Unique Preference Aggregation in Design and Decision Making

Addendum to Chapter 5 of

Wolfert, A.R.M. (2023). ***Open Design Systems (Odesys).*** **IOS Press.**

ISBN 978-1-64368-416-1 (print)
ISBN 978-1-64368-417-8 (online)
DOI 10.3233/RIDS10
Library of Congress Control Number: 2023942638
Compiled January 26, 2026

Preference aggregation is a core operation in multi-objective design optimisation and group decision-making, as it determines the **best-fit-for-common-purpose** alternative within complex socio-technical contexts. Since preferences are intrinsically linked to choice, they are subjective, inherently contextual, and reflect humans' free will to relatively order their alternatives. Therefore, their aggregation requires a rigorous measurement-theoretic foundation to ensure mathematical validity, interpretability, and uniqueness. PFM establishes the principal axioms of **unique preference aggregation**, providing a rigorous basis on which aggregation can be demonstrated.
In this Addendum it is shown that commonly used aggregation approaches in MCDM—such as weighted arithmetic and geometric means, as well as weighted distance-based optimisation methods—often fail to produce consistent rankings and are therefore unsuitable for pure MCDM. In contrast, the unique preference aggregation presented here clarifies the mathematical limits of valid aggregation and provides a principled, implementable foundation for robust multi-criteria decision analysis (MCDA) and multi-objective design optimisation (MODO) in confronting complexity.

Introduction

Preference is the **decisive quantity** in engineering design optimisation and management science for multi-criteria decision-making (MCDM). A preference expresses the relative desirability, *value*, or *utility* of a design alternative or decision option A_i with respect to a criterion C_j. Everything of value is relative. Each alternative or option is intuitively evaluated against one's conscious lived experience — a relative, subjective, and open-ended human perception arising from all the outer and inner senses. Preference is not a physical property but a subjective construct of the mind. It represents an individual's choice — free will — within the set of available options, defining the decision space from which selections are made. Free will cannot be absolutely measured, because it is not an object of thought but a reality expressed through human willing; what is possible is ordering and comparison, in which preference emerges as a relative expression of value. Preference is inherently contextual — it reflects the free ordering of alternatives within a given situation. It is therefore individual, relational and situation-dependent.

Consequently, **preference is 'synonymous' with choice**: a binary relation that induces choice, as one selects the alternative one prefers over another: i.e., $A_1 \succ A_2$. Without difference, no decision — only difference sets willing into motion. Preference scores or ratings are points whose meaning is inherently relational, as they have no absolute zero. Thus, they are elements of a one-dimensional affine space rather than a ratio or absolute scale (e.g., like time without a fixed origin, a rod without a defined length unit, or a road without kilometre markers — in all cases, only differences and ratios of differences carry meaning). They represent orderings and relative magnitudes of preference differences, but are not physical or absolute measurements. Thus, even a seemingly absolute 0–100 scale expressing raw preference ratings for a single criterion remains a **one-dimensional affine preference scale** (Barzilai, 2022), whose numerical values are defined only up to an affine transformation.
Only ratios of differences between preference values are meaningful. Therefore, the only

numerically invariant relation on a preference scale is the **k-ratio**:

