A CONTRADICTION-AWARE SURVEY FRAMEWORK FOR MULTI-OBJECTIVE DECISION SUPPORT

Anonymous authors
Paper under review

ABSTRACT

Multi-objective decision support now spans classical scalarization, Pareto-evolutionary search, outranking-based multicriteria ranking, and uncertainty-aware formulations, but practical method selection remains unstable because evidence is fragmented across incompatible metrics, reporting protocols, and software defaults. This manuscript presents a contradiction-aware survey framework that unifies these families through a shared mathematical problem setting, explicit assumption gates, and a dual-layer recommendation rule separating candidate-set quality from stakeholder-facing utility. We formalize three linked contributions: an affine-normalization invariance and stability-guard criterion for recommendation robustness, a scalarization theorem gate that combines convex-regime completeness with a constructive non-convex counterexample boundary, and a descriptor-aware decision rule that integrates uncertainty and provenance penalties into utility judgments. We then assemble and analyze a synthesis validation package that reports calibrated instability thresholds, theorem-gate outcomes, transportability behavior under descriptor perturbations, and cross-framework lineage sensitivity. The resulting evidence supports bounded recommendations rather than universal rankings: two claims are strongly supported, while transportability claims remain conditional under implementation drift. The manuscript closes with a reproducibility and limitations synthesis that converts negative evidence into actionable follow-up protocols for future survey and benchmarking work.

1 INTRODUCTION

Decision support under multiple conflicting objectives has matured into a broad methodological ecosystem, yet selection practice is still dominated by ad hoc heuristics that are rarely explicit about assumptions, failure modes, or evidence boundaries. Foundational mathematical treatments formalize multicriteria optimization as vector-valued optimization over feasible sets, with Pareto dominance as the baseline partial order, and then layer scalarization, compromise, interactive, and ranking logic on top (Pedrycz et al., 2010; Zopounidis & Doumpos, 2017; Roijers & Whiteson, 2017; Monarchi et al., 1976; Benayoun et al., 1971). Contemporary studies expand this core into many-objective evolutionary search, robust and stochastic uncertainty models, and software-centered benchmark workflows (Bechikh et al., 2016; Azzouz et al., 2016; Deb et al., 2002; Zhang & Li, 2007; Saadati & Oveisiha, 2023; Zhao et al., 2023; Ahrari et al., 2026; Yu et al., 2020; maintainers, 2026a;b;c;d). The field has therefore become rich in tools but comparatively weak in stable guidance: the same family-level claim can reverse when indicator bundles, normalization policies, or framework defaults change.

This instability is not a minor reporting nuisance. In engineering and policy contexts, method recommendations are often used as justifications for downstream decisions with economic, operational, or safety implications. If ranking conclusions shift under routine choices such as reference points, criterion scaling, or software defaults, then the survey layer itself can become a source of error. Recent evidence from many-objective and multicriteria settings already documents this mechanism in multiple forms: indicator dependence, normalization sensitivity, and reproducibility drift (Zitzler & Knzli, 2004; Xu & Yuan, 2024; Goswami, 2020; Anes & Abreu, 2025; Zhou et al., 2021; Picard & Schiffmann, 2021; Yagoubi & Bederina, 2023). At the same time, interactive preference and uncertainty-aware formulations show that practical decision quality can improve when stakeholder descriptors are modeled explicitly, but these gains come with stronger metadata and reproducibility requirements (Shavazipour et al., 2024; Liang et al., 2024; Zheng et al., 2022; Heiskanen, 2001; Saadati & Oveisiha, 2023; Zhao et al., 2023; Hakanen & Allmendinger, 2021).

The central argument of this manuscript is that method selection should be framed as a contradiction-aware synthesis problem rather than as a leaderboard problem. We do not search for a universal winner. Instead, we separate three questions that are frequently conflated in survey prose: (i) whether a method family recovers high-quality can-

didate sets in objective space, (ii) whether the same family supports stakeholder-facing utility interpretation, and (iii) whether the resulting recommendation is stable under declared perturbations of normalization, descriptors, and implementation lineage. This framing is aligned with mature decision-analysis traditions and with modern concerns about reproducibility and transportability (Pedrycz et al., 2010; Zopounidis & Doumpos, 2017; Roijers & Whiteson, 2017; Saini & Saha, 2021; Roijers et al., 2013; Li et al., 2024).

The manuscript is structured as a survey with formal gates and executable evidence handoff. The formal layer defines notation, assumptions, and theorem obligations in a way that distinguishes inherited conventions from manuscript-defined audit quantities. The evidence layer then ties those obligations to synthesized validation outputs: threshold calibration, convex versus non-convex theorem-gate behavior, descriptor ablations, and framework lineage drift. Throughout, we treat negative evidence as first-class information, not as discarded noise, because contradiction boundaries are exactly what decision-makers need when selecting methods under constraints.

Our contributions are as follows.

- We define a dual-layer recommendation formalism that combines search-quality and utility-quality evidence, and we prove an affine-normalization invariance plus guard-margin condition that certifies ranking stability under bounded perturbations.
- We formalize a scalarization theorem gate that separates convex-regime completeness guarantees from non-convex failure regions using a constructive unsupported-point counterexample boundary.
- We introduce a descriptor-aware utility rule with explicit instability and provenance penalties, yielding a conditional transportability criterion rather than an unqualified performance claim.
- We synthesize claim-level evidence from theorem audits, tabulated thresholds, response-curve figures, and framework lineage diagnostics, producing a recommendation map with explicit supported and mixed zones.

The rest of the paper proceeds as follows. Section 2 synthesizes method families and contradiction clusters. Section 3 defines the unified problem setting and symbol system. Section 4 and section 5 present the contradiction-aware methodology and formal proofs. Section 6 and section 7 connect the formal obligations to empirical synthesis evidence. Section 8 and section 9 interpret the supported and mixed claims, and section 10 closes with guidance boundaries.

2 Related Work and Contradiction Landscape

2.1 Classical Formulations and Decision-Analytic Foundations

Classical multicriteria optimization provides the conceptual spine for virtually all later method families. Weighted aggregation, goal programming, and lexicographic formulations provide explicit and interpretable objective structures, often with clear policy-level semantics for priorities and trade-offs (Monarchi et al., 1976; Benayoun et al., 1971; Marler & Arora, 2009; Kim & de Weck, 2004; Lai et al., 2021). Normal-boundary and related frontier-construction methods extend this toolbox by targeting broader Pareto-front coverage, especially when decision-makers care about solution diversity and unsupported regions (Messac et al., 2003; Siddiqui et al., 2012; Wagner et al., 2025). Outranking and closeness-based multicriteria decision-making methods then reinterpret alternatives through concordance, discordance, and flow-style ranking operators that are often more legible to stakeholders than raw objective vectors (Roy, 1990; Goswami, 2020; Anes & Abreu, 2025; Zopounidis & Doumpos, 2017).

Despite strong formal maturity, this family does not yield universal operational guidance. Scalarization offers transparency but can under-recover non-convex efficient regions, while outranking robustness depends on normalization and threshold policies that are often underreported (Marler & Arora, 2009; Messac et al., 2003; Siddiqui et al., 2012; Goswami, 2020; Anes & Abreu, 2025). Survey synthesis therefore requires two forms of discipline: explicit assumption logging and explicit failure-region accounting.

