

B ANALYTICAL DETAILS AND BENCHMARK-SPECIFIC CHECKS

Unless stated otherwise, the benchmark reductions below assume that J is exactly the true position-coordinate set, so all distance coordinates are scanned. They do not assert that the initialization screen recovers this set. Coordinates are normalized to their respective bounds before applying the shared q. The deterministic source check uses $D = 1000$, three fixed anchors per configuration and twelve scalar values. It performs 18 formula checks and 24 conditional dominance checks across 30 cases. Maximum relative formula error is 2.38×10^{-14}. These deterministic evaluations are separate from optimizer runs.

B.1 COMMON OPTIMA AND REPRESENTATION MISMATCH

DTLZ2. For fixed position vector $u \in [0,1]^{M-1}$, let $h(u)$ be the nonnegative spherical shape vector. Then

$$f_m(u,q) = h_m(u)\{1 + (D - M + 1)(q - 0.5)^2\}. \tag{13}$$

If extra distance coordinates are preserved, replace their squared terms by constants; 0.5 remains optimal on the remaining scanned coordinates. The extra-fixed-coordinate check has relative error 2.39×10^{-14}. This proves a candidate-family midpoint optimum when all position coordinates are retained. It makes no statement when a position coordinate is overwritten by q, or about terminal population quality.

DTLZ7. With positions fixed, $f_i = u_i$ for $i < M$ and

$$f_M = M(2 + 9q) - \sum_{i=1}^{M-1} u_i[1 + \sin(3\pi u_i)]. \tag{14}$$

Thus $q = 0$ is componentwise optimal. The adaptive grid can evaluate zero if its coarse center is 0.1. The source-derived condition identifies an endpoint optimum, without guaranteeing that a particular scan discovers it.

WFG1 and WFG6. The explicit position parameter is $K = 4$ in both objective settings. For fixed normalized positions, the distance transformation $s_{\text{linear}}(q, 0.35) = |q - 0.35|/|\lfloor 0.35 - q \rfloor + 0.35|$ is nonnegative and zero at 0.35. WFG1's flat/polynomial transformations and weighted reduction retain a nonnegative distance term minimized there; WFG6's nonseparable distance reduction is also minimized when all transformed distances are zero. Their shape coordinates are independent of this term because the relevant $A_i = 1$. Hence 0.35 is a common optimum when exactly the four position coordinates are retained and all distance coordinates are scanned. For WFG6, fixing extra distance coordinates can change the nonseparable reduction, so retaining all four positions alone is insufficient. Twelve deterministic anchor/configuration checks verify this dominance. The preferred $M - 1$ screen cardinality is insufficient, although its fallback is allowed to return more coordinates.

UF1. The only true position coordinate is $u = x_1$. For distance coordinates, physical $x_j = 2q - 1$ and $c_j(u) = [1 + \sin(6\pi u + j\pi/D)]/2$. With groups $J_1 = \{3, 5, \ldots\}$ and $J_2 = \{2, 4, \ldots\}$,

$$f_m = h_m(u) + 8\{(q - \mu_m(u))^2 + \sigma_m^2(u)\}, \qquad h(u) = (u, 1 - \sqrt{u}). \tag{15}$$

The means and variances are taken over the corresponding group's c_j. Each coordinate's optimal target varies with the anchor and index, so a finer shared-scalar grid cannot remove the variance component.

UF8. For positions $u = (x_1, x_2)$, physical distance coordinates are $x_j = 4q - 2$ and normalized targets are $c_j(u) = 0.5 + 0.5u_2 \sin(2\pi u_1 + j\pi/D)$. Groups are $J_1 = \{4, 7, \ldots\}$, $J_2 = \{5, 8, \ldots\}$ and $J_3 = \{3, 6, \ldots\}$. With spherical shape vector $h(u)$,

$$f_m = h_m(u) + 32\{(q - \mu_m(u))^2 + \sigma_m^2(u)\}. \tag{16}$$

The residual vanishes at some special anchors, such as $u_2 = 0$; positive variance is not a uniform lower bound over all positions. The reported checks use three interior anchors, not an assertion about every Pareto point.

B.2 Offering Loss and Population-State Counterexamples

Arithmetic for Figure 1(b). Use shape vectors $h_1 = (0.2, 0.8)$, $h_2 = (0.8, 0.2)$, targets $u_1 = 0.2$, $u_2 = 0.8$, objectives $f_a(q) = [1 + (q - u_a)^2]h_a$, and reference $R = \{h_1, h_2\}$. The $q = 0.2$ block contains $(0.2, 0.8)$ and $(1.088, 0.272)$; the $q = 0.8$ block contains $(0.272, 1.088)$ and $(0.8, 0.2)$. Each block has ordinary IGD and IGD$^+$ equal to $\sqrt{0.288^2 + 0.072^2}/2 = 0.1484318025\ldots$; their union has zero. These values compare probe blocks alone; they also apply to the corresponding initial-plus-offered pools when initialization does not improve either block's nearest-reference distances. The construction illustrates block complementarity before truncation and does not guarantee that a finite-population selector retains both reference points.

Exact radii for Section 3.4. For its four decisions, objective vectors and single-niche selector, the nondominated candidates have radii

$$s = (2, 2): \quad (\rho(a), \rho(b), \rho(c)) = (1, 1/\sqrt{2}, 1),$$
$$s = (100, 2): \quad (\rho(a), \rho(b), \rho(c)) = (1, \sqrt{0.2501}, 0.02).$$

Both ideals are $(0, 0)$. The first normalization selects $\{a, b\}$; adding dominated d yields the second and selects $\{b, c\}$, discarding d. Sharing the second normalization also selects $\{b, c\}$ from the smaller pool. For $R = \{(1, 1)\}$, both selections and the full paid pool have IGD$^+ = 0$. These calculations verify the stated finite-reference counterexample; they do not identify population or optimizer state from the retention gap.

B.3 Reflection and the LSMOP1/GLSMOP1 Contrast

For $d = D - |J|$, the scanned family is $\mathcal{D} = \{q\mathbf{1} : q \in [0, 1]\}$. If the reflected and unreflected scanned-coordinate sets J_{refl} and J_{unref} are both nonempty, its reflected image is $T(\mathcal{D}) =$

Table 3: Stationary-point range of each actual GLSMOP1 distance group over all first-role values $u \in [0, 1]$, for $D = 1000$ and the default subcomponent count. Strict containment in $(0.475, 0.525)$ is checked by exact integer cross multiplication.

M	Group	Active coordinates	Minimum $q_i^\star$	Maximum $q_i^\star$
2	1	285	0.498721	0.501772
2	2	710	0.500117	0.500300
3	1	200	0.499548	0.499958
3	2	510	0.499660	0.499893
3	3	280	0.499731	0.499876

$\{q\mathbf{1}_{J_{\mathrm{unref}}} + (1-q)\mathbf{1}_{J_{\mathrm{refl}}} : q \in [0, 1]\}$, and

$$\mathcal{D} \cap T(\mathcal{D}) = \{\tfrac{1}{2}\mathbf{1}\}. \tag{17}$$

Indeed, membership in both families forces the unreflected and reflected coordinate values to satisfy $q = 1 - q$. For a common original target c_0 with reflected fraction ρ, Equation (2) reduces to the variance floor $\rho(1-\rho)(1-2c_0)^2$ under equal normalized weights. This is a representation statement, not a terminal-IGD bound.

The vendored GLSMOP functions reflect alternating distance coordinates of the corresponding LSMOP functions: $x_i \leftarrow \ell_i + u_i - x_i$ for $i = M, M+2, \ldots, D$. This is a specific reparameterization, not an independent family of variable-role definitions.

Source distance components and the GLSMOP1 grid optimum. Assume $D = 1000$, $M = 2/3$, the default five subcomponents, and exactly $J = \{1, \ldots, M-1\}$. For each actual active distance group H_i, write $\alpha_j = 1 + j/D$, let u be the anchor's first coordinate, and let b_j indicate reflection: $b_j = 1$ exactly when $j - M$ is even. The active groups are $H_1 = \{2, \ldots, 286\}$ and $H_2 = \{287, \ldots, 996\}$ for $M = 2$, and $H_1 = \{3, \ldots, 202\}$, $H_2 = \{203, \ldots, 712\}$ and $H_3 = \{713, \ldots, 992\}$ for $M = 3$. Remaining trailing coordinates do not enter the distance component. With these explicit groups,

$$\begin{aligned} G_i(q, u) &= \frac{100}{|H_i|} \sum_{j \in H_i} [\alpha_j((1-2b_j)q + b_j) - u]^2, \\ q_i^\star(u) &= \frac{\sum_{j \in H_i} \alpha_j(1-2b_j)(u - \alpha_j b_j)}{\sum_{j \in H_i} \alpha_j^2}. \end{aligned} \tag{18}$$

The strictly convex quadratic has the displayed stationary point; objectives multiply $1 + G_i$ by nonnegative anchor-dependent shape factors. For original LSMOP1, all $b_j = 0$, so the optimum varies with the anchor and group. Alternating reflection in GLSMOP1 nearly cancels this slope. Since $q_i^\star$ is affine in u, checking $u = 0, 1$ bounds its full range. Replacing α_j by $(D+j)/D$ permits exact integer cross multiplication, with products below 2^{53}; Table 3 gives all five intervals, strictly inside $(0.475, 0.525)$.

Under the grid and scoring rules in Appendix C.2, every possible original coarse/fine value lies on the 0.05 lattice. Thus $q = 0.5$ uniquely minimizes every G_i on that lattice and weakly minimizes each anchor's objective vector. Positive-weight scalarization, clipping, minimum across anchors and averaging are coordinatewise nondecreasing, giving $R(0.5) \leq R(q)$ and zero gain γ in Equation (30) in Appendix F.1. Ties can change the returned block but cannot produce positive gain over the midpoint. This conditional GLSMOP1 explanation does not cover all GLSMOP functions or certify screen recovery; the routed sets are checked separately. It also neither makes midpoint injection useless nor implies equal terminal performance under different FE schedules.

C Shared Search Backbone and Implementation

This Appendix specifies the fixed backbone used in Section 4. The four broad-study policies use two independent proposals per winner–loser pair and differ in their queries and offered candidates. The

Table 4: Shared backbone parameters. Matched-action parameters apply only to the component controls.

Component	Parameter	Value
Budget	Main-search reserve	$0.60F_{\max}$
Grouping	Maximum scanned groups	6
Anchors	Count rule	$\min\{N, 20, 4(M+1)\}$
Scan	Coarse table	$0.10, 0.30, 0.50, 0.70, 0.90$
Scan	Fine offsets	$-0.10, -0.05, 0, 0.05, 0.10$
Scan	Maximum rounds	2
Coordinate screen	Minimum valid rows	40
Coordinate screen	Quality strata	4
Coordinate screen	Null permutations	19
Coordinate screen	Minimum half-score	0.15
Coordinate screen	Selected/unselected ratio	1.50
Coordinate screen	Nominal cap fraction	$0.10D$
Matched action	Initial/final scale	$0.055/0.010$
Matched action	Groups per pair	1
Mutation	Site probability	$1/D$
Mutation	Distribution index	20
Archive	Capacity	$\min\{N-1, \lceil\sqrt{N}\rceil\}$
Archive	Time to live	$\max\{2, \lceil\sqrt{F_{\max}/N}\rceil\}$

targeted controls additionally change the stated block rule, query placement or injection normalization. The matched-action component controls are defined separately. The optimizer uses evaluated decisions and objectives, never benchmark identities, true position sets or reference fronts.

C.1 Overall Procedure, Parameters, and FE Accounting

Algorithm 1 Controlled probing within the shared search backbone

Require: Population size N, FE budget $F_{\max}$, reference directions V, probe policy
1: Evaluate the shared initialization I; fix its normalization; screen coordinates and choose anchors
2: Evaluate the policy's ordered query blocks, subject to the reserve rule
3: Form its offered subset O; select N objects from $I \cup O$ (retain I if no probe is applied)
4: **while** at least two FE remain **do**
5: Set $P_t = \min\{\lfloor N/2 \rfloor, \lfloor (F_{\max} - F_t)/2 \rfloor\}$ and pair winners and losers
6: Draw two independent competitive-update and mutation proposals per pair, and pre-evaluation tie coins
7: Evaluate the $2P_t$ proposals; retain one per pair under the fixed parent-conditioned comparator
8: Select from parents and retained proposals; apply the common online elite-archive gate
9: **end while**
10: **return** Full population and paid candidates; compute Native ND and Archive ND offline

The strict coordinate count is prescribed as $r = \min\{D, \max(1, M-1)\}$; it is motivated by the evaluated LSMOP construction rather than inferred as a latent dimension. The nominal fallback cap is $\min\{D, \max(r, \lceil 0.10D \rceil)\}$. The scan uses a reference-weight floor of 10^{-8}. Initial objective scales use the relative 10^{-12} floor in Equation (20); other floors are specified below.

Evaluation and cardinality accounting. Let C_{probe} count all paid probe rows, including duplicates. Then

$$F_{\text{final}} = N + C_{\text{probe}} + 2\sum_t P_t \le F_{\max}. \tag{19}$$

The two-stage scan has $C_{\text{probe}} = 10A$; Midpoint has $C_{\text{probe}} = A$. Every paid array is checked against its FE increment. Screening, comparison, population selection, archive refresh and gate evaluation consume no FE; the latter operations also preserve RNG state. Inductively, each generation starts with N objects, evaluates $2P_t$ children and retains P_t. Selection from at least N objects, with explicit filling/truncation and conditional deduplication, returns N again. Both protected and unprotected shadows have this cardinality. The formula for P_t allows another pair whenever at least

two FE remain, so budget-only termination leaves at most one unused FE. Internal runs in the original 1000-FE panels have even allocations and finish at exactly 1000 FE; follow-up runs finish at their separately stated budgets. Random plans and tie coins precede the child outcomes to which they apply.