$$k = \frac{p_a - p_b}{p_c - p_d} \quad (1)$$

where p_a, p_b, p_c, p_d are preference scores forming two differences, whose ratio defines a relative scaling factor. Only affine transformations $p_i \mapsto ap_i + b$ (i.e., scaling and translation) applied uniformly to all preference scores preserve preference differences. Any other mathematical operation (e.g., absolute values or squaring) applied separately to individual scores is not meaningful, because it alters difference ratios and thus changes the preference meaning. Linearity is a property of a vector space, not of human preferences themselves as such. Preference scores are points in a one-dimensional affine space, where only differences between scores carry meaning. Therefore, a preference difference is **not a real distance**, and preference aggregation **cannot** be based on metric distance measures or on the direct aggregation of absolute numbers. This reinforces the contemporary recognition in MCDM research (see e.g., Pajasmaa et al., 2025; French, 2023; Figueira et al., 2016) that robust multi-objective design and decision modelling requires close alignment between qualitative problem-structuring and mathematically sound preference measurement theory. This distinction between ordering and measurement has direct consequences for formal design and decision methodologies. Notably, Barzilai's preference function modelling (PFM) theory (Barzilai, 2010) had already provided a formal mathematical foundation for preference-based decision-making by establishing the measurement-theoretic conditions for meaningful preference modelling (for further engagement with PFM, the reader is referred to the first example in the Appendix).

To address this rigorously, Barzilai defined PFM axioms[1] governing PFM-consistent preference aggregation, ensuring that preference scales and their combinations remain consistent and mathematically valid (Barzilai, 2005, 2006). These axioms specify the mathematical structure that preferences must satisfy to be meaningfully combined into a **single, consistent, and unique** (up to affine transformation) aggregated preference score, thereby establishing the formal conditions under which linear algebra and calculus (e.g. Strang, 2006) can be rigorously applied to preference and other subjective quantities. Within this framework, the alternative with the highest aggregated preference represents the best-fit for common purpose, given the set of alternatives and the decision makers' criteria with their relative importance. The PFM-consistent conditions for preference representation, expressed as axioms by Barzilai, form the theoretical foundation for pure MCDM. For the purpose of aggregation in this work, the following four PFM-based axioms are outlined below. A unique and consistent decision outcome can be produced if and only if preferences satisfy these axioms:

Axiom 1 – Preference Preservation (Δ-meaningfulness)

Preference information is meaningful only in terms of **differences** between alternatives on an interval scale. These differences are relative, not absolute: distances as geometric constructs have an absolute zero, which does not exist in subjective preference measurement; zero-preference is always relative. Therefore, valid preference scales are only defined up to

[1] These are best interpreted as **formal conditions** (not axioms in the mathematical sense, which are unprovable and self-evident truths, such as Euclid's postulates) that any valid preference representation and aggregation must satisfy.

affine transformations (affine-invariant) applied uniformly:

$$p'_i = ap_i + b, \qquad a > 0, \quad \forall i, \tag{2}$$

which **preserve all meaningful preference information**. NOTE: Substituting Equation (2) into Equation (1) leaves the k-ratio unchanged, thus preserving relative preference differences between alternatives.

Axiom 2 - Comparable Criteria
Preferences from different criteria can only be aggregated if they are measured on a common, valid, preference-based interval scale with commensurate units, ensuring that no criterion dominates due to its scale rather than its assigned weight. Thus, aggregation is only valid when **preference differences are commensurate**—i.e., measured in equivalent units across all criteria—ensuring that equal marginal differences carry equal preference meaning and no criterion dominance occurs.

Axiom 3 — Meaningful Zero-Reference
All criteria must share a **common, stable and meaningful zero-reference point**, and all differences must be measured relative to this point. Only then, **linear aggregation** of these differences is mathematically valid and consistently interpretable across all alternatives and criteria; other non-linear aggregations — such as multiplicative, power-based, logarithmic, or distance-based optimisation methods — are therefore not allowed.

Axiom 4 — Uniqueness
Two preference systems producing identical judgments must correspond to the same underlying preference structure. Preference representation is therefore **unique up to affine transformations**, ensuring that criterion differences remain interval-invariant and equivalent preference information **cannot lead to conflicting aggregated rankings**.

