https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 72, No. 4, April 2026, pp. 3175–3203
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Relatively Robust Multicriteria Decisions

Thomas A. Weber[a]

[a] Chair of Operations, Economics and Strategy, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
Contact: thomas.weber@epfl.ch, https://orcid.org/0000-0003-2857-4754 (TAW)

Received: February 7, 2025
Revised: May 26, 2025
Accepted: May 28, 2025
Published Online in Articles in Advance: August 14, 2025

https://doi.org/10.1287/mnsc.2025.00510

Abstract. For a general multicriteria decision problem with linear scalarization and unknown weights, we propose relatively robust decisions, which are Pareto-efficient and at the same time maximize a performance index. The latter measures the worst-case ratio, attained by the weighted objective relative to its maximum value, with respect to all possible weights. The main results include a simple boundary representation of the performance index as the minimum of criterion-specific performance ratios, and a computationally simple method of determining a relatively robust decision up to any prespecified performance tolerance by maximizing an ε-augmented performance index. The proposed method relies merely on the continuity of all criterion functions and the compactness of the set of feasible decisions which may be nonconvex. This imposes no restrictions at all for any finite action set. A notable feature of our method is that it endogenously yields the tradeoffs between all criteria, including a performance guarantee relative to decisions justified by any other weighting. A number of structural results, examples, and applications are provided, as well as generalizations to allow for limited weight ambiguity, criterion ambiguity, and generalized aggregation of criteria based on an axiomatic foundation.

History: Accepted by Peng Sun, optimization and decision analytics.

Keywords: criterion ambiguity • multicriteria decisions • relative regret • robust optimization • weight ambiguity

1. Introduction

In real-world decision making, evaluating alternatives often involves multiple, sometimes conflicting criteria. Whether considering investment portfolios under Environmental, Social, and Governance (ESG) parameters, selecting products based on bundles of attributes, or valuing companies for both profitability and sustainability, decision makers must navigate tradeoffs between competing objectives. Similarly, lifecycle environmental impact assessments—such as comparing vehicles with electric, combustion, or hybrid engines—require reconciling diverse metrics like emissions, cost, and resource consumption. These complex but often inevitable comparisons highlight the critical need for robust multicriteria optimization frameworks.

Multicriteria optimization involves identifying solutions that balance conflicting objectives in a manner consistent with the decision maker's priorities. Traditional approaches often rely on scalarization techniques, where multiple criteria are combined into a single weighted objective function using weighted sums. However, these methods depend heavily on the precise specification of weights, which are rarely known a priori and can be difficult to justify. This uncertainty complicates the search for decisions that are both Pareto-efficient and robust to variations in tradeoff preferences. To address these challenges, we specialize the concept of relatively robust decisions by Weber (2023) so as to refer to decisions that achieve Pareto-efficiency among all relevant criteria while also maximizing a performance index designed to account for the ambiguity in weights. Specifically, the performance index measures the worst-case (WC) ratio attained by the weighted objective relative to its maximum value over all possible weight configurations. By focusing on worst-case performance, this framework provides guarantees of robustness and tradeoff transparency, which are particularly valuable in high-stakes or uncertain decision contexts.

1.1. Practical Examples

The approach developed here can be used for virtually all multicriteria decision problems, such as the following three example applications.[1]

- ESG Investing: Investors face the challenge of balancing financial returns with social and environmental impact. For example, a fund manager might evaluate portfolios based on criteria such as profitability, carbon

footprint, and diversity inclusion. Weighting these criteria is inherently subjective, and the optimal portfolio might vary widely depending on the chosen weights. The relatively robust optimization framework proposed here enables the identification of investment strategies that remain robust across different weight configurations, offering a performance guarantee regardless of the specific preferences.[2]

- Lifecycle Assessment of Vehicles: Consider evaluating the environmental impact of electric, combustion, and hybrid cars. Criteria might include greenhouse gas emissions, energy efficiency, and resource use (see, e.g., Hawkins et al. 2012). A relatively robust decision could pinpoint vehicle types or designs that perform well across a broad range of plausible weightings, resulting in a balanced and defensible choice.
- Product Evaluation and Design: Companies frequently design products to optimize attributes such as cost, durability, and aesthetic appeal. For example, in designing a smartphone, decision makers must weigh the importance of battery life, screen quality, and price.[3] Relatively robust multicriteria optimization helps determine design specifications that ensure competitive performance across various market segments with diverse preferences, valuing the availability of different attributes with different weights.

For the application of our method, one only needs that there exists a default action, that is, a baseline alternative that performs adequately across all criteria (which can always be achieved by reindexing evaluation scales), together with the technical assumption that criteria are continuous in actions, and that the finite-dimensional action set is closed and bounded (i.e., compact).

1.2. Literature

1.2.1. Origins of Multicriteria Decision Making. The idea of multicriteria optimization in Economics can be traced back to the distribution of resources among different individuals, leading to a set of undominated solutions such as the "contract curve" proposed by Edgeworth (1881, p. 21) for a simple exchange economy, and more generally a set of efficient outcomes as implied by Pareto (1894; 1897, sections 721–723),[4] which cannot be improved upon for one agent without making another agent worse off. The latter avoids a direct interpersonal comparison of individuals' utilities (or "ophelimities" in Pareto's terminology); see also Harsanyi (1955) as well as Keeney and Raiffa (1993, chapter 10) who explore group utility functions. The drawback of such an agnostic approach to optimality with multiple criteria is that the set of Pareto-optimal allocations, because of its typically large cardinality, offers only imperfect guidance about which solution should actually be implemented. For example, in the two-agent exchange economy, the contract curve usually includes allocations that attribute all resources to any single individual, which almost completely undermines the notion of *multi*criteria optimization.

In Operations Research, "goal programming" refers to the notion of minimizing the weighted deviation from criterion-specific targets (Charnes and Cooper 1961). The technique was first employed in the context of executive compensation based on different "factors" (i.e., criteria) using linear programming techniques (Charnes et al. 1955). This basic approach is usually applied to a "utopian" (or "ideal") target point, which corresponds to the (generically infeasible) vector of individually maximized criteria. Depending on the distance measure (e.g., a weighted Chebyshev distance; see, e.g., Steuer 1986, chapter 14),[5] the corresponding solutions trade off among criteria according to the specified weights, and they are naturally Pareto-efficient.[6] Instead of minimizing the distance to the ideal point (in the outcome space), it is also possible to maximize the distance to a (generically infeasible) "nadir" point, which contains the minimum value of each individual criterion on the Pareto-efficient set.[7] In this approach, known as "compromise programming" (Zeleny 1974), the appropriate choice of the weights for the different criteria remains the critical point, and in our view, very little satisfying progress has been made in the assignment of weights without imposing subjectivity, which arguably amounts to picking a solution from the Pareto-efficient set. For instance, based on an exogenous ranking of the criterion importance, it is possible to apply an ordered weighted average (Yager 1988) which in turn can be related to compromise programming (Zarghami and Szidarovszky 2010, Wang and Fu 2020).[8] Besides being subjective from the start by requiring the imposition of a preference order on criteria, this method does not provide any nontrivial performance guarantee relative to other choices of weights and/or importance rankings that might be plausible for other decision makers.[9]

1.2.2. Relative Robustness. By contrast, we approach the selection of weights from the standpoint of relative robustness (Weber 2023), involving only comparisons between feasible points, thus avoiding fictitious targets such as the aforementioned utopian and nadir points. Rather than minimizing a fixed distance metric with specific weights, the proposed method evaluates performance in terms of the worst-case tradeoffs among criteria, ensuring a solution that remains defensible regardless of the exact weighting chosen ultimately. The underlying robustness measure, equivalent to relative regret, has been used in computer science to evaluate the relative performance of algorithms (Sleator and Tarjan 1985, Ben-David and Borodin 1994), for the scenario-based evaluation of operational decisions (Kouvelis and Yu 1997), in price discrimination (Han and Weber 2023), robust optimization (Weber 2024), and fair resource allocation (Goel et al. 2009). The idea of absolute regret (AR)

is due to Savage (1951), based on the maximin robustness approach by Wald (1945) in his general treatment of sequential decision problems. The main issue with absolute regret is that it is inherently sensitive to the scale of the reference point. This sensitivity may make it impossible to derive reasonable performance guarantees, such as ensuring positive profits in a monopoly pricing problem with unknown demand (Weber 2025), because a zero-profit reference point is intrinsically small-scale. A *relative* robustness approach, we argue, yields more acceptable results, particularly in the context of multicriteria optimization, where changing the units of any given criterion would usually affect solutions that are based on absolute performance measures.

1.2.3. Connection to Distributionally Robust Optimization. Because one can reinterpret a (normalized) weight vector as a probability distribution, our approach is naturally related to distributionally robust optimization (DRO), where the true probability distribution governing uncertain parameters is unknown but assumed to lie within a known ambiguity set. Important early contributions include Delage and Ye (2010), who studied DRO with moment-based sets, and Ben-Tal et al. (2013), who developed tractable reformulations for DRO with Wasserstein ambiguity sets. More recently, Blanchet and Murthy (2019) and Mohajerin Esfahani and Kuhn (2018) developed general formulations based on Wasserstein balls, offering strong out-of-sample guarantees. Whereas DRO typically focuses on expectations or risk measures over stochastic uncertainty, our approach generalizes worst-case robustness to a multicriteria setting without requiring a probabilistic model. This positions our proposed framework as a deterministic analogue to DRO, where ambiguity arises from unknown relative preferences and state-dependent criteria rather than unknown probability distributions.

1.2.4. Relative Evaluations. The proposed approach to robust multicriteria decisions is entirely relative, in the sense that the question underlying all of our developments is "How well am I doing relative to how well I could be doing?" Indeed, the idea that the size of an object can be judged only relative to other objects goes back at least to the Taoist writings of Zhuang Zhou in the fourth century BC.[10] In Economics, Cournot (1838) was among the first to note that there is no absolute value ("*Il n'y a pas de valeurs absolues*," p. 22) and that inference from a social system can be likened to the observation of astronomical objects and their relative positions to each other (pp. 15–16), concluding by analogy that the concept of value is fundamentally relative ("*Il y en a en ce sense que des valeurs relatives*," p. 18). In fact, human perception is inherently relative, as demonstrated by Weber (1846) and his student Fechner (1860) in extensive experiments which showed that across different senses (e.g., hearing, touch, and vision) the minimum perceptible difference is proportional to the current level of the stimulus, giving rise to the Weber-Fechner law of psychophysics. Similarly, in an economic context, it is often relative reference points such as one's current wealth level (Kahneman and Tversky 1979) or the outcomes experienced by others (Loewenstein et al. 1989) that tend to determine human behavior. For example, when pondering whether to purchase a product from a cheaper store within walking distance, humans base decisions less on absolute gains than on the prospective relative savings in expenditure (Kahneman and Tversky 1984).

Beyond the aforementioned congruence with human perception, there are other practical arguments for relative evaluations. First, there is the *insensitivity to scale*, common to all relative measures such as internal rate of return (IRR), demand elasticity, profit margin, or relative regret, which allows for a direct comparison across different sizes. For example, with IRR one can readily benchmark projects of different financial magnitudes against outside investment options (of comparable risk), whereas the equivalent absolute indicator of net present value (NPV) remains silent about the required absolute investment (see, e.g., Weber 2014).[11] Second, normalization facilitates *fairness and equity*. When stakeholders have differing capacities or baselines, a relative comparison may help to ensure fairness (Goel et al. 2009). Third, relative evaluation criteria allow for *comparability across contexts*. For example, demand elasticity (as introduced by Marshall 1890, p. 162) is a unit-free relative measure that allows one to compare the changes of demand relative to price changes across widely differing goods, irrespective of the underlying unity of measurement (e.g., units of cars versus units of electric power). This last point is especially salient for multicriteria decision making, as the units (and inherent scale) for different criteria generally differ, so that a robustness criterion (and robust decision) should remain unaffected if the values of a given criterion are all multiplied by 10, for example. The relatively robust framework developed here uses a relative worst-case performance perspective. This methodology not only identifies solutions with reliable trade-off characteristics, but also provides a transparent representation of the tradeoffs themselves. These tradeoffs are reflected in a robust weight vector consistent with a robust solution.

1.3. Outline

The remainder of this paper is organized as follows: Section 2 discusses the multicriteria optimization problem and associated comparative statics, together with the performance index for a robust evaluation of different decisions. Section 3 introduces pseudo-robustness and Pareto-efficiency, which together characterize relatively robust decisions. Here we also provide a computational

approach for determining a relatively robust decision up to any given performance tolerance, and we allow for the possibility of close-to-arbitrary restrictions in the set of admissible weights, for example, based on a priori knowledge about physical constraints or a given priority ranking of criteria. In addition, there are extensions to criterion ambiguity and general aggregation of criteria, followed by a practical guide for how to apply the method. Section 4 focuses on discrete applications where our framework relies on virtually no assumptions, so the approach can be entirely data-driven. Section 5 concludes.

2. Basic Framework

Let $\mathcal{X} \subset \mathbb{R}^m$ be a nonempty, compact *action set*, for a given integer $m \geq 1$. Consider a decision maker who faces the *multicriteria optimization problem* of having to select an *action* (or *decision*, or *point*) $\boldsymbol{x} \in \mathcal{X}$ so as to "simultaneously maximize" the continuous functions $f_i : \mathcal{X} \to \mathbb{R}_+$, for $i \in \mathcal{N} = \{1, \ldots, n\}$, each of which is referred to as a *criterion*, where $n \geq 1$ is a given integer.

Remark 1. (i) The continuity of each criterion f_i ensures that small perturbations in the action $\boldsymbol{x}$ yield correspondingly small changes in the objective value. Notably, this assumption is trivially satisfied at any *isolated* point of $\mathcal{X}$.[12] In particular, this means that there is no imposed regularity requirement when the action set is finite. (ii) The requirement that each f_i be nonnegative is without loss of generality: any real-valued (continuous) criterion $\hat{f}_i$ can be translated as $f_i = \hat{f}_i - \hat{f}_i^{\bullet} \geq 0$, where $\hat{f}_i^{\bullet} = \min_{\boldsymbol{x} \in \mathcal{X}} \hat{f}_i(\boldsymbol{x})$ denotes the minimum value of $\hat{f}_i$.[13]

2.1. Decision Problem

To evaluate goal achievement for any decision $\boldsymbol{x} \in \mathcal{X}$, the decision maker considers a scalarization of his multicriteria optimization problem by means of a *weighted objective*,

$$F(\boldsymbol{x}|\boldsymbol{\lambda}) = \sum_{i=1}^{n} \lambda_i f_i(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{X}, \tag{1}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \Delta = \{\boldsymbol{w} \in \mathbb{R}_+^n : \|\boldsymbol{w}\|_1 = 1\}$ is a (*normalized*) *vector of weights* (with $\lambda_i \geq 0$ for all $i \in \mathcal{N}$, and $\lambda_1 + \cdots + \lambda_n = 1$). Maximizing the weighted objective yields the *set of ex post optimal decisions*,

$$\mathcal{X}(\boldsymbol{\lambda}) = \arg\max_{\boldsymbol{x} \in \mathcal{X}} F(\boldsymbol{x}|\boldsymbol{\lambda}), \quad \boldsymbol{\lambda} \in \Delta, \tag{2}$$

which by the extreme-value theorem (Rudin 1976, theorem 4.16, p. 89) is nonempty. Moreover, by the maximum theorem (Berge 1963, p. 116) the (set-valued) mapping $\mathcal{X} : \Delta \rightrightarrows \mathcal{X}$ is compact-valued and upper semicontinuous.

Remark 2. The weighted objective is homogeneous of degree one, that is, for all $\alpha > 0$ and $\boldsymbol{\lambda} \in \Delta$, it is $F(\cdot|\alpha\boldsymbol{\lambda}) = \alpha F(\cdot|\boldsymbol{\lambda})$, whereas the set of ex post optimal decisions is homogeneous of degree zero, in the sense that $\mathcal{X}(\alpha\boldsymbol{\lambda}) = \mathcal{X}(\boldsymbol{\lambda})$, for all $\alpha > 0$ and $\boldsymbol{\lambda} \in \Delta$. It is therefore possible to extend the definition of $F(\boldsymbol{x}|\cdot)$ in Equation (1) and the definition of $\mathcal{X}(\cdot)$ in Equation (2) to a domain containing any nonzero weight vector $\boldsymbol{w} \in \mathbb{R}_+^n \setminus \{\boldsymbol{0}\}$, because a *unique* normalized weight $\boldsymbol{\lambda} = \alpha\boldsymbol{w} \in \Delta$, with $\alpha = 1/(\sum_{i=1}^n w_i) > 0$, is always available:[14]

$$F(\boldsymbol{x}|\boldsymbol{w}) \triangleq (1/\alpha) F(\boldsymbol{x}|\boldsymbol{\lambda}), \quad \boldsymbol{x} \in \mathcal{X}, \tag{1'}$$

and

$$\mathcal{X}(\boldsymbol{w}) \triangleq \mathcal{X}(\boldsymbol{\lambda}) = \arg\max_{\boldsymbol{x} \in \mathcal{X}} (1/\alpha) F(\boldsymbol{x}|\boldsymbol{\lambda}). \tag{2'}$$

Remark 3. Limited ambiguity, that is, allowing for weights in a (nonempty, compact) subset of Δ, is discussed in Section 3.7. Criterion ambiguity is treated in Section 3.8, and Section 3.9 investigates the use of general multicriteria objectives. All three generalizations can be treated, after suitable adjustment, within the basic framework.

To keep matters nontrivial, we assume that there exists a (feasible) *default decision* ($\boldsymbol{x}_d$) such that the decision maker's weighted objective is positive, that is,

$$\exists \boldsymbol{x}_d \in \mathcal{X} : \; F(\boldsymbol{x}_d|\boldsymbol{\lambda}) > 0, \; \boldsymbol{\lambda} \in \Delta. \tag{N}$$

The nontriviality condition (N) ensures that the *ex post optimal objective* (or *value function*) is positive:

$$F^*(\boldsymbol{\lambda}) = \max_{\boldsymbol{x} \in \mathcal{X}} F(\boldsymbol{x}|\boldsymbol{\lambda}) = F(\hat{\boldsymbol{x}}|\boldsymbol{\lambda}) \geq F(\boldsymbol{x}_d|\boldsymbol{\lambda}) > 0, \quad \hat{\boldsymbol{x}} \in \mathcal{X}(\boldsymbol{\lambda}), \; \boldsymbol{\lambda} \in \Delta.$$

The sign-definiteness of the ex post optimal objective is critical for its role as a reference, against which the goal achievement of any feasible decision can be compared.

Remark 4. The nontriviality condition (N) can be satisfied without affecting $\mathcal{X}(\cdot)$, that is, without changing any set of ex post optimal decisions. It is sufficient to consider the translated criterion $\hat{f}_i = f_i + s_i$ instead of f_i, for all $i \in \mathcal{N}$, using a suitable shift $\boldsymbol{s} = (s_1, \ldots, s_n) \in \mathbb{R}_{++}^n$, in which case $\hat{F}(\boldsymbol{x}|\boldsymbol{\lambda}) = F(\boldsymbol{x}|\boldsymbol{\lambda}) + c_0(\boldsymbol{\lambda}) > 0$, because $c_0(\boldsymbol{\lambda}) = \boldsymbol{\lambda} \cdot \boldsymbol{s}$ is strictly positive (as it is bounded from below by $\min\{s_1, \ldots, s_n\} > 0$).

2.2. Comparative Statics

What happens to the criteria at the optimum when shifting weight from one criterion to another? At the optimum, one would naturally expect that a criterion which receives relatively more weight than before cannot decrease, whereas a criterion that receives relatively less weight cannot increase. The following result formalizes this intuition for a weight shift from one criterion to another.

Proposition 1. *Let $\delta_{ij} = \boldsymbol{e}_i - \boldsymbol{e}_j$, where $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ denote the i-th and j-th Euclidean unit vectors, respectively. Consider $\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}} \in \Delta$ such that $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda} + \varepsilon\delta_{ij}$ for some $\varepsilon > 0$ and some*

$i,j \in \mathcal{N}$ with $i \neq j$. Then for any $(\boldsymbol{x},\hat{\boldsymbol{x}}) \in \mathcal{X}(\boldsymbol{\lambda}) \times \mathcal{X}(\hat{\boldsymbol{\lambda}})$ it is $f_i(\boldsymbol{x}) \leq f_i(\hat{\boldsymbol{x}})$ and $f_j(\hat{\boldsymbol{x}}) \leq f_j(\boldsymbol{x})$.

A transfer from the j-th criterion weight λ_j to the i-th criterion weight λ_i augments the optimal value of criterion i and lowers the optimal value of criterion j (at least weakly). Criteria other than i and j, whose weights remain constant but whose values are affected when decisions change, may go either way as a result of the weight shift. Similarly, the value function $F^*(\boldsymbol{\lambda})$ could go up or down, depending primarily on the difference between f_i and f_j at the optimum.

Remark 5. The conclusion of Proposition 1 can be applied multiple times. In particular, it can also be used for *shifts in nonnormalized weights* $\boldsymbol{w} = (w_j, \boldsymbol{w}_{-j}) \in \mathbb{R}^n_+ \setminus \{\boldsymbol{0}\}$; see Remark 2. Thus, increasing w_j is equivalent to (at most) $n-1$ successive weight transfers in the normalized weight from the components of $\boldsymbol{\lambda}_{-j}$ to λ_j, resulting in an increase of the j-th criterion at the optimum.

Remark 6. For small shifts of weight from criterion j to criterion i, the value of the ex post optimal objective $F^*(\boldsymbol{\lambda}) = F(\boldsymbol{x}^*(\boldsymbol{\lambda})|\boldsymbol{\lambda})$ goes up (resp., down) when the corresponding score difference, $f_i(\boldsymbol{x}^*(\boldsymbol{\lambda})) - f_j(\boldsymbol{x}^*(\boldsymbol{\lambda}))$, is positive (resp., negative), at the selection $\boldsymbol{x}^*(\boldsymbol{\lambda}) \in \mathcal{X}(\boldsymbol{\lambda})$.

The following result states that for a simple nonnormalized increase of the i-th weight, the value function goes up, as long as the i-th criterion is always positive at an optimum.

Lemma 1. *Let $\boldsymbol{w}, \hat{\boldsymbol{w}} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ be nonnormalized weights, such that $\hat{\boldsymbol{w}} = \boldsymbol{w} + \delta \boldsymbol{e}_i$, for a given $\delta > 0$ and a given $i \in \mathcal{N}$. Provided that $f_i(\boldsymbol{x}) > 0$, for any $\boldsymbol{x} \in \mathcal{X}(\boldsymbol{w})$, it is $F^*(\hat{\boldsymbol{w}}) > F^*(\boldsymbol{w})$.*[15]

2.3. Performance Index

The decision maker may a priori have no knowledge about which weight $\boldsymbol{\lambda} \in \Delta$ should be used to compute the weighted objective $F(\cdot|\boldsymbol{\lambda})$ in Equation (1).[16] To deal with this *ambiguity*, the decision maker evaluates any feasible decision $\boldsymbol{x} \in \mathcal{X}$ with respect to any given weight $\boldsymbol{\lambda} \in \Delta$ by the *performance ratio*,[17]

$$\varphi(\boldsymbol{x}|\boldsymbol{\lambda}) = \frac{F(\boldsymbol{x}|\boldsymbol{\lambda})}{F^*(\boldsymbol{\lambda})} \in [0,1], \tag{3}$$

which is continuous on $\mathcal{X} \times \Delta$ and naturally bounded from above by one. Its minimum with respect to all possible weights (cf. Endnote 13),

$$\rho(\boldsymbol{x}) = \min_{\boldsymbol{\lambda} \in \Delta} \varphi(\boldsymbol{x}|\boldsymbol{\lambda}) \in [0,1], \tag{4}$$

is called the *performance index* (*evaluated at* $\boldsymbol{x}$). The performance index measures the performance of the action $\boldsymbol{x}$ relative to all possible weighted-sum scalarizations of the decision maker's multicriteria decision problem. It turns out that this performance criterion depends only on the *maximized individual criteria*,

$$f_i^* = \max_{\boldsymbol{x} \in \mathcal{X}} f_i(\boldsymbol{x}), \ \ i \in \mathcal{N}. \tag{5}$$

By the nontriviality condition (N), it is $f_i^* = F^*(\boldsymbol{e}_i) \geq F(\boldsymbol{x}_d|\boldsymbol{e}_i) > 0$, for all $i \in \mathcal{N}$. Thus, all maximized individual criteria are strictly positive. The next result provides an important representation of the performance index, in terms of relative goal achievement of a decision, relative to the various maximized individual criteria.

Proposition 2. *The performance index is equal to*

$$\rho(\boldsymbol{x}) = \min_{i \in \mathcal{N}} \phi_i(\boldsymbol{x}), \ \ \boldsymbol{x} \in \mathcal{X}, \tag{6}$$

where $\phi_i(\boldsymbol{x}) = f_i(\boldsymbol{x})/f_i^* \in [0,1]$, for all $(\boldsymbol{x}, i) \in \mathcal{X} \times \mathcal{N}$.

The representation in Equation (6) expresses the performance index as the minimum of the criterion-specific performance ratios ϕ_i. Hence, to compute $\rho(\boldsymbol{x})$ as the minimum of $\varphi(\boldsymbol{x}|\boldsymbol{\lambda})$ over all weights $\boldsymbol{\lambda} \in \Delta$, it is sufficient to restrict attention to the Euclidean unit vectors $\boldsymbol{e}_i \in \Delta$, as $\phi_i(\cdot) = F(\cdot|\boldsymbol{e}_i)/F^*(\boldsymbol{e}_i)$, for all $i \in \mathcal{N}$. This simplification arises because $\varphi(\boldsymbol{x}|\cdot)$ is quasiconcave for fixed $\boldsymbol{x}$, implying that its minimum over Δ is attained at a vertex. This highlights a "perfect complementarity" among the criterion-specific performance ratios in the determination of the performance index.

Remark 7. The idea of perfect complementarity, discussed by Cournot (1838) and Edgeworth (1897), describes elements contributing to a common goal in fixed proportions. This is equivalent to evaluating the criterion (i.e., the performance index) using the Leontief production function,[18] that is, taking the minimum among the inputs ϕ_i, resulting in $\rho(\boldsymbol{x}) = \min_{i \in \mathcal{N}} \phi_i(\boldsymbol{x})$, for all $\boldsymbol{x} \in \mathcal{X}$, as in Proposition 2.