C.2 Probe Construction and Scoring

Coordinate screening. For initial objective proportions $p_{im} = f_m(x_i)/\sum_{m'} f_{m'}(x_i)$, the shared composition features are $\phi_i = [p_{i,1:M-1}, p_{i,1:M-1}^{\odot 2}, \mathrm{Re}\log((p_{i,1:M-1}+10^{-6})/(p_{iM}+10^{-6}))]$. The screen defines fixed initial normalization and radial quality

$$\begin{aligned} z_m^0 &= \min_i f_m(x_i), \qquad \eta_m^0 = 10^{-12}\max\{1, \max_i |f_m(x_i)|\}, \\ s_m^0 &= \max\{\max_i f_m(x_i) - z_m^0, \eta_m^0\}, \qquad h_i = \|[(f(x_i) - z^0)/s^0]_+\|_2. \end{aligned} \tag{20}$$

Rows are randomized within four h_i strata and alternated between halves. Let $r_{jk}^{(1)}, r_{jk}^{(2)}$ be the sample Pearson correlations between normalized coordinate j and composition feature k within each half, with numerical safeguards for zero variance. The screen scores

$$s_j = \max_k \min\{|r_{jk}^{(1)}|, |r_{jk}^{(2)}|\}\mathbf{1}\{r_{jk}^{(1)} r_{jk}^{(2)} > 0\}. \tag{21}$$

A 19-permutation maximum, minimum response and separation checks route the preferred $M-1$ coordinates; a capped sign-stable fallback can supply a larger set. This cardinality prior comes from the evaluated LSMOP construction. The routing safeguards are specified below.

Numerical safeguards and routing boundaries. Composition features emphasize relative objective contributions but are not invariant to translations, independent rescaling or signed objectives; initialization-based screening alone does not remove confounding. The routine requires the requested initialization cardinality. For each objective row it divides by the component sum, substitutes the absolute sum when that denominator is numerically zero, and then substitutes one if needed. Nonfinite proportions become zero. Features use the first $M-1$ proportions, their squares and the real part of log ratios to the last proportion, with 10^{-6} added to both terms. Rows with nonfinite decisions, features or quality are discarded, and normalized decisions are clamped to $[0,1]$. Screening stops with fewer than 40 valid rows, fewer than 10 in either half, or constant composition features.

Quality-sorted rows use four strata with rounded equally spaced index boundaries; within each stratum, a random permutation is alternated between halves. The shuffled-label diagnostic shares this split and null plan but does not route the main comparisons. Both observed and shuffled labels use Equation (21) and descending score order. For $r = M-1$, the strict route requires r sign-stable coordinates, each above 0.15 and above $\max_{1\le b\le 19}\max_j s_j^{(b)}$, where each permutation repeats the complete split-score calculation. The minimum retained score divided by the larger of machine epsilon and the largest remaining score must exceed 1.5. An empty strict route delegates to the first sign-stable coordinates with $s_j > 0.15$, up to the nominal fallback cap. Fallback need not retain r coordinates or pass the null/separation gates; an empty fallback disables probing.

Split agreement concerns quality-balanced halves of one initialization, not distinct intervention environments. Under exchangeable, untied observed and 19 permuted maxima, the observed maximum ranks first with probability $1/20$, and strict-gate passage is a subset of that event. The implemented stratification does not establish exchangeability; this conditional argument covers neither fallback nor variable-wise errors in the presence of signal, and exact score ties have no such interpretation.

Anchor selection and scalar blocks. The normalized mapping is Equation (1); denormalization uses $x_j = \ell_j + (u_j - \ell_j)y_j$. The main reference matrix requests N directions from PlatEMO's `UniformPoint(N,M,'MUD')`; the anchor routine requests A directions separately. The anchor routine generates A reference directions, normalizes the initial objective matrix by its coordinatewise minimum and range, and computes every solution–direction angle. For direction a, candidates are sorted by angle and then radial norm, and the first unused solution is selected. This is a deterministic coverage heuristic conditional on the reference-vector generator and input order.

Let J be the routed coordinate set and $G_1, \ldots, G_{\widetilde{G}}$ the deterministic partition of its complement, where $\widetilde{G} = \min\{6, D-|J|\}$ when the complement is nonempty. For each ordered scan value q_k and anchor a, candidate-table row $(k-1)A+a$ starts at $0.5\mathbf{1}$, receives q_k on every G_g, and copies anchor a's normalized values on J. Thus the group partition is preserved for later action selection, but the scan itself tests a common q_k across all scanned coordinates. The ordered rows preserve anchorwise block comparisons and paid object multiplicities. Specifically, list the $K = D - |J|$ complement coordinates in their original order and set $e_g = \mathrm{round}(gK/\widetilde{G})$ for $g = 0, \ldots, \widetilde{G}$. Group g contains entries $e_{g-1}+1$ through e_g of that list. These consecutive intervals cover every complement coordinate exactly once, including when K is not divisible by six; an empty complement produces no groups.

Scan scoring and query policies. Each scalar evaluates all A anchors. With positive normalized reference weights w_v and $\kappa = \max\{1, \max_{x\in I} \mathrm{CV}(x)\}$, the block score is

$$R(q) = \frac{1}{|V|} \sum_{v\in V} \min_a \left[\max_m w_{vm} \left[\frac{f_m(x^{a,q}) - z_m^0}{s_m^0} \right]_+ + \frac{\mathrm{CV}(x^{a,q})}{\kappa} \right]. \tag{22}$$

The inner maximum is a weighted-Tchebycheff scalarization, as used in search by decomposition (Zhang & Li, 2007). Scan evaluates $Q_c = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and then

$$\begin{aligned} q_c &= \arg\operatorname*{first\,min}_{q\in Q_c} R(q), \\ Q_f &= \mathrm{clip}_{[0,1]}(q_c + \{-0.10, -0.05, 0, 0.05, 0.10\}), \qquad q^\star = \arg\operatorname*{first\,min}_{q\in Q_f} R(q). \end{aligned} \tag{23}$$

Midpoint evaluates and offers the $q = 0.5$ block once. No-scan evaluates none. Scan and All-scan pay for all ten query records: repeated scalars or decisions are evaluated again. All-scan uses Scan's identical records and scores but offers every paid coarse and fine block, including repeated objects. Offering is not unique-point deduplication or optimal subset selection. Safe and fixed-placement policies are specified in Appendix E.1. The initialization normalization, constraint scale and weights remain fixed throughout scan scoring; the reference-weight floor is 10^{-8}. A first minimum is selected in ordered query order. The fine-only Scan rule offers a first minimizer from its fine table; it does not apply dominance pruning or optimize a set-level quality metric. No extra FE are spent on offering or selection.

C.3 Downstream Proposal Generation and Parent-Conditioned Comparison

Winner–loser pairing and independent proposals. The backbone combines competitive updates and shift-based density estimation, with reference-vector-based environmental selection. On the current population, shift-density fitness is

$$d_i = \min_{j\neq i} \left\| \widetilde{f}_i - \max(\widetilde{f}_i, \widetilde{f}_j) \right\|_2. \tag{24}$$

Here ϵ denotes machine epsilon where an absolute floor is specified. The normalized vector $\widetilde{f}$ uses the current population's componentwise minimum and range, floored at machine epsilon. A random permutation selects $2P_t$ distinct population indices, the first and second halves form pairs, and a swap makes the member with larger shift-density fitness the winner. If d_w, d_l are their fitness values, the clipped relative gap is

$$\Delta_i = \mathrm{clip}_{[0,1]} \left(\frac{d_w - d_l}{\max\{|d_w|, \epsilon\}} \right). \tag{25}$$

For three rowwise uniform draws r_1, r_2, r_3, each replicated across coordinates, the competitive update is

$$\begin{aligned} v_i' &= r_1 v_i + (1 - r_1)\Delta_i(x_i^w - x_i^l), \\ \bar{x}_i &= x_i^l + r_2(x_i^w - x_i^l) + r_3 v_i'(1 + \Delta_i). \end{aligned} \tag{26}$$

Each proposal is normalized, clipped, and polynomially mutated with index 20 and independent site probability $1/D$. In the main comparisons, the two proposals use separate competitive-update

draws, mutation draws and velocities. Their realized decisions can coincide; both evaluations are still charged. Initial and probe objects have no stored velocity and use a zero vector when first selected as a loser. Each new proposal stores its own pre-mutation v_i'; clipping and mutation do not recompute that state. Parent-conditioned comparison and environmental selection retain whole objects, so surviving objects keep their stored velocities. The matched-action controls instead attach the common base velocity to both arms. Rowwise scalar coefficients do not restrict the update to the winner–loser line: a non-collinear inherited velocity supplies another direction even before clipping and mutation. Thus each retained object carries one velocity state. When it next acts as a loser, both independent proposals start from that same inherited state and form their own updates; there are no two persistent per-arm velocity histories attached to a single retained object.

Parent-conditioned comparison. The reference systems serve different purposes. Scan scoring and the online elite-archive gate use the initial ideal, objective scale, and $\kappa = \max\{1, \max_{x\in I} \mathrm{CV}(x)\}$. Proposal comparison uses the selected loser and winner parents in the current batch; its scale floor is $64\,\mathrm{eps}(\max\{1, |z_m|, |n_m|\})$ for their ideal z and nadir n. Environmental selection uses the current candidate pool, with range floor ϵ. Shift-density fitness uses the current population's minimum and range with the same ϵ floor. None of these latter systems is frozen for the entire run.

The comparator context is built from all loser and winner parent objectives before either proposal is evaluated. Its ideal is the parent componentwise minimum; its scale is the parent range with a positive epsilon floor. Winner objectives are normalized under this context and associated to reference directions by maximum cosine similarity. Reference vectors are clipped at zero, row normalized into weights, floored at 10^{-12}, and normalized again. For pair i, let v_i be its winner's associated direction and let $z^{\mathrm{par}}, s^{\mathrm{par}}$ be the frozen parent ideal and scale. Both proposals use the same score

$$T_i(x) = \max_m w_{v_i m} \left[\frac{f_m(x) - z_m^{\mathrm{par}}}{s_m^{\mathrm{par}}} \right]_+ .$$

The positive-part truncation also applies when a proposal improves below the parent ideal; weights and normalization are not recomputed from the proposal outcomes.

For each evaluated pair, feasibility is equivalent to zero total positive CV. If feasibility differs, the feasible proposal is selected. Otherwise CVs are compared using $64\,\mathrm{eps}(\max\{1, |v_1|, |v_2|\})$. A CV tie delegates to the weighted-Tchebycheff scores, which use an analogous tolerance. Only then does the pre-generated common coin decide. Reapplying the selector to the same paid array and fixed parent context reproduces the same indices without FE or RNG. Decision-identical paid candidates are retained with multiplicity; their ambiguous proposal identity is explicitly recorded.

C.4 Environmental Selection and Online Elite-Archive Protection

Environmental selection. Before objective-space selection, decisions are normalized by a coordinate scale, rounded to 10^{-10}, and deduplicated in stable order only when at least N unique vectors remain. This conditional rule preserves enough objects to fill the target. Objectives are translated by the current pool minimum and divided by their current pool ranges. Constraint violation is the sum of positive constraints, and Pareto-front indices are obtained from constrained nondominated sorting.

For the normalized objective $\bar{f}_i$, nearest angular direction $v(i)$, and minimum reference separation Γ_v, the angle-penalized distance is

$$\mathrm{APD}_i = \left(1 + \frac{M\theta_t \alpha_{i,v(i)}}{\max\{\Gamma_{v(i)}, \epsilon\}}\right) \|\bar{f}_i\|_2. \tag{27}$$

Angles use epsilon-normalized vectors:

$$\alpha_{iv} = \arccos\left(\mathrm{clip}_{[-1,1]} \frac{\bar{f}_i^\top v}{\max\{\|\bar{f}_i\|_2, \epsilon\} \max\{\|v\|_2, \epsilon\}}\right).$$

A zero normalized objective vector has angle $\pi/2$ to every direction and APD zero; the first direction resolves the association tie. Reference vectors are unit normalized. Every solution is associated with

the direction having minimum angle. The implementation sets the diagonal cosine to zero before taking the minimum angle, so $\Gamma_v = \min\{\pi/2, \min_{w \neq v} \angle(v, w)\}$. For P_t pairs starting at F_t, APD uses $\theta_t = ((F_t + 2P_t)/F_{\max})^2$, fixed before evaluating the batch. For each occupied direction, the selector first takes one object by ascending CV, front number, and APD. It then fills by ascending CV, front number, niche occupancy, and APD. If the initial set exceeds N, it is truncated by CV, front number, and APD.

At probe injection, the angular penalty is $\theta = 0$. In the single-niche counterexample, equal-rank feasible candidates are ordered by normalized radius and then input order.

Online elite archive and protected shadows. The archive, ages and diagnostic admission history start empty after probe injection. The first refresh follows the first evolutionary proposal comparison, using the current population and the P_t retained children; subsequent refreshes use the same current-pool rule and surviving archive entries. Paid-query logs are never a refresh source, so discarded probe rows and rejected proposals are not restored from that history. At refresh t, every previous archive age is incremented and an entry survives only when its new age is strictly less than the TTL. A decision key present anywhere in the complete previous archive cannot reset its age in the same refresh, even if the decision occurs among current parents or new evaluations. Fresh current candidates and surviving historical entries are merged, filtered to feasible solutions, deduplicated in stable order, and restricted to nondominated front one. When more than the capacity remain, the same reference-vector selector truncates the archive.

The unprotected shadow is ordinary size-N selection from the actual raw pool. At-risk entries are archive decisions absent from that shadow. The protected shadow forces one occurrence of every distinct at-risk decision and uses environmental selection for the remaining $N - |A_t^{\text{risk}}|$ slots. The capacity rule leaves at least one competitive slot. The archive, at-risk set, and protected shadow never introduce an unevaluated decision.

For consensus scoring, both shadows use the initial ideal and scale rather than population-specific renormalization. For reference v, the comparison records the best weighted-Tchebycheff value in the shadow; their mean and maximum are the mean R^2-style score and worst-reference scalar. Niche coverage uses maximum cosine association under the same fixed normalized objectives and reference matrix. Feasible count, sum and maximum normalized CV, mean R^2-style score, worst-reference scalar, and occupied niches supply the six non-worse predicates of the routing gate. The protected population is routed only if all six quantities are non-worse, at least one scalarization or coverage measure is strictly better, and the decision multisets differ. This online elite archive is part of the common backbone; it is distinct from the retrospective Archive ND endpoint, which never enters search.

Numerical comparison convention. For finite scalars, numerical comparisons use

$$\tau(a, b) = 64\,\mathrm{eps}(\max\{1, |a|, |b|\}), \qquad a \preceq_\tau b \iff a \leq b + \tau(a, b). \tag{28}$$

Strict improvement means $a < b - \tau(a, b)$. This pairwise convention need not be a transitive global ordering. The gate's local guarantee follows directly from its conjunction of six non-worse predicates and its strict-improvement test. It concerns the two current shadows under common normalization, not IGD, Pareto-set recovery, or future performance.