1. Constructing the Linear Preference Space (LPS)

The objective of this section is to define a unified, interval-invariant scale of preference scores that fully complies with the PFM axioms, while supporting basic mathematical operations necessary for meaningful aggregation across criteria. For this, consider a set of alternatives A_i $(i = 1, \dots, I)$ evaluated against criteria C_j $(j = 1, \dots, J)$, with raw preference scores $p_{i,j}$ and associated non-negative weights w_j $(w_j \geq 0)$ such that $\sum_{j=1} w_j = 1$.
To construct the **Linear Preference Space (LPS)**, the standard z-score method is applied:

$$z_{i,j} = \frac{p_{i,j} - \mu_j}{\sigma_j}, \tag{3}$$

where μ_j and σ_j are the arithmetic mean and the standard deviation of the p-scores for criterion j respectively.
Importantly, μ_j and σ_j are purely arithmetic constructs used for centering and scaling; they carry no intrinsic preference meaning and so any mathematical operations may be applied to compute them. Only then the resulting $z_{i,j}$ values retain full preference meaning, maintaining all ratios of differences (k-ratios) across alternatives and criteria. Explicitly the z-transformation is affine and can be expressed as $z = a_j p + b_j$ with $a_j = 1/\sigma_j$ and $b_j = -\mu_j/\sigma_j$.

Moreover, by definition, z-normalization ensures that, for each criterion $\mu_J = 0$ and $\sigma_J = 1$, where μ_J and σ_J are the arithmetic mean and the standard deviation of the z-scores respectively. These normalized values do not acquire intrinsic preference meaning; $\mu_J = 0$ serves as the arithmetic mean of the z-scores and provides a common and stable zero-reference point, so an alternative with $z_{i,j} = 0$ can be interpreted as being "average" relative to the other alternatives on that criterion, while $\sigma_J = 1$ defines the unit of measurement ("standard deviation") for the normalized differences, ensuring that preference differences are commensurate across all criteria.

In summary, the LPS construction is defined by the mapping $T_j : p_{i,j} \mapsto z_{i,j}$, using μ_j and σ_j to set a consistent origin and commensurate unit across all criteria. Z-normalization is a purely affine transformation of the raw scores that preserves all preference differences and establishes a common linear coordinate system.

This transformation provides a stable reference frame for aggregation fully consistent with the PFM axioms. The resulting z-scores retain the meaning of all relative differences while avoiding the notion of an absolute zero, ensuring that preferences are treated as interval differences. After normalization, preferences reside in a linear preference space where aggregation is meaningful exclusively through linear operations on differences. Linearity is a property of the constructed LPS, not of human preferences.

2. Aggregated Preference Ranking

Given the same set of alternatives A_i, criteria C_j, criterion weights w_j, and normalized preference scores $z_{i,j}$ residing in the LPS as introduced in Section 1, the next step is to construct a single representative aggregated preference score P_i^* for each alternative A_i. Formally, this representative score is defined as a unique scalar value that provides a best-fit to the weighted preference differences $z_{i,j}$ of alternative A_i across the criteria. From a purely mathematical perspective, such a best-fit in a vector space is obtained by minimizing the total weighted least squared distance (WLSD) relative to a scalar representative value F, which yields the weighted centroid:

$$\min_F \sum_j w_j (f_{i,j} - F)^2 \quad \Longleftrightarrow \quad F^* = \sum_j w_j \cdot f_{i,j}, \quad \text{where } \sum_j w_j = 1. \tag{4}$$

Here, the solution F^* is the **linear weighted centroid** of the points $f_{i,j}$, obtained via a distance-based minimization. However, preferences are **differences** and **not distances**; therefore, preference aggregation cannot be derived from distance operations. In the context of preferences, where $F = P$ and $f_{i,j} = z_{i,j}$, the LPS is not a full vector space but an affine linear space of preference differences, a subset of a vector space (LPS $\subset V$), in which only linear operations on differences—scalar multiplication and addition—preserve preference meaning. Accordingly, while the WLSD formulation in Equation (4) remains mathematically valid as an auxiliary derivation, neither the squared differences nor the (distance) minimization criterion as intermediate results carry preference meaning or ordering information. They serve solely as a mathematical device to expose the unique linear outcome admissible within the LPS. Hence, the only preference-theoretically valid result is its **linear solution**, the weighted centroid.