Example 1. Let $\mathcal{X} \subset \mathbb{R}^m_+$ be a compact action set with $\mathcal{X} \cap \mathbb{R}^m_{++} \neq \emptyset$, where $m \geq 2$ is an integer. Consider a decision problem with $n = 2$ criteria (containing an egalitarian and a utilitarian evaluation),

$$f_1(\boldsymbol{x}) = \min\{x_1, \ldots, x_m\} \text{ and}$$
$$f_2(\boldsymbol{x}) = (1/m)(x_1 + \cdots + x_m),$$

for all $\boldsymbol{x} = (x_1, \ldots, x_m)$, with maximized values $f_i^* = \max f_i(\mathcal{X}) > 0$, for $i \in \mathcal{N}$. The corresponding weighted objective becomes

$$F(\boldsymbol{x}|\ell) = (1-\ell) f_1(\boldsymbol{x}) + \ell f_2(\boldsymbol{x}), \ \ \ell \in [0,1],$$

where, for simplicity, the parameter ℓ replaces the normalized weight $\boldsymbol{\lambda} = (1-\ell, \ell)$. The optimal value, $F^*(\ell) = \max F(\mathcal{X}|\ell)$, can be obtained by solving a two-stage maximization problem,

$$F^*(\ell) = \max_{t \in \mathcal{T}} \{(1-\ell)t + \ell \mu(t)\}, \ \ \ell \in [0,1].$$

Here $\mathcal{T} = [t_1, t_2]$ is a compact interval, with $0 < t_1 \leq t_2 < \infty$, and the best average coordinate (subject to all

coordinates being at least of size t),

$$\mu(t) = (1/m)\max_{x\in\mathcal{X}}\{x_1 + \cdots + x_m : t \le x_1, \ldots, t \le x_m\},\ t \in \mathcal{T},$$

is a (weakly) decreasing function on $\mathcal{T}$, independent of ℓ. Because $t < \mu(t)$, for all $t \in \mathcal{T}$, it is clear that $F^*(\ell)$ must be increasing in ℓ. The interval $\mathcal{T}$ is adapted to the problem in the sense that at t_1 no constraint is imposed on the computation of the mean $\mu(t)$, and at t_2 the maximum of the smallest coordinate of any point in $\mathcal{X}$ is reached. Accordingly, one can verify that $t_2 = F^*(0) = f_1^* < f_2^* = F^*(1) = \mu(t_1)$, where $\mathcal{X}(\ell) = \arg\max_{x\in\mathcal{X}} F(x|\ell)$ and $0 < t_1 = \min_{\xi\in\mathcal{X}(1)}\{\xi_1, \ldots, \xi_m\} \le t_2$, where $\xi = (\xi_1, \ldots, \xi_m)$; see Figure 1. By means of Proposition 2 the performance index can then be written in the form

$$\rho(x) = \min\left\{\frac{f_1(x)}{f_1^*}, \frac{f_2(x)}{f_2^*}\right\} = \min\left\{\frac{f_1(x)}{t_2}, \frac{f_2(x)}{\mu(t_1)}\right\},\ x \in \mathcal{X}.$$

Here the criterion-specific performance ratios are $\phi_1 = f_1/t_2$ and $\phi_2 = f_2/\mu(t_1)$.

3. Robust Multicriteria Decision Making

3.1. Pseudo-Robustness

We refer to a decision $\hat{x} \in \mathcal{X}$ which maximizes the performance index $\rho(\cdot)$ as *pseudo-robust*. The corresponding (compact, nonempty) *set of pseudo-robust decisions*,[19]

$$\Psi = \arg\max_{x\in\mathcal{X}} \rho(x), \tag{7}$$

is not necessarily a singleton, as illustrated by the following example.

Example 2. Optimizing the performance index determined in Example 1, using the same two-stage maximization procedure, yields the optimal performance index,

$$\rho^* = \max_{x\in\mathcal{X}} \rho(x) = \max_{t\in\mathcal{T}} \min\left\{\frac{t}{t_2}, \frac{\mu(t)}{\mu(t_1)}\right\}.$$

Here, the first term in the minimand increases continuously in t, whereas the second is (weakly) decreasing and may be discontinuous; see Figure 2. As a result, at the optimal value $\hat{t} \in \mathcal{T} = [t_1, t_2]$ the two terms are about to cross, resulting in a "balancedness condition,"

$$\hat{t} = \sup\left\{t \in \mathcal{T} : \frac{t}{t_2} < \frac{\mu(t)}{\mu(t_1)}\right\}, \tag{8}$$

so that $\rho^* = \hat{t}/t_2 = \hat{t}/f_1^*$, as shown in Figure 2. Because for any $t \in \mathcal{T}$ it is $\mu(t) \ge t$, one obtains $\mu(\hat{t}) \ge \mu(t_2) \ge t_2 = f_1^*$. This implies that the optimal performance index,

$$\rho^* = \frac{\hat{t}}{f_1^*} \in \left[\frac{t_1}{f_1^*}, \frac{\mu(\hat{t})}{f_2^*}\right],$$

is always *nontrivial* (i.e., strictly positive). It attains its maximum possible value (of one) if and only if $\mu(\cdot)$ is constant on $\mathcal{T}$, that is, when $\mu(t_1) = \mu(t_2)$. The set of pseudo-robust actions, $\Psi = \{x \in \mathcal{X} : \min\{x_1, \ldots, x_m\} = \hat{t}/\rho^*\}$, may generally contain multiple elements.

3.2. Efficiency

When Ψ is not a singleton, the decision maker may have good reason to prefer one pseudo-robust decision over another, based on "efficiency" (or a lack thereof). Specifically, given two feasible actions $x, x' \in \mathcal{X}$, we say that decision x' is *more efficient* than decision x with respect to the vector of criteria $f = (f_1, \ldots, f_n)$, if and only if x'

Figure 1. Nonconvex Action Set $\mathcal{X} \subset \mathbb{R}^2_{++}$ in Example 1

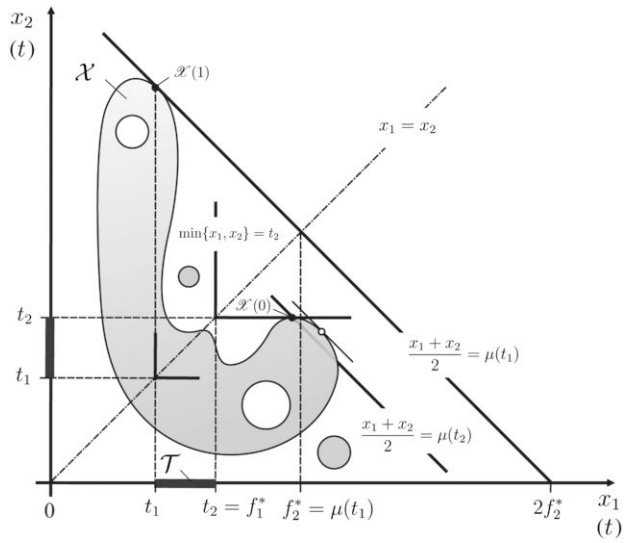
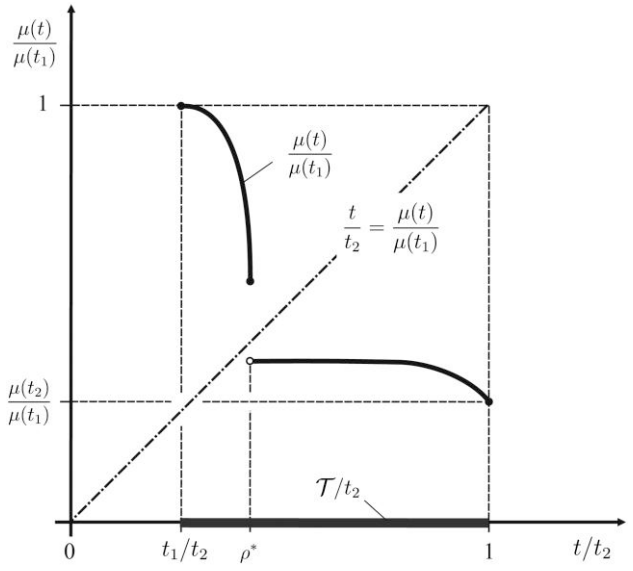


Figure 2. Qualitative Behavior of Performance Ratio $\mu(t)/\mu(t_1)$ as a Function of t/t_2 in Example 2, for the Nonconvex Action Set in Figure 1



Note. $f_1^* = t_2 > t_1$ and $f_2^* = \mu(t_1) > \mu(t_2)$.

strictly improves on at least one criterion while weakly improving on all other criteria (over their values achieved at $\boldsymbol{x}$). The corresponding preference relation $\succ$ on $\mathcal{X}$ is defined by[20]

$$\boldsymbol{x}' \succ \boldsymbol{x} \iff (\exists\, i \in \mathcal{N} :\ f_i(\boldsymbol{x}') > f_i(\boldsymbol{x}),\ \boldsymbol{f}(\boldsymbol{x}') \geq \boldsymbol{f}(\boldsymbol{x})), \quad (9)$$

for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$. The resulting set of *efficient* (or *Pareto-optimal*) decisions is

$$\mathcal{P} = \{\boldsymbol{x} \in \mathcal{X} : (\boldsymbol{f}(\boldsymbol{x}) \leq \boldsymbol{f}(\boldsymbol{x}') \Rightarrow \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}')),\ \boldsymbol{x}' \in \mathcal{X}\}. \quad (10)$$

It is easy to see that a pseudo-robust decision $\hat{\boldsymbol{x}} \in \Psi$ is not necessarily efficient, in the sense that it may be possible to find a different decision which strictly improves on at least one criterion-specific performance ratio while maintaining the optimal performance index achieved by $\hat{\boldsymbol{x}}$.

Example 3. Following up on our analysis in Example 1 and Example 2, consider the (nonconvex, compact) action set $\mathcal{X} = [0,1]^m \setminus (1/2,1]^m$, for some integer $m \geq 2$. Because $\mu(t) = 1 - 1/(2m)$, for all $t \in \mathcal{T} = [t_1, t_2]$, where $t_1 \leq t_2 = f_1^* = 1/2$ and $f_2^* = \mu(t_1)$, the optimality condition $\hat{t}/t_2 = \mu(\hat{t})/\mu(t_1) = 1$ yields $\hat{t} = 1/2$. Hence, the optimal performance index attains its maximum possible value, $\rho^* = 1$. This makes sense, because it is feasible (in $\mathcal{X}$) to attain simultaneously the highest possible minimum coordinate and the highest possible average coordinate, regardless of the weights assigned to the two objectives. Note also that the set of pseudo-robust actions,

$$\begin{aligned}\Psi &= \{\boldsymbol{x} \in \mathcal{X} : \min\{x_1, \ldots, x_m\} = \hat{t}/\rho^*\} \\ &= \{\boldsymbol{x} \in [1/2,1]^m : \exists\, i \in \mathcal{N} \text{ s.t. } x_i = 1/2\},\end{aligned}$$

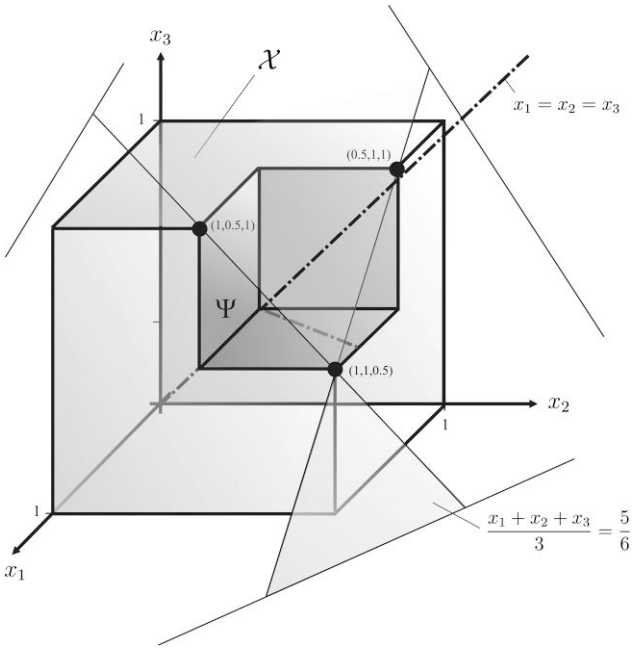
has a continuum of elements. For instance, compared with the pseudo-robust action $(1,\ldots,1)/2$, all other actions in Ψ are more efficient, particularly those in the set of Pareto-optimal decisions, $\mathcal{P} = \{(1,\ldots,1) - (\boldsymbol{e}_i/2) : i \in \mathcal{N}\}$, which is a subset of Ψ. For $m = 2$, we obtain that $\mathcal{P} = \{(0.5,1),(1,0.5)\}$. Figure 3 depicts the situation, including the set of Pareto-optimal decisions, for $m = 3$.

Remark 8. It is well known that for any weight vector $\boldsymbol{\lambda}$ with strictly positive components (ensuring all criteria are considered) an ex post optimal decision must also be efficient (see, e.g., Ehrgott 2010, proposition 3.9, p. 71). That is, for any $\boldsymbol{\lambda} \in \text{int}(\Delta)$, the corresponding optimal decision set satisfies $\mathcal{X}(\boldsymbol{\lambda}) \subset \mathcal{P}$.

3.3. Robust Decision Set

Requiring efficiency in addition to pseudo-robustness is important, because, by avoiding unnecessary shortfalls, it can only improve the weighted objective in Equation (1)—at least weakly. A decision $\hat{\boldsymbol{x}}^* \in \mathcal{X}$ is called *robust* if

Figure 3. Nonconvex Action Set in Example 3, for $m = 3$



Note. The set of Pareto-optimal decisions is $\mathcal{P} = \{(1,1,0.5),(1,0.5,1),(0.5,1,1)\}$.

it is pseudo-robust and efficient. Thus, our goal becomes to examine the properties of the *robust decision set*,

$$\mathcal{R} = \Psi \cap \mathcal{P}, \quad (11)$$

which contains all available robust decisions in $\mathcal{X}$, and to then determine a direct method for the computation of a robust decision (or an arbitrarily close approximation thereof).

Example 4. Following up on our analysis in Example 1 and Example 2, consider the (convex, compact) action set $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x}\|_2 \leq 1\}$, which is equal to the unit ball in the standard Euclidean distance, where the dimension of the underlying space is given by some integer $m \geq 2$. By direct computation, $\mu(t) = 1/\sqrt{m}$ for all $t \in \mathcal{T} = [t_1, t_2]$, and $t_1 = f_1^* = 1/\sqrt{m} = f_2^* = \mu(t_2) = t_2$. As in Example 3, the balancedness condition (8) yields $\hat{t} = t_2$, and thus an optimal performance index of $\rho^* = 1$. The set of pseudo-robust actions,

$$\begin{aligned}\Psi &= \{\boldsymbol{x} \in \mathcal{X} : \min\{x_1, \ldots, x_m\} = \hat{t}/\rho^*\} \\ &= \{(1/\sqrt{m}, \ldots, 1/\sqrt{m})\},\end{aligned}$$

is a singleton. On the other hand, the set of Pareto-optimal actions,

$$\mathcal{P} = \{\boldsymbol{x} \in \mathbb{R}^m_+ : \|\boldsymbol{x}\|_2 = 1\},$$

contains a continuum of elements, including the only pseudo-robust decision, $(1,\ldots,1)/\sqrt{m}$.

In Example 3, we examined a problem where $\mathcal{P} \subsetneq \Psi$ (so $\mathcal{R} = \mathcal{P}$), whereas in Example 4, for the same

multicriteria decision problem with a different action set, it was $\Psi \subsetneq \mathcal{P}$ (so $\mathcal{R} = \Psi$). Neither of these two extremes might apply, in which case $\mathcal{R} \notin \{\Psi, \mathcal{P}\}$, as illustrated next.

Example 5. Mixing and matching features from Example 3 and Example 4, let us consider the (nonconvex, compact) action set $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\|_2 \leq 1\} \setminus (1/2, 1]^2$ in the Euclidean plane. The corresponding set of Pareto-optimal actions is given by the intersection of the unit circle with both the action set $\mathcal{X}$ and the positive quadrant $\mathbb{R}^2_+$, so $\mathcal{P} = \{\boldsymbol{x} \in \mathcal{X} \cap \mathbb{R}^2_+ : \|\boldsymbol{x}\|_2 = 1\}$. Meanwhile, the set of pseudo-robust actions is $\Psi = \{\boldsymbol{x} \in \mathcal{X} : \min\{x_1, x_2\} = 1/2\}$. Thus, by Equation (11) it is $\mathcal{R} = \Psi \cap \mathcal{P} = \{(1/2, \sqrt{3}/2), (\sqrt{3}/2, 1/2)\}$, which in this setting means $\mathcal{R} \notin \{\Psi, \mathcal{P}\}$.

The existence of robust actions is ensured by the next result, together with the fact that the robust decision set $\mathcal{R}$ must be closed and bounded (i.e., compact).

Lemma 2. *The robust decision set $\mathcal{R}$ is nonempty and compact.*

The proof of Lemma 2 starts by noting that the set $\mathcal{P}$ of efficient actions is compact because it must be bounded (by the boundedness of the encompassing action set $\mathcal{X}$) and closed (by the continuity of the criteria). One can then construct a sequence of pseudo-robust actions in the (nonempty) compact set Ψ which might be inefficient (or else $\mathcal{R} \neq \emptyset$ holds true immediately). But each inefficient pseudo-robust action suggests the existence of a more efficient action, which incidentally must also be pseudo-robust. Given that Ψ is compact, the Bolzano-Weierstrass theorem (Berge 1963, p. 67) then guarantees the existence of a converging subsequence of pseudo-robust actions, with a limit that must be a feasible decision which is both pseudo-robust and efficient, so $\mathcal{R} \neq \emptyset$. Compactness of the robust decision set then follows, as it has to be both closed and bounded.

Lemma 3. *The optimal performance index ρ^* is such that* $\max \rho(\mathcal{P}) = \rho^* = \max \rho(\mathcal{X})$ *and* $\rho(\mathcal{R}) = \rho(\Psi) = \{\rho^*\}$.

The preceding result (re)states the fact that the decision maker can restrict attention to efficient actions when maximizing the performance index, meaning that there always exists an efficient action which attains the optimal performance index ρ^*; this action is by definition robust. Conversely, any robust action necessarily achieves a performance index of ρ^*, which is quite straightforward in light of both Lemma 2 and the definition of the robust decision set in Equation (11).

3.4. Robust Decisions: Computation

How can one determine a (relatively) robust decision? By Lemma 3 we can limit attention to maximizing the performance index over all efficient decisions in our search for robust decisions. Thus, to be guaranteed an efficient decision which is also approximately pseudo-robust, by virtue of Proposition 2 we introduce the *ε-augmented performance index*,

$$\Phi_\varepsilon(\boldsymbol{x}) = (1-\varepsilon)\min_{i\in\mathcal{N}} \phi_i(\boldsymbol{x}) + (\varepsilon/n)\sum_{i\in\mathcal{N}} \phi_i(\boldsymbol{x}),\ (\boldsymbol{x}, \varepsilon) \in \mathcal{X} \times [0,1], \tag{12}$$

so that $\Phi_0 = \rho = \min_{i\in\mathcal{N}} \phi_i$, and $\Phi_1 = (1/n)\sum_{i\in\mathcal{N}} \phi_i$. Because $\Phi_\varepsilon(\cdot)$ is a continuous function, its maximizer, referred to as the *set of ε-robust actions*,

$$\mathscr{R}_\varepsilon = \arg\max_{\boldsymbol{x}\in\mathcal{X}} \Phi_\varepsilon(\boldsymbol{x}),\ \varepsilon \in [0,1], \tag{13}$$

is nonempty and compact, again by virtue of the extreme-value theorem and the maximum theorem. It is useful, for our further analysis, to shift the perspective from the available decisions in $\mathcal{X}$ to their respective consequences (or "outcomes") in terms of their robustness performance, given the prevailing weight ambiguity.

Remark 9. The (nonempty, compact) outcome space,

$$\mathcal{Y} = \{\boldsymbol{y} \in [0,1]^n : y_i = \phi_i(\boldsymbol{x}),\ (\boldsymbol{x}, i) \in \mathcal{X} \times \mathcal{N}\}, \tag{14}$$

contains the vectors $\boldsymbol{y}$ of criterion-specific performance ratios, achieved by the available decisions $\boldsymbol{x}$, so $\mathcal{Y} = \phi(\mathcal{X})$, where $\phi = (\phi_1, \ldots, \phi_n)$. As several feasible decisions might result in the same score vector, the mapping $\phi : \mathcal{X} \to \mathcal{Y}$ may not be one-to-one. Consider now the ε-augmented performance index in the outcome space,

$$\hat{\Phi}_\varepsilon(\boldsymbol{y}) = (1-\varepsilon)\min_{i\in\mathcal{N}} y_i + (\varepsilon/n)\sum_{i\in\mathcal{N}} y_i,\ (\boldsymbol{y}, \varepsilon) \in \mathcal{Y} \times \mathbb{R}_+.$$

Using ideas from Example 1, maximization of this weighted objective yields

$$\hat{\Phi}^*_\varepsilon = \max_{\boldsymbol{y}\in\mathcal{Y}} \Phi_\varepsilon(\boldsymbol{y}) = \max_{t\in\mathcal{T}} \{(1-\varepsilon)t + \varepsilon\hat{\mu}(t)\},\ \varepsilon \in [0,1],$$

where $\mathcal{T} = [t_1, t_2]$ is a suitable compact interval, and

$$\hat{\mu}(t) = \max\left\{(1/n)\sum_{i\in\mathcal{N}} y_i : t \leq y_i,\ (\boldsymbol{y}, i) \in \mathcal{Y} \times \mathcal{N}\right\},\ t \in \mathcal{T},$$

denotes the *average* (*criterion-specific*) *performance ratio*. The interval boundaries t_1, t_2 of $\mathcal{T}$, with $0 < t_1 \leq t_2 \leq 1$, are given by

$$t_1 = \min_{\boldsymbol{y}\in\mathcal{Y}} \min_{i\in\mathcal{N}} y_i \quad \text{and} \quad t_2 = \max_{\boldsymbol{y}\in\mathcal{Y}} \min_{i\in\mathcal{N}} y_i.$$

In addition, one can easily verify that

$$\hat{\Phi}^*_0 = t_2 = \rho^* = \max \rho(\mathcal{X}) \text{ and}$$
$$\hat{\Phi}^*_1 = \hat{\mu}(t_1) = (1/n)\max_{\boldsymbol{y}\in\mathcal{Y}} \sum_{i\in\mathcal{N}} y_i \geq t_2.$$

Proposition 1 implies that any selection $t_\varepsilon \in \mathscr{T}_\varepsilon = \arg\max_{t\in\mathcal{T}}\{(1-\varepsilon)t + \varepsilon\hat{\mu}(t)\}$ must be decreasing in ε, and similarly, $\hat{\mu}(t_\varepsilon)$ must be increasing in ε, at least

weakly. The latter also follows directly from the monotonicity of t_ε, because $\hat{\mu}(t)$ must be nonincreasing in t.

Lemma 4. *Let* $\varepsilon \in (0,1]$. (i) $\mathscr{R}_0 = \Psi$. (ii) $\mathscr{R}_\varepsilon \subset \mathcal{P}$. (iii) *If* $\boldsymbol{x} \in \mathscr{R}_\varepsilon \setminus \Psi$ *and* $\hat{\boldsymbol{x}} \in \Psi$, *then there exists* $\varepsilon' \in (0,\varepsilon)$ *such that* $\Phi_{\hat{\varepsilon}}(\boldsymbol{x}) < \Phi_{\hat{\varepsilon}}(\hat{\boldsymbol{x}})$, *for all* $\hat{\varepsilon} \in [0,\varepsilon']$.

Part (i) of the preceding result notes that without ε-augmentation (i.e., for $\varepsilon = 0$) the weighted objective Φ_ε specializes to the performance index (via Proposition 2), the maximization of which produces the set of pseudo-robust decisions, as in Equation (7). Part (ii) stipulates that whenever the ε-augmentation is *nontrivial* (i.e., for $\varepsilon > 0$), maximization yields efficient actions. Finally, part (iii) means that if a decision $\boldsymbol{x}$ maximizes the ε-augmented performance index in Equation (12), for some nontrivial $\varepsilon \in (0,1]$, without being pseudo-robust, then any pseudo-robust decision $\hat{\boldsymbol{x}} \in \Psi$ would strictly improve upon $\boldsymbol{x}$ in terms of any $\hat{\varepsilon}$-augmented performance index, as long as $\hat{\varepsilon}$ (smaller than ε) lies in a sufficiently small right-neighborhood of the origin.

We are now able to establish a cornerstone property for the practice of robust multicriteria optimization, in the sense that a robust decision (i.e., an element of $\mathcal{R}$) can be obtained as a lower limit of the set of ε-robust actions, for $\varepsilon \to 0^+$.[21] In Section 3.5, we then show that $\varepsilon > 0$ can be chosen so as to guarantee an approximation of the optimal performance ρ^* up to any desired precision.

Proposition 3. *Let* $\mathcal{Q} = \underline{\mathrm{Lim}}_{\varepsilon \to 0^+} \mathscr{R}_\varepsilon$. *Then* $\mathcal{Q} \subset \mathcal{R}$ *and* $\mathcal{Q} \neq \emptyset$.

The following example shows that it is possible that $\mathcal{Q} \subsetneq \mathcal{R}$. In other words, some points in the robust decision set might not be reached using the proposed approximation procedure.

Example 6. Consider a (nonconvex, compact) action set as shown in Figure 4(a), specified by

$$\mathcal{X} = \{\boldsymbol{x} \in [0,1] \times \mathbb{R}_+ : x_2 \leq 2 + \sqrt{1 - x_1}\} \cup \{(2, \tfrac{3}{2})\}.$$

There are $n = 2$ criteria that simply measure the coordinate achievement, so $f_i(\boldsymbol{x}) = x_i$, for all $(\boldsymbol{x}, i) \in \mathcal{X} \times \mathcal{N}$. The set of efficient actions is

$$\mathcal{P} = \{\boldsymbol{x} \in [0,1] \times \mathbb{R}_+ : x_2 = 2 + \sqrt{1 - x_1}\} \cup \{(2, \tfrac{3}{2})\},$$

whereas the set of pseudo-robust actions is given by $\Psi = \{(1, x_2) : 1 \leq x_2 \leq 2\} \cup \{(2, \tfrac{3}{2})\}$. By its definition in Equation (11) the set of robust actions can be obtained as the intersection of the preceding sets:

$$\mathcal{R} = \Psi \cap \mathcal{P} = \{(1,2), (2, \tfrac{3}{2})\}.$$

At this point, let us consider the ε-augmented performance index in Equation (12). By Lemma 4 we have $\mathscr{R}_0 = \Psi$, and for $\varepsilon \in (0,1]$ the maximizer of Φ_ε is efficient, so

$$\begin{aligned}\Phi_\varepsilon^* &= \max_{\boldsymbol{x} \in \mathcal{P}} \Phi_\varepsilon(\boldsymbol{x}) \\ &= \max\left\{ \max_{x_1 \in [0,1]} \left\{ \frac{(1-\varepsilon)x_1}{2} + \frac{\varepsilon}{2}\left(\frac{x_1}{2} + \frac{2 + \sqrt{1 - x_1}}{3}\right)\right\}, \right. \\ &\qquad \left. \Phi_\varepsilon\left(2, \tfrac{3}{2}\right) \right\},\end{aligned}$$

where $\phi_i = f_i / f_i^*$, for $i \in \mathcal{N}$, with $(f_1^*, f_2^*) = (2,3)$, and $\Phi_\varepsilon(2, \tfrac{3}{2}) = (2 + \varepsilon)/4$. By direct computation one finds that the maximizer of the (nontrivially) ε-augmented performance index,[22]

$$\mathscr{R}_\varepsilon = \left\{ \left(2, \frac{3}{2}\right) \right\}, \quad \varepsilon \in (0,1],$$

is a singleton. Proposition 3 then guarantees that the lower limit of this maximizer is robust:

$$\mathcal{Q} = \underline{\mathrm{Lim}}_{\varepsilon \to 0^+} \mathscr{R}_\varepsilon = \left\{ \left(2, \frac{3}{2}\right) \right\} \subsetneq \left\{ (1,2), \left(2, \frac{3}{2}\right) \right\} = \mathcal{R}.$$

The fact that $(2, \tfrac{3}{2})$ is an isolated point (cf. Endnote 12) is not important, because one could easily connect it to $\mathcal{X}$ by adding points to the action set that are always suboptimal.[23]

3.5. Approximation Error of ε-Robust Decisions

Maximizing the ε-augmented performance index Φ_ε in Equation (12) enables us to approximate a robust decision to an arbitrary prespecified precision, as a function of $\varepsilon \in (0,1]$. The quality of any approximate decision $\boldsymbol{x}_\varepsilon \in \mathscr{R}_\varepsilon$ in Equation (11) is gauged by its *approximation error*,

$$\psi_\varepsilon = \rho^* - \rho_\varepsilon \in [0,1], \quad \varepsilon \in [0,1], \tag{15}$$

that is, the difference between the optimal performance index, $\rho^* = \max \rho(\mathcal{X})$, and the achieved performance index, $\rho_\varepsilon = \rho(\boldsymbol{x}_\varepsilon)$. This quantity measures the loss in robustness from using the ε-approximation rather than the true robust decision. For the remainder of this discussion, we keep the selection $\boldsymbol{x}_\varepsilon \in \mathscr{R}_\varepsilon$ fixed, for all $\varepsilon \in [0,1]$. If we denote by $\mu_\varepsilon = \mu(\boldsymbol{x}_\varepsilon) = (1/n)\sum_{i \in \mathcal{N}} \phi_i(\boldsymbol{x}_\varepsilon)$ the average performance ratio achieved by $\boldsymbol{x}_\varepsilon$, the *optimal ε-augmented performance index* becomes

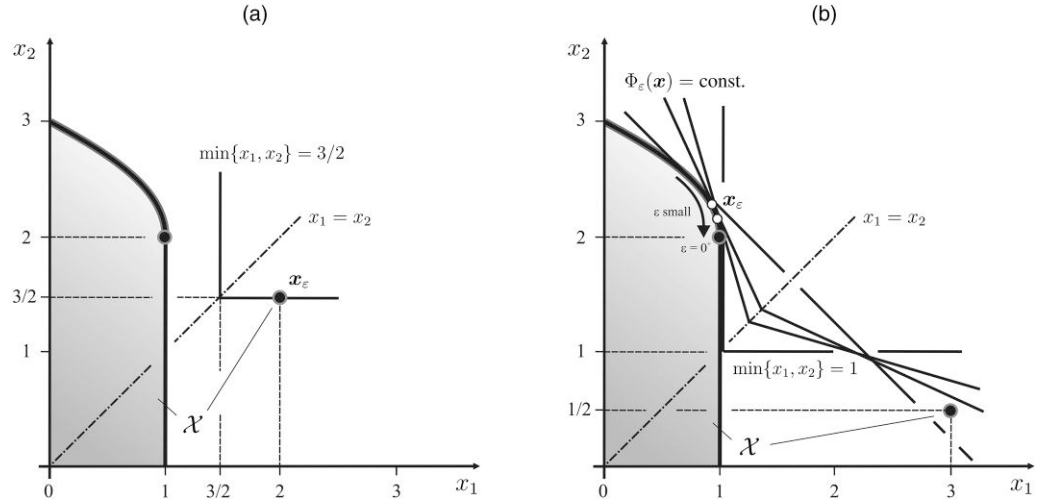
$$\Phi_\varepsilon^* = \max \Phi_\varepsilon(\mathcal{X}) = (1-\varepsilon)\rho_\varepsilon + \varepsilon\mu_\varepsilon, \quad \varepsilon \in [0,1]. \tag{16}$$

The following result guarantees continuity, as well as first- and second-order monotonicity of the optimal ε-augmented performance index.