C.5 Computational and Storage Costs

These costs exclude black-box objective evaluations. With $K_\phi = 3(M - 1)$ and 19 null permutations, screening uses $O(20NDK_\phi)$ correlation arithmetic plus stratification and ranking. For A anchors, K_q paid scalar records and R reference vectors, scan construction and scoring cost $O(AK_qD + AK_qRM)$. Per generation, density costs $O(N^2M)$, proposals $O(P_tD)$, and reference association $O((N+P_t)RM)$, in addition to decision processing, platform nondominated sorting and archive selection. The listed search arrays require $O(ND + AK_qD + (N + P_t)R + 19N + DK_\phi)$ storage, excluding cumulative archive-history diagnostics and paid-query logging. Retaining all paid decisions and objectives adds $O(F_{\max}(D+M))$ storage; hash-based ledgers and disk-streamed records have their own storage and I/O costs. Equal FE does not imply equal wall-clock cost, and no blanket linear-time claim is made.

D BROAD STUDY AND BUDGET SENSITIVITY

The four-group benchmark matrix has 1280 paired initialization/configuration/run settings. The task matrix, probe rules, transformations, statistical families, complete comparisons and sensitivity analyses are reported together below. All listed configurations enter the analysis.

D.1 BENCHMARK MATRIX AND SHARED PROTOCOL

Original LSMOP comprises LSMOP1–9 with $M = 2/3$, giving 18 configurations. *Transformed LSMOP* comprises LSMOP1/6/9 with $M = 2/3$ under fixed reflection, power and positive objective-affine maps, giving 18 configurations. The maps are defined below; each fixed map is shared across all policies and paired runs. *GLSMOP* comprises GLSMOP1–9 with $M = 2/3$, giving 18 configurations in an LSMOP-derived reflection suite. *DTLZ/WFG/UF* comprises ten configurations: DTLZ2/7 and WFG1/6 with $M = 2/3$, UF1 with $M = 2$, and UF8 with $M = 3$. WFG uses $K = 4$; UF uses $D = 1000$ rather than the original 30-variable competition setting. All groups use $D = 1000$, $N = 100$ and exactly 1000 FE. Only the transformed-LSMOP comparisons use reference-front ranges for objective normalization; the other three groups use native objective units. These constructions cover the structural premises in Section 3, with twenty paired runs per configuration.

Shared policies and budget. The four policies use the shared backbone. All routed outcomes enter the analysis, including zero score gains and fallback routes. The anchor count is $A = \min\{N, 20, 4(M+1)\}$; probing requires a nonempty route and at least $0.60F_{\max}$ reserved for evolution. Every run satisfies $100 + C_{\text{probe}} + 2\sum_t P_t = 1000$. Different probe costs change the remaining search FE and FE-indexed schedule positions. The component controls instead use their separately defined matched-action proposal construction.

D.2 TRANSFORMATIONS, OUTPUT RECONSTRUCTION, AND METRICS

Fixed transformations. Normalize decisions to $y \in [0,1]^D$. A random permutation of all coordinate indices defines a fixed mask containing half of the coordinates. Reflection maps masked coordinates to $1 - y_j$ and leaves others unchanged. Power maps masked coordinates to $y_j^{1/2}$ and the remainder to y_j^2. The canonical objective is evaluated at the transformed coordinates, with the same box bounds and reference front. The fixed maps consume no optimizer random draws. Power transformations also change the canonical initialization distribution; this test does not isolate a causal effect of scanning from all other consequences of reparameterization.

The objective-affine view is $f'_m = a_m f_m + b_m$, where $a = 10^{\text{linspace}(-1,1,M)}$ and $b = (1, \ldots, M)$. The reference points undergo the same transformation. For both transformed-LSMOP primary outputs, subtract the reference minimum in each objective and divide by its range, replacing an exactly zero range by one. For nonzero ranges this cancels direct positive-affine unit changes in distances; it does not undo changes in the algorithm's composition features or selected solutions. Each problem/objective/transform setting has one fixed map, shared by all methods and runs. No general uncertainty over maps is estimated.

Output reconstruction and metrics. Repeated decisions remain separate paid queries. Objective and constraint row counts must equal the FE counter. Internal methods finish at their declared budget B with 100 population objects; $B = 1000$ throughout this broad study, with follow-up budgets stated separately.

For a budget prefix B, keep the first occurrence of each distinct decision. Apply PlatEMO's constraint-aware nondominated rank and descending within-front crowding distance, following the rank-and-crowding principle of NSGA-II (Deb et al., 2002); break remaining ties by ascending query index. Retain at most $N = 100$ objects without padding. Denote this set by $\mathcal{A}_B$ and its feasible nondominated subset by $\text{ND}(\mathcal{A}_B)$. The primary archive endpoint is

$$\text{IGD}_{\text{archive,ND}}(B) = \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \min_{a \in \text{ND}(\mathcal{A}_B)} \|f(a) - r\|_2, \tag{29}$$

with reference-range normalization for transformed LSMOP. Here $\mathcal{R}$ is the same finite reference set returned by the problem's `GetOptimum` routine. Each benchmark reference front is the finite set returned by its `GetOptimum(10000)` routine; the returned cardinality, which can differ from the requested size, is retained. This fixed set is shared across paired policies and budgets, with the stated affine transformation applied to both objectives and reference points. Reference construction is offline and uses no paid optimizer queries. The secondary all-object archive endpoint replaces $\mathrm{ND}(\mathcal{A}_B)$ by $\mathcal{A}_B$. Native ND instead uses the feasible nondominated subset of the final algorithm population. All controls use the same terminal budget, so the native comparisons have no FE overshoot. The prefix archive is an offline output rule and never influences the optimizer's decisions.

An archive capped by front rank and crowding is not a monotonically growing set. Its IGD, including the ND endpoint, can therefore worsen at a later prefix. Native and archive fronts can also have different numbers of distinct nondominated members, which are recorded.

D.3 Statistical Families and Matched-Query Audit

Inference and sensitivity families. The six contrasts per configuration are Scan–Midpoint, All-scan–Scan, All-scan–Midpoint, Scan–No-scan, Midpoint–No-scan and All-scan–No-scan. Both primary outputs share a 216-test family in each 18-configuration LSMOP group and a 120-test family for DTLZ/WFG/UF. Secondary archive-all families have 108/108/108/60 tests; secondary Archive ND and Native ND IGD^+ jointly use 216/216/216/120. Exact-sign and asymptotic-rank sensitivities repeat the same grouping; endpoints are not independently corrected within a primary group.

Matched-query audit. All 1280 paired settings have identical Scan/All-scan pre-injection queries, objectives, scores, anchors and random states. Each Midpoint block matches the paid 0.5 block of its paired Scan. Every run finishes with exactly 1000 FE and 100 population objects.

Geometric and metric definitions. With a fixed reference set and fixed objective scaling, both IGD and IGD^+ are nonincreasing under set inclusion: adding candidates cannot increase any reference point's nearest distance. Nonnegative offering/selection losses and their additive decomposition therefore do not rely exclusively on IGD^+. The additional reason for using IGD^+ in the dominance arguments is that a dominating candidate is no farther from any reference point in its one-sided distance. Ordinary Euclidean distance lacks this property, so standard IGD need not be preserved when dominated candidates are removed. Neither inclusion argument applies automatically to separately truncated archives or selected populations. The quadratic identity in Equation (2) is independent of either metric. For DTLZ7, the adaptive fine grid includes zero when its coarse center is 0.1, consistent with the lower-endpoint optimum. The discovery and retention gaps use the uncapped paid pool, as required for the set-inclusion inequality in Section 3.2.

D.4 Additional Diagnostics for the Broad Study

Discovery and routed-coordinate premises. The extra-query discovery fraction $[\overline{I^+(R, S_{\mathrm{mid}})} - \overline{I^+(R, S)}]/\overline{I^+(R, S_{\mathrm{mid}})}$ compares the uncapped initial-plus-midpoint prefix with all paid scan queries. Its median is 0.753 on original LSMOP and zero on GLSMOP. All true position coordinates are retained in 360/360 shared original-LSMOP initializations and 360/360 on GLSMOP, excluding wholesale position-coordinate loss as an explanation for this contrast. The LSMOP1/GLSMOP1 calculation in Appendix B.3 connects it to anchor-dependent quadratic means versus a common nearest grid point under the stated reflection and routing premises. In the main GLSMOP1 study, the routed set is exactly $\{1, \ldots, M-1\}$ in 20/20 shared initializations for $M = 2$ and 17/20 for $M = 3$. These 37 cases satisfy the stated routing condition; the remaining three retain extra distance coordinates and also have zero score gain, but this is an empirical observation outside that proof's premises. All-scan also reduces the mean uncapped-prefix IGD^+ retention gap on every original-LSMOP configuration. These are explanatory measurements, not additional primary tests.

Native failures despite a common paid prefix. For reflected LSMOP6 ($M = 2$), the native Scan/All-scan mean-IGD ratio is 0.104 [0.0360,0.229] ($p_H = 0.00310$), while the archive ratio is 0.926 [0.874,0.976] ($p_H = 0.103$). For GLSMOP6 ($M = 2$), the corresponding ratios are 0.0907 [0.0170,0.231] ($p_H = 0.00460$) and 1.007 [0.924,1.099] ($p_H = 1$). Thus the native losses

coexist with unresolved archive comparisons. The common paid prefix remains available to archive reconstruction, but subsequent queries and archive truncation can differ. These contrasts neither bound retention loss mathematically nor identify whether diversity, ordering or another state change causes the native failure.

DTLZ and WFG source checks. DTLZ2 preserves its exact positions in all 40 paired settings and has zero additional-query IGD^+ discovery gain. Three two-objective scans choose a block anchor-wise dominated by the already-paid midpoint under tied internal scores, and the sensitivity analysis examines the resulting tail-sensitive terminal effect. For $M = 3$, both immediate retention gaps are zero despite the terminal All-scan advantage discussed in Section 4.2. DTLZ7 has its distance optimum at zero; additional queries reduce initial-plus-midpoint IGD^+ by 99.6% and 99.0% for $M = 2/3$.

WFG retains all four true position coordinates in 21/80 paired settings: 1/20 and 9/20 for WFG1 ($M = 2/3$), and 0/20 and 11/20 for WFG6. All 21 cases use fallback, since the strict $M - 1$ route cannot retain four positions. WFG6 additionally requires every distance coordinate to be scanned; retaining its positions alone does not certify that premise. Each recorded route must therefore be checked against the source conditions.

D.5 Sensitivity, Outliers, and Complete Comparison Panels

Tail-sensitive examples and rank tests. On DTLZ2 with $M = 2$, the Native ND Scan/All-scan mean ratio is 22.0 with pointwise interval [1.45,45.1]; large errors amplify its magnitude. Across all 768 primary comparisons, the sole significant rank/mean direction disagreement occurs on original LSMOP7 with $M = 3$ for Archive ND. Here No-scan has lower IGD in 19/20 pairs and is favored by the signed-rank test ($p_H = 0.0205$), although the No-scan/Scan mean ratio is 70.25 owing to one large No-scan error. These descriptive cases distinguish arithmetic-mean magnitude from paired rank-test direction.

IGD^+ sensitivity. Matching the existing IGD and IGD^+ effect summaries by configuration, contrast and output gives 768 corresponding tests with the stated joint families. Their W/T/L classifications agree in 746 cases. Seven Archive ND and fifteen Native ND classifications change: twelve T become W, three T become L, five W become T and two L become T. No significant win changes to a significant loss or vice versa. Thus the classified evidence is largely stable, but not identical across metrics. For example, original LSMOP5 with $M = 3$ changes from T to W for All-scan versus Scan on Archive ND, whereas power-transformed LSMOP1 with $M = 2$ changes from W to T for the same contrast and output. These are summaries of the existing sensitivity results, not new optimizer runs or a claim of metric equivalence. For the query-allocation contrast Scan versus Midpoint, every classification is unchanged on original LSMOP, transformed LSMOP and GLSMOP, including the two transformed archive losses. In DTLZ/WFG/UF, Archive ND changes from 7/1/2 to 7/0/3 and Native ND from 6/1/3 to 6/2/2. For the retention contrast All-scan versus Scan, original-LSMOP Archive ND changes from 12/6/0 to 13/5/0, while transformed-LSMOP Archive ND changes from 7/11/0 to 6/12/0 and Native ND from 7/10/1 to 6/11/1. Its other group/output classifications are unchanged. These distinctions preserve the mixed evidence for extra adaptive queries without asserting that every retention benefit is metric-invariant.

Complete comparison panels. Table 1 reports all group summaries; Figures 4–6 and 7 retain all 64 configurations, six contrasts and both outputs. Table 5 additionally reports the intervals and adjusted p values for the query and retention contrasts on DTLZ/WFG/UF. The transformed-LSMOP Scan/Midpoint comparison includes two archive losses and nine unresolved native cases; it does not support uniformly spending $10A$ probe FE.

Column labels give the numerical numerator/denominator. Cells in the complete panels show comparator/reference arithmetic-mean IGD ratios and W/T/L under the group's joint exact signed-rank Holm family. Colors use a common scale and saturate outside 10^{-4}–10^{4}. Row suffixes R/P/A denote reflection, power and objective-affine transformations. The panels report all primary effect ratios and corrected decisions; Table 5 gives numerical intervals and adjusted p values for the two focused contrasts on DTLZ/WFG/UF.