Weighted centroid $P_i^*(z)$

Thus, in the LPS, the aggregated preference of an alternative is defined exclusively by linear operations on normalized preference differences. Accordingly, for an alternative A_i

with criterion scores $z_{i,j}$ and weights w_j, the representative aggregated preference score $\boldsymbol{P_i^*}$ is uniquely defined as the **weighted centroid** of its z-scores:

$$P_i^*(z) = \sum_j w_j \cdot z_{i,j} \quad \text{where} \ \sum_j w_j = 1 \tag{5}$$

This operator is a linear combination of preference differences (i.e., differences relative to the stable zero-reference point, obtained by centering the preference scores via z-normalization) and is therefore fully compatible with the affine preference structure of the LPS. The resulting $P_i^*(z)$ resides in the same affine space as the underlying z-scores, preserves all ratios of preference differences, and remains invariant under affine transformations of the original preference scales. NOTE: Within the multi-objective optimisation method IMAP (see see Wolfert, 2023) this $P_i^*(z)$-based preference function aggregation operator is named A, the *a-fine-aggregator* (see Teuber et al., 2025), which computes an aggregated preference score for candidate alternatives across multiple objectives.

$\boldsymbol{P_i^*}(z)$ Ranking

Given the aggregated preference per alternative, the resulting **ranking** is obtained by ordering the alternatives according to their aggregated preference scores $P_i^*(z)$ in descending order, i.e.,

$$\arg\max_i \ P_i^*(z).$$

The alternative with the highest aggregated preference P_i^* is therefore most preferred relative to the other alternatives and taking into account all criteria and their associated weights. For interpretability, it is often convenient to rescale the aggregated P_i^* values to the interval $[0, 100]$, so that the relatively worst (min: least preferred) and best (max: most preferred) alternatives correspond to 0 and 100, respectively. This can be achieved using the following linear min–max transformation, which preserves the relative ordering:

$$P_i^{*\ \text{scaled}} = \frac{P_i^* - \min(P^*)}{\max(P^*) - \min(P^*)} \cdot 100, \tag{6}$$

where $\min(P^*)$ and $\max(P^*)$ denote the minimum and maximum aggregated preference values across all alternatives. The endpoints 0 and 100 thus represent the relative minimum and maximum within the alternative set, with the zero point being purely relational rather than an absolute preference origin.

NOTE (1): The aggregated P_i^* values may also be shifted or rescaled using any affine transformation (cf. Equation (2)) without affecting the ranking. For instance, the interval $[0, 100]$ could equivalently be mapped to $[100, 200]$, preserving both preference meaning and ordering. Because the min–max transformation is linear, it preserves all relative preference differences, keeping the k-ratios from Equation (1) fully consistent and thus maintaining affine invariance.

NOTE (2): The aggregated P_i^* values may also be graphically depicted in Figure 1 below. The figure illustrates the aggregated preference P_i^* for each alternative, showing how the weighted criteria combine into a single representative value per alternative. Further use and interpretation of this figure will be provided in the next section.