Lemma 5. Φ_ε^* *is continuous, increasing, and convex in* $\varepsilon \in [0,1]$.

The proof of the first-order monotonicity uses the fact that $\mu_\varepsilon \geq \rho_\varepsilon$ (for all $\varepsilon \in [0,1]$), together with natural properties of an optimal solution to Equation (13) in order to establish that Φ_ε^* in Equation (16) must increase

Figure 4. Action Sets in Examples 6 and 7



Notes. (a) Action set in Example 6 with $\underline{\text{Lim}}_{\varepsilon\to 0^+}\mathscr{R}_\varepsilon = \{(2, \frac{3}{2})\} \subsetneq \mathcal{R} = \{(1,2),(2,\frac{3}{2})\}$. (b) Action set in Example 7 with $\underline{\text{Lim}}_{\varepsilon\to 0^+}\mathscr{R}_\varepsilon = \{(1,2)\} = \mathcal{R}$.

in ε (at least weakly). That the optimal ε-augmented performance index also exhibits a second-order monotonicity means that tightening the approximation parameter further and further leads to progressively slower decreases of Φ^*_ε toward $\Phi^*_0 = \rho^*$ (as $\varepsilon \to 0^+$). The convexity of the optimal ε-augmented performance index also implies its smoothness (almost everywhere), as pointed out next.

Remark 10. By the Rademacher theorem (Villani 2008, theorem 10.8, p. 222), the convexity of Φ^*_ε in $\varepsilon \in [0,1]$, which implies Lipschitz continuity, guarantees that the optimal ε-augmented performance index is differentiable almost everywhere (a.e.) on [0, 1]. The Alexandrov theorem (Villani 2008, theorem 14.25, p. 402) goes even further by establishing its second-order differentiability a.e., in the sense of having a Taylor expansion with a smaller-than-quadratic local error at almost every point $\varepsilon \in [0,1]$. By the envelope theorem (see, e.g., Mas-Colell et al. 1995, theorem M.L.1, pp. 965–966), at points of differentiability one therefore obtains:

$$\frac{d\Phi^*_\varepsilon}{d\varepsilon} = \mu_\varepsilon - \rho_\varepsilon \geq 0, \quad \dot\forall \varepsilon \in [0,1].$$

In other words, the gradient of the optimal ε-augmented performance index is (a.e.) equal to the difference between the average performance ratio (μ_ε) and the minimum performance ratio (ρ_ε) attained by the approximately robust decision $\boldsymbol{x}_\varepsilon$.

The maximized weighted objective in Equation (16) is a convex combination of ρ_ε and μ_ε; the latter both exhibit "natural" comparative statics as implied by the reweighting result in Proposition 1.

Lemma 6. (i) ρ_ε *is decreasing in* $\varepsilon \in [0,1]$. (ii) $\rho^* \geq \rho_\varepsilon$, *for all* $\varepsilon \in [0,1]$. (iii) $\lim_{\varepsilon\to 0^+}\rho_\varepsilon = \rho^*$. (iv) μ_ε *is increasing in* $\varepsilon \in [0,1]$. (v) $\mu_\varepsilon \geq \rho^*$, *for all* $\varepsilon \in [0,1]$.

Parts (i) and (iv) of Lemma 6 assert that decreasing the augmentation parameter ε can only decrease the average criterion-specific performance ratio μ_ε and at the same time increase the performance index ρ_ε, where both are achieved at a given ε-robust selection $\boldsymbol{x}_\varepsilon$. Meanwhile, by parts (ii) and (v) the optimal performance ratio $\rho^* = \max\rho(\mathcal{X})$ is bracketed by these two values, in the sense that $\rho_\varepsilon \leq \rho^* \leq \mu_\varepsilon$, for all $\varepsilon \in [0,1]$. Finally, parts (i) and (iii) establish the monotone convergence of ρ_ε to ρ^*, as $\varepsilon \to 0^+$.

The following result provides a *performance guarantee* for any ε-approximation of our robust multicriteria decision problem, alluded to at the outset of our discussion.

Proposition 4. *Fix any* $\delta > 0$. *If* $\varepsilon \leq \delta/\mu_1$, *then* $\psi_\varepsilon \leq \delta$.

A special case of the preceding result is that *for any* $\delta > 0$ *the approximation error* ψ_{δ/μ_1} *cannot exceed* δ. In other words, any ε-robust decision $\boldsymbol{x}_\varepsilon$, for $\varepsilon = \delta/\mu_1$, attains a performance index $\rho_\varepsilon \in [\rho^* - \delta, \rho^*]$, where ρ^* is the optimal performance index. The following example applies this result in an already familiar context.

Example 7. Somewhat similar to Example 6, we consider the (nonconvex, compact) action set

$$\mathcal{X} = \{\boldsymbol{x} \in [0,1] \times \mathbb{R}_+ : x_2 \leq 2 + \sqrt{1 - x_1}\} \cup \{(3, \tfrac{1}{2})\},$$

together with the same $n = 2$ criteria $f_i(\boldsymbol{x}) = x_i$, for all $(\boldsymbol{x}, i) \in \mathcal{X} \times \mathcal{N}$. Consequently, the set of efficient actions is $\mathcal{P} = \{\boldsymbol{x} \in [0,1] \times \mathbb{R}_+ : x_2 = 2 + \sqrt{1 - x_1}\} \cup \{(3, \tfrac{1}{2})\}$, and the set of pseudo-robust actions is $\Psi = \{(1, x_2) : 1 \leq x_2 \leq 2\}$. Thus, by Equation (11) the set of robust actions amounts to $\mathcal{R} = \Psi \cap \mathcal{P} = \{(1,2)\}$, which is a

singleton. By Lemma 4 the ε-augmented performance index Φ_ε in Equation (12) is efficient (i.e., $\mathscr{R}_\varepsilon \subset \mathcal{P}$), for all $\varepsilon \in (0,1]$, and $\mathscr{R}_0 = \Psi$. In particular, the optimal value of the ε-augmented performance index is

$$\Phi^*_\varepsilon = \max_{x\in\mathcal{P}} \Phi_\varepsilon(\boldsymbol{x}) = \max\left\{ \max_{x_1\in[0,1]} \left\{ \frac{(1-\varepsilon)x_1}{3} + \frac{\varepsilon}{2}\left(\frac{x_1}{3} + \frac{2+\sqrt{1-x_1}}{3}\right)\right\}, \Phi_\varepsilon(3,\tfrac{1}{2})\right\},$$

where $\phi_i = f_i/f_i^*$, for $i \in \mathcal{N}$, with $(f_1^*, f_2^*) = (3,3)$, and $\Phi_\varepsilon(3,\frac{1}{2}) = (2+5\varepsilon)/12$. For sufficiently small ε, the maximizer,

$$\mathscr{R}_\varepsilon = \left\{ \left(1 - \left(\frac{\varepsilon/2}{2-\varepsilon}\right)^2, 2 + \frac{\varepsilon/2}{2-\varepsilon}\right)\right\},$$

$$0 < \varepsilon < \frac{4}{7}\left(2 - \frac{1}{\sqrt{2}}\right) \approx 0.7388,$$

is single-valued, achieving $\Phi^*_\varepsilon = (1/24)(16 - 3\varepsilon^2)/(2-\varepsilon)$.[24] By Proposition 3 the lower limit of this maximizer for $\varepsilon \to 0^+$ is robust, and in our case:

$$\mathcal{Q} = \underline{\mathrm{Lim}}_{\varepsilon\to 0^+} \mathscr{R}_\varepsilon = \{(1,2)\} = \mathcal{R};$$

see Figure 4(b). Note also that $\mathscr{R}_1 = \{(3,\frac{1}{2})\}$, which in turn yields the average criterion-specific performance ratio achieved by the decision $\boldsymbol{x}_\varepsilon \in \mathscr{R}_\varepsilon$, for $\varepsilon = 1$:

$$\mu_1 = \frac{1}{2}\left(\frac{x_1}{f_1^*} + \frac{x_2}{f_2^*}\right)\Big|_{(x_1,x_2)=(3,\frac{1}{2})} = \frac{7}{12} \approx 0.5833.$$

Thus, to guarantee that the approximation error ψ_ε in Equation (15) cannot exceed $\delta = 5\%$, it is by Proposition 4 enough to find an action $\boldsymbol{x}_\varepsilon$ that maximizes the ε-augmented performance index Φ_ε in Equation (12) for some $\varepsilon \in (0, \delta/\mu_1] \approx (0, 8.6\%]$. Incidentally, for $\varepsilon = 8\%$, we find $\rho_\varepsilon \approx 33.32\%$, so that the realized approximation error of $\psi_\varepsilon \approx 0.01\%$ (with respect to $\rho^* = 1/3$) is actually much smaller than the prespecified 5% approximation-error bound.

It is also possible to derive a priori performance estimates without knowledge of the optimal performance index (ρ^*) just by solving the ε-approximation problem in Equation (13) for some admissible ε. Indeed, Lemma 6 (ii) and Lemma 5 together imply that

$$\rho_\varepsilon \le \rho^* \le \Phi^*_\varepsilon, \quad \varepsilon \in [0,1]. \tag{17}$$

For any $\varepsilon \in [0,1]$, let us now consider the midpoint estimator,

$$\hat{\rho}_\varepsilon = \frac{\rho_\varepsilon + \Phi^*_\varepsilon}{2} = \rho_\varepsilon + \varepsilon\,\frac{\mu_\varepsilon - \rho_\varepsilon}{2} = \rho(\boldsymbol{x}_\varepsilon) + \varepsilon\,\frac{\mu(\boldsymbol{x}_\varepsilon) - \rho(\boldsymbol{x}_\varepsilon)}{2}, \quad \varepsilon \in [0,1]. \tag{18}$$

The latter means that $\hat{\rho}_\varepsilon$ effectively approximates ρ^*, for $\varepsilon \to 0^+$.

Lemma 7. *$|\hat{\rho}_\varepsilon - \rho^*| \le \varepsilon(\mu_\varepsilon - \rho_\varepsilon)/2$, for all $\varepsilon \in [0,1]$.*

Because $\mu_\varepsilon - \rho_\varepsilon \in [0,1]$, a somewhat simpler (though generally less precise) approximation inequality than the one given in the preceding Lemma 7 is $|\hat{\rho}_\varepsilon - \rho^*| \le \varepsilon/2$, for all $\varepsilon \in [0,1]$.

Remark 11. For a robust decision $\hat{\boldsymbol{x}} \in \mathcal{R}$ let $\hat{\boldsymbol{y}} = (f_1(\hat{\boldsymbol{x}}), \ldots, f_n(\hat{\boldsymbol{x}}))$; in addition, let $\boldsymbol{y}^* = (f_1^*, \ldots, f_n^*)$ be the utopian point in the outcome space. Because by construction $\hat{y}_i \ge \rho^* y_i^*$, for all $i \in \mathcal{N}$, it is

$$\rho^* \ge \underline{\rho} = \sup\{r \in [0,1] : r\boldsymbol{y}^* \in \mathcal{Y}\},$$

where $\sup \emptyset = 0$. In the *balanced case*, where $f_i(\hat{\boldsymbol{x}}) = \rho^* f_i^*$, for all $i \in \mathcal{N}$, the lower bound becomes tight (i.e., $\rho^* = \underline{\rho}$). However, in general $\underline{\rho}$ may not be very reliable (e.g., generically worse than the performance index $\rho_d = \rho(\boldsymbol{x}_d)$, attained by the default decision $\boldsymbol{x}_d$).

3.6. Robust Weights

3.6.1. General Case. An interesting and useful byproduct of a robust decision $\hat{\boldsymbol{x}} \in \mathcal{R}$ is its associated *robust* (*normalized*) *weight*,

$$\hat{\boldsymbol{\lambda}} = \left(\frac{\hat{h}}{f_1(\hat{\boldsymbol{x}})}, \ldots, \frac{\hat{h}}{f_n(\hat{\boldsymbol{x}})}\right) \in \Delta, \tag{19}$$

where the positive constant

$$\hat{h} = \left(\sum_{i\in\mathcal{N}} \frac{1}{f_i(\hat{\boldsymbol{x}})}\right)^{-1} \tag{20}$$

denotes the harmonic mean of the criterion scores at the robust decision. By construction one obtains that the contribution to the robustly weighted objective $F(\cdot|\hat{\boldsymbol{\lambda}})$, evaluated at the robust decision $\hat{\boldsymbol{x}}$, is equal for all criteria, in the sense that

$$\max_{x\in\mathcal{X}} \min_{i\in\mathcal{N}} \hat{\lambda}_i f_i(\boldsymbol{x}) = \hat{h} = \hat{\lambda}_i f_i(\hat{\boldsymbol{x}}), \quad i \in \mathcal{N}. \tag{21}$$

When the action set is convex, then the chosen robust action naturally also maximizes the robustly weighted criterion, that is, $\hat{\boldsymbol{x}} \in \mathscr{X}(\hat{\boldsymbol{\lambda}})$. One can think of the robust weight as an endogenous belief that can be used to evaluate the expected criterion achievement. It defines the tradeoffs compatible with the robust decision, where the latter was found while remaining agnostic over all possible weights (in Δ).

3.6.2. Balanced Case. In the case where $f_i(\hat{\boldsymbol{x}}) = \rho^* f_i^*$, for all $i \in \mathcal{N}$ (cf. Remark 11), we have

$$\hat{\boldsymbol{\lambda}} = \left(\frac{h^*}{f_1^*}, \ldots, \frac{h^*}{f_n^*}\right), \tag{19′}$$

where

$$h^* = \left(\sum_{i\in\mathcal{N}} \frac{1}{f_i^*}\right)^{-1} \tag{20′}$$

is the harmonic mean of the maximum criterion scores.

The balancedness condition then becomes

$$\rho^* = \frac{f_i(\hat{x})}{f_i^*} = \frac{\hat{h}}{h^*}, \quad i \in \mathcal{N}.$$

In the balanced case, the robust weight can therefore be determined based on the maximized individual criteria alone. If all maximized individual criteria are equal, the robust weight becomes uniform. Indeed, $f_i^* = c > 0$, for all $i \in \mathcal{N}$, implies that $h^* = c/n$, so $\hat{\boldsymbol{\lambda}} = (1/n, \ldots, 1/n)$.[25]

Example 8. In Example 1, we determined the value function $F^*(\ell) = (1-\ell)\hat{t} + \ell\mu(\hat{t})$, so that by Equations (19′) and (20′) in this balanced case (as established in Example 2):

$$\hat{\boldsymbol{\lambda}} = (1-\hat{\ell}, \hat{\ell}) = \left(\frac{\mu(\hat{t})}{\hat{t}+\mu(\hat{t})}, \frac{\hat{t}}{\hat{t}+\mu(\hat{t})}\right) = \left(\frac{f_2^*}{f_1^*+f_2^*}, \frac{f_1^*}{f_1^*+f_2^*}\right).$$

Generically, it is $F^*(\hat{\ell}) \neq F(\hat{x}|\hat{\ell})$, even when the decision set is convex. To see this, let $m = n = 2$ and consider the convex domain, $\mathcal{X} = \{x \in \mathbb{R}_2^+ : v_1x_1 + v_2x_2 \leq 1\}$, where the constants v_1, v_2 are such that $0 < v_1 < v_2 < \infty$. Then $\mu(t) = (1 - v_2t)/v_1 + t$, so that $\mu'(t) = 1 - (v_2/v_1) < 0$. Meanwhile, $F(t|\ell) = (1-\ell)t + \ell\mu(t)$, for $t \in [0, f_1^*] = [0, 1/(v_1+v_2)]$, and

$$F'(t|\ell) = (1-\ell) + \ell\mu'(t) = 1 - \frac{\ell v_2}{v_1} \geq 0 \iff \ell \leq \frac{v_1}{v_2}.$$

Therefore,
$F^*(\ell) = F(t^*(\ell)|\ell) = (\ell/v_1) + \max\{0, 1 - (\ell v_2/v_1)\}f_1^*$,
where

$$t^*(\ell) \in \begin{cases} \{f_1^*\}, & \text{if } \ell < v_1/v_2, \\ [0, f_1^*], & \text{if } \ell = v_1/v_2, \\ \{0\}, & \text{if } \ell > v_1/v_2, \end{cases}$$

and $f_1^* = 1/(v_1+v_2)$. Because $f_2^* = 1/v_1$, we find $v_1/v_2 = f_1^*/(f_2^* - f_1^*)$. By virtue of the fact that $\hat{\ell} = f_1^*/(f_1^*+f_2^*) < v_1/v_2$, it is $t^*(\hat{\ell}) = f_1^*$. The balancedness condition,

$$\frac{\hat{t}}{f_1^*} = \frac{\mu(\hat{t})}{f_2^*},$$

is equivalent to $\hat{t}/f_1^* = 1 - (\hat{t}/f_1^*)$, so $\hat{t} = f_1^*/2$. Hence, we find that $\hat{t} \neq t^*(\hat{\ell})$. Let us check the corresponding performance index. Indeed,

$$\rho(\hat{t}) = 0.5 > 0 = \min\left\{\frac{f_1^*}{f_1^*}, 1 - \frac{f_1^*}{f_1^*}\right\} = \rho(t^*(\hat{\ell})).$$

Note also,

$$\hat{\ell} = \frac{f_1^*}{f_1^*+f_2^*} = \frac{1}{2 + (v_2/v_1)} < \frac{1}{3}.$$

That is, the robust weight puts more than twice as much emphasis on the Rawlsian (or egalitarian) objective as on the utilitarian objective. One can conclude that in general there is no "robustness equivalence principle," in the sense that substituting the robust parameter into the original scalarization (via maximization of the corresponding weighted objective) might *not* lead to a robust decision.[26]

Example 9. Consider the robust allocation of two resources to $n = 2$ agents. The total amount of each resource has been normalized to one. Allocations are determined as decisions $\boldsymbol{x} = (x_1, x_2) \in [0,1]^2$, under which agent 1 obtains x_1 of resource 1 and x_2 of resource 2, whereas agent 2 obtains $1 - x_1$ of resource 1 and $1 - x_2$ of resource 2. Agent 1's utility is $f_1(\boldsymbol{x}) = x_1^\alpha x_2^{1-\alpha}$, and agent 2's utility is $f_2(\boldsymbol{x}) = (1-x_1)^\beta(1-x_2)^{1-\beta}$, where $\alpha, \beta \in (0,1)$ are given scalars. Figure 5 provides an illustration in the corresponding Edgeworth box (see, e.g., Pareto 1906, p. 187). Because a robust allocation is necessarily Pareto-optimal, both agents' marginal rates of substitution for the two goods must be equal, so $\boldsymbol{x} \in \mathcal{P}$ if and only if

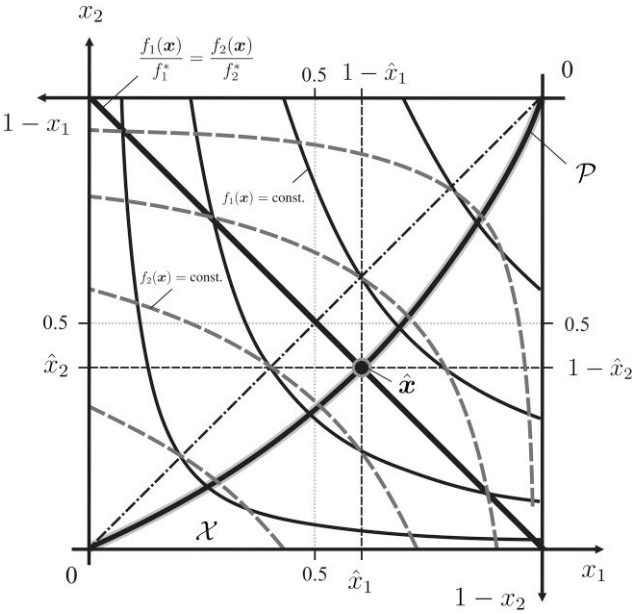
$$\frac{\partial f_1(\boldsymbol{x})/\partial x_1}{\partial f_1(\boldsymbol{x})/\partial x_2} = \frac{\alpha}{1-\alpha}\frac{x_2}{x_1} = \frac{\beta}{1-\beta}\frac{1-x_2}{1-x_1} = \frac{\partial f_2(\boldsymbol{x})/\partial x_1}{\partial f_2(\boldsymbol{x})/\partial x_2}. \quad (22)$$

In addition, a robust allocation must maximize the performance ratio, and one can verify that this requires the boundary performance ratios to be equal (i.e., a balancedness condition), so

$$f_1(\boldsymbol{x}) = x_1^\alpha x_2^{1-\alpha} = (1-x_1)^\beta(1-x_2)^{1-\beta} = f_2(\boldsymbol{x}), \quad (23)$$

because $f_1^* = f_2^* = 1$. In the interesting special case where

Figure 5. Robust Allocation of Resources $\hat{\boldsymbol{x}} = (\hat{x}_1, \hat{x}_2)$ in Edgeworth Box $\mathcal{X} = [0,1] \times [0,1]$



Note. For two agents with utility functions (f_1 and f_2) specified in Example 9 (for $\alpha > \beta$), the unique robust allocation satisfies Pareto-efficiency in Equation (22) and the balancedness condition in Equation (23), under full ambiguity.

$\alpha+\beta=1$, Equations (22) and (23) together yield the unique robust solution $\hat{\boldsymbol{x}}=(\alpha,\beta)$, resulting in an optimal performance index of $\rho^*=\alpha^\alpha\beta^\beta=\alpha^\alpha(1-\alpha)^{1-\alpha}$, with the last expression being a strictly convex function in $\alpha\in(0,1)$ achieving its minimum (of 1/2) at $\alpha=1/2$ and ρ^* going to one as $\alpha\to\{0^+,1^-\}$.[27]

3.7. Limited Ambiguity

Consider the (nonempty, compact) subset $\Omega\subset\mathbb{R}^n_+\setminus\{\mathbf{0}\}$ which may reflect the decision maker's a priori knowledge about the relevant weights for the problem at hand, and let

$$\rho(\boldsymbol{x}|\Omega)=\inf_{\boldsymbol{w}\in\Omega}\varphi(\boldsymbol{x}|\boldsymbol{w}),\ \boldsymbol{x}\in\mathcal{X}, \tag{24}$$

be the corresponding Ω-*conditioned performance index*. The different weight combinations considered feasible may not necessarily be normalized, especially at an initial stage in practice when information about reasonable weight ranges is being compiled. The function $\pi:\mathbb{R}^n_+\setminus\{\mathbf{0}\}\to\Delta$, with $\pi(\boldsymbol{w})=\boldsymbol{w}/\|\boldsymbol{w}\|_1$, describes the *radial projection* of any nonzero weight $\boldsymbol{w}$ in $\mathbb{R}^n_+$ onto the unit simplex Δ. It allows for the normalization under weight ambiguity; see Remark 2. The radial projection is a nonlinear function which is homogeneous of degree zero (i.e., $\pi(\alpha\boldsymbol{w})=\pi(\boldsymbol{w})$, for all $\alpha>0$ and all $\boldsymbol{w}\in\mathbb{R}^n_+\setminus\{\mathbf{0}\}$). It can be represented using the conical closure (see, e.g., Berge 1963, p. 14), $\mathcal{C}(\Omega)=\{\hat{\boldsymbol{w}}\in\mathbb{R}^n_+:\hat{\boldsymbol{w}}=\alpha\boldsymbol{w},\boldsymbol{w}\in\Omega,\alpha\in\mathbb{R}_+\}$, as described, among other useful properties of $\pi(\cdot)$, in the following auxiliary result.