	Archive ND						Native ND					
LSMOP1, M2	3.6 W	1.2 W	4.3 W	6 W	1.6 W	6.9 W	2.9 W	1.5 W	4.2 W	4.7 W	1.6 W	6.9 W
LSMOP1, M3	1.8 W	1.2 W	2.2 W	3.8 W	2.1 W	4.7 W	1.5 W	1.5 W	2.2 W	3.2 W	2.1 W	4.8 W
LSMOP2, M2	1.4 W	1.4 W	2 W	2.2 W	1.6 W	3.1 W	1.3 W	1.3 W	1.7 W	1.9 W	1.5 W	2.6 W
LSMOP2, M3	1.2 W	1.1 T	1.2 W	1.2 W	1 T	1.3 W	1 T	0.97 T	0.98 T	1 T	1 T	1 T
LSMOP3, M2	18 W	1 T	18 W	22 W	1.2 W	22 W	84 W	1 T	84 W	24 W	0.28 T	24 W
LSMOP3, M3	1.9 W	1.9 W	3.6 W	3.6 W	1.8 W	6.7 W	2.1 T	3.2 W	6.8 W	2.8 W	1.3 T	9.1 W
LSMOP4, M2	1.4 W	1.6 W	2.1 W	1.6 W	1.2 W	2.5 W	1 T	1.9 W	2 W	1.2 W	1.2 W	2.3 W
LSMOP4, M3	1.1 W	1.3 W	1.5 W	1.3 W	1.1 W	1.7 W	1 T	1.2 W	1.2 W	1.1 W	1.1 W	1.3 W
LSMOP5, M2	3.5 W	1 T	3.5 W	6.8 W	1.9 W	6.8 W	3.5 W	1 T	3.5 W	6.8 W	1.9 W	6.8 W
LSMOP5, M3	5.6 W	1 T	5.8 W	12 W	2.1 W	12 W	5.5 W	1.1 W	5.9 W	12 W	2.1 W	12 W
LSMOP6, M2	1.7 W	1.3 W	2.3 W	2 W	1.2 W	2.7 W	2.8e2 W	4.7 W	1.3e3 W	0.51 L	0.0019 L	2.4 W
LSMOP6, M3	1.8e2 W	1.4 W	2.4e2 W	3.3e3 W	19 W	4.6e3 W	1.8e2 W	1.4 W	2.5e2 W	3.1e3 W	17 W	4.4e3 W
LSMOP7, M2	3.1e2 W	1 T	3.1e2 W	1.5e3 W	4.9 W	1.5e3 W	3.1e2 W	1 T	3.1e2 W	1.5e3 W	4.9 W	1.5e3 W
LSMOP7, M3	0.73 L	1.1 W	0.82 L	70 L	96 W	79 T	1.2e4 W	1.1 W	1.3e4 W	1.7e4 W	1.5 T	2e4 W
LSMOP8, M2	2.7 W	1 T	2.7 W	5 W	1.8 W	5 W	2.7 W	1 T	2.7 W	5 W	1.8 W	5 W
LSMOP8, M3	3.4 W	1.4 W	5 W	4.6 W	1.3 W	6.7 W	2.5 W	1.1 T	2.6 W	2.8 W	1.1 T	3 W
LSMOP9, M2	1.9 W	1.4 W	2.6 W	69 W	37 W	95 W	1.3 T	2.5 W	3.2 W	34 W	27 W	86 W
LSMOP9, M3	19 W	1.5 W	28 W	54 W	2.9 W	80 W	9.6 W	3 W	29 W	28 W	2.9 W	85 W
	M/S	S/A	M/A	N/S	N/M	N/A	M/S	S/A	M/A	N/S	N/M	N/A

Figure 4: Original LSMOP comparisons. Cells show comparator/reference mean IGD ratios and W/T/L. Column labels give the numerical numerator/denominator.

	Archive ND						Native ND					
LSMOP1, M2 R	1 L	1 W	1 T	1.6 W	1.6 W	1.6 W	1 T	1 W	1 T	1.6 W	1.6 W	1.6 W
LSMOP1, M2 P	11 W	1 W	11 W	12 W	1.1 W	12 W	11 W	1 W	11 W	12 W	1.1 W	12 W
LSMOP1, M2 A	3.6 W	1.2 W	4.3 W	6 W	1.6 W	6.9 W	2.9 W	1.5 W	4.2 W	4.7 W	1.6 W	6.9 W
LSMOP1, M3 R	1 T	1 T	1 T	2.1 W	2.1 W	2.2 W	1 T	1 T	1 T	2.1 W	2.1 W	2.2 W
LSMOP1, M3 P	2.1 W	1.3 W	2.7 W	2.2 W	1.1 T	2.9 W	1.3 W	1.8 W	2.4 W	1.4 W	1.1 W	2.5 W
LSMOP1, M3 A	1.1 T	1.1 T	1.2 W	1.5 W	1.4 W	1.8 W	1 T	1.2 T	1.2 W	1.6 W	1.5 W	1.8 W
LSMOP6, M2 R	1 T	0.93 T	0.94 T	1.2 W	1.2 W	1.1 W	2.2 T	0.1 L	0.23 L	1.5 T	0.66 L	0.15 L
LSMOP6, M2 P	0.54 L	1.3 T	0.7 T	0.54 L	1 W	0.71 T	36 T	1.2 T	42 T	0.51 L	0.014 T	0.6 L
LSMOP6, M2 A	1.7 W	1.3 W	2.3 W	2 W	1.2 W	2.7 W	2.8e2 W	4.7 W	1.3e3 W	0.51 L	0.0019 L	2.4 W
LSMOP6, M3 R	0.57 T	2.3 T	1.3 T	17 W	30 W	39 W	0.63 T	2.3 T	1.4 T	17 W	27 W	39 W
LSMOP6, M3 P	1.1e4 W	1 T	1.2e4 W	1.8e4 W	1.6 W	1.9e4 W	1e4 W	1 T	1.1e4 W	1.8e4 W	1.8 W	1.9e4 W
LSMOP6, M3 A	1 T	1.1 T	1.1 T	2.9 W	2.8 W	3.1 W	1.1 T	1.1 T	1.2 T	2.8 W	2.6 W	3.1 W
LSMOP9, M2 R	0.98 T	1 T	0.98 T	19 W	20 W	19 W	0.95 T	1 T	0.97 T	17 W	18 W	18 W
LSMOP9, M2 P	37 W	1 T	37 W	44 W	1.2 W	44 W	37 W	1 T	37 W	45 W	1.2 W	45 W
LSMOP9, M2 A	2.4 W	1.4 W	3.3 W	53 W	22 W	73 W	1.3 W	2.6 W	3.4 W	25 W	19 W	66 W
LSMOP9, M3 R	0.99 T	1 T	1 T	2.8 W	2.8 W	2.8 W	0.99 T	0.99 T	0.98 T	2.8 W	2.9 W	2.8 W
LSMOP9, M3 P	48 W	1 T	50 W	59 W	1.2 W	61 W	42 W	1.2 T	49 W	53 W	1.3 W	61 W
LSMOP9, M3 A	1.6 W	1.2 W	1.9 W	3.5 W	2.2 W	4.2 W	1.8 W	1.3 W	2.3 W	3.7 W	2.1 W	4.8 W
	M/S	S/A	M/A	N/S	N/M	N/A	M/S	S/A	M/A	N/S	N/M	N/A

Figure 5: Transformed LSMOP comparisons. Cells show comparator/reference mean IGD ratios and W/T/L. Column labels give the numerical numerator/denominator. Distances use the study normalization.

Sensitivity and reproducibility. Across 768 primary comparisons, asymptotic signed-rank and exact sign-test sensitivities change 12 and 26 decisions. Secondary metric checks cover 768 IGD^+ and 384 archive-all effects under the families defined above. DTLZ/WFG/UF All-scan–Scan IGD^+ outcomes remain 6/4/0 on both outputs; all-object archive IGD adds one original LSMOP loss for this contrast.

D.6 Budget Sensitivity

The five selected cases are LSMOP1/$M = 2$, GLSMOP1/$M = 2$, reflected LSMOP6/$M = 2$, GLSMOP6/$M = 2$, and DTLZ2/$M = 3$. Each of No-scan, Midpoint, Scan and All-scan is run separately at $B = 1000, 5000, 10000$, with twenty paired runs. Initialization costs 100 FE and probing costs $0, A, 10A, 10A$, respectively, with $A = 4(M+1)$; saved FE return to evolution. Initial decisions match across policies and budgets, Scan/All-scan share their paid pre-injection state, and Midpoint matches the paid $q = 0.5$ block. Original FE-indexed schedules remain in use, so these

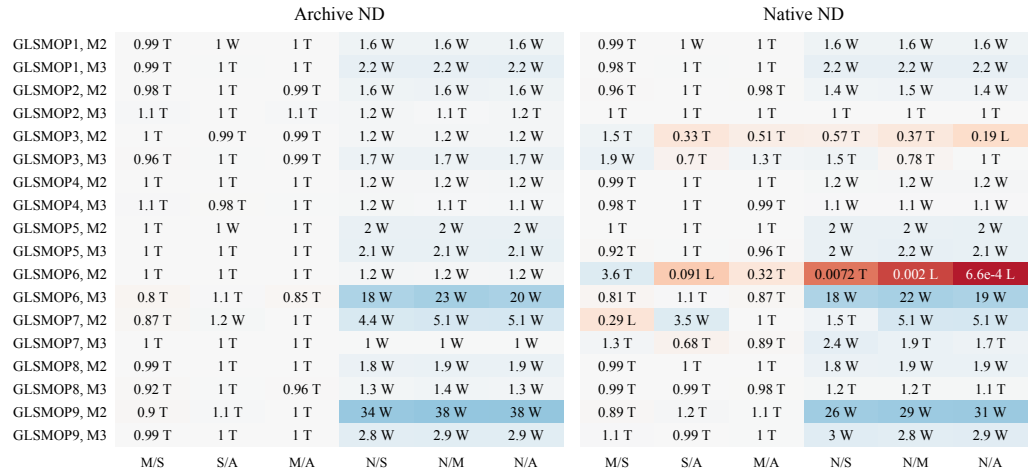

Figure 6: GLSMOP comparisons. Cells show comparator/reference mean IGD ratios and W/T/L. Column labels give the numerical numerator/denominator.

Table 5: Complete query and retention contrasts on DTLZ/WFG/UF. The table reports comparator/reference mean IGD ratios, pointwise paired 95% bootstrap intervals, Holm-adjusted p values from the original 120-test family, and W/T/L. M, S, and A denote Midpoint, Scan, and All-scan, respectively. Each comparison uses twenty paired runs.

		Archive ND			Native ND		
Case	Ratio	Estimate [95% CI]	p_H	W/T/L	Estimate [95% CI]	p_H	W/T/L
DTLZ2, $M = 2$	M/S	0.794 [0.707, 0.881]	2.29×10^{-4}	L	0.0594 [0.0297, 0.89]	2.29×10^{-4}	L
	S/A	1.73 [1.46, 2.03]	4.01×10^{-4}	W	22 [1.45, 45.1]	2.29×10^{-4}	W
DTLZ2, $M = 3$	M/S	0.892 [0.865, 0.919]	3.05×10^{-4}	L	0.911 [0.876, 0.95]	0.0179	L
	S/A	1.14 [1.1, 1.18]	0.00283	W	1.15 [1.09, 1.21]	0.0111	W
DTLZ7, $M = 2$	M/S	779 [665, 899]	2.29×10^{-4}	W	382 [330, 438]	2.29×10^{-4}	W
	S/A	1.01 [0.851, 1.18]	1	T	0.92 [0.728, 1.13]	1	T
DTLZ7, $M = 3$	M/S	127 [120, 133]	2.29×10^{-4}	W	70 [63.8, 76.4]	2.29×10^{-4}	W
	S/A	0.963 [0.895, 1.05]	1	T	1.06 [0.947, 1.19]	1	T
UF1, $M = 2$	M/S	3.67 [3.36, 3.94]	2.29×10^{-4}	W	3.34 [2.96, 3.67]	2.29×10^{-4}	W
	S/A	1.23 [1.16, 1.32]	2.29×10^{-4}	W	1.35 [1.24, 1.49]	2.29×10^{-4}	W
UF8, $M = 3$	M/S	1.04 [0.991, 1.09]	0.306	T	0.712 [0.693, 0.731]	2.29×10^{-4}	L
	S/A	1.19 [1.13, 1.25]	2.29×10^{-4}	W	1.73 [1.62, 1.84]	2.29×10^{-4}	W
WFG1, $M = 2$	M/S	1.93 [1.69, 2.22]	2.29×10^{-4}	W	1.93 [1.69, 2.24]	2.29×10^{-4}	W
	S/A	1 [0.854, 1.17]	1	T	0.965 [0.823, 1.12]	1	T
WFG1, $M = 3$	M/S	1.53 [1.41, 1.67]	2.29×10^{-4}	W	1.53 [1.4, 1.67]	2.29×10^{-4}	W
	S/A	1.05 [0.971, 1.15]	1	T	1.05 [0.963, 1.16]	1	T
WFG6, $M = 2$	M/S	2.23 [1.94, 2.56]	2.29×10^{-4}	W	0.929 [0.819, 1.06]	1	T
	S/A	1.39 [1.24, 1.57]	3.05×10^{-4}	W	3.07 [2.68, 3.47]	2.29×10^{-4}	W
WFG6, $M = 3$	M/S	1.24 [1.14, 1.34]	0.00653	W	1.17 [1.1, 1.25]	0.00546	W
	S/A	1.12 [1.05, 1.18]	0.0156	W	1.22 [1.15, 1.29]	2.29×10^{-4}	W

are budget-dependent policies, not prefixes or continuations of a fixed trajectory. Both outputs use exactly B FE; reflected LSMOP uses reference-range normalization and the other cases native units. With paired difference $d_r(B)$ defined as comparator minus reference IGD, six pairs, five cases, three budgets and two outputs give 180 endpoint tests. Another 120 test $\delta_r(B) = d_r(B) - d_r(1000)$ at the larger budgets; all 300 share one Holm family. Table 6 reports every contrast and interaction.

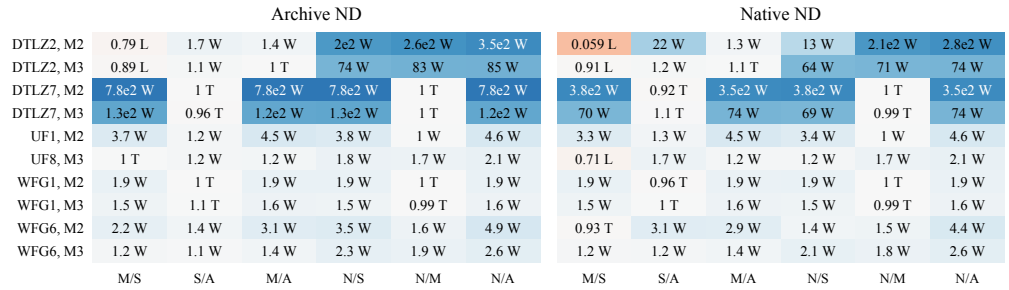

Figure 7: DTLZ / WFG / UF comparisons. Cells show comparator/reference mean IGD ratios and W/T/L. Column labels give the numerical numerator/denominator.

Table 6: All six policy contrasts and both outputs, aggregated as counts across the five selected cases. The three budget columns show W/T/L. The last two columns show increase/unresolved/decrease of the paired contrast relative to 1k, not optimizer wins. Every cell sums to five; all 300 tests share one Holm family.