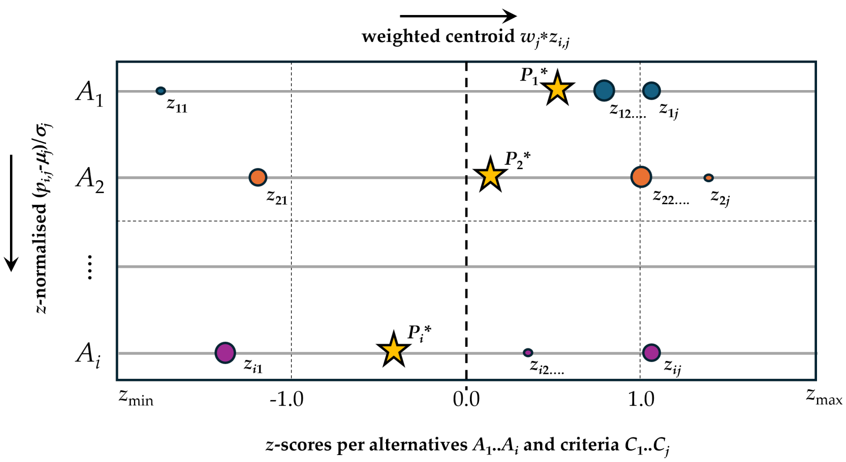

Figure 1: P_i^* as aggregated preference per alternative

3. Demonstrative Example

Consider a 4×3 ranking example[2] : i.e., four alternatives A_i and three criteria C_j with weights w_j, and 12 given preference scores p_{ij}. Here, C_1, C_2, and C_3 represent functionality, footprint, and costs, respectively.

Alternative A_i	C_1 (functionality)	C_2 (footprint)	C_3 (costs)
A_1	100	0	90
A_2	0	100	100
A_3	20	45	55
A_4	85	60	0
Weights w_j	0.4	0.1	0.5

The preference scores range from 0 to 100, where 0 represents the 'worst' performance among the alternatives (i.e., the least preferred alternative relative to the others), and 100 represents the 'best' performance among the alternatives (i.e., the most preferred alternative relative to the others). The value 0 therefore denotes the lowest relative score on the preference scale and does not imply the absence of performance on the corresponding criterion (i.e, no zero-performance).
Our goal is now to compute a mathematically meaningful aggregated preference score for each alternative, allowing us to Fthe alternatives from highest to lowest. To this end, the p-scores are first transformed into z-scores using Equation (3), yielding the corresponding **z-scores** for the four alternatives (A_1–A_4) and three criteria (C_1–C_3), as reported in the table below.

[2] In the Odesys book (see Wolfert, 2023) similar MCDA examples can be found in Chapter 5, where the alternatives are referred to as variants. The example used here is taken from that Chapter.

	C_1	C_2	C_3
A_1	1.1556	−1.4327	0.7351
A_2	−1.2148	1.3628	0.9908
A_3	−0.7408	−0.1747	−0.1598
A_4	0.8000	0.2446	−1.5660
w_j	0.4	0.1	0.5

For these criteria, it can be verified : **(1)** that $\mu_{C_1} = \mu_{C_2} = \mu_{C_3} = 0$ and $\sigma_{C_1} = \sigma_{C_2} = \sigma_{C_3} = 1$; **(2)** using Equation (1) that the k-factors in both p- and z-scores are affine invariant across the criteria: e.g., $k_{C1}(p) = k_{C1}(z) = (A_1 - A_2)/(A_4 - A_3) \approx 1.53846$; **(3)** using Equation (2) that the p- and z-scores are affinely related and therefore affine invariant: e.g., for criterion C_1 the p-values are mapped to the corresponding z-scores by $z_{i,1} = 0.02370\, p_{i,1} - 1.21485$ (all numbers rounded to 5 decimal places for verification).
Now we can compute the aggregated P_i^* per alternative using the weighted centroid from Equation (5) and the affinely invariant scaling to $[0, 100]$ using Equation (6). The final ranking is determined from highest to lowest P_i^*, resulting in the following table:

Alternative	P_i^*	scaled [0-100]	Ranking
A_1	0.68651	100	1
A_2	0.14572	52	2
A_3	-0.39368	4	3
A_4	-0.43855	0	4

Lastly, the z-scores, along with their weight contributions and the aggregated preference P_i^*, are graphically depicted below in Figure 2.