Lemma 8. *Let Ω be a (nonempty, compact) subset of $\mathbb{R}^n_+\setminus\{\mathbf{0}\}$. Then* (i)(a) *$\pi(\Omega)=\mathcal{C}(\Omega)\cap\Delta$ is nonempty and compact, and* (b) *$\pi(\Omega)=\pi(\partial\Omega)$.* (ii) *If $\Omega'\subset\Omega$ is open (in $\mathbb{R}^n_+$), then $\pi(\Omega')$ is open (in Δ).* (iii) *If $\Omega'\subseteq\Omega$ and $\Omega'\neq\emptyset$, then $\partial\pi(\Omega')=\partial\pi(\partial\Omega')$ is nonempty and compact.* (iv) *If Ω is convex, then $\pi(\Omega)$ is convex.* (v) *If $\Omega'\subset\Omega$ is a straight line segment, then $\pi(\Omega')$ is a straight line segment (or a point).* (vi) *$\pi(\mathrm{co}(\Omega))=\mathrm{co}(\pi(\partial\Omega))$.* (vii) *If $\Omega=\Omega'\cup\Omega''$ for some $\Omega',\Omega''\subset\mathbb{R}^n_+$, then $\pi(\Omega)=\pi(\Omega')\cup\pi(\Omega'')$.* (viii) *If $\Omega=\Omega'\cap\Omega''$ for some $\Omega',\Omega''\subset\mathbb{R}^n_+$, then $\pi(\Omega)=\mathcal{C}(\Omega'\cap\Omega'')\cap\Delta$.*

Part (i) of Lemma 8 notes that (a) the radial projection of Ω onto Δ can be obtained by simply intersecting the conical closure of Ω with Δ, and (b) one can limit attention to the boundary $\partial\Omega$ (ignoring interior points of Ω). Although a continuous function generally does not map open sets to open sets (e.g., a constant function would not), part (ii) guarantees that $\pi(\cdot)$ does exactly that, provided the "interiority" of a point in Ω is assessed in $\mathbb{R}^n_+$ and then, after its projection, in the lower-dimensional Δ. Part (iii) ensures that weight ambiguity in *any* (nonempty) subset Ω' of Ω leads to a compact domain $\partial\pi(\partial\Omega')$ for a robust decision according to Proposition 5 below. Part (iv) establishes that convex sets are projected to convex sets.[28] Following (v) and (vi), the radial projection leaves straight-line geometries intact, thus, for example, converting (bounded) polyhedra in $\mathbb{R}^+_n$ to polytopes in Δ. Finally, parts (vii) and (viii) note that the radial projection of a union of sets is the union of the corresponding single-set projections, but the same does generally not apply to an intersection.[29]

Example 10. Assume the (nonnormalized) weights $\boldsymbol{w}$ considered reasonable by the decision maker are such that each of its components w_i, for $i\in\mathcal{N}$, is known to lie in some interval, resulting in a rectangular ambiguity set,

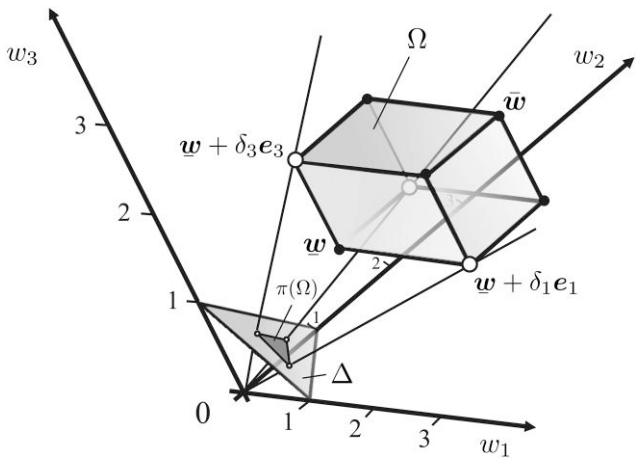
$$\Omega=[\underline{\boldsymbol{w}},\overline{\boldsymbol{w}}]=\prod_{i\in\mathcal{N}}[\underline{w}_i,\overline{w}_i],$$

with bounds $\underline{\boldsymbol{w}}=(\underline{w}_1,\ldots,\underline{w}_n)\neq\mathbf{0}$ and $\overline{\boldsymbol{w}}=(\overline{w}_1,\ldots,\overline{w}_n)\geq\underline{\boldsymbol{w}}$. Using Lemma 8, leveraging the preservation of straight-line geometries, it can be verified that the radial projection of Ω onto Δ has the form

$$\pi(\Omega)=\mathrm{co}\left(\left\{\frac{\underline{\boldsymbol{w}}+\delta_1\boldsymbol{e}_1}{\|\underline{\boldsymbol{w}}\|_1+\delta_1},\ldots,\frac{\underline{\boldsymbol{w}}+\delta_n\boldsymbol{e}_n}{\|\underline{\boldsymbol{w}}\|_1+\delta_n}\right\}\right), \tag{25}$$

where $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_n)=\overline{\boldsymbol{w}}-\underline{\boldsymbol{w}}$ and $\|\underline{\boldsymbol{w}}\|_1=\sum_{l=1}^n\underline{w}_l$. As an example, Figure 6 illustrates that the radial projection of a box in $n=3$ dimensions is a triangle in Δ. Thus, the shape of $\pi(\Omega)$ depends on (at most) n of the original 2^n vertices of Ω. Moreover, if the modulus of the lower bound of the (nontrivial) box-shaped ambiguity set Ω becomes very small, then the situation tends to become equivalent to full ambiguity. To see this, consider a $\overline{\boldsymbol{w}}$ with only positive components and let $\underline{\boldsymbol{w}}=\varepsilon\boldsymbol{e}_1$, for a sufficiently small $\varepsilon>0$. Then, by Equation (25) it is $\pi(\Omega)=\mathrm{co}(\{\boldsymbol{\lambda}^{(1)},\ldots,\boldsymbol{\lambda}^{(n)}\})$, where $\boldsymbol{\lambda}^{(1)}=\boldsymbol{e}_1$ and $\boldsymbol{\lambda}^{(i)}=(\varepsilon\boldsymbol{e}_1+\overline{w}_i\boldsymbol{e}_i)/(\varepsilon+\overline{w}_i)$, for $i\in\{2,\ldots,n\}$. Hence, $\lim_{\varepsilon\to0^+}\boldsymbol{\lambda}^{(i)}=\boldsymbol{e}_i$, for all $i\in\mathcal{N}$,

Figure 6. Radial Projection of Box-Shaped Ambiguity Set $\Omega=[\underline{\boldsymbol{w}},\overline{\boldsymbol{w}}]$



Note. With $\underline{\boldsymbol{w}}=(1,1,1)$ and $\overline{\boldsymbol{w}}=(3,2,2)$, a radial projection onto Δ yields $\pi(\Omega)=\mathrm{co}(\{(3/5,1/5,1/5),(1/4,1/2,1/4),(1/4,1/4,1/2)\})$, as described in Example 10.

implying full ambiguity in the limit, as $\Delta = \mathrm{co}(\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}) = \pi(\mathbb{R}^n_+ \setminus \{\mathbf{0}\})$.

Under limited ambiguity, a coordinate-wise decomposition is no longer available. However, because $\varphi(\boldsymbol{x}|\cdot)$ is quasiconcave, for any given $\boldsymbol{x} \in \mathcal{X}$, as established in the proof of Proposition 2, the upper contour sets of the performance ratio in the space of weights are necessarily convex. This in turn allows restricting attention, for the representation of the performance index, to the extreme points of the convex hull of the ambiguity set, which yields a representation much in the same spirit as before.

Proposition 5. *The Ω-conditioned performance index in Equation (24) is such that*

$$\rho(\boldsymbol{x}|\Omega) = \min_{\boldsymbol{\lambda} \in \partial\pi(\partial\Omega)} \varphi(\boldsymbol{x}|\boldsymbol{\lambda}) = \min_{\boldsymbol{\lambda} \in \Lambda} \varphi(\boldsymbol{x}|\boldsymbol{\lambda}), \quad \boldsymbol{x} \in \mathcal{X},$$

for some "extremal base" Λ which is a (compact) subset of $\partial\pi(\partial\Omega)$, such that $\mathrm{co}(\Lambda) = \mathrm{co}(\pi(\Omega))$.

Although choosing $\Lambda = \partial\pi(\Omega) = \partial\pi(\partial\Omega)$ (cf. Lemma 8 (i) (b)) is always possible, it is usually advantageous to opt for the smallest possible extremal base Λ, so it consists only of the "extreme points" of $\mathrm{co}(\pi(\Omega))$, where the latter cannot be represented as convex combinations of other points in Λ (see, e.g., Rockafellar 1970, section 18, p. 162).[30] If Ω is finite or a (finite) polytope, then the smallest extremal base Λ of extreme points of $\mathrm{co}(\pi(\Omega))$ is also finite.[31] Because any bounded convex set can be approximated by a finite polytope (Bronstein 2008), a finite extremal base can be used to represent the convex hull of the given ambiguity set (or its radial projection onto Δ) up to any desired precision. That $\pi(\Omega)$ may be *nonconvex* is not important because the level sets of the performance ratio are convex, so that the minima of $\varphi(\boldsymbol{x}|\cdot)$ on $\pi(\Omega)$ are attained on the boundary of its convexification, $\mathrm{co}(\pi(\Omega))$.

Example 11. Consider the robust allocation of resources to $n = 2$ agents as in Example 9, given the box-shaped ambiguity set Ω as in Example 11, with its radial projection specified by Equation (25),

$$\pi(\Omega) = \mathrm{co}(\Lambda) = \mathrm{co}\left(\left\{\frac{(\overline{w}_1, \underline{w}_2)}{\overline{w}_1 + \underline{w}_2}, \frac{(\underline{w}_1, \overline{w}_2)}{\underline{w}_1 + \overline{w}_2}\right\}\right) \subsetneq \Delta,$$

where $\Lambda = \{\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)}\}$. Thus, by virtue of Proposition 5 the Ω-conditioned performance index becomes

$$\rho(\boldsymbol{x}|\Omega) = \min\left\{\frac{\overline{w}_1 f_1(\boldsymbol{x}) + \underline{w}_2 f_2(\boldsymbol{x})}{\max_{\hat{\boldsymbol{x}} \in \mathcal{X}}\{\overline{w}_1 f_1(\hat{\boldsymbol{x}}) + \underline{w}_2 f_2(\hat{\boldsymbol{x}})\}}, \frac{\underline{w}_1 f_1(\boldsymbol{x}) + \overline{w}_2 f_2(\boldsymbol{x})}{\max_{\hat{\boldsymbol{x}} \in \mathcal{X}}\{\underline{w}_1 f_1(\hat{\boldsymbol{x}}) + \overline{w}_2 f_2(\hat{\boldsymbol{x}})\}}\right\}, \quad \boldsymbol{x} \in \mathcal{X}.$$

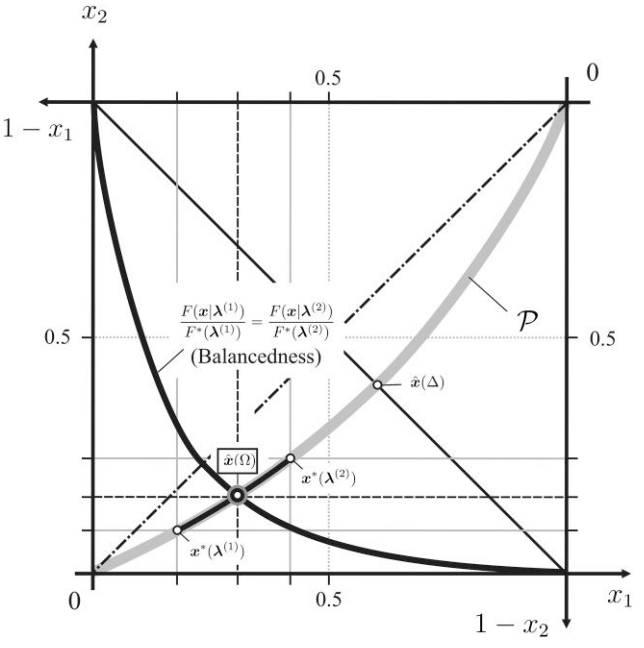
For a pseudo-robust decision $\boldsymbol{x}$ (in Ψ), balancedness must hold, so that, by substituting f_1, f_2 from Example 9,

$$\frac{F(\boldsymbol{x}|\boldsymbol{\lambda}^{(1)})}{F(\boldsymbol{x}|\boldsymbol{\lambda}^{(2)})} = \frac{\overline{w}_1 x_1^{\alpha} x_2^{1-\alpha} + \underline{w}_2 (1-x_1)^{\beta}(1-x_2)^{1-\beta}}{\underline{w}_1 x_1^{\alpha} x_2^{1-\alpha} + \overline{w}_2 (1-x_1)^{\beta}(1-x_2)^{1-\beta}} = \frac{F^*(\boldsymbol{\lambda}^{(1)})}{F^*(\boldsymbol{\lambda}^{(2)})}. \tag{26}$$

Meanwhile, a feasible allocation decision $\boldsymbol{x} = (x_1, x_2)$ is Pareto-optimal (in $\mathcal{P}$) if and only if it satisfies Equation (22), as before. When the box-shaped ambiguity set becomes smaller, the elements of the extremal base Λ become more similar and coincide in the limit (i.e., $\boldsymbol{\lambda}^{(1)} - \boldsymbol{\lambda}^{(2)} \to \mathbf{0}$ as $\|\overline{\boldsymbol{w}} - \underline{\boldsymbol{w}}\|_1 \to 0^+$). Figure 7 shows the robust allocation $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}(\Omega)$ on the contract curve between the Pareto-optimal allocations $\boldsymbol{x}^*(\boldsymbol{\lambda}^{(i)})$ that maximize the weighted objective $F(\boldsymbol{x}|\boldsymbol{\lambda}^{(i)})$ for $i \in \mathcal{N}$. Under full ambiguity (i.e., for $\Omega = \Delta$), it is $\boldsymbol{\lambda}^{(i)} \equiv \boldsymbol{e}_i$; see Example 9.

Example 12. In some applications, the weights are naturally ranked by importance (see, e.g., Wang and Fu 2020). The ordered set of weights, $\Omega' = \{(w_1, \ldots, w_n) \in \mathbb{R}^n_+ : w_1 \geq w_2 \geq \cdots \geq w_n\} = \mathcal{C}(\Omega')$, has a radial projection onto Δ of the form $\pi(\Omega') = \mathrm{co}(\{\boldsymbol{\lambda}'^{(i)} : i \in \mathcal{N}\})$, where $\boldsymbol{\lambda}'^{(i)} = \frac{1}{i}\sum_{j=1}^{i} \boldsymbol{e}_j$, for all $i \in \mathcal{N}$. Combining this importance-ranking with a box-shaped ambiguity set $\Omega'' = [\underline{\boldsymbol{w}}, \underline{\boldsymbol{w}} + \boldsymbol{\delta}]$ as in Example 10, with $\pi(\Omega'') = \mathrm{co}(\{\boldsymbol{\lambda}''^{(i)} : i \in \mathcal{N}\})$, where $\boldsymbol{\lambda}''^{(i)} = (\underline{\boldsymbol{w}} + \delta_i \boldsymbol{e}_i)/(\|\underline{\boldsymbol{w}}\|_1 + \delta_i)$,

Figure 7. Robust Allocation of Resources, $\{\hat{\boldsymbol{x}}(\Omega)\} = \arg\max_{\boldsymbol{x} \in \mathcal{P}} \rho(\boldsymbol{x}|\Omega)$, to Two Agents in Example 11 (for $\alpha > \beta$)



Note. Under limited ambiguity $\Omega = [\underline{\boldsymbol{w}}, \overline{\boldsymbol{w}}]$ as in Example 10, the robust allocation $\hat{\boldsymbol{x}}(\Omega)$ satisfies Pareto-efficiency in Equation (22) and balancedness in Equation (26); it differs from the allocation $\hat{\boldsymbol{x}}(\Delta)$ under full ambiguity, which maximizes $\rho(\boldsymbol{x}) = \rho(\boldsymbol{x}|\Delta)$ on $\mathcal{P}$.

for all $i \in \mathcal{N}$, leads to

$$\Lambda = \left\{ \mathbb{1}_{\{\underline{\boldsymbol{w}}+\delta_i \boldsymbol{e}_i \notin \Omega'\}} \boldsymbol{\lambda}'^{(i)} + \mathbb{1}_{\{\underline{\boldsymbol{w}}+\delta_i \boldsymbol{e}_i \in \Omega'\}} \boldsymbol{\lambda}''^{(i)} : i \in \mathcal{N} \right\}.$$

This extremal base (of cardinality n) suits practical applications where a decision maker disposes of plausible ranges for the weights to be placed on the criteria, together with a ranking of their importance (cf. Section 4.3.2).

Remark 12. (i) Limiting ambiguity can only increase robustness performance. That is, if $\Omega, \hat{\Omega} \subset \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$ are nonempty and compact and satisfy $\pi(\hat{\Omega}) \subseteq \pi(\Omega)$, then $\rho(\cdot|\Omega) \leq \rho(\cdot|\hat{\Omega})$, which follows directly from the definition of the Ω-conditioned performance index in Equation (24). (ii) In the absence of ambiguity, the optimal robustness performance has to be maximal. That is, if $\hat{\Omega} = \{\boldsymbol{w}\}$, then $\rho^*(\hat{\Omega}) = \max_{\boldsymbol{x} \in \mathcal{X}} \rho(\boldsymbol{x}|\{\boldsymbol{w}\}) = F^*(\boldsymbol{w})/F^*(\boldsymbol{w}) = 1$.

3.8. Criterion Ambiguity

Given a (nonempty, compact) action set $\mathcal{X} \subset \mathbb{R}^m$ as in Section 2, we now allow for ambiguity in each of the n' criteria, given by the continuous functions $g_i : \mathcal{X} \times \Theta \to \mathbb{R}_+$, for $i \in \mathcal{N}' = \{1, \ldots, n'\}$, which map $(\boldsymbol{x}, \boldsymbol{\theta})$ to real numbers, where $\boldsymbol{x} \in \mathcal{X}$ is a feasible action and $\boldsymbol{\theta}$ is an ex ante unknown state in the (nonempty, compact) *state space* $\Theta \subset \mathbb{R}^l$, and where $l, m, n' \geq 1$ are given integers. The unknown state $\boldsymbol{\theta}$ represents the decision maker's uncertainty about the exact value that each of his n' objectives may attain under a chosen action $\boldsymbol{x}$.

Remark 13. (i) Continuity of the criteria in the state implies that a small perturbation in the state can have only a small impact on the decision maker's objective. This continuity is automatically satisfied when the state space Θ is finite.[32] (ii) The unknown state may introduce dependencies between the different g_i. Indeed, even if $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{n'})$ (for $l = n'$) and criteria are such that $g_i(\boldsymbol{x}, \boldsymbol{\theta}) \equiv g_i(\boldsymbol{x}, \theta_i)$, for all $i \in \mathcal{N}'$, then the different criteria's ambiguity may still be linked via common constraints in Θ.[33] But if in addition the state space is a Cartesian product, so $\Theta = \Theta_1 \times \cdots \Theta_{n'}$, and $\theta_i \in \Theta_i$, for all $i \in \mathcal{N}$, then the decoupling of ambiguity across the different criteria is complete, in the sense that observing the part of the state that determines one criterion does not reduce ambiguity for any other criterion.

To assess goal achievement of an action in a given state, the decision maker considers a scalarization of his multicriteria optimization problem by means of a weighted objective,

$$G(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}') = \sum_{i=1}^{n'} \lambda'_i \, g_i(\boldsymbol{x}, \boldsymbol{\theta}), \quad (\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\lambda}') \in \mathcal{X} \times \Theta \times \Delta',$$

where $\Delta' = \{\boldsymbol{w}' = (w_1, \ldots, w_{n'}) \in \mathbb{R}^{n'}_+ : w_1 + \cdots + w_{n'} = 1\}$ denotes the set of all (normalized) weights. As in the nontriviality condition (N) (cf. Section 2), we assume that there exists a default decision $(\boldsymbol{x}_d)$ such that the decision maker's weighted objective is positive across all states, that is,

$$\exists \boldsymbol{x}_d \in \mathcal{X} : \, G(\boldsymbol{x}_d, \boldsymbol{\theta}|\boldsymbol{\lambda}') > 0, \quad (\boldsymbol{\theta}, \boldsymbol{\lambda}') \in \Theta \times \Delta'. \tag{N'}$$

The modified nontriviality condition (N′) ensures that the ex post optimal objective is always positive, so

$$G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}') = \max_{\boldsymbol{x} \in \mathcal{X}} G(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}') \geq G(\boldsymbol{x}_d, \boldsymbol{\theta}|\boldsymbol{\lambda}') > 0,$$

for all $(\boldsymbol{\theta}, \boldsymbol{\lambda}') \in \Theta \times \Delta'$.

Remark 14. The modified nontriviality condition (N′) can always be satisfied, without altering the set of ex post optimal decisions, by using a (positive, translated) criterion $\hat{g}_i = g_i + \varepsilon$, for some $\varepsilon > 0$; see also Remark 4.

Given a (nonempty, compact) ambiguity set $\Omega' \subset \mathbb{R}^{n'}_+$, the decision maker's robustness objective, as in Section 3.7, is to maximize the (Ω', Θ)-*conditioned performance index*,

$$\rho(\boldsymbol{x}|\Omega', \Theta) = \min_{(\boldsymbol{\lambda}', \boldsymbol{\theta}) \in \pi(\Omega') \times \Theta} \frac{G(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}')}{G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}')}, \quad \boldsymbol{x} \in \mathcal{X}.$$

The following result recasts the robustness objective into a by-now-familiar representation.

Proposition 6. *Let $\Lambda' \subset \Delta'$ be an extremal base of* $\mathrm{co}(\pi(\Omega'))$. *Then*

$$\rho(\boldsymbol{x}|\Omega', \Theta) = \min_{\boldsymbol{\lambda}' \in \Lambda'} \left\{ \min_{\boldsymbol{\theta} \in \Theta} \frac{G(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}')}{G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}')} \right\}, \quad \boldsymbol{x} \in \mathcal{X}. \tag{27}$$

Based on Proposition 6, if the (smallest) extremal base is finite, we can reduce the general multicriteria decision problem to our basic framework in Section 2, with full weight ambiguity and no criterion ambiguity.

Corollary 1. *Assume that there exists a (finite, smallest) integer $n \geq 1$ so that $\Lambda' = \{\boldsymbol{\lambda}'^{(1)}, \ldots, \boldsymbol{\lambda}'^{(n)}\}$ and $\mathrm{co}(\Lambda') = \mathrm{co}(\pi(\Omega'))$, and let $f_i(\boldsymbol{x}) = \min_{\boldsymbol{\theta} \in \Theta} \{G(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}'^{(i)})/G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}'^{(i)})\}$, for all $(\boldsymbol{x}, i) \in \mathcal{X} \times \mathcal{N}$. Then*

$$\rho(\boldsymbol{x}) = \min_{i \in \mathcal{N}} \left\{ \frac{f_i(\boldsymbol{x})}{f_i^*} \right\}, \quad \boldsymbol{x} \in \mathcal{X}, \tag{28}$$

where $f_i^* = \max_{\boldsymbol{x} \in \mathcal{X}} f_i(\boldsymbol{x}) = 1$, for all $i \in \mathcal{N} = \{1, \ldots, n\}$, represents $\rho(\cdot|\Omega', \Theta)$ in Equation (27).

Example 13. (i) In the case of full ambiguity, it is $\Lambda' = \Delta'$, so that $n = n'$ and $\Delta' = \mathrm{co}(\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}) = \Delta$, with $f_i(\boldsymbol{x}) = \min_{\boldsymbol{\theta} \in \Theta} \{g_i(\boldsymbol{x}, \boldsymbol{\theta})/G^*(\boldsymbol{\theta}|\boldsymbol{e}_i)\}$, for all $(\boldsymbol{x}, i) \in \mathcal{X} \times \mathcal{N}$. (ii) Consider now a multicriteria decision problem under full ambiguity, with criteria of the form $g_i(\boldsymbol{x}, \boldsymbol{\theta}) = (\hat{g}_i(\boldsymbol{x}))^{\theta_i}$, for all $i \in \mathcal{N}$, where the state $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ is only known to lie in the (nonempty,

compact) set $\Theta \subset \mathbb{R}^n_+$ (with $\Theta \neq \{\mathbf{0}\}$ to avoid trivialities), and where the functions $\hat{g}_i : \mathcal{X} \to \mathbb{R}_{++}$ are continuous. Let $\hat{g}_i^* = \max \hat{g}_i(\mathcal{X}) > 0$ and $\overline{\theta}_i = \max_{\boldsymbol{\theta}\in\Theta} \theta_i$; furthermore, set

$$f_i(\boldsymbol{x}) = \left(\frac{\hat{g}_i(\boldsymbol{x})}{\hat{g}_i^*}\right)^{\overline{\theta}_i} = \min_{\boldsymbol{\theta}\in\Theta} \left(\frac{\hat{g}_i^{\theta_i}(\boldsymbol{x})}{\max_{\boldsymbol{x}\in\mathcal{X}} \hat{g}_i^{\theta_i}(\boldsymbol{x})}\right), \quad \boldsymbol{x} \in \mathcal{X},$$

for all $i \in \mathcal{N}$, taking into account that ξ^θ decreases in θ, for all $\xi \in [0,1]$. The extreme state $\overline{\boldsymbol{\theta}} = (\overline{\theta}_1, \ldots, \overline{\theta}_n)$ determines the robust choice, as it leads ceteris paribus to the smallest criterion values—thus following a precautionary principle. Then, by Corollary 1 we follow our basic framework, maximizing the performance index in Equation (28) with respect to all Pareto-optimal decisions in Equation (10). For instance, if $\Theta = \Delta$, the reduction to the weighted objective in Equation (1) obtains with $f_i = \hat{g}_i / \hat{g}_i^*$, for all $i \in \mathcal{N}$.

Remark 15. (i) The minimization over the state space in Example 13 reflects a robust, *precautionary* stance: the decision maker evaluates actions under the most adverse plausible realization of the state with respect to the achievable performance *ratio*, consistent with a notion of *relative* worst-case robustness. (ii) The treatment of criterion ambiguity in this section also bears a conceptual resemblance to models of DRO, where decisions are evaluated against worst-case distributions within a specified ambiguity set. In our framework, the uncertainty is not over probability distributions but over states $\boldsymbol{\theta} \in \Theta$, with performance assessed under the worst-case realization of both the state and the weights. This can be viewed as a nonprobabilistic analogue of DRO, where the ambiguity set Ω' over the weights plays a role similar to ambiguity sets over distributions in DRO models. In particular, the two-layer minimization in Equation (27)—over weights and states—mirrors the inner DRO minimization over distributions, highlighting a parallel structure between worst-case evaluation across distributional and multicriteria settings.

3.9. General Multicriteria Objectives

In certain practical applications, it may seem appealing to consider alternatives to the arithmetic mean in Equation (1), such as a harmonic or geometric mean, with suitable weights. One could even think of using a (weighted) power mean which would accommodate each of the earlier options as a special case; see Example 14 below, which illustrates *aggregation ambiguity*. Rather than commit to a particular functional form for aggregating multiple criteria, however, we propose a general class of multicriteria objectives that adhere to a set of reasonable axioms. These axioms reduce (in the case of uniform weights) to those known to characterize the quasiarithmetic mean. A key finding, notably, is that maximizing this general weighted objective is entirely equivalent to maximizing the arithmetic mean objective in Equation (1), provided the criteria are suitably transformed.

Let $H : \mathcal{Y} \times \Delta \to \mathbb{R}$ denote a (continuous) multicriteria objective, where $\mathcal{Y} \subseteq \mathbb{R}^n$ is a (nonempty) domain such that $\hat{\boldsymbol{f}}(\mathcal{X}) \subset \mathcal{Y}$, with the vector of criteria $\hat{\boldsymbol{f}} = (\hat{f}_1, \ldots, \hat{f}_n)$, and where $\Delta \subset \mathbb{R}^n_+$ is the unit simplex; see Section 2.1. For any decision $\boldsymbol{x} \in \mathcal{X}$ and weight $\boldsymbol{\lambda} \in \Delta$, the multicriteria objective produces an overall score $\hat{H}(\boldsymbol{x}|\boldsymbol{\lambda}) = H(\hat{\boldsymbol{f}}(\boldsymbol{x})|\boldsymbol{\lambda})$, which the decision maker would like to maximize by choosing an appropriate decision—in the presence of (possibly limited) ambiguity about $\boldsymbol{\lambda}$ (and possibly also about $\hat{\boldsymbol{f}}$) as discussed in Section 3.7 (and Section 3.8). To guide the form of a sufficiently flexible and interpretable objective, we posit five axioms (Axioms 1–5), which nest those proposed by Kolmogorov (1930) in his seminal work "On the notion of mean."