Reference / comparator	Output	1k	5k	10k	$\delta(5k)$	$\delta(10k)$
Scan / Midpoint	Archive	1/3/1	1/4/0	1/4/0	1/4/0	1/4/0
Scan / Midpoint	Native	1/4/0	1/4/0	1/4/0	1/4/0	1/4/0
All-scan / Scan	Archive	3/2/0	0/5/0	0/5/0	0/4/1	0/4/1
All-scan / Scan	Native	3/1/1	1/4/0	1/4/0	0/3/2	0/3/2
All-scan / Midpoint	Archive	1/4/0	1/4/0	1/4/0	1/4/0	1/4/0
All-scan / Midpoint	Native	1/4/0	1/4/0	1/4/0	0/5/0	0/4/1
Scan / No-scan	Archive	5/0/0	4/0/1	4/1/0	0/3/2	1/3/1
Scan / No-scan	Native	3/2/0	3/2/0	3/2/0	0/2/3	0/2/3
Midpoint / No-scan	Archive	5/0/0	3/0/2	3/2/0	0/3/2	0/4/1
Midpoint / No-scan	Native	3/1/1	3/1/1	3/2/0	0/2/3	0/2/3
All-scan / No-scan	Archive	5/0/0	4/0/1	4/1/0	0/3/2	1/3/1
All-scan / No-scan	Native	3/0/2	3/1/1	3/2/0	0/2/3	0/2/3

Results and scope. Scan/Midpoint favors Scan on LSMOP1 for both outputs at every budget; its Midpoint/Scan ratios increase from 3.56 to 8.91 for Archive ND and 2.50 to 3.83 for Native ND between 1000 and 10000 FE. This does not extend uniformly across cases: DTLZ2 Archive ND favors Midpoint at 1000 FE. Retention support is more budget-dependent. All-scan/Scan changes from 3/2/0 to 0/5/0 on Archive ND and from 3/1/1 to 1/4/0 on Native ND at both larger budgets. LSMOP1 retains a native benefit, with Scan/All-scan ratios 1.690, 1.139 and 1.127. Its paired archive contrast decreases relative to 1000 FE at 5000 and 10000 FE ($p_H = 0.0431$ and 0.000572); native contrasts also decrease for LSMOP1 and GLSMOP1. These direct interactions, rather than changes in significance labels, support attenuation in those cases.

GLSMOP6 Native ND favors Scan over All-scan at 1000 FE (ratio 0.100); ratios 0.590 and 0.675 at larger budgets have wide intervals and unresolved tests, as do the reflected-LSMOP6 retention comparisons. This establishes neither recovery nor equivalence. Native quality can deteriorate, and neither a budget-dependent policy nor the capped archive is guaranteed to improve monotonically with B. The complete IGD^+ and exact-sign sensitivities change six and seventeen of the 300 classifications, respectively. The reported effect sizes, direct budget-interaction tests and all-contrast counts limit generalization of the short-budget retention finding to these selected cases.

E Targeted Controls and Population-State Analysis

This appendix groups single-block repair, equal-cost grid placement, injection normalization and subsequent population/optimizer-state interventions. The first two experiments use twenty paired runs per configuration, distinct from the broad study, at $D = 1000$, $N = 100$ and 1000 FE. Later

Table 7: Function-evaluation allocation for the targeted controls at a total budget of 1000 FE. All controls spend 100 FE on the common initialization; paired entries correspond to $M = 2/3$.

Control	Probe FE, $M = 2/3$	Evolution FE	Offered candidates
No-scan	$0/0$	$900/900$	none
Midpoint	$12/16$	$888/884$	midpoint block
Scan	$120/160$	$780/740$	selected fine block
Safe / Fixed-safe	$120/160$	$780/740$	one undominated block
All-scan / Fixed-all	$120/160$	$780/740$	all paid scan blocks

follow-ups state their own pairing and budgets. Inference follows the statistical procedures described in Section 4.1, with the study-specific test families specified below.

E.1 Experimental Design and Interventions

Configurations and FE allocation. The eight configurations are LSMOP1 ($M = 2$), LSMOP6 ($M = 3$), GLSMOP1 ($M = 2$), DTLZ2 ($M = 2/3$), DTLZ7 ($M = 3$), UF8 ($M = 3$) and WFG6 ($M = 3$, $K = 4$). Together they examine repair, representation alignment and retention across different benchmark constructions.

Fixed-safe and Fixed-all replace the five adaptive fine query values by $\{0, 0.25, 0.5, 0.75, 1\}$, retaining the original five coarse values, roles and anchors. Fixed-safe applies the Safe single-block rule; Fixed-all offers every paid block. The repeated 0.5 query is paid in both stages, so each policy spends $10A$ probe FE. The common initialization costs 100 FE; Table 7 lists the remaining allocations. No-scan and Midpoint are included for budget context and are not members of the five-policy repair/placement comparison. The 120/160 probe FE consume 12%/16% of the total budget. With at most 100 evaluations per full evolution batch, the 780/740 FE allocations allow seven full batches and one partial batch; Midpoint allows eight full batches and one partial batch, while No-scan allows nine full batches. Initialization and injection can therefore occupy a substantial part of this short search horizon. These allocations do not establish how the comparisons would change at larger budgets.

Dominance-safe and shared-normalization interventions. The block dominance test concatenates the objectives in fixed anchor order. It removes a block only if another evaluated block is no worse at every entry and strictly better at one. Exact stored floating-point comparisons define a transitive partial order; a tolerance-based pairwise relation could lose that property. The present benchmark suite is unconstrained, so every candidate is feasible. A constrained extension would need to define and justify its block order separately. Equal-score undominated blocks use the first query index. Both coarse and fine blocks are eligible, including paid duplicates, without additional evaluations or reference-front access. Among the surviving blocks, Safe minimizes the unchanged clipped score and offers one block of A candidates. Safe/Scan therefore repairs both fine-only eligibility and dominated choices; it does not optimize a set-level quality metric. These controls test whether the retention advantage survives this dominance-safe rule; they do not establish that the chosen block is optimal among all paid single-block alternatives. Since $q_c \in Q_f$, deterministic reevaluation under the common score context gives

$$\min_{q \in Q_c \cup Q_f} R(q) = \min_{q \in Q_f} R(q).$$

Expanding eligibility alone therefore cannot supply a strictly lower raw score. Dominance pruning and the ordering of tied minima can still change the offered block. Terminal equality or an unresolved test does not measure how often either event occurred.

Recorded block choices. A post-outcome audit compares the stored dominance matrices, paid decision blocks and actual offered decisions across all eight configurations. Each run contains ten paid block records, including repeated coarse/fine queries. Safe prunes 889 of 1600 block records across 160 runs; Fixed-safe prunes 948 of 1600. These totals count paid occurrences, not distinct blocks. The legacy choice is dominated in 26 Safe runs and 30 Fixed-safe runs. The respective

legacy choices are the first fine-table minimum and the first minimum over all ten unpruned fixed-placement records.

Safe changes the selected record index in 119/160 runs, but 78 changes select another occurrence of exactly the same ordered candidate block. Only 41 changes alter the candidate set: 20 on LSMOP6 ($M = 3$), one on DTLZ2 ($M = 2$), and 20 on UF8 ($M = 3$). None changes only the order within an otherwise identical candidate set. Direct comparison with the paired Scan records confirms that the remaining 119/160 pairs have exactly identical offered decisions, ordered post-injection decision arrays and ordered final decision arrays. Thus frequent endpoint equality has direct record-level support here, even though pruning itself is common. For Fixed-safe, all 30 index changes alter the candidate set relative to its unpruned fixed-grid choice; none is merely a repeated-block switch. These descriptive checks add no hypothesis tests and do not imply that different offered blocks must produce significantly different endpoints.

The shared-normalization intervention changes only the objective ideal and range passed to the injection selector. For a common paid pool S, these are $z_m = \min_{s\in S} f_m(s)$ and $s_m = \max\{\max_{s\in S} f_m(s) - z_m, \epsilon\}$. All other selection rules remain unchanged, including decision deduplication and Pareto ranking. Reference associations and niche occupancies are recomputed under the shared scaling, using the same reference vectors and tie-breaking rules. All later normalization is still population dependent.

Statistical families. The five-policy repair/placement comparison includes all ten unordered pairs across eight cases and both primary outputs, forming a 160-test family. IGD^+ uses a separate 160-test family and archive-all IGD an 80-test family. The normalization comparison uses forty default-normalization records from the five-policy comparison and forty shared-normalization runs; record reuse does not add independent runs. Its six pairwise contrasts on two outputs and two paired log-effect interactions share one fourteen-test family. Each interaction tests whether sharing injection normalization changes the terminal All-scan/Safe IGD ratio. The intervention does not fix normalization during later evolution.

Computational boundary. For K paid blocks, Safe adds $O(K^2AM)$ objective comparisons without additional FE or set-level metric optimization. Equal FE therefore does not imply equal wall-clock cost. External methods have the separate resource boundaries in Appendix F.4.

Run and matching audit. All 340 Safe-policy runs across the repair and normalization comparisons pass the anchorwise dominance check. The 160 paired settings have identical adaptive prefixes across Scan, Safe, All-scan and, where present, shared-normalization controls; both fixed policies have identical fixed prefixes. Query counts, scores, anchors, pre-injection RNG states and the exact 1000-FE accounting are checked.

E.2 Single-Block Repair and Equal-Cost Placement

The repair comparison's DTLZ2 two-objective Scan/Safe native mean ratio is 5.87, but nineteen of twenty differences are zero; its corrected comparison is unresolved. The Safe/All-scan native mean ratio is instead 1.31 [1.22,1.41], with $p_H = 7.21 \times 10^{-4}$. This distinction prevents a rare repair event from being represented as a typical multi-block gain. All configuration intervals, including null and unfavorable effects, follow below.

The main All-scan/Safe pattern persists under the sign-test sensitivity, although some significance decisions change. Table 8 reports all ten primary contrasts and both outputs, including null and unfavorable results.

Table 8: Repair and placement effects for eight configurations and twenty paired runs. Each output gives comparator/reference mean IGD [pointwise 95% paired bootstrap interval], joint 160-test Holm p_H, and W/T/L.

	Archive ND			Native ND		
Ref. vs. comp.	Ratio [95% CI]	p_H	W/T/L	Ratio [95% CI]	p_H	W/T/L
LSMOP1, $M = 2$						
Safe vs. Scan	1 [1, 1]	1	T	1 [1, 1]	1	T
All-scan vs. Scan	1.15 [1.13, 1.17]	0.000305	W	1.51 [1.38, 1.65]	0.000305	W
All-scan vs. Safe	1.15 [1.13, 1.17]	0.000305	W	1.51 [1.38, 1.65]	0.000305	W
Safe vs. F-safe	1.16 [1.1, 1.23]	0.0115	W	1.08 [1.02, 1.14]	1	T
All-scan vs. F-all	1 [0.991, 1.01]	1	T	1 [0.993, 1.01]	1	T
F-all vs. F-safe	1.34 [1.28, 1.4]	0.000305	W	1.62 [1.51, 1.73]	0.000305	W
Scan vs. F-safe	1.16 [1.1, 1.23]	0.0115	W	1.08 [1.02, 1.14]	1	T
Scan vs. F-all	0.872 [0.855, 0.89]	0.000305	L	0.666 [0.611, 0.723]	0.000305	L
Safe vs. F-all	0.872 [0.854, 0.89]	0.000305	L	0.666 [0.611, 0.723]	0.000305	L
All-scan vs. F-safe	1.34 [1.27, 1.41]	0.000305	W	1.62 [1.51, 1.75]	0.000305	W
LSMOP6, $M = 3$						
Safe vs. Scan	0.973 [0.889, 1.08]	1	T	0.992 [0.885, 1.13]	1	T
All-scan vs. Scan	1.28 [1.18, 1.39]	0.000305	W	1.31 [1.2, 1.45]	0.000305	W
All-scan vs. Safe	1.31 [1.17, 1.48]	0.0171	W	1.32 [1.18, 1.48]	0.0391	W
Safe vs. F-safe	1.01 [1, 1.03]	1	T	1 [1, 1]	1	T
All-scan vs. F-all	1.06 [0.953, 1.18]	1	T	1.05 [0.951, 1.16]	1	T
F-all vs. F-safe	1.25 [1.16, 1.36]	0.00114	W	1.26 [1.16, 1.37]	0.00114	W
Scan vs. F-safe	1.04 [0.934, 1.14]	1	T	1.01 [0.883, 1.13]	1	T
Scan vs. F-all	0.831 [0.74, 0.917]	0.0521	T	0.802 [0.713, 0.89]	0.000305	L
Safe vs. F-all	0.809 [0.74, 0.885]	0.0322	L	0.796 [0.731, 0.862]	0.00114	L
All-scan vs. F-safe	1.33 [1.19, 1.49]	0.0143	W	1.32 [1.18, 1.49]	0.0391	W
GLSMOP1, $M = 2$						
Safe vs. Scan	1 [1, 1]	1	T	1 [1, 1]	1	T
All-scan vs. Scan	1 [1, 1]	0.147	T	1.01 [1, 1.01]	0.0143	W
All-scan vs. Safe	1 [1, 1]	0.147	T	1.01 [1, 1.01]	0.0143	W
Safe vs. F-safe	1 [1, 1]	1	T	1 [1, 1]	1	T
All-scan vs. F-all	1.01 [1, 1.01]	0.0115	W	1.01 [1.01, 1.01]	0.00152	W
F-all vs. F-safe	0.998 [0.995, 1]	1	T	0.996 [0.992, 1]	1	T
Scan vs. F-safe	1 [1, 1]	1	T	1 [1, 1]	1	T
Scan vs. F-all	1 [1, 1.01]	1	T	1 [0.999, 1.01]	1	T
Safe vs. F-all	1 [1, 1.01]	1	T	1 [0.999, 1.01]	1	T
All-scan vs. F-safe	1 [1, 1]	0.147	T	1.01 [1, 1.01]	0.0143	W
DTLZ2, $M = 2$						
Safe vs. Scan	1.05 [1, 1.15]	1	T	5.87 [1, 16.1]	1	T
All-scan vs. Scan	1.48 [1.32, 1.67]	0.000305	W	7.7 [1.24, 21]	0.000721	W
All-scan vs. Safe	1.41 [1.29, 1.54]	0.000305	W	1.31 [1.22, 1.41]	0.000721	W
Safe vs. F-safe	1 [1, 1]	1	T	1 [1, 1]	1	T
All-scan vs. F-all	1.27 [1.17, 1.37]	0.00214	W	1.2 [1.12, 1.28]	0.0171	W
F-all vs. F-safe	1.11 [1.03, 1.21]	1	T	1.1 [1.02, 1.18]	1	T
Scan vs. F-safe	0.955 [0.871, 1]	1	T	0.17 [0.062, 1]	1	T
Scan vs. F-all	0.857 [0.768, 0.951]	1	T	0.155 [0.0576, 0.956]	0.606	T
Safe vs. F-all	0.897 [0.827, 0.973]	1	T	0.911 [0.848, 0.977]	1	T
All-scan vs. F-safe	1.41 [1.29, 1.54]	0.000305	W	1.31 [1.22, 1.41]	0.000721	W