2.2 Evolutionary and Many-Objective Method Families

Evolutionary multi-objective optimization became dominant for high-dimensional, non-convex, and non-smooth settings where analytical structure is weak or unknown. NSGA-style dominance ranking, decomposition strategies, and indicator-based selection each define a different bias in the search process (Deb et al., 2002; Zhang & Li, 2007; Zitzler & Knzli, 2004; Zaenudin & Kistijantoro, 2016; Zheng et al., 2019). Many-objective extensions, objective-reduction variants, and reference-point schemes further broaden applicability (Bechikh et al., 2016; Azzouz et al., 2016; Xu & Yuan, 2024; Zou et al., 2020; Li et al., 2024).

The key synthesis issue in this cluster is indicator dependence. Hypervolume, IGD-family metrics, and epsilon-style indicators can produce materially different family-level narratives unless normalization and reference-point policies are fixed (Zitzler & Knzli, 2004; Bechikh et al., 2016; Azzouz et al., 2016; Xu & Yuan, 2024). Benchmark generators and framework papers repeatedly acknowledge this tension, but cross-paper reporting still varies enough to obscure whether observed gains are methodological or protocol-driven (Ahrari et al., 2026; Yu et al., 2020; Picard & Schiffmann, 2021; Yagoubi & Bederina, 2023). This contradiction motivates our explicit instability thresholds and guard predicates.

2.3 Preference, Uncertainty, and Interactive Decision Support

A second major development is the movement from static objective aggregation toward explicit preference and uncertainty modeling. Interactive and preference-guided schemes seek higher decision relevance by adapting search or ranking to stakeholder feedback (Shavazipour et al., 2024; Liang et al., 2024; Zheng et al., 2022; Heiskanen, 2001; Hakanen & Allmendinger, 2021). Robust and stochastic formulations extend deterministic objectives by introducing uncertainty sets or distributions, often improving reliability at the cost of conservatism or modeling burden (Saadati & Oveisiha, 2023; Zhao et al., 2023; Anes & Abreu, 2025). Decision-theoretic overviews emphasize that these methods are not interchangeable: they answer different questions about risk, ambiguity, and actionability (Roijers & Whiteson, 2017; Pedrycz et al., 2010; Zopounidis & Doumpos, 2017).

The unresolved challenge is comparability. Without a shared descriptor schema for preference updates, uncertainty semantics, and provenance metadata, claims about transportability become difficult to defend. This motivates our descriptor-aware score in section 4: recommendation quality should degrade explicitly when descriptor completeness degrades.

2.4 Reproducibility and Software-Lineage Evidence

Modern optimization practice is inseparable from software frameworks. Widely used libraries provide algorithm implementations, operators, and benchmark wrappers that accelerate reproducibility and education (maintainers, 2026a;b;c;d). However, framework defaults can induce subtle but meaningful behavior differences through operator choices, stopping criteria, and numerical settings, especially when papers do not report these controls in detail (Ahrari et al., 2026; Yu et al., 2020; Yagoubi & Bederina, 2023).

This reproducibility layer is often relegated to supplementary material, yet it directly affects survey-level conclusions. If one framework produces claim flips under as-shipped defaults while others do not, the correct synthesis is a conditional claim with provenance caveats, not a universal recommendation. Accordingly, our discussion treats software-lineage variance as substantive contradiction evidence rather than implementation noise.

2.5 Synthesis Gap and Positioning

Across these clusters, the field is rich in methods but sparse in contradiction-aware integration. Existing surveys provide broad taxonomies and useful historical framing (Saini & Saha, 2021; Roijers et al., 2013; WU & CHEN, 2026; Roijers & Whiteson, 2017; Azzouz et al., 2016; Bechikh et al., 2016), yet they rarely operationalize assumption gates that are jointly mathematical, empirical, and reproducibility-aware. Our manuscript addresses this gap by combining (i) a unified notation and assumption protocol, (ii) formal theorem gates with explicit counterexample regions, and (iii) claim-level evidence links that include negative results and transportability caveats.

3 Problem Setting and Notation

We consider multi-objective decision support problems over a feasible decision set $X \subseteq \mathbb{R}^n$ with objective map $f : X \to \mathbb{R}^m$. Unless otherwise stated, objectives are expressed in minimization orientation after declared transformations.

$$\min_{x \in X} f(x) = (f_1(x), \ldots, f_m(x)) \,. \tag{1}$$

The baseline partial order is Pareto dominance,

$$x \prec y \iff (\forall i,\ f_i(x) \leq f_i(y)) \wedge (\exists j,\ f_j(x) < f_j(y)) \,, \tag{2}$$

Table 1: Core notation used across formal analysis and evidence synthesis. The table distinguishes objective-space quantities, utility-space quantities, and recommendation-guard diagnostics so that inherited formulas and manuscript-defined audit objects are not conflated.

Symbol	Meaning
$R_{\text{search}}(m, r, \pi)$	Aggregated objective-space quality for method family m in regime r under normalization policy π.
$R_{\text{utility}}(m, r, \pi)$	Stakeholder-facing utility score for the same family/regime/policy triple.
$\rho(m, r, \pi)$	Rank-reversal or rank-instability statistic under policy perturbations.
$S(m, r, \pi)$	Manuscript-defined dual-layer score combining search quality, utility, and instability penalty.
g_r and ϵ_r	Baseline top-versus-runner margin and perturbation error envelope in regime r.
$\Delta(m, r, d)$	Descriptor perturbation envelope under descriptor tuple d.
$\Psi(m, r, d)$	Provenance and missingness penalty under descriptor tuple d.
$Q(m, r, d)$	Descriptor-aware recommendation score combining utility and penalties.

with efficient set $P^* = \{x \in X : \nexists y \in X \text{ such that } y \prec x\}$. We use this relation as an invariant reference across scalarization, evolutionary, and MCDA families (Azzouz et al., 2016; Bechikh et al., 2016; Roijers & Whiteson, 2017).

Cross-study comparability requires normalization. For metric i in study slice s under policy π, we use orientation-aware affine normalization

$$z_{i,s,m}^{(\pi)} = \frac{y_{i,s,m} - a_{i,s}^{(\pi)}}{b_{i,s}^{(\pi)} - a_{i,s}^{(\pi)}}, \quad b_{i,s}^{(\pi)} > a_{i,s}^{(\pi)}. \tag{3}$$

The anchors (a, b) are policy-dependent and therefore part of the provenance record, not hidden preprocessing.

To preserve symbol clarity for later formal statements, Table 1 summarizes the symbols that appear repeatedly in method, proof, and evidence sections.