Axiom 1 (Monotonicity). *For all* $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}$, $\hat{\boldsymbol{y}} = (\hat{y}_1, \ldots, \hat{y}) \in \mathcal{Y}$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \Delta$, *and* $i \in \mathcal{N}$:

$$\hat{\boldsymbol{y}} - \boldsymbol{y} = (\hat{y}_i - y_i)\boldsymbol{e}_i, \hat{y}_i > y_i \Rightarrow \begin{cases} H(\hat{\boldsymbol{y}}|\boldsymbol{\lambda}) > H(\boldsymbol{y}|\boldsymbol{\lambda}), & \text{if } \lambda_i > 0, \\ H(\hat{\boldsymbol{y}}|\boldsymbol{\lambda}) = H(\boldsymbol{y}|\boldsymbol{\lambda}), & \text{if } \lambda_i = 0. \end{cases}$$

Axiom 2 (Symmetry). *For all* $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \Delta$, *and* $i, j \in \mathcal{N}$:

$$(\hat{\boldsymbol{y}}, \hat{\boldsymbol{\lambda}}) = (\boldsymbol{y}, \boldsymbol{\lambda}) + (y_i - y_j, \lambda_i - \lambda_j)(\boldsymbol{e}_j - \boldsymbol{e}_i) \Rightarrow H(\hat{\boldsymbol{y}}|\hat{\boldsymbol{\lambda}}) = H(\boldsymbol{y}|\boldsymbol{\lambda}).$$

Axiom 3 (Reflexivity). $H(\alpha \mathbf{1}_n|\boldsymbol{\lambda}) = \alpha$, *for all* $\alpha > 0$ *(with* $\alpha \mathbf{1}_n \in \mathcal{Y}$*) and* $\boldsymbol{\lambda} \in \Delta$.[34]

Axiom 4 (Associativity). *For all* $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \Delta$, *and* $m \in \mathcal{N} \setminus \{n\}$:[35]

$$(\lambda_1, \ldots, \lambda_m) \neq \mathbf{0}, H\left(\sum_{i=1}^m y_i \boldsymbol{e}_i \middle| \frac{\sum_{i=1}^m \lambda_i \boldsymbol{e}_i}{\sum_{i=1}^m \lambda_i}\right) = \alpha \Rightarrow H(\alpha \mathbf{1}_m, y_{m+1}, \ldots, y_n|\boldsymbol{\lambda}) = H(\boldsymbol{y}|\boldsymbol{\lambda}).$$

Axiom 5 (Coordinate Filter). $H(\boldsymbol{y}|\boldsymbol{e}_i) = y_i$, *for all* $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}$ *and* $i \in \mathcal{N}$.

The significance of these five basic requirements is as follows: *Monotonicity* (Axiom 1) means that as long as the weight component λ_i is positive, increasing the value y_i of criterion $i \in \mathcal{N}$ must also increase the weighted objective, whereas for $\lambda_i = 0$ the weighted objective becomes insensitive to criterion i; *symmetry* (Axiom 2) requires that the weighted objective is invariant with respect to any joint permutation of indices belonging to criteria and their associated weights; *reflexivity* (Axiom 3) imposes score-consistency in the sense that if all criterion scores are identical, then that should also be the value of the weighted objective, no matter

what (normalized) weights are applied; in a similar vein, *associativity* (Axiom 4) postulates that replacing a group of inputs with their internal weighted average leaves the overall aggregation unchanged; finally, in order to guarantee that the weights provide a homotopic relation between all criteria in isolation, the weighted objective is a *coordinate filter* (Axiom 5) if it yields the i-th component of $\boldsymbol{y}$ when putting all weight on the i-th coordinate (i.e., for $\boldsymbol{\lambda} = \boldsymbol{e}_i$).

Based on Kolmogorov's result,[36] we define the *general h-mean*,

$$H(\boldsymbol{y}|\boldsymbol{\lambda}) = h^{-1}\left(\sum_{i=1}^{n} \lambda_i h(y_i)\right), \tag{29}$$

where the *kernel* $h : \mathcal{D} \to \mathbb{R}$ is a continuous strictly monotonic function, defined on a suitable domain $\mathcal{D} \subset \mathbb{R}$.

Lemma 9. *The general h-mean* $H : \mathcal{Y} \times \Delta \to \mathbb{R}$ *in Equation* (29) *satisfies* Axioms 1–5.

The following example illustrates the flexibility afforded by the general h-mean.

Example 14. Consider two well-known averages.

i. Let $\mathcal{Y} = \mathbb{R}^n_{++}$. For any weight $\boldsymbol{\lambda} \in \Delta$ and power parameter $p \in \mathbb{R}$, define the *power mean*:

$$M_p(\boldsymbol{y}|\boldsymbol{\lambda}) = \begin{cases} \left(\sum_{i=1}^n \lambda_i y_i^p\right)^{1/p}, & p \neq 0, \\ \prod_{i=1}^n y_i^{\lambda_i}, & p = 0, \end{cases}$$

which corresponds to Equation (29) with $h(y) = y^p$ for $p \neq 0$, and $h(y) = \log y$ for $p = 0$. This recovers the harmonic ($p = -1$), geometric ($p = 0$), and arithmetic ($p = 1$) means, among others. In the limit, the power mean approaches the minimum (for $p \to -\infty$) or maximum (for $p \to +\infty$) of all y_i with positive weights. Importantly, M_p is increasing in p (see, e.g., Hardy et al. 1934, theorem 16, p. 26) and uniquely satisfies homogeneity: $M_p(\alpha \boldsymbol{y}|\boldsymbol{\lambda}) = \alpha M_p(\boldsymbol{y}|\boldsymbol{\lambda})$ for all $\alpha > 0$ (see, e.g., Hardy et al. 1934, theorem 84, p. 68).[37]

ii. For $h(\cdot) = \exp(\cdot)$, the general h-mean in Equation (29) becomes a weighted mean in the log-semiring, related to the LogSumExp (LSE) function, because $H(\boldsymbol{y}|\boldsymbol{\lambda}) = \mathrm{LSE}(y_1 + \log\lambda_1, \ldots, y_n + \log\lambda_n)$, for all $(\boldsymbol{y}, \boldsymbol{\lambda}) \in \mathbb{R}^n \times \mathrm{int}(\Delta)$, where $\mathrm{LSE}(\boldsymbol{y}) = \log(\sum_{i=1}^n \exp(y_i))$, for all $\boldsymbol{y} \in \mathbb{R}^n$. Such a formulation carries fruit in probabilistic modeling and neural networks, where LSE and softmax appear naturally because of their smoothness properties.[38]

3.9.1. Reduction to Basic Framework. Crucially, the general h-mean in Equation (29) reduces to our standard weighted objective in Equation (1) under a transformation of the criteria, by setting $f_i = h \circ \hat{f}_i$, so that

$$\hat{H}(\boldsymbol{x}|\boldsymbol{\lambda}) = h^{-1}(F(\boldsymbol{x}|\boldsymbol{\lambda})), \ (\boldsymbol{x}, \boldsymbol{\lambda}) \in \mathcal{X} \times \Delta.$$

If the kernel h is increasing (resp., decreasing), then maximizing $\hat{H}(\cdot|\boldsymbol{\lambda})$ is equivalent to maximizing (resp., minimizing) $F(\cdot|\boldsymbol{\lambda})$. Thus, our earlier results carry over directly—with minor adjustments in the decreasing case, as detailed in Section 3.9.2.

Example 15. Consider Example 13 (ii) for the "diagonal" state space $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^n_+ : \theta_1 = \cdots = \theta_n \in [\underline{p}, \overline{p}]\}$, where the constants $\underline{p}, \overline{p}$ are such that $0 < \underline{p} < \overline{p} < \infty$. By Example 14 (i), this represents a situation in which the decision maker is uncertain about the appropriate aggregation method, except that a (homogeneous) power mean M_p should be used, for some $p \in [\underline{p}, \overline{p}]$. The result in Example 13 (ii) suggests as robust choice the largest available option: the power mean $M_{\overline{p}}$.

Remark 16. (i) As Example 15 suggests, the insights about criterion ambiguity in Section 3.8 may sometimes be combined with the general representation of multicriteria objectives in Section 3.9 to handle *aggregation ambiguity*, that is, uncertainty over which aggregation rule or kernel should be used. (ii) For the general h-mean in Equation (29), we define $H = h^{-1} \circ F_h$ with $F_h = \sum_{i=1}^n \lambda_i (h \circ f_i)$. However, in general, $h^{-1} \circ (F_h / F_h^*) \neq (h^{-1} \circ F_h)/(h^{-1} \circ F_h^*)$ with $F_h^*(\cdot) = \max F_h(\mathcal{X}|\cdot)$, so our optimization focuses on F_h directly. (iii) For the practically very important power mean in Example 14 (i) and Example 15 (as long as $p \neq 0$), the *robust* optimization of F_h is equivalent to the *robust* optimization of M_p, because $h(\cdot) = (\cdot)^p$ and $h^{-1}(\cdot) = (\cdot)^{1/p}$ both feature multiplicative separability.

3.9.2. Special Case: Minimization Under Decreasing Kernel. For a decreasing kernel, $f_i = h \circ \hat{f}_i$ transforms larger $\hat{f}_i$ into smaller f_i, so that the objective reflects a *smaller-is-better* interpretation (akin to optimizing a loss function). Assuming $\mathbb{R}^n_{++} \subset \mathcal{Y}$, by Axiom 3 it is $\lim_{\alpha \to 0^+} H(\alpha \mathbf{1}_n|\boldsymbol{\lambda}) = H(\mathbf{0}^+|\boldsymbol{\lambda}) = 0$. Because in most practical applications (e.g., for logarithmic or inverse transformations in multiobjective loss minimization) it is $h(0^+) = \infty$, we obtain $h^{-1}(\infty) = 0^+$, so that on the compact action set $\mathcal{X}$ it is $f_i^\bullet = \min f_i(\mathcal{X}) = h(\hat{f}_i^*) > 0$, for all $i \in \mathcal{N}$, as $\hat{f}_i^* = \max \hat{f}_i(\mathcal{X})$ is necessarily finite. Hence, we can consider the *adjusted performance ratio*,

$$\varphi(\boldsymbol{x}|\boldsymbol{\lambda}) = \frac{F^\bullet(\boldsymbol{\lambda})}{F(\boldsymbol{x}|\boldsymbol{\lambda})} \in (0, 1], \ (\boldsymbol{x}, \boldsymbol{\lambda}) \in \mathcal{X} \times \Delta, \tag{30}$$

where $F^\bullet(\boldsymbol{\lambda}) = \min F(\mathcal{X}|\boldsymbol{\lambda})$, for all $\boldsymbol{\lambda} \in \Delta$. A representation of the performance index $\rho(\boldsymbol{x})$ in Equation (4) then obtains, analogous to Equation (6), for all $\boldsymbol{x} \in \mathcal{X}$, as the minimum of the (adjusted) criterion-specific performance ratios $\phi_i(\boldsymbol{x}) = f_i^\bullet / f_i(\boldsymbol{x})$ with respect to $i \in \mathcal{N}$. This defines the (nonempty, compact) set of pseudo-robust actions,

$$\Psi = \arg\max_{x \in \mathcal{X}} \left\{ \min_{i \in \mathcal{N}} \frac{f_i^\bullet}{f_i(\boldsymbol{x})} \right\}.$$

The set of Pareto-optimal actions $\mathcal{P}$ needs to be based on the original vector of criteria $\hat{\boldsymbol{f}}$, so that

$$\mathcal{P} = \{\boldsymbol{x} \in \mathcal{X} : (\hat{\boldsymbol{f}}(\boldsymbol{x}) \leq \hat{\boldsymbol{f}}(\boldsymbol{x}') \Rightarrow \hat{\boldsymbol{f}}(\boldsymbol{x}) = \hat{\boldsymbol{f}}(\boldsymbol{x}')), \ \boldsymbol{x}' \in \mathcal{X}\},$$

analogous to Equation (10). The robust decision set is then $\mathcal{R} = \Psi \cap \mathcal{P}$ as before, in Equation (11).

3.10. Applying the Method

"What do I do? What do I get? How do I adapt it to my case?"—We now address these practitioner questions by outlining the method's core, the interpretation of its results, and important extensions for customization, as a navigation device for approaching the multicriteria decision problem introduced in Section 2.1 and its generalizations.

I. (*Core*) To determine an approximately robust decision $\hat{\boldsymbol{x}}_\varepsilon$ (at any prespecified precision ε) with respect to n criteria, one needs to solve just $n+1$ optimization problems: one to maximize the ε-augmented performance index Φ_ε in Equations (12) and (13) and n to compute the normalization constants f_i^* in Equation (5) for the criterion-specific performance ratios ϕ_i in Equation (6). The associated normalized robust weight $\hat{\boldsymbol{\lambda}} \in \Delta$, which encapsulates the tradeoffs between the criteria embedded in both the action set and the shape of the criterion functions (independent of any scaling), is then obtained (approximately) from Equations (19) and (20) by setting $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}_\varepsilon$.

II. (*Interpretation*) The performance index $\rho(\hat{\boldsymbol{x}}_\varepsilon) \in [0,1]$, with $\rho(\cdot)$ given in Equation (6), guarantees a minimum percentage that the robust solution $\hat{\boldsymbol{x}}_\varepsilon$ achieves of the optimal weighted objective in Equation (1), for *any* weight in Δ. This means that the performance of $\hat{\boldsymbol{x}}_\varepsilon$ is guaranteed to be no worse than $\rho(\hat{\boldsymbol{x}}_\varepsilon)$ times the best achievable performance under any possible weight vector, ensuring robustness to unknown or contested preferences. In this manner, the weight-induced subjectivity in multicriteria optimization can be removed. The robust weight rationalizes the robust decision as an optimum of the weighted objective in Equation (1) for $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}$. The vector $\hat{\boldsymbol{\lambda}}$ can be interpreted as the revealed weight structure that best justifies the robust solution, based on the problem's internal tradeoffs.

III. (*Customization*) The method accommodates several practically important extensions:

– *Partial weight information:* Prior knowledge or constraints on weights can be imposed by requiring weight vectors (not necessarily normalized) to belong to a suitable subset (cf. Section 3.7).

– *State-dependent criteria:* Uncertain or context-dependent criteria f_i can be captured by worst-case performance ratios, effectively reducing the problem to the core framework (cf. Section 3.8).

– *Generalized aggregation:* Rather than relying on the arithmetic mean in Equation (1), one may adopt alternative scalarization functions consistent with an axiomatic foundation (cf. Section 3.9).

Overall, the relatively robust methodology provides a computationally tractable and conceptually transparent toolkit for balancing multiple, possibly ambiguous objectives in diverse real-world settings (cf. Section 4.3).

4. Discrete Applications

4.1. Finite Action Set

4.1.1. Relative Robustness Criterion. Consider a finite set of alternatives $\mathcal{X} = \{1,\dots,J\}$, evaluated according to n positive criteria $f_1,\dots,f_n : \mathcal{X} \to \mathbb{R}_{++}$. Each alternative $j \in \mathcal{X}$ receives a score $f_i(j) = s_{ij} > 0$, for all $i \in \mathcal{N}$. If we set $\hat{s}_i = \max_{j\in\mathcal{X}} s_{ij}$, then the performance index for the j-th option becomes

$$\rho_j = \min_{i\in\mathcal{N}} \frac{s_{ij}}{\hat{s}_i},\ j \in \mathcal{X}. \tag{31}$$

By Proposition 3, a *robust decision* $j^* \in \mathcal{R}$ can be found by computing the lower limit, as follows:

$$j^* \in \underline{\mathrm{Lim}}_{\varepsilon\to 0^+} \arg\max_{j\in\mathcal{X}} \left\{ (1-\varepsilon)\rho_j + \frac{\varepsilon}{n} \sum_{i\in\mathcal{N}} \frac{s_{ij}}{\hat{s}_i} \right\}. \tag{32}$$

This robust decision achieves the optimal performance index,

$$\rho^* = \max_{j\in\mathcal{X}} \rho_j = \rho_{j^*} = \min_{i\in\mathcal{N}} \frac{s_{ij^*}}{\hat{s}_i}. \tag{33}$$

As discussed in Section 3, simply maximizing the performance index ρ_j yields pseudo-robust solutions, which are generically inefficient; see, for instance, Example 6. We now compare the proposed approach with several alternative robustness criteria.

4.1.2. Alternative Robustness Criteria. The following three robustness criteria, defined in the present context of discrete action sets, are frequently used in the literature for dealing with parameter ambiguity.

i. The *Laplace criterion* corresponds to a weighted objective $F(\cdot|\boldsymbol{\lambda}_{\mathrm{Laplace}})$ in Equation (1) with a uniform weight, $\boldsymbol{\lambda}_{\mathrm{Laplace}} = (1,\dots,1)/n \in \Delta$, leading to the "Laplace solution,"

$$j_{\mathrm{Laplace}} \in \arg\max_{j\in\mathcal{X}} F(j|\boldsymbol{\lambda}_{\mathrm{Laplace}}) = \arg\max_{j\in\mathcal{X}} \sum_{i=1}^{n} s_{ij}. \tag{34}$$

ii. The *worst-case criterion* is defined as the minimum payoff across the admissible weights, which yields a so-called "maximin solution,"

$$j_{\mathrm{WC}} \in \arg\max_{j\in\mathcal{X}} \min_{\boldsymbol{\lambda}\in\Delta} F(j|\boldsymbol{\lambda}) = \arg\max_{j\in\mathcal{X}} \min_{i\in\mathcal{N}} s_{ij}. \tag{35}$$

iii. The *absolute-regret criterion* evaluates absolute regret, that is, the maximum difference ex post between what is and the best that could have been, resulting in the "(minimax) absolute-regret solution,"

$$\begin{aligned} j_{\mathrm{AR}} &\in \arg\min_{j\in\mathcal{X}} \max_{\boldsymbol{\lambda}\in\Delta} \{F^*(\boldsymbol{\lambda}) - F(j|\boldsymbol{\lambda})\} \\ &= \arg\min_{j\in\mathcal{X}} \max_{i\in\mathcal{N}} \{\hat{s}_i - s_{ij}\}. \end{aligned} \tag{36}$$

4.1.3. Comparison. The following example, which features a finite action set, illustrates the proposed robust

solution in Equation (32) against decisions recommended by alternative robustness criteria, notably the Laplace criterion in Equation (34), the WC (or maximin) solution in Equation (35), and the solution minimizing the (maximum) AR in Equation (36).

Example 16. Consider a discrete-choice situation for $J = 5$ options, for which the scores s_{ij} across $n = 3$ criteria are recorded in Table 1. In the context of relative robustness, options 2 and 5 are tied for the highest performance index in Equation (31) and are thus both pseudo-robust, so $\Psi = \{2,5\}$. At the same time, option 5 Pareto-dominates option 2 (with $\mathcal{P} = \{1,3,4,5\}$), so that $j^* = 5 \in \mathcal{R} = \Psi \cap \mathcal{P}$ is the unique robust choice. This solution can also be obtained directly from Equation (32), written in the form

$$j^* \in \underline{\mathrm{Lim}}_{\varepsilon \to 0^+} \arg\max_{j \in \mathcal{X}} \Phi_\varepsilon(j),$$

where $\Phi_\varepsilon(j) = (1-\varepsilon)\rho_j + (\varepsilon/n)\sum_{i=1}^n (s_{ij}/\hat{s}_i)$, for all $j \in \mathcal{X}$ and all $\varepsilon \in [0,1]$. Indeed, because $\rho_2 = \rho_5 = 3/11 > \rho_1 = \rho_4 = 4/15 > \rho_3 = 1/24$, for sufficiently small $\varepsilon \in (0,1]$ the ε-augmented performance index is

$$\max_{j \in \{2,5\}} \Phi_\varepsilon(j) > \Phi_\varepsilon(4) > \Phi_\varepsilon(3).$$

Meanwhile, it is

$$\Phi_\varepsilon(5) - \Phi_\varepsilon(2) = \frac{\varepsilon}{3}\left(\frac{12-7}{24} + \frac{9-8}{15}\right) = \frac{\varepsilon}{3}\left(\frac{11}{40}\right) > 0, \ \varepsilon \in (0,1].$$

Therefore, $\mathcal{R}_\varepsilon = \{5\}$, for all sufficiently small $\varepsilon > 0$, so $j^* \in \underline{\mathrm{Lim}}_{\varepsilon \to 0^+} \mathcal{R}_\varepsilon = \{5\}$, and one finds $j^* = 5$ as the unique robust option. Regarding alternative robustness evaluation, both the Laplace criterion in Equation (34) and the absolute-regret criterion in Equation (36) produce the solution $j_{\mathrm{Laplace}} = j_{\mathrm{AR}} = 3$ with very poor performance in at least one criterion. In fact, changing the payoff s_{13} from one to zero would produce even a zero performance index and zero worst-case performance guarantee for that same decision (still optimal under these criteria),[39] which may be rather difficult to justify in any real-world scenario. Finally, maximizing the worst-criterion performance in Equation (35) leads to indifference between options 1 and 4, at a suboptimal performance index and largest absolute regret. By contrast, the proposed robust solution $j^* = 5$ provides (by construction) the best performance index and a reasonable compromise solution in terms of the other criteria. It guarantees a strictly positive performance across all criteria. From Equations (19) and (20) we find that the corresponding (normalized) robust weight vector is

$$\hat{\boldsymbol{\lambda}} = \left(\frac{1}{12} + \frac{1}{6} + \frac{1}{36}\right)^{-1}\left(\frac{1}{12}, \frac{1}{6}, \frac{1}{36}\right) = (0.3, 0.6, 0.1).$$

Consistent with Equation (21), it is such that

$$\hat{\lambda}_i s_{ij^*} = \left(\sum_{i \in \mathcal{N}} \frac{1}{s_{ij^*}}\right)^{-1} = \left(\frac{1}{12} + \frac{1}{6} + \frac{1}{36}\right)^{-1} = 3.6 \geq \max_{j \in \mathcal{X} \setminus \{j^*\}} \min_{i \in \mathcal{N}} \hat{\lambda}_i s_{ij}, \ \ i \in \{1,2,3\},$$

where 3.6 corresponds to the harmonic mean of the criterion scores for the robust option j^*. Consider now the robustly reweighted scores $\sigma_{ij} = \hat{\lambda}_i s_{ij}$, for all $(i,j) \in \mathcal{N} \times \mathcal{X}$, displayed in Table 2, together with the associated robustness evaluations. It is interesting that after reweighting, option 3 goes from lowest to largest absolute regret, which highlights the sensitivity of the AR criterion to uncertainty about the relative importance of the criteria. The maximin payoff (WC) solution shifts from options 1 and 4 to option 5, whereas the Laplace solution shifts from option 3 to option 4. By contrast, the performance index remains unaffected by the robust reweighting (or any other reweighting), so option 5 is still the unique robust solution.

4.2. Data-Driven Approach

Consider any (nonempty, compact) action set $\mathcal{X} \subset \mathbb{R}^m$ for some integer $m \geq 1$, as in Section 2. Assume further that the decision maker does not know the functional form of the multicriterion $\boldsymbol{f} = (f_1, \ldots, f_n)$, but is still able to observe its value $\boldsymbol{s}^k = (s_1^k, \ldots, s_n^k)$ for different decisions $\boldsymbol{x}^k \in \mathcal{X}$ over the course of $K \geq 1$ experiments, where $k \in \mathcal{K} = \{1, \ldots, K\}$. Indeed, given the *realized score set* $\hat{\mathcal{S}} = \{\boldsymbol{s}^k : k \in \mathcal{K}\}$ and the *realized action set* $\hat{\mathcal{X}} = \{\boldsymbol{x}^k : k \in \mathcal{K}\} \subset \mathcal{X}$, the *average criterion scores* are[40]

$$\hat{f}_i(\boldsymbol{x}) = \frac{\sum_{k \in \mathcal{K}} \mathbf{1}_{\{\boldsymbol{x}^k = \boldsymbol{x}\}} s_i^k}{\sum_{k \in \mathcal{K}} \mathbf{1}_{\{\boldsymbol{x}^k = \boldsymbol{x}\}}}, \ \ (\boldsymbol{x}, i) \in \hat{\mathcal{X}} \times \mathcal{N},$$

for all $(\boldsymbol{x}, i) \in \hat{\mathcal{X}} \times \mathcal{N}$. With this, the (*data-driven*) *criterion-*

Table 1. Discrete Decision Options in Example 16, Evaluated with Different Robustness Criteria

	Criteria			Performance ratios/index				Alternative evaluations		
Option (j)	s_{1j}	s_{2j}	s_{3j}	$s_{1j}/\hat{s}_1$	$s_{2j}/\hat{s}_2$	$s_{3j}/\hat{s}_3$	ρ_j	Laplace	WC	AR
1	24	20	16	1	10/11	4/15	4/15	20	**16**	44
2	7	6	32	7/24	3/11	8/15	**3/11**	15	6	28
3	1	2	60	1/24	1/11	1	1/24	**21**	1	**23**
4	22	22	16	11/12	1	4/15	4/15	20	**16**	44
5	12	6	36	1/2	3/11	3/5	**3/11**	18	6	24

Note. Values in **bold** indicate optimality for the corresponding robustness criterion.

Table 2. Discrete Decision Options in Example 16, Evaluated after Robust Reweighting

	Criteria			Performance ratios/index				Alternative evaluations		
Option (j)	σ_{1j}	σ_{2j}	σ_{3j}	$\sigma_{1j}/\hat{\sigma}_1$	$\sigma_{2j}/\hat{\sigma}_2$	$\sigma_{3j}/\hat{\sigma}_3$	$\hat{\rho}_j$	Laplace	WC	AR
1	7.2	12	1.6	1	10/11	4/15	4/15	20.8	1.6	**4.4**
2	2.1	3.6	3.2	7/24	3/11	8/15	**3/11**	8.9	2.1	9.6
3	0.3	1.2	6	1/24	1/11	1	1/24	7.5	0.3	12
4	6.6	13.2	1.6	11/12	1	4/15	4/15	**21.4**	1.6	**4.4**
5	3.6	3.6	3.6	1/2	3/11	3/5	**3/11**	10.8	**3.6**	9.6

Note. Values in **bold** indicate optimality for the corresponding robustness criterion.

specific performance ratios are given by

$$\hat{\phi}_i(\boldsymbol{x}) = \frac{\hat{f}_i(\boldsymbol{x})}{\hat{f}_i^*}, \ (\boldsymbol{x}, i) \in \hat{\mathcal{X}} \times \mathcal{N},$$

where $\hat{f}_i^* = \max \hat{f}_i(\hat{\mathcal{X}})$ is the i-th maximum average criterion score. The (*data-driven*) *performance index*,

$$\hat{\rho}(\boldsymbol{x}) = \min_{i \in \mathcal{N}} \hat{\phi}_i(\boldsymbol{x}), \ \boldsymbol{x} \in \hat{\mathcal{X}},$$

is defined for all observed decisions. A (*data-driven*) *robust decision* $\hat{\boldsymbol{x}}^*$ (among all sampled actions in $\hat{\mathcal{X}}$) can then be determined in the standard way by applying Proposition 3 to the observational analogues of our standard approach:

$$\hat{\boldsymbol{x}}^* \in \underline{\mathrm{Lim}}_{\varepsilon \to 0^+} \arg\max_{\boldsymbol{x} \in \hat{\mathcal{X}}} \left\{ (1-\varepsilon)\hat{\rho}(\boldsymbol{x}) + \frac{\varepsilon}{n} \sum_{i \in \mathcal{N}} \hat{\phi}_i(\boldsymbol{x}) \right\}. \quad (37)$$

In other words, $\hat{\rho}(\boldsymbol{x})$ captures the worst-case relative performance of action $\boldsymbol{x}$ across all observed criteria, and $\hat{\boldsymbol{x}}^*$ identifies a robust decision, which is efficient and has the best-possible worst-case performance. Note that instead of going through the motions of actually taking the limit, it is sufficient to maximize the ε-augmented performance index, that is, the maximand in Equation (37), for a sufficiently small $\varepsilon \in (0, 1]$. This approximates the optimal robustness performance, captured by the optimal performance index ρ^*, arbitrarily closely by virtue of the ε-performance guarantee provided in Proposition 4. Finally, as in Equations (19) and (20), the (*data-driven*) *robust weight* is

$$\hat{\boldsymbol{\lambda}} = \left(\sum_{i \in \mathcal{N}} \frac{1}{\hat{f}_i(\hat{\boldsymbol{x}}^*)} \right)^{-1} \left(\frac{1}{\hat{f}_i(\hat{\boldsymbol{x}}^*)} \right)_{i=1}^{n} \in \Delta. \quad (38)$$

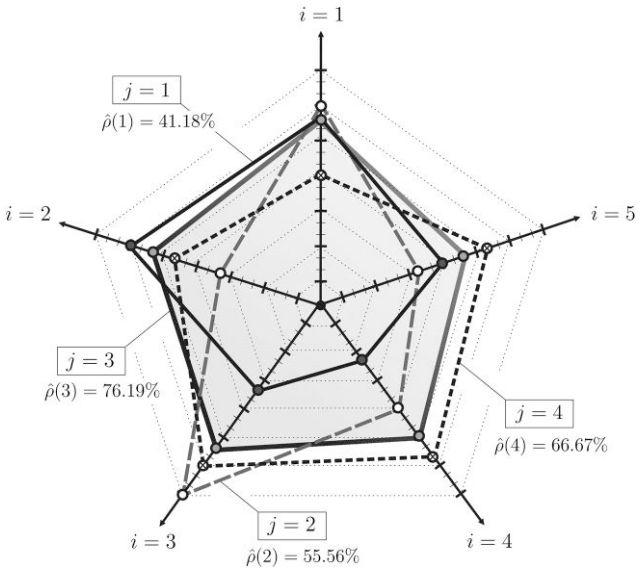
Based on the observed data, this vector provides a robust estimate of the tradeoffs between the different criteria from the vantage point of relative robustness.