Continued on next page

Table 8 continued

	Archive ND			Native ND		
Ref. vs. comp.	Ratio [95% CI]	p_H	W/T/L	Ratio [95% CI]	p_H	W/T/L
DTLZ2, $M = 3$						
Safe vs. Scan	1 [1, 1]	1	T	1 [1, 1]	1	T
All-scan vs. Scan	1.11 [1.06, 1.15]	0.0449	W	1.13 [1.07, 1.19]	0.0391	W
All-scan vs. Safe	1.11 [1.06, 1.16]	0.0449	W	1.13 [1.07, 1.19]	0.0391	W
Safe vs. F-safe	1 [1, 1]	1	T	1 [1, 1]	1	T
All-scan vs. F-all	1.06 [1.03, 1.09]	0.0391	W	1.09 [1.05, 1.13]	0.0749	T
F-all vs. F-safe	1.04 [1, 1.09]	1	T	1.04 [1, 1.08]	1	T
Scan vs. F-safe	1 [1, 1]	1	T	1 [1, 1]	1	T
Scan vs. F-all	0.96 [0.922, 0.998]	1	T	0.963 [0.93, 0.999]	1	T
Safe vs. F-all	0.96 [0.922, 0.997]	1	T	0.963 [0.929, 0.999]	1	T
All-scan vs. F-safe	1.11 [1.06, 1.15]	0.0449	W	1.13 [1.07, 1.19]	0.0391	W
DTLZ7, $M = 3$						
Safe vs. Scan	1 [1, 1]	1	T	1 [1, 1]	1	T
All-scan vs. Scan	0.991 [0.943, 1.04]	1	T	1.01 [0.922, 1.12]	1	T
All-scan vs. Safe	0.991 [0.943, 1.04]	1	T	1.01 [0.922, 1.13]	1	T
Safe vs. F-safe	1 [1, 1]	1	T	1 [1, 1]	1	T
All-scan vs. F-all	0.988 [0.93, 1.05]	1	T	1.02 [0.904, 1.14]	1	T
F-all vs. F-safe	1 [0.944, 1.07]	1	T	0.995 [0.884, 1.11]	1	T
Scan vs. F-safe	1 [1, 1]	1	T	1 [1, 1]	1	T
Scan vs. F-all	0.996 [0.939, 1.06]	1	T	1 [0.904, 1.13]	1	T
Safe vs. F-all	0.996 [0.938, 1.06]	1	T	1 [0.903, 1.13]	1	T
All-scan vs. F-safe	0.991 [0.943, 1.04]	1	T	1.01 [0.922, 1.13]	1	T
UF8, $M = 3$						
Safe vs. Scan	0.983 [0.956, 1.01]	1	T	1.26 [1.2, 1.32]	0.000305	W
All-scan vs. Scan	1.14 [1.08, 1.19]	0.0143	W	1.67 [1.58, 1.77]	0.000305	W
All-scan vs. Safe	1.15 [1.11, 1.21]	0.000721	W	1.32 [1.23, 1.42]	0.000305	W
Safe vs. F-safe	0.966 [0.947, 0.983]	0.147	T	1.09 [1.04, 1.16]	1	T
All-scan vs. F-all	0.947 [0.903, 0.995]	1	T	0.96 [0.919, 1]	1	T
F-all vs. F-safe	1.18 [1.12, 1.24]	0.00402	W	1.51 [1.4, 1.61]	0.000511	W
Scan vs. F-safe	0.983 [0.964, 1]	1	T	0.867 [0.827, 0.903]	0.000305	L
Scan vs. F-all	0.834 [0.795, 0.876]	0.00152	L	0.575 [0.545, 0.606]	0.000305	L
Safe vs. F-all	0.82 [0.778, 0.866]	0.00114	L	0.725 [0.674, 0.782]	0.000511	L
All-scan vs. F-safe	1.12 [1.06, 1.17]	0.0449	W	1.45 [1.36, 1.54]	0.000305	W
WFG6, $M = 3$						
Safe vs. Scan	1 [1, 1]	1	T	1 [1, 1]	1	T
All-scan vs. Scan	1.14 [1.1, 1.19]	0.000511	W	1.14 [1.1, 1.19]	0.000721	W
All-scan vs. Safe	1.14 [1.1, 1.19]	0.000511	W	1.14 [1.1, 1.19]	0.000721	W
Safe vs. F-safe	0.998 [0.979, 1.02]	1	T	1 [0.974, 1.03]	1	T
All-scan vs. F-all	0.978 [0.949, 1.01]	1	T	1.01 [0.961, 1.05]	1	T
F-all vs. F-safe	1.17 [1.12, 1.21]	0.00114	W	1.14 [1.09, 1.19]	0.00114	W
Scan vs. F-safe	0.998 [0.979, 1.02]	1	T	1 [0.974, 1.03]	1	T
Scan vs. F-all	0.856 [0.823, 0.891]	0.000511	L	0.88 [0.849, 0.91]	0.000305	L
Safe vs. F-all	0.856 [0.823, 0.892]	0.000511	L	0.88 [0.849, 0.91]	0.000305	L
All-scan vs. F-safe	1.14 [1.1, 1.18]	0.000511	W	1.14 [1.11, 1.18]	0.000511	W

E.3 Nonrepeated Fixed Grids at Equal Cost

All grids use coarse positions $(0.1, 0.3, 0.5, 0.7, 0.9)$. Repeat uses fine positions $(0, 0.25, 0.5, 0.75, 1)$; UniqueL and UniqueR replace the second paid 0.5 with 0.4 or 0.6, respectively. Thus both preserve the original nine distinct positions and pay the same $10A$ FE for ten records. Each grid is crossed with dominance-safe single-block offering (Safe) and all-paid offering (All). The eight cases are LSMOP1/$M = 2$, LSMOP6/$M = 3$, GLSMOP1/$M = 2$, DTLZ2/$M = 2, 3$, DTLZ7/$M = 3$, UF8/$M = 3$, and WFG6/$M = 3$ (WFG parameter 4), all

Table 9: All seven contrasts across eight cases (W/T/L). Each cell sums to eight; all 112 tests share one Holm family.

Reference / comparator	Archive ND	Native ND
UniqueLSafe / RepeatSafe	0/8/0	0/8/0
UniqueLAll / RepeatAll	0/8/0	0/8/0
UniqueRSafe / RepeatSafe	0/8/0	0/8/0
UniqueRAll / RepeatAll	0/8/0	0/8/0
RepeatAll / RepeatSafe	3/5/0	2/6/0
UniqueLAll / UniqueLSafe	3/5/0	2/6/0
UniqueRAll / UniqueRSafe	3/5/0	3/5/0

at $B = 1000$. Initialization, screening, anchors, independent proposals and disabled downstream directional actions are unchanged. Each Safe/All pair has identical paid queries before injection.

Results. The new grids remove the exact paid scan duplicates in every run: Repeat has 108 distinct scan decisions out of 120 for $M = 2$, or 144 out of 160 for $M = 3$; both unique grids have 120 or 160 distinct decisions, respectively. This is a measured result for these cases, not an assumption that distinct scalar positions always imply distinct decisions. Every record remains charged; no refund or free replacement query is used.

None of the 64 unique-versus-repeat endpoint tests is supported after Holm correction (minimum adjusted $p = 0.352$). Repeat/unique-grid mean ratios range from 0.950 to 1.058. In contrast, All/Safe support remains on LSMOP1 and UF8 for both outputs under every grid, and on WFG6 for Archive ND under every grid; WFG6 Native ND is additionally supported under UniqueR. Table 9 summarizes 16 supported All advantages and no supported Safe advantages across the 48 All/Safe tests. Thus the tested retention benefits survive removal of the repeated midpoint, while the extra distinct position itself has no resolved benefit at this sample size and budget. This is not an equivalence test. The replacement jointly changes placement, spacing and multiplicity; it does not isolate a pure deduplication effect. Adaptive policies were not rerun, so no new paired claim against adaptation is made.

IGD$^+$ preserves all fixed-grid classifications; exact-sign sensitivity changes three tests in the complete 112-test family. Table 9 summarizes all seven contrasts; the ratio range and minimum adjusted p above specify the scale and unresolved status of the unique-versus-repeat comparisons. No equivalence margin was specified, so these null tests do not establish interchangeability.

E.4 Injection Normalization and Population State

Default and shared-normalization All-scan match exactly in query ledger, terminal decisions/objectives and final random state in all twenty runs: their injection pools already contain initialization and every paid block. This is a negative control, not additional independent evidence. Sharing normalization changes Safe's selection in sixteen runs (median 1.5, maximum five objects), and changes reference association for a median 10.5 offered candidates.

For one run, use policy names to denote default endpoint values and subscript s for their shared-normalization counterparts. Since All-scan$_s$ = All-scan,

$$\log(\text{All-scan}_s/\text{Safe}_s) - \log(\text{All-scan}/\text{Safe}) = \log(\text{Safe}/\text{Safe}_s).$$

Thus the interaction measures Safe's response to injection scaling. Exponentiating its mean paired log change gives a geometric effect, not a ratio of arithmetic means. Table 11 reports the six pairwise contrasts and interactions; unresolved interactions do not quantify mediation or exclude normalization effects during later search.

The immediate nondominated decision sets match in all twenty Safe/All-scan pairs under either normalization. Complete populations differ by 37–49 members under default scaling and 37–47 under shared scaling (median 44 for both). Each condition retains a mean 16.05 nondominated and 83.95 dominated population members. These injection counts and the dominated-member g summaries in Table 2 are descriptive; the source audit below adds provenance, and Appendix E.5 reports separate state interventions.

Table 10: Post-injection provenance on DTLZ2 ($M = 3$). Counts are means over 20 runs per policy and normalization. The entries hold separately under both pool-based and shared normalization; the reused runs are not additional independent samples.

Source and status	Safe	All-scan
Initial, nondominated	0.05	0.05
Initial, dominated	83.95	40.50
Probe, nondominated	16.00	16.00
Probe, dominated	0.00	43.45
Total	100.00	100.00

Table 11: Normalization effects under one fourteen-test Holm family. Pair entries give comparator/reference mean IGD, pointwise 95% CI, adjusted p_H, and W/T/L. Interactions exponentiate paired log effects.

	Archive ND			Native ND		
Ref. vs. comp.	Ratio [95% CI]	p_H	W/T/L	Ratio [95% CI]	p_H	W/T/L
All-scan vs. Safe	1.11 [1.06,1.16]	0.00395	W	1.13 [1.07,1.19]	0.00395	W
All-scan$_s$ vs. Safe$_s$	1.12 [1.08,1.16]	0.000435	W	1.16 [1.11,1.20]	0.000187	W
Safe$_s$ vs. Safe	0.987 [0.947,1.02]	1	T	0.98 [0.945,1.01]	1	T
All-scan$_s$ vs. All-scan	1 [1,1]	1	T	1 [1,1]	1	T
All-scan vs. Safe$_s$	1.12 [1.08,1.16]	0.000435	W	1.16 [1.11,1.20]	0.000187	W
All-scan$_s$ vs. Safe	1.11 [1.06,1.15]	0.00395	W	1.13 [1.07,1.19]	0.00395	W
Interaction	0.987 [0.949,1.022]	1	T	0.979 [0.945,1.011]	1	T

Provenance of population members. Stored post-injection decisions are matched exactly to initial/probe decisions and checked against selected input indices. All 80 populations are fully accounted for, with no member matching both sources. Table 10 gives the complete decomposition: every Safe population has 84 initial and 16 probe members, whereas All-scan's additional probe members are dominated.

The mean within-run g among All-scan's dominated initial members is 82.53, versus 4.62 among its dominated probe members. This identifies a composition difference behind the aggregate 42.22, while Safe's dominated members all originate in initialization. Under shared normalization, every Safe-only member is initial and every All-scan-only member is a probe candidate. Under default normalization, All-scan also retains a mean 0.45 initial members absent from Safe; the median 44 differing members should therefore not be interpreted as exactly 44 replacements of one source by the other in every run. These are descriptive source matches, not an intervention holding order and all other population properties fixed.

For an independently uniform, unselected initialization point,

$$\mathbb{E}[(X_j - 0.5)^2] = \int_0^1 (x - 0.5)^2 \, dx = \frac{1}{12}, \qquad \mathbb{E}[g] = \frac{998}{12} \simeq 83.17.$$

This is an unconditional scale reference for Safe's observed mean 82.80, not a prediction after selection. The median 44 differing members counts between-population differences, not empty slots. Provenance follows from the direct matches above; neither source counts nor lower g alone identify a cause of terminal improvement.

E.5 Population- and Optimizer-State Follow-up

To probe the earlier Safe/All population differences without attributing them to one factor, this follow-up uses DTLZ2/$M = 3$, $D = 1000$, $N = 100$, two budgets $B \in \{1000, 5000\}$ and twenty paired runs. Each run shares 100 initial and 160 scan evaluations. Safe and All use paid-pool

Table 12: State branches. Every population branch pays the same 100-FE refresh; velocity branches intervene later at FE 560.

Branch	Assignment
A00, A10, A01, A11	Safe base with Q0/C0, Q1/C0, Q0/C1, Q1/C1.
AllBridge	Canonically ordered All population from the same paid pool.
OrderSlot	Whole-row permutation after refresh; future random draws remain slot-indexed.
OrderTransport	Same permutation with transported computations, draws and ties; exact A00 replay control.
VKeep	Unchanged later state; exact A00 replay control.
VShuffle	Permute dominated-member velocities, preserving their multiset.
VZero	Zero dominated-member velocities; removal of momentum, not an equal-magnitude control.

injection normalization and must have the same exact nondominated decision set S. The Safe base orders S before dominated members, each by first query index; velocities are zero and elite/history states empty. Reference vectors, initial ideal/scale, FE and generation match; downstream scan-conditioned actions are disabled and the common protection rule is unchanged. Every population branch then pays the same 100-FE refresh and continues from FE 360 with independent proposals. Its archive contains only the common prefix and its own refresh/later queries. The refresh and standardized order distinguish this protocol from the earlier Safe/All trajectories.