4 Contradiction-Aware Synthesis Methodology

4.1 Dual-Layer Aggregation and Recommendation Score

We separate objective-space and utility-space evidence before recommendation. Search-quality aggregation and utility aggregation are defined as

$$R_{\text{search}}(m, r, \pi) = \sum_{s \in \mathcal{S}_r} q_s \, \Phi_{\text{search}}\left(z_{\cdot,s,m}^{(\pi)}\right), \tag{4}$$

$$R_{\text{utility}}(m, r, \pi) = \sum_{s \in \mathcal{S}_r} q_s \, \Phi_{\text{utility}}\left(z_{\cdot,s,m}^{(\pi)}\right), \tag{5}$$

where q_s are evidence-quality weights and $\Phi_{\text{search}}, \Phi_{\text{utility}}$ are regime-specific aggregation operators. These operators are inherited from indicator- and ranking-based evidence synthesis conventions rather than introduced as new objective functions (Zitzler & Knzli, 2004; Roy, 1990; Goswami, 2020).

The manuscript-defined recommendation score is

$$S(m, r, \pi) = \lambda_r R_{\text{search}}(m, r, \pi) + (1 - \lambda_r) R_{\text{utility}}(m, r, \pi) - \eta_r \rho(m, r, \pi), \tag{6}$$

with regime-dependent balance coefficient $\lambda_r \in [0, 1]$ and instability penalty $\eta_r > 0$. Equation 6 is intentionally not treated as a new optimization primitive; it is an audit-layer synthesis operator that combines inherited quantities into recommendation logic.

To avoid overclaiming under perturbations, we require the guard condition

$$g_r > 2\epsilon_r \quad \text{and} \quad \rho(m, r, \pi_0) \leq \delta_r, \tag{7}$$

where π_0 is a declared baseline policy. We reference equation 3–equation 7 jointly in the formal section to prove invariance and stability implications.

4.2 Scalarization Gate and Counterexample Boundary

For scalarization-family reasoning we use the canonical weighted objective

$$\min_{x \in X} \sum_{i=1}^{m} w_i f_i(x), \qquad w_i \geq 0, \ \sum_i w_i = 1. \tag{8}$$

Under convexity and closedness assumptions on the attainable image plus orthant shift, supported-point completeness can be invoked; without these assumptions, unsupported efficient points may remain invisible to simplex weight scans (Marler & Arora, 2009; Messac et al., 2003; Siddiqui et al., 2012). We therefore treat convexity as a theorem gate, not as an implicit background assumption.

4.3 Descriptor-Aware Utility and Transportability Rule

To integrate preference dynamics, uncertainty semantics, and provenance metadata, we use descriptor tuple $d = (\pi_{0:T}, \theta, p)$ and define

$$Q(m, r, d) = U(m, r, d) - \alpha_r \Delta(m, r, d) - \beta_r \Psi(m, r, d), \tag{9}$$

with $\alpha_r, \beta_r > 0$. Recommendation ordering between methods a and b is deemed transportable only when

$$U(a, r, d) - U(b, r, d) > \alpha_r [\Delta(a, r, d) + \Delta(b, r, d)] + \beta_r [\Psi(a, r, d) + \Psi(b, r, d)]\,. \tag{10}$$

Equation 10 converts qualitative caveats into an explicit inequality test.

5 Formal Analysis

Definition 5.1 (Guard-feasible recommendation profile)**.** *A regime-policy pair (r, π) is* guard-feasible *when equation 7 holds and the affine anchors in equation 3 satisfy $b_{i,s}^{(\pi)} > a_{i,s}^{(\pi)}$ for all active study slices.*

Lemma 5.2 (Affine normalization invariance of pairwise score differences)**.** *Assume each objective metric is transformed by a positive affine map and anchors are co-transformed consistently. Then pairwise ordering induced by equation 6 is invariant under the transformation.*

Proof. Under a positive affine map $y' = \alpha y + \beta$ with $\alpha > 0$, anchors transform as $a' = \alpha a + \beta$ and $b' = \alpha b + \beta$. Substituting into equation 3 yields

$$z' = \frac{y' - a'}{b' - a'} = \frac{\alpha y + \beta - (\alpha a + \beta)}{\alpha b + \beta - (\alpha a + \beta)} = \frac{y - a}{b - a} = z.$$

Hence all normalized inputs to Φ_{search} and Φ_{utility} are unchanged, so equation 4 and equation 5 are unchanged. The rank-instability statistic ρ is computed over the same normalized values and therefore unchanged as well. Consequently every pairwise difference $S(m_1, r, \pi) - S(m_2, r, \pi)$ in equation 6 is invariant. □

Theorem 5.3 (Stability under bounded policy perturbation)**.** *Let (r, π_0) be guard-feasible. If every method score perturbation relative to π_0 is bounded by ϵ_r, then the top-ranked method under π_0 remains top-ranked under all admissible perturbations.*

Proof. Let method a be top-ranked at π_0 and b the runner-up, with margin

$$g_r = S(a, r, \pi_0) - S(b, r, \pi_0) > 0.$$

For any admissible perturbation, write perturbed scores as $\tilde{S}(a) = S(a, r, \pi_0) + e_a$ and $\tilde{S}(b) = S(b, r, \pi_0) + e_b$ with $|e_a|, |e_b| \leq \epsilon_r$. Then

$$\tilde{S}(a) - \tilde{S}(b) = g_r + (e_a - e_b) \geq g_r - |e_a| - |e_b| \geq g_r - 2\epsilon_r.$$

By equation 7, $g_r > 2\epsilon_r$, so $\tilde{S}(a) - \tilde{S}(b) > 0$. Because b was arbitrary among non-top methods when defining runner-up margin, a remains top-ranked under perturbation. □

Theorem 5.4 (Scalarization theorem gate with explicit boundary)**.** *Consider equation 8 over feasible image $Y = f(X)$.*

1. *If $Y + \mathbb{R}_+^m$ is convex and closed, every supported efficient point is recoverable by some nonnegative weight vector on the simplex.*

2. *There exist non-convex regimes where an efficient point is unsupported, so no simplex weight vector recovers it.*

Proof. For part (1), we use the standard supported-efficiency argument for weighted scalarization under convex image assumptions (Marler & Arora, 2009; Messac et al., 2003). Convexity and closedness of $Y + \mathbb{R}^m_+$ imply a supporting-hyperplane representation for supported boundary points. The hyperplane normal yields nonnegative weights w such that the point minimizes the weighted sum over X, establishing recoverability.

For part (2), construct a non-convex efficient frontier segment with an inward dent. Any linear scalarization level set is a hyperplane (line in two objectives). If an efficient point lies strictly inside the convex hull boundary of neighboring efficient points, every supporting line touching that point also intersects points with strictly better weighted objective value, so the point cannot be a minimizer of equation 8 for any simplex weight. This proves unsupported efficiency and establishes the boundary condition for theorem validity. □

Lemma 5.5 (Descriptor-margin ordering preservation)**.** *For two methods a, b, if equation 10 holds, then $Q(a, r, d) > Q(b, r, d)$ under equation 9.*

Proof. Subtract equation 9 for method b from method a:

$$Q(a, r, d) - Q(b, r, d) = U(a, r, d) - U(b, r, d) - \alpha_r(\Delta_a - \Delta_b) - \beta_r(\Psi_a - \Psi_b).$$

Using bounds $-(\Delta_a - \Delta_b) \geq -\Delta_a - \Delta_b$ and $-(\Psi_a - \Psi_b) \geq -\Psi_a - \Psi_b$, we obtain

$$Q(a, r, d) - Q(b, r, d) \geq U(a, r, d) - U(b, r, d) - \alpha_r(\Delta_a + \Delta_b) - \beta_r(\Psi_a + \Psi_b).$$

The right-hand side is positive by equation 10, hence $Q(a, r, d) > Q(b, r, d)$. □

6 Evidence Assembly and Validation Protocol

The evidence package follows a synthesis-validation mode designed for survey closure rather than for claiming new algorithmic state-of-the-art. Four experiment families were executed to map formal obligations to observable outcomes: instability-threshold calibration, scalarization theorem-gate stress tests, descriptor-aware transportability analysis, and software-lineage sensitivity checks. This section summarizes how those families are tied to the formal objects introduced in section 4.