Example 17. Consider the joint evaluation of human performance on a certain task (such as giving a university lecture) by K evaluators who score J individuals in N different performance dimensions (or criteria) on a Likert scale (from one to seven),[41] so $\mathcal{K} = \{1, \ldots, K\}$, $\mathcal{N} = \{1, \ldots, N\}$, and $\mathcal{X} = \{1, \ldots, J\}$. Figure 8 shows a spider plot comparing $J = 4$ different individuals. Assuming that all evaluators scored all the individuals using a seven-point Likert scale, the realized score set $\hat{\mathcal{S}}$ has elements $\boldsymbol{s}^k \in \{1, \ldots, 7\}^N$, for all $k \in \mathcal{K}$, whereas the realized action set $\hat{\mathcal{X}}$ is equal to $\mathcal{X}$. Individual 3, although never achieving a highest criterion-specific average score, exhibits the best data-driven performance index ($\hat{\rho}^* = \hat{\rho}(3) = 76.19\% = \max_{j \in \{1,2,3,4\}} \hat{\rho}(j)$). Because in a noisy observation environment ties in the performance index are fairly unlikely, the set of pseudo-robust alternatives Ψ tends to be a singleton, thus also resulting in a singleton set of robust options (e.g., $\mathcal{R} = \{3\}$). This example illustrates how the data-driven approach identifies robust performance through balanced tradeoffs rather than peak performance in any one dimension, thus favoring individual 3, whose weakest dimension is relatively strong.

4.3. Real-World Applications

Relatively robust multicriteria optimization provides a flexible and transparent framework for decision-making under uncertainty. As developed in Sections 2

Figure 8. Average Scores for $N = 5$ Criteria (Evaluated by $K = 3$ Experts), and Resulting Data-Driven Performance Indices for $J = 4$ Options in Example 17



Note. All options are Pareto-efficient (i.e., $\mathcal{P} = \{1, 2, 3, 4\}$); only option 3 is also pseudo-robust (i.e., $\Psi = \{3\}$), so $\mathcal{R} = \Psi \cap \mathcal{P} = \{3\}$.

and 3, it is especially well suited for complex problems where the relative importance of multiple criteria is ambiguous or hard to justify. The method systematically balances competing objectives using only minimal assumptions on the decision maker's preferences, and delivers solutions that are both Pareto-efficient and robust (cf. Proposition 3). We illustrate its practical relevance in three diverse contexts: the energy trilemma, quality-adjusted life year (QALY)-based health evaluations, and corporate resource allocation. In each case, the relatively robust methodology facilitates structured tradeoff management and data-informed robustness, whether operating in an abstract policy space or on a finite empirical action set (cf. Sections 4.1 and 4.2).

4.3.1. Energy Trilemma: Balancing Energy Security, Equity, and Sustainability. The World Energy Council's Energy Trilemma framework evaluates nations on three key dimensions, namely "energy security" (i.e., the reliability and resilience of energy supply), "energy equity" (i.e., the accessibility and affordability of energy for all citizens), and "environmental sustainability" (i.e., the reduction of greenhouse gas emissions and environmental impact). Nations face challenges in improving one dimension without compromising others (World Energy Council 2024).[42] For instance, increasing energy equity by subsidizing fossil fuels may undermine sustainability, while prioritizing environmental sustainability through renewable energy investment might initially reduce energy security or equity. An optimal strategy depends on the relative importance attributed to each of the three criteria. In this setting, a relatively robust approach can

- Identify Robust Policies: By modeling each nation's energy policies and outcomes as feasible decisions, the method identifies those policies that achieve strong tradeoffs across all three dimensions. For example, a relatively robust policy might balance investment in renewable energy, grid modernization, and subsidies for low-income households, ensuring consistent performance regardless of variations of the relative importance across dimensions, which might still be importance-ranked (cf. Section 3.7).
- Quantify Tradeoffs: The approach provides a clear representation of tradeoffs, such as how much equity might need to be sacrificed for a given improvement in sustainability under different weighting scenarios.
- "Robustify" Strategic Goals: Policymakers can develop long-term strategies that are resilient to evolving societal priorities, such as shifts toward greater emphasis on the precautionary principle (cf. Section 3.9).

4.3.2. QALY Impact of Diseases: Evaluating Quality-Adjusted Life Years. QALYs measure the impact of diseases and medical treatments by combining longevity and quality of life into a single index (Pliskin 1974, Pliskin and Beck 1976, Miyamoto et al. 1998).[43] The impact of different impairments—such as mobility loss, chronic pain, or cognitive decline—depends on how these are weighted relative to each other in calculating overall health outcomes. Subject to restrictions on weights (e.g., a priority ranking; cf. Example 12) and with a flexible aggregation of criteria (cf. Section 3.9) the method can

- Handle Uncertain Weightings: When precise utility-weightings for different impairments are unavailable (as one would generally assume), the method identifies health interventions or treatment plans that remain effective across a wide range of subjective assessments. For example, it might suggest treatments that optimize outcomes for both pain relief and mobility restoration, ensuring robust quality-of-life improvements.
- Optimize Resource Allocation: Health authorities can use the method to prioritize interventions that deliver the highest QALY improvements per dollar spent, even when the relative importance of various health dimensions (e.g., physical versus mental health) is uncertain.[44]
- Support Evidence-Based Policy: The method provides a performance guarantee for proposed policies, demonstrating their effectiveness regardless of the relative emphasis placed on specific impairments.

4.3.3. Resource Allocation in Companies: Balancing Risk, Resources, and Rival Assets. In organizations, resource allocation involves deciding how to distribute limited resources—financial, human, or physical—across competing projects. Each project is evaluated based on multiple criteria, such as "risk of noncompletion" (i.e., the probability that a project fails because of delays or unforeseen issues), "use of human resources" (i.e., the availability and workload of skilled personnel required for the project), and "use of rival assets" (i.e., the occupation of assets that cannot be used by multiple projects simultaneously, for example, specialized machinery). Applying the framework of relatively robust decision making allows organizations to[45]

- Identify Balanced Portfolios: By modeling projects and their criteria as feasible decisions $\boldsymbol{x} \in \mathcal{X}$, with performance criteria $f_1(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})$, the method selects a portfolio of projects that balances competing objectives, ensuring efficient resource use even when preferences over criteria are unclear.
- Manage Tradeoffs: By aggregating criterion-specific performance ratios $\phi_i(\boldsymbol{x})$ into a performance index $\rho(\boldsymbol{x})$, the method reveals, for example, how prioritizing low-risk projects impacts resource utilization and asset deployment, informing decisions about a robust balance between risk and resource efficiency.
- Adapt to Uncertainty: As organizational priorities shift (e.g., toward innovation or risk aversion), the method ensures that allocation strategies can remain robust across changing criteria (cf. Section 3.8) and

weight configurations for the criteria, and possibly across different aggregation methods (cf. Section 3.9).

In settings with discrete project sets and empirical observations, a data-driven approach can be applied (cf. Section 4.2).

5. Conclusion

Multicriteria optimization aims to resolve conflicts between competing objectives by finding Pareto-efficient decisions for which improving one criterion necessarily degrades at least one other. Although Pareto-efficiency sets a minimum standard of nonwastefulness, practical decision making often requires selecting a single compromise solution from the available Pareto frontier. Scalarization, such as assigning weights to criteria, is a common method for operationalizing this selection, but it assumes knowledge of the weights, which may not be available or easily justifiable, and it may also exclude Pareto-efficient solutions when the action set is nonconvex.

Here, we seek to mitigate the dependence on precise weight specifications by introducing a performance index that evaluates the worst-case weighted performance of a decision relative to its maximum potential. A Pareto-efficient decision that maximizes this index is viewed as relatively robust, balancing competing criteria in a way that offers resilience to weight uncertainty. A critical feature of the method is its computational simplicity, relying only on the compactness of the feasible set and the continuity of criterion functions to guarantee the existence and basic regularity of the solutions. This ensures broad applicability, even for complex, nonconvex problems. Criterion ambiguity and more general aggregation methods can be accommodated. In the case of finite sets, this method is completely general and can be operationalized through a data-driven approach under virtually no assumptions.

Appendix. Proofs

Proof of Proposition 1. The proof proceeds by analyzing what happens when either one of the two inequalities is violated, and subsequently when both are violated. This yields three cases, each of which leads to a contradiction.

- **Case 1:** If $f_i(\hat{x}) < f_i(x)$ and $f_j(\hat{x}) \le f_j(x)$, then

$$F^*(\hat{\boldsymbol{\lambda}}) = F(\hat{x}|\hat{\boldsymbol{\lambda}}) = \sum_{l=1}^{n} \hat{\lambda}_l f_l(\hat{x}) < \sum_{l=1}^{n} \hat{\lambda}_l f_l(x) = F(x|\hat{\boldsymbol{\lambda}}),$$

which is a contradiction.

- **Case 2:** If $f_i(\hat{x}) \ge f_i(x)$ and $f_j(\hat{x}) > f_j(x)$, then

$$F^*(\boldsymbol{\lambda}) = F(x|\boldsymbol{\lambda}) = \sum_{l=1}^{n} \lambda_l f_l(x) < \sum_{l=1}^{n} \lambda_l f_l(\hat{x}) = F(\hat{x}|\boldsymbol{\lambda}),$$

which is a contradiction.

- **Case 3:** If $f_i(\hat{x}) < f_i(x)$ and $f_j(\hat{x}) > f_j(x)$, then

$$\begin{aligned} F^*(\hat{\boldsymbol{\lambda}}) &= F(\hat{x}|\boldsymbol{\lambda}) + \varepsilon(f_i(\hat{x}) - f_j(\hat{x})) \ge F(x|\hat{\boldsymbol{\lambda}}) \\ &= F(x|\boldsymbol{\lambda}) + \varepsilon(f_i(x) - f_j(x)). \end{aligned}$$

But because $\varepsilon(f_i(x) - f_j(x)) > \varepsilon(f_i(\hat{x}) - f_j(\hat{x}))$, this implies

$$\begin{aligned} F(\hat{x}|\lambda) &\ge F(x|\hat{\boldsymbol{\lambda}}) - \varepsilon(f_i(\hat{x}) - f_j(\hat{x})) \\ &= \underbrace{F(x|\boldsymbol{\lambda})}_{=F^*(\boldsymbol{\lambda})} + \underbrace{\varepsilon(f_i(x) - f_j(x)) - \varepsilon(f_i(\hat{x}) - f_j(\hat{x}))}_{>0} \\ &> F^*(\boldsymbol{\lambda}), \end{aligned}$$

which is a contradiction.

Hence, $f_i(\hat{x}) \ge f_i(x)$ and $f_j(\hat{x}) \le f_j(x)$, as claimed. □

Proof of Lemma 1. Fix $\delta > 0$ and $i \in \mathcal{N}$, and let $(x, \hat{x}) \in \mathcal{X}(w) \times \mathcal{X}(\hat{w})$. Then

$$\begin{aligned} F^*(\hat{w}) - F^*(w) &= F(\hat{x}|\hat{w}) - F(x|w) \\ &= \underbrace{F(\hat{x}|\hat{w}) - F(x|\hat{w})}_{\ge 0} + \underbrace{F(x|\hat{w}) - F(x|w)}_{=\delta f_i(x)} \ge \delta f_i(x), \end{aligned}$$

because, by assumption, $\hat{w} = w + \delta e_i$. As the preceding inequality holds for any $x \in \mathcal{X}(w)$, it follows that

$$F^*(\hat{w}) - F^*(w) \ge \delta(\max f_i(\mathcal{X}(w))) > 0,$$

and thus $F^*(\hat{w}) > F^*(w)$, which completes the proof. □

Proof of Proposition 2. We first show that, for any given decision, the performance ratio is quasiconcave in the weight. To that end, fix any feasible action $x \in \mathcal{X}$, and define

$$\mathcal{U}_c(x) = \{\boldsymbol{\lambda} \in \Delta : \varphi(x|\boldsymbol{\lambda}) \ge c\} \tag{A.1}$$

as the set of (normalized) weights $\boldsymbol{\lambda}$ that yield a performance ratio $\varphi(x|\boldsymbol{\lambda})$ in Equation (3) at a level of at least $c \in [0,1]$. In fact, the upper contour sets in Equation (A.1) are nested in their level, that is,

$$c \le \hat{c} \implies \mathcal{U}_{\hat{c}}(x) \subseteq \mathcal{U}_c(x), \tag{A.2}$$

for any $\hat{c} \in [0,1]$. Furthermore, for sufficiently small values of c, we have $\mathcal{U}_c(x) = \Delta$.[46] Thus, we may choose $c \in [0,1]$ such that $\mathcal{U}_c(x)$ is nonempty, and consider any two weights $\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}} \in \mathcal{U}_c(x)$. For any $\theta \in (0,1)$, their convex combination $\boldsymbol{\lambda}_\theta = \theta\boldsymbol{\lambda} + (1-\theta)\hat{\boldsymbol{\lambda}}$ belongs to $\mathcal{U}_c(x)$ if and only if

$$\varphi(x|\boldsymbol{\lambda}_\theta) = \frac{F(x|\boldsymbol{\lambda}_\theta)}{F^*(\boldsymbol{\lambda}_\theta)} \ge c. \tag{A.3}$$

Because $\varphi(x|\boldsymbol{\lambda}) \ge c$ and $\varphi(x|\hat{\boldsymbol{\lambda}}) \ge c$, the definition of the weighted objective in Equation (1) implies

$$\begin{aligned} F(x|\boldsymbol{\lambda}_\theta) &= F(x|\theta\boldsymbol{\lambda} + (1-\theta)\hat{\boldsymbol{\lambda}}) = \theta F(x|\boldsymbol{\lambda}) + (1-\theta)F(x|\hat{\boldsymbol{\lambda}}) \\ &\ge c(\theta F^*(\boldsymbol{\lambda}) + (1-\theta)F^*(\hat{\boldsymbol{\lambda}})), \end{aligned} \tag{A.4}$$

where we use that $\min\{F^*(\boldsymbol{\lambda}), F^*(\hat{\boldsymbol{\lambda}})\} > 0$ by the nontriviality condition (N). Moreover, the convex combination of the individually optimized objectives $F^*(\boldsymbol{\lambda})$ and $F^*(\hat{\boldsymbol{\lambda}})$ (each obtained from possibly different decisions) cannot be smaller than the value of the jointly optimized objective (from a single decision):

$$\begin{aligned} \theta F^*(\boldsymbol{\lambda}) + (1-\theta)F^*(\hat{\boldsymbol{\lambda}}) &= \max_{x,\hat{x}\in\mathcal{X}} \{\theta F(x|\boldsymbol{\lambda}) + (1-\theta)F(\hat{x}|\hat{\boldsymbol{\lambda}})\} \\ &\ge \max_{x\in\mathcal{X}} \{\theta F(x|\boldsymbol{\lambda}) + (1-\theta)F(x|\hat{\boldsymbol{\lambda}})\} \\ &= \max_{x\in\mathcal{X}} \{F(x|\theta\boldsymbol{\lambda} + (1-\theta)\hat{\boldsymbol{\lambda}})\} \;=\; F^*(\boldsymbol{\lambda}_\theta). \end{aligned} \tag{A.5}$$

Therefore, as long as $c \ge 0$, the fact that $F^*(\boldsymbol{\lambda}_\theta) > 0$, together with Equations (A.4) and (A.5), implies that Equation (A.3) holds. As an immediate consequence, it is $\boldsymbol{\lambda}_\theta \in \mathcal{U}_c(x)$, which in turn means that $\mathcal{U}_c(x)$ is convex for all $c \in [0,1]$ and all $x \in \mathcal{X}$, as claimed.

The convexity of the upper contour sets $\mathcal{U}_c(\boldsymbol{x})$ implies that $\varphi(\boldsymbol{x}|\boldsymbol{\lambda})$ is quasiconcave in $\boldsymbol{\lambda}$ (see, e.g., Arrow and Enthoven 1961, p. 780). Hence, the minimum of the performance ratio over all weights in Δ is attained on the boundary:

$$\min_{\boldsymbol{\lambda}\in\Delta}\varphi(\boldsymbol{x}|\boldsymbol{\lambda}) = \min_{\boldsymbol{\lambda}\in\partial\Delta}\varphi(\boldsymbol{x}|\boldsymbol{\lambda}).$$

The same quasiconcavity argument applies to any face of Δ, so the minimum must also lie on the boundary of each face. Repeating this argument recursively over edges and vertices, we conclude that the minimum is attained in the set $\Lambda = \{\boldsymbol{e}_1,\dots,\boldsymbol{e}_n\}$ of the n vertices of Δ. Thus, we obtain a form of "perfect complementarity,"

$$\rho(\boldsymbol{x}) = \min_{\boldsymbol{\lambda}\in\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_n\}}\left\{\frac{F(\boldsymbol{x}|\boldsymbol{\lambda})}{F^*(\boldsymbol{\lambda})}\right\} = \min_{i\in\mathcal{N}}\left\{\frac{F(\boldsymbol{x}|\boldsymbol{e}_i)}{F^*(\boldsymbol{e}_i)}\right\} = \min_{i\in\mathcal{N}}\left\{\frac{f_i(\boldsymbol{x})}{f_i^*}\right\} = \min_{i\in\mathcal{N}}\phi_i(\boldsymbol{x}),$$

as claimed in Equation (6), which concludes the proof. □

Proof of Lemma 2. For any $\boldsymbol{x}\in\mathcal{X}$, Proposition 2 provides for a representation of the performance index, which can be rewritten in the form $\rho(\boldsymbol{x}) = \phi_{\iota(\boldsymbol{x})}$, where $\iota(\boldsymbol{x})\in\mathcal{I}(\boldsymbol{x}) = \arg\min_{i\in\mathcal{N}}\phi_i(\boldsymbol{x})$. The set of pseudo-robust decisions Ψ in Equation (7) is nonempty by the extreme-value theorem (Rudin 1976, theorem 4.16, p. 89), and it is compact by the maximum theorem (Berge 1963, p. 116). These two theorems also guarantee that the correspondence $\mathcal{I}(\cdot)$ is upper semicontinuous and nonempty-valued. Consider now $\mathcal{P}$, which by Equations (9) and (10) can be written as

$$\mathcal{P} = \{\boldsymbol{x}\in\mathcal{X} : \nexists\,\boldsymbol{x}'\in\mathcal{X}\text{ s.t. }\boldsymbol{x}'\succ\boldsymbol{x}\} = \{\boldsymbol{x}\in\mathcal{X} : \boldsymbol{x}'\preceq\boldsymbol{x},\,\boldsymbol{x}'\in\mathcal{X}\},$$

where $\preceq$ is the negation of $\succ$. By the continuity of $\boldsymbol{f}$, the preference $\succ$ is continuous, so that $\mathcal{P}$ is compact. We now show that it must also be nonempty. For this, start with any $\boldsymbol{\xi}^0\in\Psi$: If $\boldsymbol{\xi}^0\in\mathcal{P}$, then immediately $\boldsymbol{\xi}^0\in\mathcal{R}$, so the claim holds. Otherwise, if $\boldsymbol{\xi}^0\notin\mathcal{P}$, then there exists $\boldsymbol{\xi}^1\in\mathcal{X}$ such that $\boldsymbol{\xi}^1\succ\boldsymbol{\xi}^0$. By Equation (9) there exists $i\in\mathcal{N}$ such that $f_i(\boldsymbol{\xi}^1) > f_i(\boldsymbol{\xi}^0)$ and $\boldsymbol{f}(\boldsymbol{\xi}^1)\geq\boldsymbol{f}(\boldsymbol{\xi}^0)$. Equivalently, $\phi_i(\boldsymbol{\xi}^1) > \phi_i(\boldsymbol{\xi}^0)$ and $\phi_j(\boldsymbol{\xi}^1)\geq\phi_j(\boldsymbol{\xi}^0)$, for all $j\in\mathcal{N}$, so that $\rho(\boldsymbol{\xi}^1)\geq\rho(\boldsymbol{\xi}^0)$. Because by assumption $\boldsymbol{\xi}^0\in\Psi$, by Equation (7) it must be $\rho(\boldsymbol{\xi}^1)\geq\rho(\boldsymbol{\xi}^0) = \rho^* = \max\rho(\mathcal{X})$, so $\boldsymbol{\xi}^1\in\Psi$ as well. In this manner one can now construct a sequence of decisions $\boldsymbol{\xi}^k$, for $k\in\{0,1,\dots\}$, continuing until $\boldsymbol{\xi}^k\in\mathcal{P}$ for some $k\geq 0$, which then establishes that $\Psi\cap\mathcal{P}$ is indeed nonempty. Failing that, by the Bolzano-Weierstrass theorem (Berge 1963, p. 67) the sequence $(\boldsymbol{\xi}^k)_{k\geq 0}$ must have a convergent subsequence in the compact set $\mathcal{X}$. Because the convergence is monotone in the (continuous) preference, there can only be a single accumulation point, which is equal to the limit of the sequence: $\hat{\boldsymbol{x}} = \lim_{k\to\infty}\boldsymbol{\xi}^k$. Given that the set of all pseudo-robust decisions Ψ is compact (i.e., in particular closed) as noted earlier, it is $\hat{\boldsymbol{x}}\in\Psi$. In addition, $\boldsymbol{x}'\preceq\hat{\boldsymbol{x}}$ for all $\boldsymbol{x}'\in\mathcal{X}$, as the upper and lower contour sets of a continuous preference are closed. But this means that $\hat{\boldsymbol{x}}\in\Psi\cap\mathcal{P}$, so the robust decision set $\mathcal{R}$ is both nonempty and compact, for it is the intersection of two compact sets. □

Proof of Lemma 3. By Lemma 2, the robust decision set is nonempty and compact. Hence, by the extreme-value theorem there exists $\hat{\boldsymbol{x}}^*\in\mathcal{R} = \Psi\cap\mathcal{P}$, and Equation (7) yields $\rho(\mathcal{R}) = \rho(\Psi\cap\mathcal{P}) = \{\rho^*\} = \rho(\Psi)$, as well as $\max\rho(\mathcal{P}) = \rho^* = \max\rho(\mathcal{X})$, establishing the claim. □

Proof of Lemma 4. (i) The claim follows by combining Equations (7) and (12). (ii) Fix $\varepsilon\in(0,1]$, and—following the outcome-based logic discussed in Remark 9—consider a selection

$$\boldsymbol{y}_\varepsilon^*\in\mathcal{Y}_\varepsilon = \arg\max_{\boldsymbol{y}\in\mathcal{Y}}\left\{\sum_{i\in\mathcal{N}} y_i : t_\varepsilon\leq y_1,\dots,t_\varepsilon\leq y_n,\, t_\varepsilon\in\mathcal{T}_\varepsilon\right\}.$$

We now show that $\boldsymbol{y}_\varepsilon^*$ is efficient (with respect to the coordinates of points in $\mathcal{Y}$).[47] Suppose it is not; then there exists a feasible $\hat{\boldsymbol{y}}\succ\boldsymbol{y}_\varepsilon^*$ which achieves a strictly greater payoff, a contradiction, which in turn establishes the claimed efficiency of $\boldsymbol{y}_\varepsilon^*$. Moreover, it is clear that $\mathcal{Y}_\varepsilon = \phi(\mathcal{R}_\varepsilon)$, so that

$$\mathcal{R}_\varepsilon = \phi^{-1}(\mathcal{Y}_\varepsilon) = \{\boldsymbol{x}\in\mathcal{X} : \phi(\boldsymbol{x})\in\mathcal{Y}_\varepsilon\}.$$

As a result, $\mathcal{R}_\varepsilon\subset\mathcal{P}$, for all $\varepsilon\in(0,1]$, as claimed. (iii) Let $\varepsilon\in(0,1]$, and consider any $(\boldsymbol{x},\hat{\boldsymbol{x}})\in\mathcal{R}_\varepsilon\times\Psi$. If $\boldsymbol{x}\notin\Psi$, then

$$\rho = \min_{i\in\mathcal{N}}\phi_i(\boldsymbol{x}) < \min_{i\in\mathcal{N}}\phi_i(\hat{\boldsymbol{x}}) = \rho^* = \max\rho(\mathcal{X}).$$

On the other hand, $\boldsymbol{x}\in\mathcal{R}_\varepsilon$ implies that $\Phi_\varepsilon(\boldsymbol{x})\geq\Phi_\varepsilon(\hat{\boldsymbol{x}})$. Let $\kappa = \sum_{i\in\mathcal{N}}(\phi_i(\boldsymbol{x}) - \phi_i(\hat{\boldsymbol{x}}))$ be the total coordinate-wise difference in performance. Then for any $\hat{\varepsilon} > 0$,

$$\Phi_{\hat{\varepsilon}}(\boldsymbol{x})\geq\Phi_{\hat{\varepsilon}}(\hat{\boldsymbol{x}}) \iff \hat{\varepsilon}\cdot\kappa\geq\rho^*-\rho > 0 \iff \hat{\varepsilon}\geq\frac{\rho^*-\rho}{\kappa} > 0.$$

This establishes the claim in part (iii), for any $\varepsilon'\in(0,\min\{\varepsilon,(\rho^*-\rho)/\kappa\})$ and $\hat{\varepsilon}\in(0,\varepsilon')$. For $\hat{\varepsilon} = 0$, the claim follows from part (i). □

Proof of Proposition 3. We first note that $\mathcal{R}_\varepsilon$ is upper semicontinuous in $\varepsilon\in[0,1]$, and set $\mathcal{Q} = \underline{\mathrm{Lim}}_{\varepsilon\to 0^+}\mathcal{R}_\varepsilon$. The proof proceeds in two steps. We first show that $\mathcal{Q}\subset\mathcal{R}$ and then $\mathcal{Q}\neq\emptyset$.