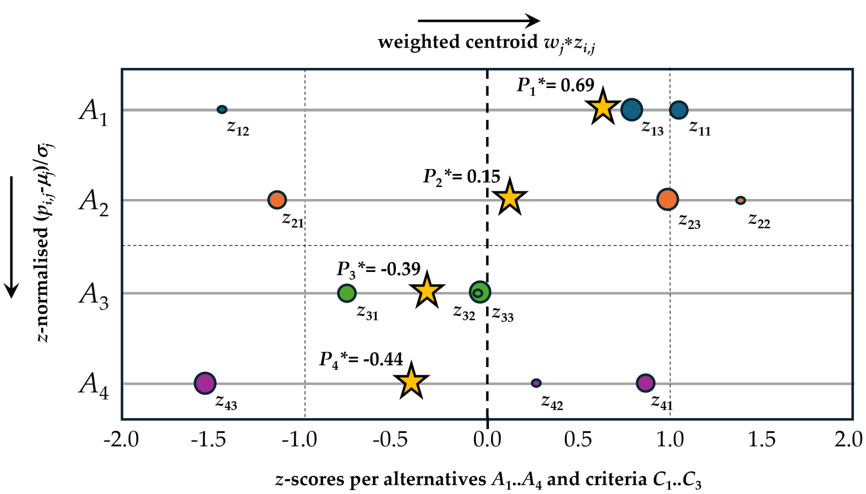

Figure 2: P_i^* as aggregated preference per alternative

The figure illustrates the aggregated preference P_i^* for each alternative, showing how the weighted criteria combine into a single representative value per alternative. Higher P_i^* values indicate stronger aggregated preference and a better overall fit. This visual summary reflects the ranking in Table and clearly illustrates the relative preference among alternatives, with Alternative 1 being the most preferred, as indicated by its position furthest to the right.

Intuitive Meaning

In the normalized LPS (defined by z-scores), P_i^* can be interpreted as the **barycentre** of a set of points in an affine (linear) space. Graphically, this corresponds to the center of mass of the points, where each point contributes proportionally to its criterion weight. Intuitively, P_i^* identifies the single point per alternative that best balances all weighted criterion preferences. NOTE: the center of mass per alternative across the weighted criteria, including P_i^*, equals zero (i.e., "horizontal" equilibrium), just as the centroid of the alternatives for each criterion also equals zero (i.e., "vertical" equilibrium).

Although preferences are *not* physical distances and have no absolute zero, the barycentre (center of mass) analogy remains helpful. To illustrate this intuition, consider a horizontal beam with three point-masses placed at different positions, as shown in the Table below. Here, the position values x_j correspond to the z-scores of Alternative 1 across the three criteria. The support must be located at the center of mass to keep the beam in equilibrium. Each point-mass represents a single preference-point contribution. The equilibrium position of the beam is given by $x_{\text{eq}} = (\sum_j m_j x_j)/(\sum_j m_j) \approx 0.6865$, which exactly equals the aggregated preference P_1^*. Even with a negative-position mass, the weighted positions combine into a single balance point. This illustrates the analogy between P_1^* and x_{eq}: although not physical distances, these values provide an intuitive "center of mass" for multiple weighted preferences, just as P_i^* represents the aggregated preference in the normalized linear preference space.

Mass m_j	Weight	Position x_j
1	0.40	1.1556
2	0.10	-1.4327
3	0.50	0.7351

NOTE: The weighted center of mass is fully symmetric in the z-scores. Scores per alternative are horizontally centered, including the aggregated preference P_i^*, while scores per criterion are vertically centered. For example, for criterion C_1 with four normalized scores and equal masses ($m_j = 1$), the centroid is $\bar{x} = \frac{1}{4}(1.1556 - 1.2148 - 0.7408 + 0.8000) = 0$, demonstrating that the weighted center of mass is symmetric: horizontally for alternative scores (including P_i^*) and vertically for criterion scores.

Weighted-mean and Distance-based rankings fail

Let us now repeat the 4×3 example from the start of this section, but using the following preferences scores p_{ij}.