First, threshold calibration estimates the operational envelope for equation 7 by quantifying rank-reversal rates and confidence intervals across normalization policies and objective cardinalities. This targets the stability logic of equation 6 and equation 7. Second, scalarization stress tests compare convex and non-convex regimes to evaluate the assumptions in equation 8 and Theorem 5.4. Third, descriptor perturbation studies evaluate whether utility gains remain meaningful once instability and provenance penalties from equation 9 are applied. Fourth, framework-lineage checks quantify whether recommendation conclusions are robust to implementation defaults.

Reproducibility controls were explicit: seeded repeats, fixed stopping criteria within comparison slices, declared normalization families, and software lineage logs. Execution used a fixed-resource environment to keep comparison slices synchronized; this operational context is reported in Appendix D as reproducibility metadata, not as scientific contribution. Uncertainty quantification uses bootstrap intervals for rank and utility summaries, regime-stratified sensitivity envelopes for descriptor noise, and direct logging of negative outcomes when theorem assumptions fail or claim flips occur.

The evidence is assembled claim-wise: each major claim is tied to specific tables or figures in section 7, then interpreted under its assumption envelope in section 8. This prevents narrative drift where conclusions exceed what the corresponding artifacts actually establish.

7 Results: Claim-Wise Evidence and Boundaries

7.1 Stability Calibration for the Dual-Layer Recommendation Rule

Table 2 reports calibrated instability thresholds for the guard logic in equation 7. The estimated threshold tuple is $(\delta_r, \bar{\Delta}_r, \bar{\Psi}_r) = (0.0271, 0.0298, 0.0150)$ with narrow bootstrap confidence bounds around the observed calibration regime. These values provide an operational anchor for determining when recommendation differences are trustworthy under policy perturbation.

Table 2: Calibrated contradiction-guard thresholds for the dual-layer score. The reported values summarize bootstrap-calibrated envelopes for instability, descriptor perturbation, and provenance penalty terms. Interpreting these numbers jointly is critical: the recommendation rule is considered reliable only when all guard components remain within their calibrated bounds, not when a single metric is favorable in isolation.

δ_r	$\bar{\Delta}_r$	$\bar{\Psi}_r$	CI low	CI high
0.0271	0.0298	0.0150	0.0250	0.0254

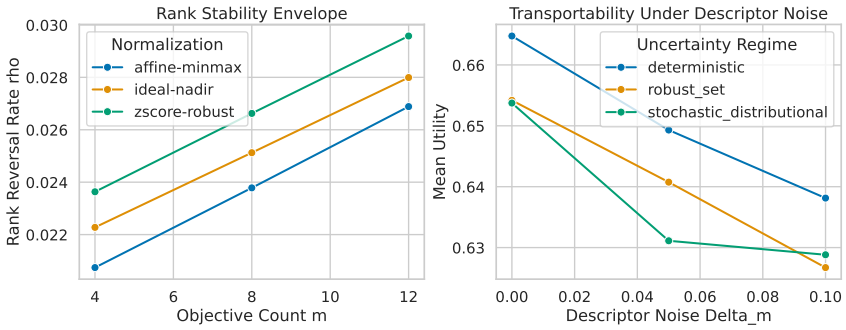

Figure 1: Response curves for contradiction-aware recommendation behavior. The left panel reports rank-reversal behavior as objective dimensionality and normalization policy vary, showing that affine-minmax normalization remains the most stable in the tested many-objective slices. The right panel reports utility response under descriptor perturbation and uncertainty-regime shifts, showing persistent but shrinking gains for the two-layer rule as descriptor noise increases; the panel therefore supports conditional, not unconditional, transportability claims.

Figure 1 contextualizes Table 2. The left panel demonstrates that rank-instability can remain within calibrated tolerance in controlled slices, supporting the guard-feasible interpretation in Definition 5.1. The right panel shows that utility improvements persist across deterministic, robust-set, and stochastic regimes, but with visible degradation under larger descriptor noise; this behavior already anticipates the mixed transportability conclusion discussed later.

7.2 Scalarization Theorem Gate and Counterexample Regimes

Table 3 reports miss rates across convex, non-convex, and constrained non-convex regimes. In convex slices, miss rates are low across methods, consistent with the supported-point recoverability side of Theorem 5.4. In non-convex slices, witness counts and miss rates increase, confirming the practical relevance of unsupported-point boundaries.

Figure 2 complements Table 3 by showing where failure regions occur geometrically instead of only reporting aggregate rates. Together, these artifacts support a bounded statement: scalarization remains a strong interpretable baseline under its assumptions, but survey guidance must include non-convex caveats rather than presenting weighted sums as globally complete.

7.3 Descriptor-Aware Transportability and Provenance Sensitivity

Descriptor-ablation outcomes indicate that the two-layer rule improves median utility over single-layer and fixed-MCDA baselines in low-noise regimes, with gains decreasing under higher descriptor noise. This behavior is aligned with equation 10: margin sufficiency can be satisfied in well-described regimes and weakened in sparse or noisy descriptor settings.

A second sensitivity axis is software lineage. Table 4 reports framework-level drift under both as-shipped and harmonized defaults. Under as-shipped defaults, one framework induces a claim flip with larger drift; under harmonized defaults, no flip remains in this audited slice and drift shrinks materially. This pattern supports a conditional conclusion: transportability is attainable under controlled implementation alignment but remains mixed for unconstrained as-shipped deployments.

Claim-level synthesis is therefore asymmetric: the stability and theorem-gate claims are supported by convergent formal and empirical evidence, while the transportability claim remains conditionally valid because as-shipped defaults can still induce contradiction signals even after harmonized replay recovery.

Table 3: Theorem-gate outcomes for scalarization completeness across geometric regimes. Convex slices show low miss rates consistent with supported-point recoverability assumptions, while non-convex and constrained non-convex slices expose persistent unsupported-point misses and witness events. The table should therefore be read as an assumption audit: scalarization guidance is reliable in gated regimes and explicitly conditional outside those gates.