Step 1: $\mathcal{Q}\subset\mathcal{R}$. By upper semicontinuity of $\mathcal{R}_\varepsilon$ we have that $\mathcal{Q}\subset\mathcal{R}_0 = \Psi$ (see, e.g., Aubin and Frankowska 1990, p. 41), taking into account that Ψ is compact; cf. Endnote 19 for Section 3.1. Because $\mathcal{R}_\varepsilon\subset\mathcal{P}$, for all $\varepsilon\in(0,1]$, we further obtain $\mathcal{Q}\subset\mathrm{cl}\,\mathcal{P}$. Assume that there is a point $\boldsymbol{x}\in\mathcal{Q}\setminus\mathcal{P}$, which means that $\boldsymbol{x} = \lim_{\varepsilon\to 0^+}\boldsymbol{x}_\varepsilon$, for some selection $\boldsymbol{x}_\varepsilon\in\mathcal{R}_\varepsilon$, where $\varepsilon\in(0,1]$. Because $\boldsymbol{x}\notin\mathcal{P}$, there exists $\hat{\boldsymbol{x}}\in\mathcal{P}$ such that $\hat{\boldsymbol{x}}\succ\boldsymbol{x}$. That is, $\phi_j(\hat{\boldsymbol{x}}) > \phi_j(\boldsymbol{x})$ for some $j\in\mathcal{N}$, and $\phi(\hat{\boldsymbol{x}})\geq\phi(\boldsymbol{x})$. If we denote by $A = (1/n)\sum_{i\in\mathcal{N}}\phi_i$ the average performance over all criteria, then $\boldsymbol{x}\notin\mathcal{P}$ implies that $A(\boldsymbol{x}) < A(\hat{\boldsymbol{x}})$. By the continuity of $A(\cdot)$ on $\mathcal{X}$, it is therefore $\lim_{\varepsilon\to 0^+}A(\boldsymbol{x}_\varepsilon) = A(\boldsymbol{x}) < A(\hat{\boldsymbol{x}})$. Hence, there exists a $\overline{\varepsilon}\in(0,1)$ such that $A(\boldsymbol{x}_\varepsilon) < A(\hat{\boldsymbol{x}})$ for all $\varepsilon\in(0,\overline{\varepsilon})$. Thus, for any given $\varepsilon\in(0,\overline{\varepsilon})$:

$$\begin{aligned}\Phi_\varepsilon(\hat{\boldsymbol{x}}) - \Phi_\varepsilon(\boldsymbol{x}_\varepsilon) &= (1-\varepsilon)(\rho(\hat{\boldsymbol{x}}) - \rho(\boldsymbol{x}_\varepsilon)) + \varepsilon(A(\hat{\boldsymbol{x}}) - A(\boldsymbol{x}_\varepsilon))\\ &\geq \varepsilon(A(\hat{\boldsymbol{x}}) - A(\boldsymbol{x}_\varepsilon)) > 0,\end{aligned}$$

where we have taken into account that $\hat{\boldsymbol{x}}\in\Psi$, so $\rho(\hat{\boldsymbol{x}}) = \max\rho(\mathcal{X})\geq\rho(\boldsymbol{x}_\varepsilon)$. But this means $\Phi_\varepsilon(\boldsymbol{x}_\varepsilon) < \Phi_\varepsilon(\hat{\boldsymbol{x}})$, contradicting $\boldsymbol{x}_\varepsilon\in\mathcal{R}_\varepsilon$. Hence, we have shown that necessarily $\mathcal{Q}\setminus\mathcal{P} = \emptyset$. Because we already know that $\mathcal{Q}\subset\Psi\cap(\mathrm{cl}\,\mathcal{P})$, one obtains $\mathcal{Q}\subset\Psi\cap\mathcal{P} = \mathcal{R}$, as claimed.

Step 2: $\mathcal{Q}\neq\emptyset$. Consider a monotone sequence $(\varepsilon_k)_{k=0}^\infty\subset(0,1)$ with $\varepsilon_{k+1} < \varepsilon_k$, for all $k\geq 0$, and such that $\lim_{k\to\infty}\varepsilon_k = 0$. Using the same selection $\boldsymbol{x}_\varepsilon$ as in Step 1, the sequence $(\boldsymbol{x}_{\varepsilon_k})_{k=0}^\infty\subset\mathcal{X}$ is a sequence of points contained in the compact set $\mathcal{X}$, so that by the Bolzano-Weierstrass theorem (Berge 1963, p. 67) there exists a convergent subsequence $(\boldsymbol{x}_{\varepsilon_{k_j}})_{j=0}^\infty\subset$

$(\boldsymbol{x}_{\varepsilon_k})_{k=0}^{\infty}$ (with limit in $\mathcal{X}$). That is, there exists $q \in \mathcal{X}$, such that $q = \lim_{j\to\infty} \boldsymbol{x}_{\varepsilon_{k_j}}$, and by the definition of $\mathcal{Q}$ we therefore obtain that $q \in \mathcal{Q}$ and thus, $\mathcal{Q} \neq \emptyset$, as stated in the result.

This concludes the proof. □

Proof of Lemma 5. Continuity of the optimal ε-augmented performance index Φ_ε^* over $\varepsilon \in [0,1]$ follows from the maximum theorem (Berge 1963, p. 116), given the continuity of Φ_ε in both arguments and the compactness of $\mathcal{X}$. To establish monotonicity, fix $\varepsilon', \varepsilon'' \in [0,1]$ with $\varepsilon'' > \varepsilon'$. By optimality of $\boldsymbol{x}_{\varepsilon''}$ for $\Phi_{\varepsilon''}$ and feasibility of $\boldsymbol{x}_{\varepsilon'}$ at ε'', we obtain

$$\Phi_{\varepsilon''}^* = (1-\varepsilon'')\rho_{\varepsilon''} + \varepsilon''\mu_{\varepsilon''} \geq (1-\varepsilon'')\rho_{\varepsilon'} + \varepsilon''\mu_{\varepsilon'} = \Phi_{\varepsilon''}(\boldsymbol{x}_{\varepsilon'}).$$

Subtracting $\Phi_{\varepsilon'}^* = (1-\varepsilon')\rho_{\varepsilon'} + \varepsilon'\mu_{\varepsilon'}$, we find

$$\Phi_{\varepsilon''}^* - \Phi_{\varepsilon'}^* \geq \Phi_{\varepsilon''}(\boldsymbol{x}_{\varepsilon'}) - \Phi_{\varepsilon'}^* = (\varepsilon''-\varepsilon')(\mu_{\varepsilon'} - \rho_{\varepsilon'}) \geq 0,$$

because $\mu_{\varepsilon'} \geq \rho_{\varepsilon'}$ by definition. This confirms that Φ_ε^* is nondecreasing on $[0,1]$. To establish convexity, we show that regular increments in ε increase the corresponding differences of the value function. If $\check{\varepsilon} = (\varepsilon' + \varepsilon'')/2$ denotes the midpoint between ε' and ε'', it is sufficient to show that the difference of successive differences,

$$\Delta = (\Phi_{\varepsilon''}^* - \Phi_{\check{\varepsilon}}^*) - (\Phi_{\check{\varepsilon}}^* - \Phi_{\varepsilon'}^*),$$

must be nonnegative. Indeed, setting $\delta = (\varepsilon'' - \varepsilon')/2 > 0$, it is

$$\Phi_{\varepsilon''}^* - \Phi_{\check{\varepsilon}}^* \geq \delta(\mu_{\check{\varepsilon}} - \rho_{\check{\varepsilon}}) \geq \Phi_{\check{\varepsilon}}^* - \Phi_{\varepsilon'}^*,$$

where the first inequality follows from $\Phi_{\varepsilon''}^* \geq \Phi_{\varepsilon''}(\boldsymbol{x}_{\check{\varepsilon}})$, and the second from $\Phi_{\check{\varepsilon}}^* \geq \Phi_{\varepsilon'}(\boldsymbol{x}_{\check{\varepsilon}})$. Together, these imply $\Delta \geq 0$, which proves that Φ_ε^* is convex on $[0,1]$. □

Proof of Lemma 6. (i)/(iv) The function $\Phi_\varepsilon(\cdot)$ conforms to Equation (1) as a weighted objective comprising the two criteria $\rho(\cdot)$ and $\mu(\cdot)$. By Proposition 1, increasing $\varepsilon \in [0,1)$, which shifts weight from the first criterion to the second, can only decrease the optimal value of the first and increase that of the second. Thus, ρ_ε is decreasing and μ_ε is increasing in $\varepsilon \in [0,1]$, as claimed. (ii) Because $\Phi_0 = \rho$, we have that $\rho_0 = \max_{\boldsymbol{x}\in\mathcal{X}} \Phi_0(\boldsymbol{x}) = \rho^*$. Hence, by part (i) we have that $\rho^* \geq \rho_\varepsilon$, for all $\varepsilon \in [0,1]$. (iii) Because ρ_ε is bounded above by ρ^* and nonincreasing in ε, the sequence $(\rho_\varepsilon)_{\varepsilon\in[0,1]}$ converges as $\varepsilon \to 1^-$ by the monotone convergence theorem (see, e.g., Rudin 1976, theorem 3.14, p. 55). Moreover, because ρ^* is the least upper bound, the limit of ρ_ε as $\varepsilon \to 1^-$ must equal ρ^*. (v) This result follows from the monotonicity of Φ_ε^* established in Lemma 5. Because $\Phi_\varepsilon^* = (1-\varepsilon)\rho_\varepsilon + \varepsilon\mu_\varepsilon$ is nondecreasing in ε, and $\rho_\varepsilon \leq \rho^*$, by parts (i) and (ii), it must be that $\mu_\varepsilon \geq \rho^*$ for all $\varepsilon \in [0,1]$; otherwise, the convex combination would decrease, contradicting monotonicity. □

Proof of Proposition 4. By Equations (15) and (16), along with Lemma 5, we have

$$0 \leq \psi_\varepsilon = \rho^* - \rho_\varepsilon = \Phi_0^* - \frac{\Phi_\varepsilon^* - \varepsilon\mu_\varepsilon}{1-\varepsilon} \leq \Phi_0^* - \Phi_\varepsilon^* + \varepsilon\mu_\varepsilon \leq \varepsilon\mu_\varepsilon,\ \varepsilon \in [0,1].$$

Thus, given a desired approximation-error bound $\delta > 0$, for any $\hat{\varepsilon} \in (0,1]$, it is

$$\varepsilon \leq \min\{\hat{\varepsilon}, \delta/\mu_{\hat{\varepsilon}}\} \ \Rightarrow\ \psi_\varepsilon \leq \delta.$$

Because, by Lemma 6 (iv), the function $\hat{\varepsilon} \longmapsto \delta/\mu_{\hat{\varepsilon}}$ is decreasing, a simpler implication follows by choosing $\hat{\varepsilon} = 1$:

$$\varepsilon \leq \delta/\mu_1 \ \Rightarrow\ \psi_\varepsilon \leq \delta.$$

This establishes the claimed approximation guarantee. □

Proof of Lemma 7. Fix any $\varepsilon \in [0,1]$. By Equation (16), the difference between the upper and lower bounds in Equation (17) is given by

$$d_\varepsilon = \Phi_\varepsilon^* - \rho_\varepsilon = \varepsilon(\mu_\varepsilon - \rho_\varepsilon) \geq 0.$$

The midpoint estimator $\hat{\rho}_\varepsilon$ defined in Equation (18) is the arithmetic average of these bounds:

$$\hat{\rho}_\varepsilon = \frac{\Phi_\varepsilon^* + \rho_\varepsilon}{2}.$$

By construction, this implies

$$|\hat{\rho}_\varepsilon - \rho^*| \leq \frac{1}{2}(\Phi_\varepsilon^* - \rho_\varepsilon) = \frac{d_\varepsilon}{2}.$$

Therefore, $\hat{\rho}_\varepsilon$ approximates ρ^* to within $d_\varepsilon/2$, as claimed. □

Proof of Lemma 8. Let $\Omega \subset \mathbb{R}_+^n \setminus \{\mathbf{0}\}$ be a nonempty compact set, which is not reduced to the origin.

i. (a) Because Ω is nonempty, its radial projection $\pi(\Omega)$ is nonempty as well. Moreover, π is continuous on the compact set Ω, so $\pi(\Omega)$ is compact (Apostol 1974, theorem 4.25, p. 82). We now show that $\pi(\Omega) = \mathcal{C}(\Omega) \cap \Delta$. For the inclusion $\pi(\Omega) \subset \mathcal{C}(\Omega) \cap \Delta$, take any $\boldsymbol{\lambda} \in \pi(\Omega)$. Then there exists $\boldsymbol{w} \in \Omega$ such that $\pi(\boldsymbol{w}) = \boldsymbol{\lambda}$, i.e., $\boldsymbol{\lambda} = \boldsymbol{w}/\|\boldsymbol{w}\|_1$. Because $\Omega \subset \mathbb{R}_+^n \setminus \{\mathbf{0}\}$, it follows that for any $\alpha > 0$, we have $\alpha\boldsymbol{w} \in \mathcal{C}(\Omega)$, and in particular, $\boldsymbol{\lambda} = \alpha\boldsymbol{w}$ for $\alpha = 1/\|\boldsymbol{w}\|_1$. Hence $\boldsymbol{\lambda} \in \mathcal{C}(\Omega) \cap \Delta$. Conversely, suppose $\hat{\boldsymbol{\lambda}} \in \mathcal{C}(\Omega) \cap \Delta$. Then there exists $\hat{\alpha} > 0$ such that $\hat{\boldsymbol{w}} = \hat{\alpha}\hat{\boldsymbol{\lambda}} \in \Omega$. Applying the radial projection yields $\pi(\hat{\boldsymbol{w}}) = \hat{\boldsymbol{w}}/\|\hat{\boldsymbol{w}}\|_1 = \hat{\boldsymbol{\lambda}}$, because $\|\hat{\boldsymbol{\lambda}}\|_1 = 1$. Therefore, $\hat{\boldsymbol{\lambda}} \in \pi(\Omega)$, proving $\mathcal{C}(\Omega) \cap \Delta \subset \pi(\Omega)$. Together, we obtain $\pi(\Omega) = \mathcal{C}(\Omega) \cap \Delta$. (b) Because $\partial\Omega \subset \Omega$, we trivially have $\pi(\partial\Omega) \subset \pi(\Omega)$. To prove the reverse inclusion, let $\hat{\boldsymbol{\lambda}} \in \pi(\Omega)$. Then $\hat{\boldsymbol{\lambda}} = \pi(\hat{\boldsymbol{w}})$ for some $\hat{\boldsymbol{w}} = \hat{\alpha}\hat{\boldsymbol{\lambda}} \in \Omega$, with $\hat{\alpha} > 0$. Because Ω is compact and $\mathbf{0} \notin \Omega$, the ray $\alpha\hat{\boldsymbol{\lambda}}$ intersects Ω in a bounded interval of positive length. That is, there exist real numbers $\underline{\alpha}, \overline{\alpha}$, with $0 < \underline{\alpha} \leq \hat{\alpha} \leq \overline{\alpha} < \infty$, such that $\underline{\alpha}\hat{\boldsymbol{\lambda}}, \overline{\alpha}\hat{\boldsymbol{\lambda}} \in \partial\Omega$. Hence, $\hat{\boldsymbol{\lambda}} = \pi(\underline{\alpha}\hat{\boldsymbol{\lambda}}) \in \pi(\partial\Omega)$, so $\pi(\Omega) \subset \pi(\partial\Omega)$, completing the proof that $\pi(\Omega) = \pi(\partial\Omega)$.

ii. Let $\boldsymbol{w} \in \operatorname{int}(\Omega)$. Because $\Omega \subset \mathcal{C}(\Omega)$, it follows that $\boldsymbol{w} \in \operatorname{int}(\mathcal{C}(\Omega))$. Because Δ is a smooth manifold and the intersection of an open set with it is open (in the relative topology), we obtain that $\pi(\operatorname{int}(\Omega)) = \operatorname{int}(\mathcal{C}(\Omega)) \cap \Delta$ is open in Δ,[48] and thus $\pi(\boldsymbol{w}) \in \operatorname{int}(\pi(\Omega))$. This proves that the radial projection maps interior points of Ω to interior points of $\pi(\Omega)$. It follows that if $\Omega' \subset \Omega$ is open in $\mathbb{R}_+^n$, then $\pi(\Omega')$ is open in Δ.

iii. Let $\Omega' \subset \Omega$ be nonempty. Then $\pi(\Omega') \subset \Delta$ is nonempty. Its closure is compact because Ω' is contained in the compact set Ω. Using continuity of π and part (i) (b) we have that

$$\partial\pi(\Omega') = \partial(\operatorname{cl}(\pi(\Omega'))) = \partial\pi(\operatorname{cl}(\Omega')) = \partial\pi(\partial(\operatorname{cl}(\Omega'))) = \partial\pi(\partial\Omega').$$

Hence, the boundary of $\pi(\Omega')$ equals the boundary of the projection of the boundary of Ω'.

iv. Suppose Ω is convex. If $\pi(\Omega)$ were not convex, then there would exist $\boldsymbol{\lambda}', \boldsymbol{\lambda}'' \in \pi(\Omega)$ and $\mu \in (0,1)$ such that

$\boldsymbol{\lambda} = \mu\boldsymbol{\lambda}' + (1-\mu)\boldsymbol{\lambda}'' \notin \pi(\Omega)$. Let $w' \in \pi^{-1}(\{\boldsymbol{\lambda}'\})$, and $w'' \in \pi^{-1}(\{\boldsymbol{\lambda}''\})$. Then for any $\alpha > 0$,

$$\alpha\boldsymbol{\lambda} = \frac{\alpha\mu}{\|w'\|_1} w' + \frac{\alpha(1-\mu)}{\|w''\|_1} w''.$$

In particular, define

$$\alpha = \left(\frac{\mu}{\|w'\|_1} + \frac{1-\mu}{\|w''\|_1}\right)^{-1} > 0, \text{ and } \hat{\mu} = \frac{\alpha\mu}{\|w'\|_1} \in (0,1).$$

Then

$$\alpha\boldsymbol{\lambda} = \hat{\mu}w' + (1-\hat{\mu})w'' \in \Omega,$$

by convexity of Ω. Hence, $\pi(\alpha\boldsymbol{\lambda}) = \boldsymbol{\lambda} \in \pi(\Omega)$, contradicting our assumption. Therefore, $\pi(\Omega)$ must be convex.

v. We now show that the image $\pi(\Omega')$ of a straight line segment $\Omega' \subset \Omega$ is a straight line segment in Δ, with the possibility of a degenerate case when the straight line segment is projected to a single point in Δ. Indeed,

$$\begin{aligned}
&\pi(\hat{\mu}w' + (1-\hat{\mu})w'') \\
&= \frac{\hat{\mu}w' + (1-\hat{\mu})w''}{\|\hat{\mu}w' + (1-\hat{\mu})w''\|_1} = \frac{\hat{\mu}\|w'\|_1\boldsymbol{\lambda}' + (1-\hat{\mu})\|w''\|_1\boldsymbol{\lambda}''}{\|\hat{\mu}w' + (1-\hat{\mu})w''\|_1} \\
&= \frac{\hat{\mu}\|w'\|_1}{\hat{\mu}\|w'\|_1 + (1-\hat{\mu})\|w''\|_1}\boldsymbol{\lambda}' + \left(1 - \frac{\hat{\mu}\|w'\|_1}{\hat{\mu}\|w'\|_1 + (1-\hat{\mu})\|w''\|_1}\right)\boldsymbol{\lambda}'' \\
&= \mu(\hat{\mu})\boldsymbol{\lambda}' + (1-\mu(\hat{\mu}))\boldsymbol{\lambda}'',
\end{aligned}$$

where $\boldsymbol{\lambda}' = w'/\|w'\|_1, \boldsymbol{\lambda}'' = w''/\|w''\|_1$, and

$$\mu(\hat{\mu}) = \frac{\hat{\mu}\|w'\|_1}{\hat{\mu}\|w'\|_1 + (1-\hat{\mu})\|w''\|_1}, \ \hat{\mu} \in [0,1].$$

The latter function is continuously differentiable, with

$$\frac{d\mu(\hat{\mu})}{d\hat{\mu}} = \frac{\|w'\|_1\|w''\|_1}{(\hat{\mu}\|w'\|_1 + (1-\hat{\mu})\|w''\|_1)^2} > 0, \ \hat{\mu} \in [0,1],$$

which implies that $\mu(\cdot)$ is strictly increasing on [0, 1]. Thus, provided that $\boldsymbol{\lambda}' \neq \boldsymbol{\lambda}''$, the radial projection of a straight line is a bijection, which also preserves the orientation of any straight path between two different points in Ω, as long as they do not map to the same point in Δ (for $\boldsymbol{\lambda}' = \boldsymbol{\lambda}''$).

vi. Consider first the case where $\text{co}(\Omega)$ is a bounded polyhedron (i.e., a polytope; see, e.g., Nemhauser and Wolsey 1999, definition 2.2, p. 86). By part (i)(b) it is $\pi(\text{co}(\Omega)) = \pi(\partial(\text{co}(\Omega)))$. By part (iv), the set $\pi(\text{co}(\Omega))$ is convex. Part (v) then guarantees that the straight lines between any two vertices of $\text{co}(\Omega)$ remain straight lines in their radial projection onto Δ, which implies it is enough to project the vertices of $\text{co}(\Omega)$ and then take the convex hull. Because the vertices are included in $\partial\Omega$, we therefore obtain that $\pi(\text{co}(\Omega)) = \text{co}(\pi(\partial\Omega))$ as claimed. If the bounded convex set $\text{co}(\Omega)$ is not a polytope, then it can be approximated arbitrarily closely by a polytope (see, e.g., Bronstein 2008), so that the claim—omitting some of the convergence-specific details—follows in the limit.

vii. Let $\Omega', \Omega'' \subset \mathbb{R}^n_+$ such that $\Omega = \Omega' \cup \Omega''$. Then $\mathcal{C}(\Omega' \cup \Omega'') = \mathcal{C}(\Omega') \cup \mathcal{C}(\Omega'')$, which implies the claim by part (i)(a) (cf. Endnote 48): $\pi(\Omega) = (\mathcal{C}(\Omega') \cap \Delta) \cup (\mathcal{C}(\Omega'') \cap \Delta) = \pi(\Omega') \cup \pi(\Omega'')$.

viii. Let $\Omega', \Omega'' \subset \mathbb{R}^n_+$ such that $\Omega = \Omega' \cap \Omega''$. Then, by part (i)(a) we have $\pi(\Omega) = \mathcal{C}(\Omega' \cap \Omega'') \cap \Delta$. If $\Omega = \Omega' \cap \Omega''$ for some $\Omega', \Omega'' \subset \mathbb{R}^n_+$, then $\pi(\Omega) = \mathcal{C}(\Omega' \cap \Omega'') \cap \Delta$, as claimed. □

Proof of Proposition 5. Recall that the ambiguity set $\Omega \subset \mathbb{R}^n_+$ is compact, nonempty, and not reduced to the origin. For any feasible decision $\boldsymbol{x} \in \mathcal{X}$, the Ω-conditioned performance index corresponds to its worst-case relative performance. Using the radial projection of Ω and invoking Lemma 8 (i), we obtain:[49]

$$\rho(\boldsymbol{x}|\Omega) = \min_{w \in \Omega} \frac{F(\boldsymbol{x}|w)}{F^*(w)} = \min_{\boldsymbol{\lambda} \in \pi(\Omega)} \frac{F(\boldsymbol{x}|\boldsymbol{\lambda})}{F^*(\boldsymbol{\lambda})} = \min_{\boldsymbol{\lambda} \in \pi(\partial\Omega)} \frac{F(\boldsymbol{x}|\boldsymbol{\lambda})}{F^*(\boldsymbol{\lambda})}.$$

As shown in the proof of Proposition 2, the performance ratio $\varphi(\boldsymbol{x}|\boldsymbol{\lambda}) = F(\boldsymbol{x}|\boldsymbol{\lambda})/F^*(\boldsymbol{\lambda})$ is quasiconcave in $\boldsymbol{\lambda}$ for fixed $\boldsymbol{x} \in \mathcal{X}$. Because $\pi(\Omega)$ is compact, the minimum is attained on the boundary $\partial\pi(\Omega)$, which equals $\pi(\partial\Omega)$ by Lemma 8 (i)(b). Moreover, the quasiconcavity of $\varphi(\boldsymbol{x}|\cdot)$ implies that all upper contour sets are convex, so the minimum is also attained on the boundary of the convex hull $\text{co}(\pi(\partial\Omega))$. Thus, if there exists an extremal base Λ such that $\text{co}(\Lambda) = \text{co}(\pi(\Omega))$, then the minimum is attained on the compact set Λ, as claimed. □

Proof of Proposition 6. Fix any $\boldsymbol{\theta} \in \Theta$, and consider the state-contingent Ω'-conditioned performance index,

$$\phi(\boldsymbol{x}, \boldsymbol{\theta}|\Omega') = \min_{\boldsymbol{\lambda}' \in \pi(\Omega')} \frac{G(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}')}{G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}')}, \ \boldsymbol{x} \in \mathcal{X}, \tag{A.6}$$

which measures the performance of an action relative to all feasible weighted-sum scalarizations of the multicriteria decision problem, in state $\boldsymbol{\theta}$. Because Λ' is an extremal base of $\pi(\Omega')$, we have $\text{co}(\Lambda') = \text{co}(\pi(\Omega'))$. The quasiconcavity of $G(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}')/G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}')$ in $\boldsymbol{\lambda}'$ for fixed $(\boldsymbol{x}, \boldsymbol{\theta})$, which obtains as in the proof of Proposition 2, ensures that the minimum in Equation (A.6) is attained at an extreme point of $\text{co}(\pi(\Omega'))$. Thus, as in Proposition 5, we obtain an equivalent representation,

$$\phi(\boldsymbol{x}, \boldsymbol{\theta}|\Omega') = \min_{\boldsymbol{\lambda}' \in \Lambda'} \frac{G(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}')}{G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}')}, \ \boldsymbol{x} \in \mathcal{X}. \tag{A.7}$$

By Equation (27), the performance index is the minimum of the state-contingent Ω'-conditioned performance index in Equation (A.7) over all $\boldsymbol{\theta} \in \Theta$, that is,

$$\rho(\boldsymbol{x}|\Omega', \Theta) = \min_{\boldsymbol{\theta} \in \Theta} \phi(\boldsymbol{x}, \boldsymbol{\theta}) = \min_{\boldsymbol{\theta} \in \Theta} \left\{ \min_{\boldsymbol{\lambda}' \in \Lambda'} \frac{G(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}')}{G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}')} \right\}, \ \boldsymbol{x} \in \mathcal{X}. \tag{A.8}$$

Thus, reversing the order of minimization in Equation (A.8) yields Equation (27), as claimed. □

Proof of Corollary 1. Let $\Lambda' = \{\boldsymbol{\lambda}'^{(1)}, \ldots, \boldsymbol{\lambda}'^{(n)}\}$ be the (smallest) extremal base of $\pi(\Omega')$ (or, equivalently, of $\text{co}(\pi(\Omega))$), which is finite by hypothesis, with $n \geq 1$. Then Equation (27) becomes

$$\begin{aligned}
\rho(\boldsymbol{x}|\Omega', \Theta) &= \min_{\boldsymbol{\lambda}' \in \Lambda'} \left\{ \min_{\boldsymbol{\theta} \in \Theta} \frac{G(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}')}{G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}')} \right\} = \min_{i \in \mathcal{N}} \left\{ \min_{\boldsymbol{\theta} \in \Theta} \frac{G(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}'^{(i)})}{G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}'^{(i)})} \right\} \\
&= \min_{i \in \mathcal{N}} \{f_i(\boldsymbol{x})\}, \ \boldsymbol{x} \in \mathcal{X},
\end{aligned} \tag{A.9}$$

where we have set $f_i(\boldsymbol{x}) = \min_{\boldsymbol{\theta}\in\Theta}\{G(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{\lambda}'^{(i)})/G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}'^{(i)})\}$, for all $(\boldsymbol{x},i)\in\mathcal{X}\times\mathcal{N}$. Because by definition $\max_{x\in\mathcal{X}} G(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{\lambda}'^{(i)}) = G^*(\boldsymbol{\theta}|\boldsymbol{\lambda}'^{(i)})$, it is $f_i^* = \max_{x\in\mathcal{X}} f_i(\boldsymbol{x}) = 1$, for all $i\in\mathcal{N}$, so that Equation (A.9) is in fact equivalent to Equation (28), which completes the proof. □

Proof of Lemma 9. Fix $i,j,m\in\mathcal{N}$ (with $m\neq n$), $\boldsymbol{\lambda}\in\Delta$, and $\delta>0$. Then for any $\boldsymbol{y}\in\mathcal{Y}$: If $\hat{\boldsymbol{y}} = \boldsymbol{y}+\delta\boldsymbol{e}_i\in\mathcal{Y}$, then

$$\begin{aligned} H(\boldsymbol{y}|\boldsymbol{\lambda}) &= h^{-1}\left(\sum_{l=1}^{n}\lambda_l h(y_l)\right)\\ &= h^{-1}\left(\sum_{l=1}^{n}\lambda_l h(\hat{y}_l) - \lambda_i(h(\hat{y}_i)-h(y_i))\right)\\ &\begin{cases} < H(\hat{\boldsymbol{y}}|\boldsymbol{\lambda}), & \text{if } \lambda_i>0,\\ = H(\hat{\boldsymbol{y}}|\boldsymbol{\lambda}), & \text{if } \lambda_i=0. \end{cases}\end{aligned}$$

This implies Axiom 1.[50] Consider now the case where $(\hat{\boldsymbol{y}},\hat{\boldsymbol{\lambda}}) = (\boldsymbol{y},\boldsymbol{\lambda}) + (y_i-y_j,\lambda_i-\lambda_j)(\boldsymbol{e}_j-\boldsymbol{e}_i)$. Then for $i=j$, Axiom 2 holds trivially. For $i\neq j$, it is

$$H(\hat{\boldsymbol{y}}|\hat{\boldsymbol{\lambda}}) = h^{-1}\left(\sum_{l\in\mathcal{N}\setminus\{i,j\}}\lambda_l h(y_l)+\lambda_i h(y_i)+\lambda_j h(y_j)\right) = H(\boldsymbol{y}|\boldsymbol{\lambda}),$$

which also yields Axiom 2. Because $\boldsymbol{\lambda}$ is by assumption normalized, we also obtain that Axiom 3 holds, as

$$H(\alpha\mathbf{1}_n) = h^{-1}\left(\sum_{l=1}^{n}\lambda_l h(\alpha)\right) = h^{-1}(h(\alpha)) = \alpha,$$

for any $\alpha>0$. As far as associativity is concerned, let us assume that $(\lambda_1,\dots,\lambda_m)\neq\mathbf{0}$, and let

$$\alpha = H\left(\sum_{l=1}^{m} y_l\boldsymbol{e}_l\middle|\frac{\sum_{l=1}^{m}\lambda_l\boldsymbol{e}_l}{\sum_{l=1}^{m}\lambda_l}\right) = h^{-1}\left(\frac{\sum_{l=1}^{m}\lambda_l h(y_l)}{\sum_{l=1}^{m}\lambda_l}\right).$$

With this, one obtains

$$\begin{aligned} &H(\alpha\mathbf{1}_m, y_{m+1},\dots,y_n|\boldsymbol{\lambda})\\ &= h^{-1}\left(h\circ h^{-1}\left(\frac{\sum_{l=1}^{m}\lambda_l h(y_l)}{\sum_{l=1}^{m}\lambda_l}\right)\cdot\left(\sum_{l=1}^{m}\lambda_l\right)+\sum_{l=m+1}^{n}\lambda_l h(y_l)\right)\\ &= H(\boldsymbol{y}|\boldsymbol{\lambda}),\end{aligned}$$

whence Axiom 4 must hold (taking into account that $h\circ h^{-1}(\cdot)=(\cdot)$). Finally, we observe that

$$H(\boldsymbol{y}|\boldsymbol{e}_i) = h^{-1}\circ h(y_i) = y_i,$$

so the coordinate-filter requirement, which constitutes Axiom 5, is also satisfied. This concludes the proof. □

Endnotes

[1] Practical examples for multicriteria decision making are legion; see, for example, the collections of case studies by Berbel et al. (2018) in agriculture, Masri et al. (2018) in financial decision making, and Ravindran (2016) for supply chain management, to name just a few.