Interventions and checks. For a dominated member, write $F(x) = (1 + g(x))u(x_1, x_2)$ and $g(x) = \sum_{j=3}^{1000}(x_j - 0.5)^2$. Q1 halves the excess in g above a domination threshold:

$$g_{\min} = \max\{0, \min_{s\in S} \max_i F_i(s)/u_i(x) - 1\}, \qquad g_{\text{new}} = g_{\min} + \tfrac{1}{2}(g - g_{\min}).$$

It rescales $x_{3:D} - 0.5$ by $\sqrt{g_{\text{new}}/g}$, keeping angular coordinates and tail direction fixed. Rows without a finite domination witness or with $g \le g_{\min} + 10^{-10}\max(1, g)$ remain unchanged and inactive. C1 applies a nonzero cyclic shift to each dominated member's tail, preserving its objectives while changing decisions. The shift is fixed separately for each run and ancestral member identity and is shared across the Q levels and budgets; angular coordinates remain fixed. Q0/C0 leave the corresponding factor unchanged. Q and C commute, but neither holds full decision geometry fixed. Bounds and the exact set S are checked after fresh evaluation; analytic/evaluated objectives agree to relative tolerance 10^{-12}. Evaluated objectives are retained without overwriting them to impose equality. Each changed dominated member must still have a strict domination witness in S. These interventions are not an isolated abstract quality effect or a reconstruction of the original provenance change.

Row ordering uses an independent permutation, replacing the identity by a one-step cyclic shift. Refresh queries stay in A00 logical order and the physical permutation follows the refresh, so archive query-order ties are not an extra intervention. Whole-row state travels with the member; OrderTransport additionally applies the inverse mapping to computations, draws, ties and output accounting. Future pairing, coefficients, mutation, retention and group choices use a shared pre-sampled random plan.

Velocity branches start after A00's second completed selection, at FE 560, after two 100-FE generations. They deep-copy decisions, objectives, constraints, order, reference vectors, normalization, elite objects and their velocities, ages, history, FE, generation and future random plans. VShuffle permutes only currently dominated members' velocity rows, with an independent permutation and the same identity-to-cyclic fallback; nondominated velocities remain fixed. VZero removes only dominated-member momentum. These branches add no FE; inactive assignments are retained and the checkpoint is never delayed to obtain a nonzero effect. All 80 deterministic controls replay A00 exactly, and all 120 velocity branches match before intervention. Q/C changes 82–84 dominated members per run; Q1 reduces mean dominated-member g from 82.78 to 41.54, versus 42.42 for AllBridge. Every velocity assignment changes at least one row. No inactive assignment is dropped. The contraction changes mean pairwise distance from 11.55749 to 8.20527; C1 at Q0 changes it to 11.55849, so even near-equal average distance does not establish identical geometry.



Figure 8: Population- and optimizer-state contrasts on DTLZ2 with three objectives. Cells show numerator/denominator mean IGD ratios, denominator-oriented W/T/L, and pointwise paired 95% bootstrap intervals. Holm adjustment retains the original 40-test family, including the interaction reported in Table 13. Each row gives the numerical numerator/denominator.

Table 13: Q×C interaction, $(I_{01} - I_{11}) - (I_{00} - I_{10})$. Intervals are pointwise paired 95% intervals; p_{H} uses the entire 40-test primary family. Positive values mean larger radial benefit at C1.

B	Output	Mean interaction	95% interval	p_{H}
1000	Archive ND	2.11×10^{-5}	$[-0.00345, 0.00377]$	1
1000	Native ND	0.0011	$[-0.00289, 0.00548]$	1
5000	Archive ND	0.00401	$[-0.00632, 0.015]$	1
5000	Native ND	-0.000511	$[-0.103, 0.1]$	1

Inference and results. The two conditional Q contrasts, two conditional C contrasts and Q×C interaction, A00 versus OrderSlot/AllBridge, and three velocity pairs give ten contrasts. Two budgets and both primary IGD outputs form the original 40-test Holm family; replay controls are not hypothesis tests. Outputs use exactly B FE, one DTLZ2 reference front and native units. Figure 8 and Table 13 retain all primary contrasts.

Only three primary contrasts pass correction, all at 1000 FE. A00/AllBridge ratios are 1.180 [1.143,1.213] for Archive ND and 1.161 [1.118,1.204] for Native ND ($p_H = 7.63 \times 10^{-5}$ and 2.23×10^{-4}). VKeep/VZero favors zeroing on Archive ND: 1.080 [1.047,1.115], $p_H = 0.00638$; its native test is unresolved ($p_H = 0.946$). Radial, composition, interaction, slot-order and velocity-permutation contrasts remain unresolved, as do all twenty 5000-FE tests. Several native intervals are wide, and this family includes no budget-interaction test.

IGD exact-sign sensitivity supports both bridge contrasts but not velocity zeroing ($p_H = 0.0953$). IGD$^+$ signed ranks support only the 1000-FE Archive ND bridge ($p_H = 0.00191$); no IGD$^+$ exact-sign contrast passes correction. Velocity zeroing is therefore a primary-IGD finding with limited sensitivity support, and bridge evidence depends on endpoint and test. AllBridge supports a short-budget composite state effect, not a decomposition of the earlier Safe/All gain: matched mean g does not match membership or geometry, and unresolved factors are not equivalent.

F Component Controls and External Comparisons

These studies assess component sensitivity, screening failures and external performance under their own methods, runs and statistical families. They are separate from the 5960 main-study runs and are not pooled with them.

F.1 Component-Control Design and Protocol

Full is the matched-action component control, not a policy in the main four-policy independent-proposal study. Table 14 defines its comparators. The ordinary correlation screen ranks $\max_k |\operatorname{corr}(y_j, \phi_k)|$ on all initial rows, retains $M - 1$ coordinates and breaks ties by index. The wide control uses $\{0, 0.25, 0.5, 0.75, 1\}$ as its second table and selects the first best block over both paid tables; it is distinct from dominance-safe Fixed-safe. The screen and score otherwise follow the shared-backbone definitions. The relative score gain is

$$\gamma = \operatorname{clip}_{[0,1]}\left(\frac{R(0.5) - R(q^\star)}{\max\{|R(0.5)|, \epsilon\}}\right) \mathbf{1}\{R(q^\star) < R(0.5)\}. \tag{30}$$

This is neither calibrated confidence nor a condition for block injection; clipping can tie distinct objective improvements.

Matched actions and injection-only controls. After the paid block competes with initialization, Full forms density-based winner–loser pairs. One competitive update and polynomial mutation produce a common normalized base y_i^0 and velocity. A group is eligible if nonempty with a finite scan target and positive gain γ. Among eligible scanned groups, select the largest γ times the absolute winner–base mean gap; pre-sampled random priorities break ties. All unselected coordinates stay at the common base. If no group is eligible, both candidates equal that base.

For the selected group let Δ_i be its winner mean minus base mean, $s_i^{\mathrm{dir}} = \operatorname{sign}(\Delta_i)$ and $s_i^{\mathrm{rnd}} \in \{-1, +1\}$ equiprobably. With $p_t = \operatorname{clip}_{[0,1]}(F_t/F_{\max})$ before evaluation,

$$\begin{aligned}
\xi_i &= 0.5 + U_i, \quad U_i \sim \mathcal{U}[0, 1), \\
a_i &= \min\{|\Delta_i|, [0.055(1 - p_t) + 0.010p_t]\xi_i\gamma\}, \\
m_{ij} &= \max\{0, \min(a_i, \rho(y_{ij}^0, s_i^{\mathrm{dir}}), \rho(y_{ij}^0, s_i^{\mathrm{rnd}}))\}, \\
y_{ij}^{\mathrm{dir}} &= y_{ij}^0 + s_i^{\mathrm{dir}} m_{ij}, \qquad y_{ij}^{\mathrm{rnd}} = y_{ij}^0 + s_i^{\mathrm{rnd}} m_{ij}.
\end{aligned}$$

Here $\rho(y, +1) = 1 - y$, $\rho(y, -1) = y$, and $\rho(y, 0) = 0$. Signs, jitter and priorities are sampled for every eligible group before selection, avoiding conditional resampling. Both moves stay in the box and share the same absolute displacement and velocity metadata; they need not be distinct or opposite. Denormalization multiplies their magnitudes by the same coordinate range. Both candidates are evaluated, including coincidences, before the common parent-conditioned comparator, selection and guarded archive. The directed sign points toward the current winner, not the scanned scalar $q^\star$; its local comparison gives no terminal improvement guarantee.

Scan+ind. preserves the scan and its population injection but replaces matched actions by two independent update/mutation draws with their own velocities. Its No-scan partner uses the same proposal rule and returns the saved probe FE to evolution. Full/Scan+ind. therefore changes proposal dependence and repeated evaluations as well as scan-conditioned actions; it is not a pure directional-sign test. Diagnostic screening computations remain in the simple controls to match random streams, so no runtime saving is claimed.

Matrix, pairing and inference. Original and transformed LSMOP each use 18 configurations, six modes and twenty paired runs (2160 runs each); GLSMOP uses 18 configurations, two modes and twenty runs (720). All 5040 runs finish at exactly 1000 FE with $D = 1000$, $N = 100$ and finite records. The 1800 matched scan/injection boundaries agree in paid queries, scores, selected populations and random state. Outputs and transformations follow the broad-study protocol; repeated evaluations remain separately charged and recorded. Online selection uses the conditional stable deduplication rule; retrospective output reconstruction separately deduplicates the paid history. Independent draws need not yield distinct decisions and are not resampled on collision.

Table 14: Fixed controls. All probing controls in this table inject the selected paid block. Independent proposals share parents and selection, but use separate update and mutation draws.

Name	Routed screen	Second grid	Evolution proposals
Full	Split-stable	Adaptive	Matched actions
Corr	Full-sample correlation	Adaptive	Matched actions
Corr+wide	Full-sample correlation	Fixed wide	Matched actions
Scan+ind.	Split-stable	Adaptive	Independent
Corr+wide+ind.	Full-sample correlation	Fixed wide	Independent
No-scan	No applied scan	None	Independent

Table 15: Complete component-study W/T/L summaries under the primary asymptotic signed-rank convention (Asym.) and exploratory conditional exact sensitivity (Exact). Each group's family jointly covers both outputs.

			Archive ND		Native ND	
Group	Reference	Comparator	Asym.	Exact	Asym.	Exact
Original	Full	No-scan	17/1/0	17/1/0	16/1/1	17/0/1
Original	Full	Scan+ind.	2/16/0	2/16/0	0/18/0	0/18/0
Original	Full	Corr	0/18/0	0/18/0	0/18/0	0/18/0
Original	Full	Corr+wide	2/14/2	5/11/2	1/16/1	2/15/1
Original	Full	Corr+wide+ind.	3/14/1	3/13/2	0/17/1	0/17/1
Original	Scan+ind.	No-scan	16/2/0	17/1/0	15/2/1	16/1/1
Original	Corr+wide	Corr+wide+ind.	2/15/1	2/13/3	0/18/0	0/17/1
Original	Corr+wide+ind.	No-scan	17/1/0	17/1/0	16/2/0	16/1/1
Transformed	Full	No-scan	17/1/0	17/1/0	15/2/1	15/2/1
Transformed	Full	Scan+ind.	0/18/0	0/18/0	0/18/0	0/17/1
Transformed	Full	Corr	0/16/2	0/14/4	0/16/2	0/14/4
Transformed	Full	Corr+wide	2/14/2	2/12/4	2/13/3	2/11/5
Transformed	Full	Corr+wide+ind.	2/13/3	2/12/4	2/12/4	2/11/5
Transformed	Scan+ind.	No-scan	17/1/0	17/1/0	15/1/2	15/0/3
Transformed	Corr+wide	Corr+wide+ind.	0/18/0	0/16/2	0/18/0	0/17/1
Transformed	Corr+wide+ind.	No-scan	17/1/0	17/1/0	15/2/1	15/0/3
GLSMOP	Full	Scan+ind.	2/14/2	2/13/3	0/18/0	0/16/2

For original/transformed LSMOP, the eight contrasts in Table 15 and both outputs form separate 288-test families; GLSMOP Full/Scan+ind. uses a 36-test family. The primary test is the two-sided continuity-corrected normal approximation to signed ranks, excluding exact zeros and averaging tied absolute ranks; all-zero differences give $p = 1$. For each nonzero paired difference d_i, defined as comparator minus reference IGD, let r_i be the rank of $|d_i|$ among the nonzero differences, using average ranks for ties. Define $W_+ = \sum_{i:d_i>0} r_i$ and $T = \sum_i r_i$. Mean-IGD ratios, pointwise paired bootstrap intervals, and W/T/L follow the reporting conventions in Section 4.1. The variance is $\sigma^2 = \frac{1}{4}\sum_i r_i^2$ and the two-sided p-value is $2\Phi(-\max\{0, |W_+ - T/2| - 0.5\}/\sigma)$, including the tied-rank variance adjustment. Conditional exact signed ranks and exact sign tests retain the same complete families as sensitivities. Secondary all-object archives retain separate 144/144/18-test families. The summary table includes every prespecified contrast and both outputs; the following text reports the principal effect magnitudes and failure cases.

F.2 Component-Control Results and Failure Cases

Table 15 retains all contrast/output summaries. Conditional exact sensitivity changes 32 of the 612 primary decisions: seven unresolved outcomes become wins and 25 become losses, with no win/loss reversals. It does not replace the original primary analysis.

Injection and matched actions. On original LSMOP, Scan+ind./No-scan gives 16/2/0 on Archive ND and 15/2/1 on Native ND, with median No-scan/Scan+ind. mean-IGD ratios 5.45 and 4.45. The simpler Corr+wide+ind. also improves frequently. By contrast, Full/Scan+ind. gives 2/16/0 and

0/18/0; transformed-LSMOP primary tests are unresolved throughout. Injection can therefore repay its opportunity cost without establishing a general advantage for matched actions. Unresolved tests do not establish interchangeability. Injection also has failures: LSMOP6/$M = 2$ favors No-scan on Native ND, with ratio 0.5137 [0.5135,0.5140], and transformed native failures include its power and objective-affine views.

Screening failure and its follow-up. Full/Corr is unresolved on all original-LSMOP configurations; routed sets and each endpoint agree exactly in 346/360 pairs. On transformed LSMOP, Full/Corr gives 0/16/2 on each primary output, with losses on objective-affine LSMOP1 and LSMOP9 at $M = 3$. For objective-affine LSMOP6/$M = 3$, Corr/Full mean ratios are 0.000517 [0.000386,0.000746] and 0.00062 [0.000467,0.000898]. Five pairs are equal and all fifteen nonzero differences favor Corr. The primary asymptotic test remains unresolved ($p_H = 0.116$), whereas exact-rank and sign sensitivities favor Corr ($p_H = 0.00995$ and 0.0104). Large mean differences and pointwise intervals must not be substituted for the stated primary test.