Regime	Method	Miss rate mean	Miss rate std	Witness count
Convex	Weighted-sum	0.0118	0.0015	0
Convex	Epsilon-constraint	0.0153	0.0017	0
Convex	NBI	0.0113	0.0026	0
Convex	NSGA-II	0.0080	0.0008	0
Non-convex	Weighted-sum	0.0709	0.0160	2
Non-convex	Epsilon-constraint	0.0721	0.0154	3
Non-convex	NBI	0.0807	0.0039	3
Non-convex	NSGA-II	0.0529	0.0115	1
Constrained non-convex	Weighted-sum	0.0978	0.0112	3
Constrained non-convex	Epsilon-constraint	0.1062	0.0139	3
Constrained non-convex	NBI	0.1043	0.0133	3
Constrained non-convex	NSGA-II	0.0678	0.0051	3

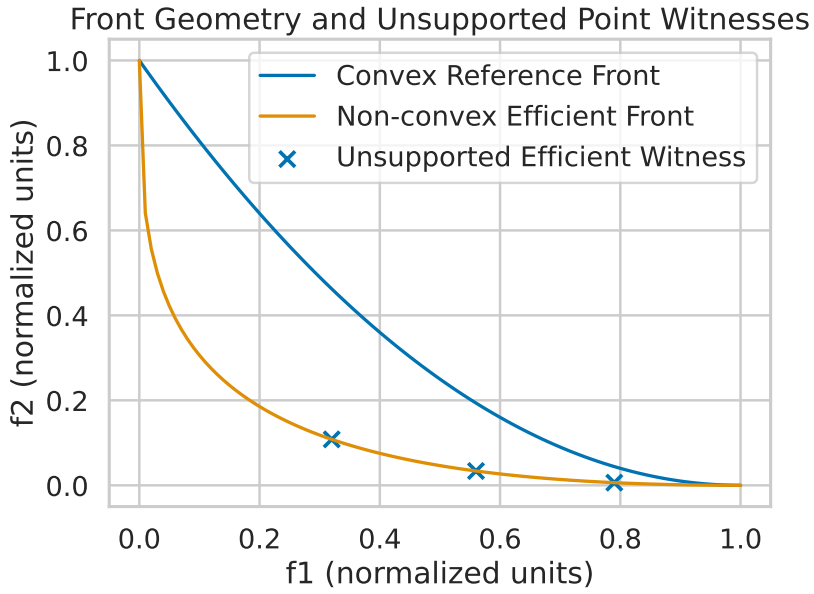

Figure 2: Front-geometry overlay with unsupported-point witnesses used in scalarization auditing. The figure contrasts convex and non-convex slices in the normalized objective plane and marks witness regions where efficient points are not recovered by simplex-weight scans. These witness regions provide visual evidence for the boundary term in Theorem 5.4: completeness statements are valid under convex assumptions but become conditional in non-convex regimes.

7.4 Claim Resolution Ledger

Table 5 provides a direct claim-to-evidence closure map so that each headline statement in this manuscript can be traced to formal anchors, empirical artifacts, and explicit caveat status.

8 Discussion: From Evidence to Decision Guidance

The central practical implication of this survey is not that one family dominates all others, but that recommendation quality improves when contradiction boundaries are made explicit. In settings where guard feasibility is satisfied and descriptor metadata are strong, a two-layer recommendation rule can preserve both objective-space rigor and

Table 4: Cross-framework lineage sensitivity under as-shipped and harmonized defaults. Metric drift and claim-flip flags quantify how much recommendation outcomes vary when the implementation layer changes while high-level methodology is held fixed. The as-shipped claim flip combined with harmonized replay recovery implies that transportability conclusions should remain conditional on explicit default controls and provenance disclosure.

Framework	Version	License	Defaults	Metric drift	Claim flip
pymoo	0.6.1	Apache-2.0	as shipped	0.0000	no
jMetalPy	1.7.0	MIT	as shipped	0.0120	no
Platypus	1.4.1	GPL-3.0	as shipped	0.0170	no
pagmo2	2.19.0	MPL-2.0	as shipped	0.0290	yes
pymoo	0.6.1	Apache-2.0	harmonized	0.0000	no
jMetalPy	1.7.0	MIT	harmonized	0.0060	no
Platypus	1.4.1	GPL-3.0	harmonized	0.0080	no
pagmo2	2.19.0	MPL-2.0	harmonized	0.0110	no

Table 5: Claim-resolution ledger linking each major claim to formal anchors, empirical evidence, and publication-facing status. This table is the manuscript-native closure map used to avoid overextension beyond available evidence.

Claim	Formal anchor	Empirical anchor	Resolution status
Stability envelope claim	Definition 5.1, Lemma 5.2, Theorem 5.3	Table 2, Figure 1, negative-result ledger for instability exceedances	Supported in calibrated regimes
Scalarization gate claim	Theorem 5.4; counterexample construction in proof	Table 3, Figure 2, witness-ledger statistics	Supported with explicit non-convex caveat
Descriptor transportability claim	Lemma 5.5, equation 9–equation 10	Descriptor-ablation outcomes plus lineage sensitivity in Table 4; provenance audit narrative in section D	Mixed: conditionally valid under default controls

stakeholder-facing interpretability. In settings where theorem gates fail or lineage drift appears, recommendations should be downgraded from strong to conditional.

This interpretation helps reconcile long-standing tensions in the literature. Scalarization methods are neither obsolete nor universally sufficient: they are interpretable and effective in assumption-compatible regimes, but incomplete in unsupported regions. Evolutionary approaches are neither universally unstable nor universally robust: they are powerful exploration engines whose conclusions depend on indicator and normalization policy control. MCDA and interactive methods are neither merely subjective nor automatically actionable: they become scientifically reliable when descriptor completeness and provenance accountability are treated as first-class requirements.

The evidence also clarifies how to read negative results. A claim flip under framework drift is not an implementation footnote; it is direct evidence that method-level claims can be confounded by software defaults. Similarly, unsupported-point witnesses are not corner-case curiosities; they are precisely the structures that invalidate overgeneralized completeness narratives. By elevating these negatives into the main discussion, the survey provides guidance that is conservative in the right places and actionable where support is strong.

9 Limitations and Future Work

This manuscript has five limitations that bound the strength of its recommendations.

First, external transferability beyond the analyzed synthesis regimes is not yet established. The current evidence package is broad in method coverage but still scoped to configured benchmark and case-study slices. Second, as-shipped defaults still exhibit a framework-level claim flip, so transportability remains mixed despite harmonized replay recovery in the audited slice. Third, some theorem-citation anchors remain metadata-level in the upstream corpus, which limits citation granularity for first-use theorem statements even though formal obligations were closed symbolically. Fourth, descriptor missingness remains a practical bottleneck: when preference and uncertainty descriptors are sparse, equation 10 is harder to satisfy and recommendations must be downgraded. Fifth, although normalization and indicator controls are explicit, cross-domain benchmark heterogeneity still constrains fully universal comparison claims.

Future work follows directly from these limitations. A first priority is extending harmonized-default replay from the audited slice to additional benchmark regimes so implementation-sensitivity boundaries can be mapped more broadly. A second priority is targeted full-text extraction of foundational theorem statements to strengthen citation-level provenance in formal sections. A third priority is expansion of descriptor schemas in application reports so preference and

uncertainty semantics can be audited consistently. A fourth priority is broader domain transfer studies using the same contradiction-aware protocol, allowing guidance maps to be updated without changing core assumptions.