[2] Extant ESG metrics differ widely. In their approach to "quantifying the impact of impact investing," Lo and Zhang (2024) remain agnostic about the impact factors to be used, taking them as given and thus keeping at bay the difficulties of attributing weights to different ESG criteria and of dealing with this model uncertainty.

[3] Lancaster (1966) already noted that a product can be viewed as a bundle of its attributes, with consumer valuations often empirically assessed using conjoint analysis (see, e.g., Green and Srinivasan 1990).

[4] Example 9 in Section 3.6 discusses robust allocation in a two-agent exchange economy using the proposed framework.

[5] This is notwithstanding the "norm equivalence" in a finite-dimensional Euclidean space, in the sense that any norm $\|\cdot\|$ can be bracketed by any other norm $\hat{\|}\cdot\hat{\|}$, so $\chi_1\hat{\|}\cdot\hat{\|}\le\|\cdot\|\le\chi_2\hat{\|}\cdot\hat{\|}$ for suitable scalars $\chi_1,\chi_2>0$.

[6] For the computation of Pareto sets, see Kung et al. (1975), Gabow et al. (1984), and Bentley et al. (1993).

[7] See Yamamoto (2002) for details about maximizing a function on a Pareto-efficient set in a polyhedral setting.

[8] A lexicographic evaluation of criteria based on perceived importance may justify an "elimination by aspects" (Tversky 1972), or more nuanced partial elimination heuristics using "attribute filters" (Kimya 2018).

[9] Numerous ad hoc methods for determining weights exist, for example, in reliability engineering based on standards (Jiang and Chen 2020).

[10] In his "Discussion on Making Things Equal," Zhuang Zhou points out that "[t]here is nothing in the world bigger than the tip of an autumn hair, and Mount T'ai is tiny" (Watson 1968, p. 44), where Mount Tai is the highest point in the Shandong province of China.

[11] For example, getting \$100 in a month, in addition to repayment of the invested principal, would be attractive if obtained by investing a principal of \$1 today, but not if it required investing a principal of \$10 million today.

[12] A point $\boldsymbol{x}\in\mathcal{X}$ is said to be *isolated* if there exists an open set $\mathcal{O}\subset\mathbb{R}^m$ such that $\mathcal{O}\cap\mathcal{X}=\{\boldsymbol{x}\}$.

[13] By the extreme-value theorem (Rudin 1976, theorem 4.16, p. 89), the minimum of a continuous function on a compact set exists and is finite.

[14] It is $\mathbb{R}^n_+\setminus\{\mathbf{0}\} = \cup_{\alpha>0}\alpha\Delta$.

[15] More precisely, as shown in the proof of Lemma 1, we have $F^*(\hat{\boldsymbol{w}})\ge F^*(\boldsymbol{w})+\delta(\max f_i(\mathcal{X}(\boldsymbol{w})))$.

[16] This "complete ignorance" (or *full ambiguity*) is relaxed in Section 3.7 where we allow the decision maker to face *limited* ambiguity.

[17] Because $F^*(\boldsymbol{\lambda})\ge\min F^*(\Delta)>0$, the function $\varphi:\mathcal{X}\times\Delta\to[0,1]$ is well defined.

[18] Leontief (1941) employed this aggregation in fixed proportions as a simplification for his analysis of a larger economy.

[19] By the extreme-value theorem, Ψ is nonempty, and by the maximum theorem, it is compact.

[20] Given two vectors $\boldsymbol{z},\hat{\boldsymbol{z}}\in\mathbb{R}^n$, where $\boldsymbol{z}=(z_1,\dots,z_n)$ and $\hat{\boldsymbol{z}}=(\hat{z}_1,\dots,\hat{z}_n)$, the standard vector inequalities are defined as follows: (i) $\boldsymbol{z}\le\hat{\boldsymbol{z}}\iff z_i\le\hat{z}_i,\ \forall i\in\mathcal{N}$; (ii) $\boldsymbol{z}\ll\hat{\boldsymbol{z}}\iff z_i<\hat{z}_i,\ \forall i\in\mathcal{N}$; (iii) $\boldsymbol{z}<\hat{\boldsymbol{z}}\iff(\boldsymbol{z}\le\hat{\boldsymbol{z}}\ \textit{and}\ \exists\, j\in\mathcal{N}\ \textit{such that}\ z_j<\hat{z}_j)$. Here, $\mathcal{N}=\{1,\dots,n\}$, with $n\ge2$.

[21] The relevant lower (set) limit is given by $\underline{\mathrm{Lim}}_{\varepsilon\to0^+}\mathscr{R}_\varepsilon = \{x\in\mathcal{X}: \lim_{\varepsilon\to0^+}(\inf_{x'\in\mathscr{R}_\varepsilon}\|x-x'\|)=0\}$ (see, e.g., Aubin and Frankowska 1990, definition 1.4.6, p. 41). Given a sequence $(\varepsilon_k)_{k=1}^{\infty}\subset(0,1]$ with $\lim_{k\to\infty}\varepsilon_k=0$, this lower limit contains the accumulation points of any sequence $(x_k)_{k=1}^{\infty}$ with elements $x_k\in\mathscr{R}_{\varepsilon_k}$, for all $k\ge1$.

[22] Note that: $\Phi_\varepsilon(2,\frac{3}{2})-\max_{x_1\in[0,1]}\left\{\frac{(1-\varepsilon)x_1}{2}+\frac{\varepsilon}{2}\left(\frac{x_1}{2}+\frac{2+\sqrt{1-x_1}}{3}\right)\right\} = \frac{\varepsilon}{3}\frac{1-(7/12)\varepsilon}{2-\varepsilon}>0$, for all $\varepsilon\in(0,1]$.

[23] The larger (path-connected, compact) action set $\mathcal{X}'=\mathcal{X}\cup\{(2,\frac{3}{2})\zeta:\zeta\in[0,1]\}$ leaves results unchanged.

[24] It is $\frac{1}{24}(16-3\varepsilon^2)/(2-\varepsilon)-\Phi_\varepsilon(3,\frac{1}{2}) = \frac{1}{24}(8-16\varepsilon+7\varepsilon^2)/(2-\varepsilon)>0$ if and only if $\varepsilon<\frac{4}{7}\left(2-\frac{1}{\sqrt{2}}\right)$.

[25] Normalizing the maximum criteria to the same score (e.g., $c \in \{1, 10, 100\}$) is natural in many applications.

[26] The term "robustness equivalence principle" is analogous to the well-known "certainty equivalence principle," for example, in linear-quadratic optimal control problems (Bertsekas 1995, p. 23), where it is optimal (i.e., maximizing the expected value of a quadratic objective functional) to replace uncertain parameters by their means.

[27] For $\alpha = \beta$, Equations (22) and (23) imply the unique robust allocation $\hat{\boldsymbol{x}} = (1/2, 1/2)$, achieving $\rho^* = 1/2$.

[28] The converse of part (iv) does not hold. That is, if Ω is nonconvex, then $\pi(\Omega)$ may still be convex.

[29] The analysis in Example 10 yields an extreme counterexample for $\Omega = \{\boldsymbol{w}\} = \Omega' \cap \Omega''$, with $\Omega' = [\underline{\boldsymbol{w}}, \boldsymbol{w}]$, $\Omega'' = [\boldsymbol{w}, \overline{\boldsymbol{w}}]$, and $\underline{\boldsymbol{w}} \ll \boldsymbol{w} \ll \overline{\boldsymbol{w}}$. Indeed, for $\|\underline{\boldsymbol{w}}\|_1 \to 0^+$ one obtains $\text{int}(\Delta) \subset \pi(\Omega')$, so $\pi(\Omega') \cap \pi(\Omega'') = \pi(\Omega'')$, whereas $\pi(\Omega) = \pi(\{\boldsymbol{w}\})$ is merely a singleton.

[30] The underlying justification is Minkowski's theorem, which states that any compact convex set is equal to the convex hull of its extreme points; see, for example, Rockafellar (1970, corollary 18.5.1, p. 167) or Schneider (2014, corollary 1.4.5, p. 17).

[31] In general, the smallest Λ may not be finite (e.g., if $\Omega \subset \mathbb{R}^3_+$ is a unit ball, then the smallest $\Lambda = \partial\pi(\Omega)$ has a continuum of elements).

[32] For analogous comments about the assumed continuity of the criteria in the action, see Remark 1 (i) in Section 2.

[33] For example, in the case where $\Theta = \{(1,0),(0,1)\}$ for $l = n' = 2$, observing $\theta_1 = 1$ implies $\theta_2 = 0$ (and vice versa).

[34] For any $m \in \mathcal{N}$, we define $\mathbf{1}_m = (1, \ldots, 1) \in \mathbb{R}^m_+$.

[35] It is $(\sum_{i=1}^m \lambda_i \boldsymbol{e}_i)/(\sum_{i=1}^m \lambda_i) = \pi(\sum_{i=1}^m \lambda_i \boldsymbol{e}_i)$, with the radial projection $\pi(\cdot)$, from $\mathbb{R}^n_+ \setminus \{\mathbf{0}\}$ onto Δ; see Section 3.8.

[36] Kolmogorov (1930) showed that if Axioms 1–4 are satisfied for $\boldsymbol{\lambda} = \mathbf{1}_n/n$ (cf. Endnote 34), then there exists a (continuous) strictly monotonic function $h : \mathcal{D} \to \mathbb{R}$, defined on a suitable domain $\mathcal{D} \subset \mathbb{R}$, such that

$$H(\boldsymbol{y}|\mathbf{1}_n/n) = h^{-1}(\textstyle\sum_{i=1}^n (1/n) h(y_i)), \quad \boldsymbol{y} \in \mathcal{Y}.$$

Nagumo (1930) obtained a similar result, and the preceding "quasiarithmetic mean" is therefore also known as the "Kolmogorov-Nagumo mean."

[37] Homogeneity is not required by Axioms 1–5, because relatively robust decisions (cf. Proposition 2) are invariant under rescaling.

[38] The gradient of LSE is the softmax function (Boltzmann 1871; Gibbs 1902, chapter XIV), widely used in machine learning (Bridle 1990).

[39] The purpose of restricting all criteria to strictly positive values was merely to skip checking the nontriviality condition (N) in Section 2.1 which is satisfied here (e.g., $\boldsymbol{x}_d = 2$). Hence, setting $s_{13} = 0$ poses no problem.

[40] Under some mild additional assumptions (which we omit), the law of large numbers would guarantee convergence of the average scores to the true criterion-function values, at least when the number of available score observations for each realized action (i.e., each element of $\hat{\mathcal{X}}$) goes to infinity, assuming that decisions can be properly recognized, as may naturally be the case on a sufficiently discrete underlying action set $\mathcal{X}$.

[41] Nathan and Lord (1983) propose "knowledge," "delivery," "relevance," "interpersonal," and "organization" as criteria for the quality of a university lecturer; see also DeNisi and Murphy (2017) for a survey.

[42] This is not the only "energy trilemma"; Tilman et al. (2009) discuss a food-energy-environment trilemma in the context of biofuels.

[43] Klarman et al. (1968) recognized the need for quality-of-life considerations in cost-effective treatment options (for chronic renal disease).

[44] Cutler and Richardson (1997), Buxton (2005), and Newhouse (2021) discuss QALY for assessing cost-effectiveness in healthcare.

[45] Baker and Freeland (1975) provide an overview of early scoring models. Stewart (1991) proposes a multicriteria decision support system for project selection using essentially equal weights for different objectives (akin to the Laplacian approach; cf. Section 4.1). Finally, Hall et al. (2015) suggest a project evaluation with an "underperformance riskiness index" relying on (typically unavailable) past statistical data.

[46] Indeed, because $\varphi(\boldsymbol{x}|\boldsymbol{\lambda}) \geq \min_{i \in \mathcal{N}}\{f_i(\boldsymbol{x})/\underline{F}(\boldsymbol{x}_d)\} = \underline{c}(\boldsymbol{x}) \in [0,1]$, for all $\boldsymbol{\lambda} \in \Delta$, Equation (A.2) implies that $c \leq \underline{c}(\boldsymbol{x}) \Rightarrow \mathcal{U}_c(\boldsymbol{x}) = \Delta$.

[47] Efficiency in the outcome set corresponds directly to efficiency in the action set (cf. Section 3.2), because the latter is evaluated via a criterion vector that produces points in the outcome set.

[48] The representation $\pi(\Omega) = \mathcal{C}(\Omega) \cap \Delta$ was given in part (i) for a compact set Ω, but it also applies to any subset of Ω.

[49] By part (a) of Lemma 8 (i), $\pi(\Omega)$ is nonempty and compact, so the minimum with respect to $\boldsymbol{\lambda}$ must be attained by the extreme-value theorem.

[50] When h is increasing, the argument of h^{-1} cannot become smaller by dropping the term $\lambda_i(h(\hat{y}_i) - h(y_i))$. Conversely, when h is decreasing, the argument of h^{-1} weakly decreases, so the value of that inverse cannot go down.

References

Apostol TM (1974) *Mathematical Analysis*, 2nd ed. (Addison-Wesley, Reading, MA).

Arrow KJ, Enthoven AC (1961) Quasi-concave programming. *Econometrica* 29(4):779–800.

Aubin J-P, Frankowska H (1990) *Set-Valued Analysis* (Birkhäuser, Boston).

Baker N, Freeland J (1975) Recent advances in R&D benefit measurement and project selection methods. *Management Sci.* 21(10): 1164–1175.

Ben-David S, Borodin A (1994) A new measure for the study of on-line algorithms. *Algorithmica* 11(1):73–91.

Ben-Tal A, den Hertog D, Waegenaere AD, Melenberg B, Rennen G (2013) Robust solutions of optimization problems affected by uncertain probabilities. *Management Sci.* 59(2):341–357.

Bentley JL, Clarkson KL, Levine DB (1993) Fast linear expected-time algorithms for computing maxima and convex hulls. *Algorithmica* 9(2):168–183.

Berbel J, Bournaris T, Manos B, Matsatsinis N, Viaggi D, eds. (2018) *Multicriteria Analysis in Agriculture: Current Trends and Recent Applications* (Springer, New York).

Berge C (1963) *Topological Spaces* (Oliver and Boyd, Edinburgh, UK).

Bertsekas DP (1995) *Dynamic Programming and Optimal Control*, vol. 1 (Athena Scientific, Belmont, MA).

Blanchet J, Murthy K (2019) Quantifying distributional model risk via optimal transport. *Math. Oper. Res.* 44(2):565–600.

Boltzmann L (1871) Über das Wärmegleichgewicht zwischen mehratomigen Gasmolekülen. *Sitzungsberichte der Mathematisch-Naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenschaften, Wien* 63(3):397–418.

Bridle JS (1990) Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. Fogelman Soulié F, Hérault J, eds. *Neurocomputing: Algorithms, Architectures and Applications*, NATO ASI Series, vol. 68 (Springer, New York), 227–236.

Bronstein EM (2008) Approximation of convex sets by polytopes. *J. Math. Sci.* 153(6):727–762. [Translation of Russian original, which appeared in Sovremennaya Matematika / Fundamental'nye Napravleniya (Contemporary Mathematics / Fundamental Directions), Vol. 22, Geometry, 2007.]

Buxton MJ (2005) How much are health-care systems prepared to pay to produce QALY? *Eur. J. Health Econom.* 6(4):285–287.

Charnes A, Cooper WW (1961) *Management Models and Industrial Applications of Linear Programming*, vol. 1 (Wiley, New York).

Charnes A, Cooper WW, Ferguson RO (1955) Optimal estimation of executive compensation by linear programming. *Management Sci.* 1(2):138–151.

Cournot A (1838) *Recherches sur les Principes Mathématiques de la Théorie des Richesses* (Hachette, Paris).

Cutler DM, Richardson E (1997) Measuring the health of the U.S. population. *Brookings Papers Econom. Activity Microeconom.* 28(1997): 217–282.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):595–612.

DeNisi AS, Murphy KR (2017) Performance appraisal and performance management: 100 years of progress? *J. Appl. Psych.* 102(3):421–433.

Edgeworth FY (1881) *Mathematical Psychics* (C. Kegan Paul, London).

Edgeworth FY (1897) La teoria pura del monopolio. *Giornale degli Economisti* 15(Anno 8):13–31.

Ehrgott M (2010) *Multicriteria Optimization*, 2nd ed. (Springer, Berlin).

Fechner GT (1860) *Elemente der Psychophysik (Part II)* (Breitkopf & Härtel, Leipzig, Germany).

Gabow HN, Bentley JL, Tarjan RE (1984) Scaling and related techniques for geometry problems. *Proc. 16th Annual ACM Sympos. Theory Comput.* (Association for Computing Machinery, New York), 135–143.

Gibbs JW (1902) *Elementary Principles in Statistical Mechanics* (General Publishing, Toronto).

Goel A, Meyerson A, Weber TA (2009) Fair welfare maximization. *Econom. Theory* 41(3):465–494.

Green PE, Srinivasan V (1990) Conjoint analysis in marketing: New developments with implications for research and practice. *J. Marketing* 54(4):3–19.

Hall NG, Long DZ, Qi J, Sim M (2015) Managing underperformance risk in project portfolio selection. *Oper. Res.* 63(3):660–675.

Han J, Weber TA (2023) Price discrimination with robust beliefs. *Eur. J. Oper. Res.* 306(2):795–809.

Hardy GH, Littlewood JE, Pólya G (1934) *Inequalities* (Cambridge University Press, Cambridge, UK).

Harsanyi JC (1955) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *J. Political Econom.* 63(4):309–321.

Hawkins TR, Gausen OM, Strømman AH (2012) Environmental impacts of hybrid and electric vehicles: A review. *Internat. J. Life Cycle Assessment* 17(8):997–1014.

Jiang R, Chen Z (2020) A standard-based approach for multi-criteria performance evaluation of engineered systems. *Reliability Engrg. System Safety* 202(107001):1–12

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291.

Kahneman D, Tversky A (1984) Choices, values, and frames. *Amer. Psychologist* 39(4):341–350.

Keeney RL, Raiffa H (1993) *Decisions with Multiple Objectives* (Cambridge University Press, Cambridge, UK).

Kimya M (2018) Choice, consideration sets, and attribute filters. *Amer. Econom. J. Microeconom.* 10(4):223–247.

Klarman HE, Francis JOS, Rosenthal GD (1968) Cost effectiveness analysis applied to the treatment of chronic renal disease. *Medical Care* 6(1):48–54.

Kolmogorov AN (1930) Sur la notion de la moyenne. *Atti della Accademia Nazionale dei Lincei, Classe di Scienze Fisiche, Matematiche e Naturali Rendiconti* 12(9):388–391.

Kouvelis P, Yu G (1997) *Robust Discrete Optimization and Its Applications* (Springer, New York).

Kung H, Luccio F, Preparata F (1975) On finding the maxima of a set of vectors. *J. ACM* 22(4):469–476.

Lancaster KJ (1966) A new approach to consumer theory. *J. Political Econom.* 74(2):132–157.

Leontief WW (1941) *The Structure of American Economy 1919–1929: An Empirical Application of Equilibrium Analysis* (Harvard University Press, Cambridge, MA).

Lo AW, Zhang R (2024) Quantifying the impact of impact investing. *Management Sci.* 70(10):7161–7186.

Loewenstein GF, Thompson L, Bazerman MH (1989) Social utility and decision making in interpersonal contexts. *J. Personality Soc. Psych.* 57(3):426–441.

Marshall A (1890) *Principles of Economics* (Macmillan, London).

Mas-Colell A, Whinston MD, Green JR (1995) *Microeconomic Theory* (Oxford University Press, Oxford, UK).

Masri H, Pérez-Gladish B, Zopounidis C, eds. (2018) *Financial Decision Aid Using Multiple Criteria: Recent Models and Applications* (Springer, New York).

Miyamoto JM, Wakker PP, Bleichrodt H, Peters HJM (1998) The zero-condition: A simplifying assumption in QALY measurement and multiattribute utility. *Management Sci.* 44(6):839–849.

Mohajerin Esfahani P, Kuhn D (2018) Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Programming* 171(1–2):115–166.

Nagumo M (1930) Über eine Klasse der Mittelwerte. *Japanese J. Math.* 7:71–79.

Nathan BR, Lord RG (1983) Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. *J. Appl. Psychol.* 68(1):102–114.

Nemhauser GL, Wolsey LA (1999) *Integer and Combinatorial Optimization* (Wiley Interscience, New York).

Newhouse JP (2021) An ounce of prevention. *J. Econom. Perspect.* 35(2):101–118.

Pareto V (1894) Il massimo di utilità dato dalla libera concorrenza. *Giornale degli Economisti* 9(2):48–66.

Pareto V (1897) *Cours d'Économie Politique*, vol. 2 (F. Rouge, Lausanne, Switzerland).

Pareto V (1906) *Manuale di Economia Politica* (Società Editrice Libraria, Milan).

Pliskin JS (1974) The management of patients with end-stage renal failure: A decision-theoretic approach. Doctoral thesis, Harvard University, Boston.

Pliskin JS, Beck CH (1976) A health index for patient selection: A value function approach with application to chronic renal failure patients. *Management Sci.* 22(9):1009–1021.

Ravindran AR, ed. (2016) *Multiple Criteria Decision Making in Supply Chain Management* (CRC Press, New York).

Rockafellar RT (1970) *Convex Analysis* (Princeton University Press, Princeton, NJ).

Rudin W (1976) *Principles of Mathematical Analysis*, 3rd ed. (McGraw-Hill, New York).

Savage LJ (1951) The theory of statistical decision. *J. Amer. Statist. Assoc.* 46(253):55–67.

Schneider R (2014) *Convex Bodies: The Brunn-Minkowski Theory*, 2nd ed. (Cambridge University Press, Cambridge, UK).

Sleator DD, Tarjan RE (1985) Amortized efficiency of list update and paging rules. *Comm. ACM* 28(2):202–208.

Steuer RE (1986) *Multiple Criteria Optimization: Theory, Computation, and Application* (Wiley, New York).

Stewart TJ (1991) A multi-criteria decision support system for R&D project selection. *J. Oper. Res. Soc.* 42(1):17–26.

Tilman D, Socolow R, Foley JA, Hill J, Larson E, Lynd L, Pacala S, et al. (2009) Beneficial biofuels: The food, energy, and environment trilemma. *Science* 325(5938):270–271.

Tversky A (1972) Elimination by aspects: A theory of choice. *Psych. Rev.* 79(4):281–299.

Villani C (2008) *Optimal Transport: Old and New*, Grundlehren der Mathematischen Wissenschaften, vol. 338 (Springer, Berlin).

Wald A (1945) Statistical decision functions which minimize the maximum risk. *Ann. Math.* 46(2):265–280.

Wang T, Fu Y (2020) Constructing composite indicators with individual judgements and best-worst method. *Soc. Indicators Res.* 149(1):1–14.

Watson B (1968) *The Complete Works of Chuang Tzu* (Columbia University Press, New York).

Weber EH (1846) Der Tastsinn und das Gemeingefühl. Wagner R, ed. *Handwörterbuch der Physiologie mit Rücksicht auf die physiologische Pathologie*, vol. 3, section 2 (Vieweg, Braunschweig, Germany), 481–588.

Weber TA (2014) On the (non-)equivalence of IRR and NPV. *J. Math. Econom.* 52:25–39.

Weber TA (2023) Relatively robust decisions. *Theory Decision* 94(1): 35–62.

Weber TA (2024) Optimal depth of discharge for electric batteries with robust capacity-shrinkage estimator. *Proc. 2024 4th Internat. Conf. Smart Grid Renewable Energy (SGRE)* (IEEE, New York), 1–5.

Weber TA (2025) Monopoly pricing with unknown demand. *Scandinavian J. Econom.* 127(1):235–285.

World Energy Council (2024) World Energy Trilemma 2024: Evolving with resilience and justice. Technical report, World Energy Council, London.

Yager RR (1988) On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Trans. Systems Man Cybernetics* 18(1):183–190.

Yamamoto Y (2002) Optimization over the efficient set: Overview. *J. Global Optim.* 22(1–4):285–317.

Zarghami M, Szidarovszky F (2010) On the relation between compromise programming and ordered weighted averaging operator. *Inform. Sci.* 180(11):2239–2248.

Zeleny M (1974) A concept of compromise solutions and the method of the displaced ideal. *Comput. Oper. Res.* 1(3):479–496.