In those fifteen pairs, Full's fallback retains 100 coordinates and Corr retains two; the other five use the same two-coordinate set and have equal endpoints. Both screens retain the true positions in every run, so this is an association with additional retained coordinates, not position loss. The frozen-role follow-up uses new paired objective views and frozen routes to test the route and its downstream consequences.

GLSMOP and inactive actions. Full/Scan+ind. gives 2/14/2 on Archive ND and 0/18/0 on Native ND, with median ratios 0.985 and 0.993. Full's score gain is zero in 359/360 runs; the exception is GLSMOP3/$M = 2$ with $q^\star = 0.45$, $\gamma = 0.00235$. For GLSMOP1, all 34/40 runs with exactly nominal roles have zero gain, consistent with the conditional GLSMOP1 analysis; the other six also have zero gain empirically, outside that proof's premises. Full has 136612/136800 coincident evolutionary pairs in GLSMOP, but 369–578 distinct evolutionary decisions per run, all absent from its paid prefix. This is within-pair redundancy, not a stationary decision sequence or evidence that scan injection is inactive. Independent proposals have only 411/1008000 coincident pairs, all charged and retained. These counts distinguish coincident action pairs from a stationary search trajectory.

F.3 Paired Screening and Frozen Roles

For LSMOP6/$M = 3$, each run supplies one initialization X and its canonical objectives Y, costing 100 FE. The affine view $Y' = Y \odot (0.1, 1, 10) + (1, 2, 3)$ is derived arithmetically without more objective calls. Full screens share the explicit split and null-permutation plan constructed from canonical radial quality. The three frozen routes are R_I (canonical Full), R_A (affine Full), and R_C (ordinary correlation on the affine features). Known position labels are used only in the audit.

Each run then supplies six separately charged 1000-FE trajectories: two objective views crossed with the three assigned routes, all starting from X. At the single pre-scan intervention point, only the routed-coordinate index is replaced. Full's matched-action construction, scoring, two-round 160-FE scan, retention and directional-action formulas remain in use. Role-dependent grouping, probe candidates, gains and actions are descendants of the assigned route. A common complete random plan matches pairing, velocity coefficients, mutation, retention and group choices across branches, including inactive choices. This controls random-number consumption without claiming to replay the earlier uncontrolled trajectories.

The three role pairs, two objective views and two outputs give twelve endpoint tests; six affine-minus-canonical interactions complete one 18-test Holm family. Both views use reference-range normalized IGD.

Screening results. Normalized radial quality agrees within 2.23×10^{-16}, with no changes in the naturally generated split ordering. Nevertheless, under the shared explicit plan, the affine Full route changes in 16/20 runs. Canonical Full passes its support gate in all twenty runs; affine Full passes in four. Both R_I and R_C are the known position set $\{1, 2\}$ in all twenty cases. In the sixteen changed cases, R_A contains those two coordinates and 98 additional coordinates (Jaccard similarity 0.02); in four it has the same two-member set. The change therefore does not arise from a different initialization or split plan, and it does not require losing the true position coordinates.

Table 16: Comparator/reference mean normalized IGD ratios with pointwise 95% paired bootstrap intervals.

View	Ratio	Archive ND	Native ND
Canonical	R_A/R_I	$2114.467\ [1587.961, 2629.114]^*$	$1916.480\ [1400.606, 2422.082]^*$
Canonical	R_A/R_C	$2114.467\ [1578.848, 2646.252]^*$	$1916.480\ [1399.541, 2414.043]^*$
Canonical	R_I/R_C	$1.000\ [1.000, 1.000]$	$1.000\ [1.000, 1.000]$
Affine	R_A/R_I	$2156.347\ [1624.573, 2700.253]^*$	$1950.130\ [1431.942, 2459.825]^*$
Affine	R_A/R_C	$2156.347\ [1621.486, 2709.269]^*$	$1950.130\ [1423.810, 2468.144]^*$
Affine	R_I/R_C	$1.000\ [1.000, 1.000]$	$1.000\ [1.000, 1.000]$

Ratios above one favor the denominator role set. * denotes Holm $p < 0.05$ in the complete 18-test family, including the interactions. All twenty paired runs are retained.

Table 17: Affine-minus-canonical changes in the paired normalized-IGD role contrasts, with pointwise 95% intervals.

Within-view contrast	Archive ND change	Native ND change
$R_A - R_I$	$64\ [-0.00722, 192]$	$57\ [-0.00722, 171]$
$R_A - R_C$	$64\ [-0.00723, 192]$	$57\ [-0.0072, 171]$
$R_I - R_C$	$0\ [0, 0]$	$0\ [0, 0]$

Entries are mean differences of paired differences, not ratios. All six tests are unresolved in the joint 18-test Holm family.

Downstream results. The R_I and R_C interventions yield exactly identical trajectories within each objective view, including the run where their input order differs. All 48 repeated-set trajectory comparisons pass. The four runs with unchanged R_A also remain in every test as exact-zero role contrasts. Under canonical objectives, mean normalized Archive/Native ND IGD is 1.370/1.515 for R_I and 2895.831/2902.931 for R_A; under affine objectives it is 1.373/1.518 versus 2959.873/2959.947. These absolute means contextualize the large ratios in Table 16, so the effect is not represented only as a relative number. All eight R_A versus R_I or R_C endpoint tests favor the two-coordinate route (joint-family Holm $p = 0.000549$). The four R_I/R_C tests have all-zero differences and $p = 1$.

The six affine-minus-canonical interaction tests are all unresolved, with adjusted $p = 1$, as reported in Table 17. Fixed-role outputs need not be identical across objective views, and some interaction mean differences have wide intervals. Thus the experiment identifies a harmful whole-route intervention under both views; it does not establish general scale invariance or isolate the effect of adding 98 coordinates separately from their identities. The canonical route also uses a known transformation for diagnosis and is not presented as a deployable screening repair.

Sensitivity and scope. IGD^+ and exact-sign sensitivity preserve all frozen-role classifications. Tables 16 and 17 give the primary endpoint ratios and interaction intervals. The findings remain specific to this backbone, dimension, case and scale intervention; they add no evidence about new dimensions or applications.

F.4 External Methods and Resource-Honest Outputs

The external implementations are WOF, LSMOF and LMOCSO, taken from the stored PlatEMO source tree (Tian et al., 2017). They cover problem transformation, low-dimensional reformulation and competitive-swarm search. Internal references are matched-action Full and independent-proposal All-scan. These selected-case comparisons do not identify external methods' internal mechanisms. LMOEA-LGCS is excluded because its implementation and training/resource protocol remain unaudited.

Frozen implementations. WOF uses the stored defaults

$$(\gamma, \text{groups}, \psi, t_1, t_2) = (4, 2, 3, 1000, 500),$$
$$(q, \delta, \text{optimiser}, \text{randomOptimisers}) = (M + 1, 0.5, 1, 1),$$

Table 18: Resource coverage over all five cases and twenty paired runs per budget. Ranges are observed minima–maxima; "core by B" counts completed first weight stages (WOF/LSMOF) or swarm generations (LMOCSO). Each method–budget cell contains 100 runs.

Method	B	Initial N	Actual FE	First core end	Core by B
Full / All-scan	all	100	B	–	–
WOF	1000	91–100	1092–1171	–	0/100
WOF	5000	91–100	5024–7382	3044–7382	92/100
WOF	10000	91–100	10002–11051	3044–7382	100/100
LSMOF	1000	100	1300	1300	0/100
LSMOF	5000	100	5000	1300	100/100
LSMOF	10000	100	10000	1300	100/100
LMOCSO	1000	91–100	1000–1062	102–145	100/100
LMOCSO	5000	91–100	5000–5078	102–145	100/100
LMOCSO	10000	91–100	10000–10081	102–145	100/100

with requested transformed population size 10 and reference-solution selection method 3. The mixed optimizer option is enabled; this is labeled *PlatEMO-WOF, default mixed kernels*, rather than a fixed-inner-optimizer variant. LSMOF uses the executable defaults $(w_D, \mathrm{SubN}, \mathrm{Operator}) = (10, 30, 2)$; its header's operator default differs from the actual parameter assignment. LMOCSO has no extra main parameters. This evaluates stored implementations, not all experiments in the original publications.

The five cases are LSMOP1/$M = 2$, GLSMOP1/$M = 2$, reflected LSMOP6/$M = 2$, GLSMOP6/$M = 2$, and DTLZ2/$M = 3$. Budgets are $B \in \{1000, 5000, 10000\}$. Budget-dependent algorithms are run separately at each B. Within each case and run, methods and budgets share an initial decision bank and pay for the prefix required by their actual initial size. WOF and LMOCSO reference-vector construction can change the requested size; actual initial and population sizes are recorded, not forced to 100. Their three-objective initial size was 91 in diagnostics and every formal run. Shared runs provide pairing blocks, not identical downstream random calls.

Why the budget boundary matters. For the stored LSMOF implementation,

$$G = \left\lceil \frac{0.05B}{\mathrm{SubN}\, 2w_D} \right\rceil = 1 \qquad (B = 1000, 5000, 10000).$$

One weight-optimization call evaluates $30 \times 20 = 600$ original decisions at initialization and another 600 during this iteration. Thus, its first complete weight stage alone costs 1200 FE, before counting original initialization. WOF performs a $t_1 = 1000$ ordinary-search block before its first weight stage, and the subsequent termination check can prevent entry into that stage at the smallest budget. Stage logs confirmed these boundaries; Table 18 reports actual first-stage completion and coverage for all formal runs. An initialization-only or ordinary-search-only prefix cannot establish the effectiveness of a framework's full reconstruction mechanism.

Every original objective evaluation is charged, including transformed subproblem calls, back-transformed solutions, population filling and repeated decisions. Parameters and batches are not shortened to fit 1000 FE. The primary Archive ND endpoint uses exactly the first B paid queries, even when that boundary falls within an unfinished batch. Native ND uses the population returned by the unchanged termination policy and its actual FE; it is not a strictly capped native comparison. Execution times include instrumentation and metric preparation and do not support a wall-time ranking.

All 1500 runs completed without failure or resource censoring. Comparing all three external methods with both internal references gives $3 \times 2 \times 5 \times 3 = 90$ primary Archive ND IGD contrasts. Reflected LSMOP uses reference-range normalization; the other cases use original objective units. No additional native significance family or all-method ranking is introduced.

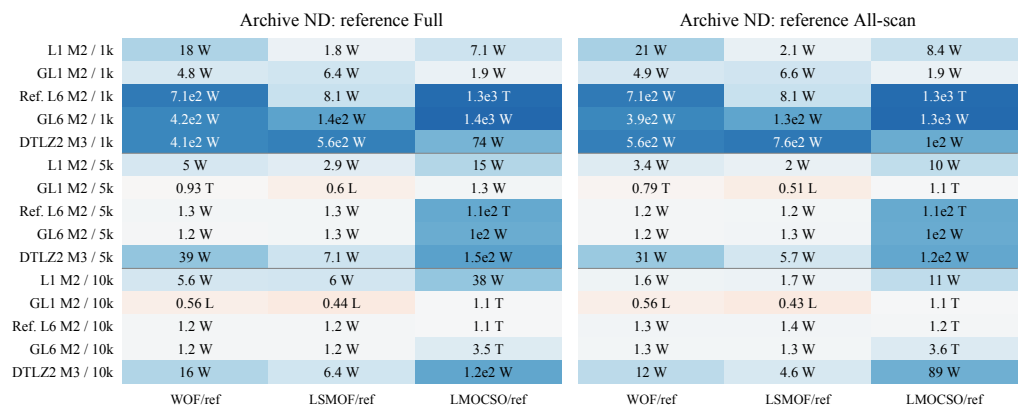

Figure 9: External-method comparisons for Archive ND at three query-prefix budgets. Cells show external/internal mean IGD ratios and internal-reference-oriented W/T/L under the original joint 90-test Holm family. The two panels use Full and All-scan as internal references. Resource and stage-completion boundaries remain as reported in Table 18.

F.4.1 Results: Coverage, Advantages, and Failure Cases

No WOF run entered a weight cycle by 1000 FE; LSMOF's first complete weight stage ended at 1300 FE. At 5000 FE, only 92/100 WOF runs completed their first weight cycle; at 10000 FE, all external runs completed a core stage. Table 18 therefore supports query-prefix comparisons, not uniform tests of fully executed mechanisms; native results and actual sizes remain descriptive.

Figure 9 displays all ninety primary Archive ND IGD comparisons. Seventy-one passed the joint Holm correction in favor of the internal reference, six favored the external method, and thirteen were unresolved. For Full the counts were 36/6/3, and for All-scan they were 35/7/3. These are counts of selected case–budget contrasts, not ninety independent benchmark samples or an average of distances across cases.

All six external advantages occurred on GLSMOP1. Against Full, LSMOF's mean-IGD ratios were 0.595 at 5000 FE (pointwise paired 95% interval [0.564, 0.630]) and 0.436 at 10000 FE ([0.430, 0.443]); both had $p_{\mathrm{H}} = 1.72\times10^{-4}$. WOF's 10000-FE ratio was 0.562 ([0.484, 0.678], $p_{\mathrm{H}} = 7.61\times 10^{-4}$). The corresponding ratios against All-scan were 0.508, 0.433 and 0.558. Thus, the strong internal results on the other selected cases do not establish that either internal method uniformly outperforms the external methods, particularly as the declared budget increases.

Arithmetic-mean ratios also need the paired uncertainty checks. For example, at 5000 FE on reflected LSMOP6, LMOCSO had normalized IGD mean 67.6, median 0.772 and range 0.451–989.7, compared with Full's mean 0.597. The resulting mean ratio of 113 was nevertheless unresolved ($p_{\mathrm{H}} = 0.538$). Large upper tails must not be read as consistent paired advantages solely from the heatmap color.

Exact-sign sensitivity supports all seventy-seven significant IGD contrasts in their original directions. IGD$^+$ signed-rank and exact-sign families each give 79 internal advantages, six external advantages and five unresolved tests. They preserve primary significant directions and resolve eight additional LMOCSO comparisons: reflected LSMOP6 at all budgets and GLSMOP6 at 10000 FE, against both references. These sensitivity findings do not replace the thirteen unresolved primary tests. Scope remains the stored implementations, five cases and $D = 1000$; LGCS, other dimensions and backbones are unevaluated.