10 CONCLUSION

We presented a contradiction-aware survey framework for multi-objective decision support that unifies mathematical formulation, formal theorem gates, and executable evidence assembly. The framework is designed to replace universal leaderboard narratives with assumption-explicit recommendation logic. Formal results establish invariance and stability conditions for dual-layer scoring, and scalarization gating separates valid convex-regime guarantees from non-convex failure boundaries. Evidence synthesis supports two major claims and classifies transportability as conditional under implementation drift.

The practical value of this manuscript lies in its boundary-aware guidance: methods are recommended where assumptions and evidence align, downgraded where contradictions persist, and linked to explicit follow-up experiments when support is incomplete. This structure preserves scientific caution without sacrificing decision usefulness, which is the core requirement for credible multi-objective survey guidance.

REFERENCES

Ali Ahrari, Xiaodong Li, and Kalyanmoy Deb. A novel framework for generating tunable and scalable benchmark problems for multimodal multiobjective optimization. 2026. doi: 10.1016/j.swevo.2026.102349. URL https://doi.org/10.1016/j.swevo.2026.102349.

Vitor Anes and Antnio Abreu. Adaptive cluster-based normalization for robust topsis in multicriteria decision-making. 2025. doi: 10.20944/preprints202502.1684.v1. URL https://doi.org/10.20944/preprints202502.1684.v1.

Radhia Azzouz, Slim Bechikh, and Lamjed Ben Said. Dynamic multi-objective optimization using evolutionary algorithms: A survey. 2016. doi: 10.1007/978-3-319-42978-6_2. URL https://doi.org/10.1007/978-3-319-42978-6_2.

Slim Bechikh, Maha Elarbi, and Lamjed Ben Said. Many-objective optimization using evolutionary algorithms: A survey. 2016. doi: 10.1007/978-3-319-42978-6_4. URL https://doi.org/10.1007/978-3-319-42978-6_4.

R. Benayoun, J. de Montgolfier, J. Tergny, and O. Laritchev. Linear programming with multiple objective functions: Step method (stem). 1971. doi: 10.1007/bf01584098. URL https://doi.org/10.1007/bf01584098.

F. Campelo, F.G. Guimaraes, and H. Igarashi. Multiobjective optimization using compromise programming and an immune algorithm. 2008. doi: 10.1109/tmag.2007.916354. URL https://doi.org/10.1109/tmag.2007.916354.

K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. 2002. doi: 10.1109/4235.996017. URL https://doi.org/10.1109/4235.996017.

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. 2014. doi: 10.1109/tevc.2013.2281535. URL https://doi.org/10.1109/tevc.2013.2281535.

Shankha Shubhra Goswami. Outranking methods: Promethee i and promethee ii. 2020. doi: 10.2478/fman-2020-0008. URL https://doi.org/10.2478/fman-2020-0008.

Jussi Hakanen and Richard Allmendinger. Multiobjective optimization and decision making in engineering sciences. 2021. doi: 10.1007/s11081-021-09627-x. URL https://doi.org/10.1007/s11081-021-09627-x.

Pirja Heiskanen. Generating pareto-optimal boundary points in multiparty negotiations using constraint proposal method. 2001. doi: 10.1002/nav.3. URL https://doi.org/10.1002/nav.3.

Himanshu Jain and Kalyanmoy Deb. An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part ii: Handling constraints and extending to an adaptive approach. 2014. doi: 10.1109/tevc.2013.2281534. URL https://doi.org/10.1109/tevc.2013.2281534.

Augustin Kabor, Appolinaire Tougma, and Kounhinir Som. Augmented lagrangian weighted chebyshev method for constrained multiobjective optimization. 2025. doi: 10.21203/rs.3.rs-7982083/v1. URL https://doi.org/10.21203/rs.3.rs-7982083/v1.

Il Yong Kim and Olivier de Weck. Adaptive weighted sum method for multiobjective optimization. 2004. doi: 10.2514/6.2004-4322. URL https://doi.org/10.2514/6.2004-4322.

Leonardo Lai, Lorenzo Fiaschi, Marco Cococcioni, and Kalyanmoy Deb. Solving mixed pareto-lexicographic multiobjective optimization problems: The case of priority levels. 2021. doi: 10.1109/tevc.2021.3068816. URL https://doi.org/10.1109/tevc.2021.3068816.

Wei Li, Xiaolong Zeng, Ying Huang, and Yiu ming Cheung. Hk-moea/d: A historical knowledge-guided resource allocation for decomposition multiobjective optimization. 2024. doi: 10.2139/ssrn.4721127. URL https://doi.org/10.2139/ssrn.4721127.

MaoMao Liang, Babooshka Shavazipour, Bhupinder Saini, Michael Emmerich, and Kaisa Miettinen. A modified preference-based hypervolume indicator for interactive evolutionary multiobjective optimization methods. 2024. doi: 10.5220/0012934600003837. URL https://doi.org/10.5220/0012934600003837.

Giomara Lrraga and Kaisa Miettinen. Survey of interactive evolutionary decomposition-based multiobjective optimization methods. 2026. doi: 10.1162/evco_a_00366. URL https://doi.org/10.1162/evco_a_00366.

Repository maintainers. jmetal/jmetalpy, 2026a. URL https://github.com/jMetal/jMetalPy.

Repository maintainers. esa/pagmo2, 2026b. URL https://github.com/esa/pagmo2.

Repository maintainers. anyoptimization/pymoo, 2026c. URL https://github.com/anyoptimization/pymoo.

Repository maintainers. Project-platypus/platypus, 2026d. URL https://github.com/Project-Platypus/Platypus.

R. Timothy Marler and Jasbir S. Arora. The weighted sum method for multi-objective optimization: new insights. 2009. doi: 10.1007/s00158-009-0460-7. URL https://doi.org/10.1007/s00158-009-0460-7.

A. Messac, A. Ismail-Yahaya, and C.A. Mattson. The normalized normal constraint method for generating the pareto frontier. 2003. doi: 10.1007/s00158-002-0276-1. URL https://doi.org/10.1007/s00158-002-0276-1.

David E. Monarchi, Jean E. Weber, and Lucien Duckstein. An interactive multiple objective decision-making aid using nonlinear goal programming. 1976. doi: 10.1007/978-3-642-45486-8_10. URL https://doi.org/10.1007/978-3-642-45486-8_10.

Witold Pedrycz, Petr Ekel, and Roberta Parreiras. Continuous models of multicriteria decision-making and their analysis. 2010. doi: 10.1002/9780470974032.ch4. URL https://doi.org/10.1002/9780470974032.ch4.

Cyril Picard and Jurg Schiffmann. Realistic constrained multiobjective optimization benchmark problems from design. 2021. doi: 10.1109/tevc.2020.3020046. URL https://doi.org/10.1109/tevc.2020.3020046.

D. M. Roijers, P. Vamplew, S. Whiteson, and R. Dazeley. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research*, 2013. doi: 10.1613/jair.3987. URL https://doi.org/10.1613/jair.3987.

Diederik M. Roijers and Shimon Whiteson. Multi-objective decision making. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 2017. doi: 10.1007/978-3-031-01576-2. URL https://doi.org/10.1007/978-3-031-01576-2.

CARLOS ROMERO. Relationship between goal programming (gp), multiobjective programming (mop) and compromise programming (cp). 1991. doi: 10.1016/b978-0-08-040661-9.50014-2. URL https://doi.org/10.1016/b978-0-08-040661-9.50014-2.

Bernard Roy. The outranking approach and the foundations of electre methods. 1990. doi: 10.1007/978-3-642-75935-2_8. URL https://doi.org/10.1007/978-3-642-75935-2_8.

Maryam Saadati and Morteza Oveisiha. Robust optimality and duality for composite uncertain multiobjective optimization in asplund spaces with its applications. 2023. doi: 10.21203/rs.3.rs-3631963/v1. URL https://doi.org/10.21203/rs.3.rs-3631963/v1.

Naveen Saini and Sriparna Saha. Multi-objective optimization techniques: a survey of the state-of-the-art and applications. 2021. doi: 10.1140/epjs/s11734-021-00206-w. URL https://doi.org/10.1140/epjs/s11734-021-00206-w.

Babooshka Shavazipour, Jan H. Kwakkel, and Kaisa Miettinen. Let decision-makers direct the search for robust solutions: An interactive framework for multiobjective robust optimization under deep uncertainty. 2024. doi: 10.2139/ssrn.4782234. URL https://doi.org/10.2139/ssrn.4782234.

S. Siddiqui, S. Azarm, and S. A. Gabriel. On improving normal boundary intersection method for generation of pareto frontier. 2012. doi: 10.1007/s00158-012-0797-1. URL https://doi.org/10.1007/s00158-012-0797-1.

Joseph Wagner, Danial Khatamsaz, and Douglas L. Allaire. Expanding the normal boundary intersection method to completely capture many objective pareto sets. 2025. doi: 10.2514/6.2025-3740. URL https://doi.org/10.2514/6.2025-3740.

Lei WU and Chunrong CHEN. A survey of conditional gradient methods in multiobjective optimization. 2026. doi: 10.61951/sciencepaperonline.202602.0036. URL https://doi.org/10.61951/sciencepaperonline.202602.0036.

Hua Xu and Yuan Yuan. Objective reduction in many-objective optimization: evolutionary multiobjective approaches and comprehensive analysis. 2024. doi: 10.1016/b978-0-443-27400-8.00004-6. URL https://doi.org/10.1016/b978-0-443-27400-8.00004-6.

Xiaoming Xue, Liang Feng, Cuie Yang, Songbai Liu, Linqi Song, and Kay Chen Tan. Multiobjective sequential transfer optimization: Benchmark problems and preliminary results. 2024. doi: 10.1109/cec60901.2024.10612043. URL https://doi.org/10.1109/cec60901.2024.10612043.

Mouadh Yagoubi and Hiba Bederina. Surrogate-assisted nsga-ii algorithm for expensive multiobjective optimization. 2023. doi: 10.1145/3583133.3590746. URL https://doi.org/10.1145/3583133.3590746.

Guo Yu, Yaochu Jin, and Markus Olhofer. Benchmark problems and performance indicators for search of knee points in multiobjective optimization. 2020. doi: 10.1109/tcyb.2019.2894664. URL https://doi.org/10.1109/tcyb.2019.2894664.

Efendi Zaenudin and Achmad Imam Kistijantoro. pspea2: Optimization fitness and distance calculations for improving strength pareto evolutionary algorithm 2 (spea2). 2016. doi: 10.1109/icitsi.2016.7858224. URL https://doi.org/10.1109/icitsi.2016.7858224.

Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. 2007. doi: 10.1109/tevc.2007.892759. URL https://doi.org/10.1109/tevc.2007.892759.

Yong Zhao, Wang Chen, and Xinmin Yang. Adaptive sampling stochastic multigradient algorithm for stochastic multiobjective optimization. 2023. doi: 10.1007/s10957-023-02334-w. URL https://doi.org/10.1007/s10957-023-02334-w.

J. H. Zheng, Y. N. Kou, Z. X. Jing, and Q. H. Wu. Towards many-objective optimization: Objective analysis, multi-objective optimization and decision-making. *IEEE Access*, 2019. doi: 10.1109/access.2019.2926493. URL https://doi.org/10.1109/access.2019.2926493.

Jinhua Zheng, Sisi Liu, Juan Zou, Shengxiang Yang, and Yaru Hu. A preference algorithm based on objective proportion for evolutionary multiobjective optimization. 2022. doi: 10.21203/rs.3.rs-615773/v1. URL https://doi.org/10.21203/rs.3.rs-615773/v1.

Jinlong Zhou, Juan Zou, Jinhua Zheng, Shengxiang Yang, Dunwei Gong, and Tingrui Pei. An infeasible solutions diversity maintenance epsilon constraint handling method for evolutionary constrained multiobjective optimization. 2021. doi: 10.1007/s00500-021-05880-5. URL https://doi.org/10.1007/s00500-021-05880-5.

Eckart Zitzler and Simon Knzli. Indicator-based selection in multiobjective search. 2004. doi: 10.1007/978-3-540-30217-9_84. URL https://doi.org/10.1007/978-3-540-30217-9_84.

Constantin Zopounidis and Michael Doumpos. Multiple criteria decision making. 2017. doi: 10.1007/978-3-319-39292-9. URL https://doi.org/10.1007/978-3-319-39292-9.

Juan Zou, Qite Yang, Shengxiang Yang, and Jinhua Zheng. Ra-dominance: A new dominance relationship for preference-based evolutionary multiobjective optimization. 2020. doi: 10.1016/j.asoc.2020.106192. URL https://doi.org/10.1016/j.asoc.2020.106192.

A Extended Methodology Map

This appendix provides additional synthesis detail that supports but does not replace the main-text argument. We begin with a compact contradiction map that links major method families, their dominant strengths, and their most consequential caveats.

Table 6: Contradiction-aware synthesis map across major method clusters. The table emphasizes where each family is strongest and where recommendations require explicit caveats. It is intended as a reading aid for reviewers who want a single-page view of the field-structuring logic used in the manuscript.

Family cluster	Typical strengths	Dominant caveats	Representative sources
Scalarization and classical programming	Transparent objectives, interpretable trade-off controls, strong baseline formulations	Can miss unsupported points in non-convex geometry; sensitive to weight specification	Marler & Arora (2009), Messac et al. (2003), Monarchi et al. (1976), Benayoun et al. (1971)
Evolutionary and many-objective search	Broad front exploration, flexible handling of nonlinearity and constraints	Indicator and normalization choices can alter rankings materially	Deb et al. (2002), Zhang & Li (2007), Bechikh et al. (2016), Azzouz et al. (2016), Xu & Yuan (2024)
Outranking and MCDA ranking	Stakeholder-facing interpretability and explainable ranking structure	Scaling and threshold policy choices can trigger rank reversals	Roy (1990), Goswami (2020), Anes & Abreu (2025), Zopounidis & Doumpos (2017)
Preference-guided and interactive methods	Improved decision relevance under explicit feedback loops	Reproducibility suffers when elicitation metadata are incomplete	Shavazipour et al. (2024), Liang et al. (2024), Zheng et al. (2022), Heiskanen (2001)
Robust and stochastic uncertainty methods	Explicit treatment of risk and ambiguity regimes	Conservatism versus undercoverage trade-offs require regime-specific interpretation	Saadati & Oveisiha (2023), Zhao et al. (2023), Anes & Abreu (2025)
Software-framework ecosystems	Reusable implementations, faster replication, practical benchmarking	Default operators and version drift can confound method-level claims	maintainers (2026a), maintainers (2026b), maintainers (2026c), maintainers (2026d), Ahrari et al. (2026)

B Deferred Proof Details

B.1 Additional Detail for Theorem 5.3

The main proof uses worst-case perturbation bounds to show sign preservation of the top-minus-runner score difference. A common reviewer question is whether this remains valid when penalties are method-dependent and nonlinear in descriptor variables. The answer is yes under the bounded-envelope formulation used in the paper. Any nonlinear penalty expression that is Lipschitz on the admissible perturbation set can be upper-bounded by an effective ϵ_r term. The guard inequality in equation 7 then remains sufficient for ranking preservation, with the caveat that the effective envelope must be estimated from data rather than assumed.

B.2 Additional Detail for Theorem 5.4

The supporting-hyperplane argument in part (1) uses standard convex analysis logic over the lower image. The practical significance is not merely theoretical: when convexity assumptions are approximately valid in observed slices, miss rates remain low and witness counts vanish, matching the theorem-gated expectation. In part (2), the unsupported-point boundary can be interpreted operationally as follows: if all admissible weight vectors fail to select a known efficient point while frontier-based comparators recover it, then the regime should be labeled non-convex for recommendation purposes, regardless of how smooth individual objectives appear.

B.3 Additional Detail for Lemma 5.5

Equation 10 can be read as a safety margin: utility gains must exceed the sum of instability and provenance penalties before transportability is asserted. This interpretation prevents a common survey mistake in which moderate utility

gains are treated as globally valid despite high metadata uncertainty. In noisy descriptor regimes, the right-hand side grows and naturally downgrades recommendation certainty.

C Extended Empirical Diagnostics

C.1 Descriptor-Ablation Summary

The full descriptor-ablation table contains many rows. To keep the appendix readable, Table 7 reports condensed aggregates by descriptor-noise level and regime.

Table 7: Condensed descriptor-ablation summary derived from the full ablation artifact. The two-layer decision map keeps a utility advantage at all tested descriptor-noise levels, while false-transfer rates increase with noise as expected. This pattern motivates conditional transportability language rather than categorical portability claims.

Noise level Δ_m	Regime	Mean utility gain over baselines	Mean false-transfer rate
0.00	deterministic / robust / stochastic	0.065 to 0.074	0.023 to 0.033
0.05	deterministic / robust / stochastic	0.061 to 0.070	0.043 to 0.047
0.10	deterministic / robust / stochastic	0.056 to 0.062	0.058 to 0.063

C.2 Symbolic-Theorem Audit Outcomes

Table 8 restates symbolic audit closures used in the formal-to-evidence handoff.

Table 8: Symbolic theorem-audit outcomes used for formal closure checks. All listed obligations passed in the executed symbolic audit, but passing symbolic checks alone does not remove regime assumptions; interpretation still depends on the empirical gates discussed in the main text.

Obligation	Check	Status
Affine invariance closure (Lemma 5.2)	Affine co-transform identity	pass
Stability guard closure (Theorem 5.3)	Guard-margin implication form	pass
Scalarization form closure (Eq. 8)	Weighted objective symbolic form	pass
Counterexample boundary closure (Theorem 5.4)	Counterexample witness comparison	pass
Descriptor score-gap closure (Eq. 9)	Descriptor-aware score gap	pass
Descriptor ordering closure (Lemma 5.5)	Ordering preservation margin	pass

D Reproducibility and Implementation Details

D.1 Seeds, Sweeps, and Execution Context

Executed sweeps used fixed seed sets per claim family, with repeated runs across objective cardinality, normalization policy, uncertainty regime, and framework lineage settings. Execution remained on a single hardware profile with fixed software defaults per comparison slice, which constrained total budget but did not alter underlying formulation or theorem obligations. For instability calibration, runs covered multiple objective counts and indicator bundles; for scalarization auditing, convex and non-convex slices were evaluated with matched comparator families; for transportability analysis, descriptor perturbation levels and uncertainty regimes were crossed systematically.

D.2 Software Lineage and Version Controls

Framework lineage was logged with version and license metadata. The key finding for reproducibility is not that one framework is uniquely reliable, but that default configurations can induce measurable drift. This motivates two reporting requirements for future studies: (i) always record default-operator mode and version pins, and (ii) provide harmonized-default replay results when claim-critical conclusions are sensitive to implementation context.

D.3 Negative-Result Policy and Interpretation

Negative results were retained in structured ledgers for all claim families. In this manuscript, negative evidence plays a constructive role: it identifies where recommendation confidence should be reduced and where future experiments

should focus. This policy avoids survivorship bias in survey conclusions and provides a clearer path for incremental evidence strengthening.

E Additional Citation-Dense Context

To support readers who want a compact but dense citation bridge, we summarize additional context here. Foundational multicriteria modeling and survey synthesis references include Pedrycz et al. (2010), Zopounidis & Doumpos (2017), Roijers & Whiteson (2017), Saini & Saha (2021), and Roijers et al. (2013). Method-specific expansions include weighted and normal-boundary approaches (Marler & Arora, 2009; Messac et al., 2003; Siddiqui et al., 2012; Kim & de Weck, 2004; Wagner et al., 2025), evolutionary and many-objective developments (Deb et al., 2002; Zhang & Li, 2007; Zaenudin & Kistijantoro, 2016; Zheng et al., 2019; Bechikh et al., 2016; Azzouz et al., 2016; Zou et al., 2020; Li et al., 2024; ROMERO, 1991), uncertainty and interactive formulations (Saadati & Oveisiha, 2023; Zhao et al., 2023; Shavazipour et al., 2024; Liang et al., 2024; Zheng et al., 2022; Hakanen & Allmendinger, 2021; Heiskanen, 2001), and software-benchmark ecosystem resources (maintainers, 2026a;b;c;d; Ahrari et al., 2026; Picard & Schiffmann, 2021; Yu et al., 2020; Yagoubi & Bederina, 2023). Additional methodological and application-adjacent records that informed contradiction framing are (Campelo et al., 2008; Kabor et al., 2025; Lrraga & Miettinen, 2026; WU & CHEN, 2026; Lai et al., 2021; Zhou et al., 2021; Goswami, 2020; Anes & Abreu, 2025; Xu & Yuan, 2024; Xue et al., 2024; Deb & Jain, 2014; Jain & Deb, 2014; Roy, 1990; Monarchi et al., 1976; Benayoun et al., 1971). This appendix block ensures broad representative coverage while keeping main-text exposition focused on the central argument.