Alternative A_i	C_1 (functionality)	C_2 (footprint)	C_3 (costs)
A_1	80	0	24
A_2	0	100	68
A_3	30	45	55
A_4	95	60	0
Weights w_j	0.4	0.1	0.5

Following a similar approach as at the start of this section, the resulting $P_i^*(z)$ ranking is $A_2\,(100) \succ A_3\,(79) \succ A_1\,(32) \succ A_4\,(0)$. In addition, we consider the well-known absolute weighted arithmetic mean (WAM). For this example, WAM leads to a full tie, $A_1 = A_2 = A_3 = A_4\,(50)$, which is not congruent with the $P_i^*(z)$ ranking. Moreover, any affine rescaling of individual criteria (e.g., C1 or C3 on the $[0, 100]$ scale) alters the WAM rankings, even changing the 'best' alternative, whereas the $P_i^*(z)$ ranking remains invariant. WAM aggregates alternatives by weighting absolute criterion scores without considering their relative positions within the set of alternatives. As a result, an extreme score on a single criterion can dominate the aggregation outcome, even if the alternative performs worse on multiple other, including heavily weighted, criteria. Like other distance-based rankings, WAM lacks a relative reference framework and therefore fails to produce unique, context-consistent, preference-stable rankings in the presence of conflicting criteria.
Using another weighted mean aggregation, the weighted geometric mean (WGM), the resulting ranking is $A_3\,(100) \succ A_1 = A_2 = A_4\,(0)$. Similar to WAM, WGM violates the PFM axioms, as its induced ranking depends on the choice of aggregation and the raw scores rather than on an affine-invariant preference structure. Consequently, neither WAM nor WGM provides a consistent or unique MCDM ranking in this case (see Appendix for further illustrative WAM/WGM examples).
It is beyond the scope of this section to provide the detailed calculations for the Euclidean and Manhattan distance-based rankings, as the approaches are outlined in the Appendix. Following this, the resulting distance-based rankings are: the Euclidean ranking (D_i^E) $A_3 \succ A_1 \succ A_2 \succ A_4$ and the Manhattan ranking (D_i^M) $A_3 \succ A_2 \succ A_1 \succ A_4$, demonstrating that non-linear distance-based aggregation fails to produce consistent rankings and also does not coincide with the unique $P_i^*(z)$ ranking. Thus, distance-based optimisations violate the PFM axioms, as their induced rankings depend on the choice of distance metric and non-linear transformations rather than being based on an affine-invariant preference structure. Consequently, distance-based methods cannot provide a unique or consistent MCDM ranking in this case (see Appendix for further illustrative examples).

4. Summary and Conclusions

A unique and consistent decision outcome can be produced if and only if preferences adhere to the **PFM axioms**, which formalize preferences as measurable, subjective differences within a decision space. These axioms ensure that preferences are represented as interval-invariant differences relative to a meaningful zero, comparable across criteria, and uniquely defined up to affine transformations. Consequently, aggregation preserves all meaningful relative differences, avoids scale- or criterion-induced distortions, and yields a consistent and unambiguous ranking of alternatives (Barzilai, 2010, 2022).
The aggregated preference $P^*(z)$, presented in this work, is **unique by construction**: it is the only aggregation operator consistent with all four PFM axioms. Any aggregation that violates these axioms fails to produce a coherent or meaningful preference-based ranking. This uniqueness follows from the fact that the PFM axioms fully characterize the admissible aggregation operators, restricting them to affine linear functionals. Under normalization, this admissible class collapses to a single form: the weighted centroid of the z-normalized preference scores.
Preferences are **differences in a one-dimensional affine space**, not distance vectors in a vector space. Aggregation therefore cannot arise from distance-based operations. Distances—whether Euclidean, Manhattan, or other norms—are arbitrary and can generate

infinitely many outcomes, failing to preserve preference meaning, ordering, and consistency. For this reason, a Linear Preference Space (LPS) must be explicitly constructed. Within the LPS, defined by a stable zero-reference and commensurate units of preference differences across criteria, aggregation of z-normalized preferences via the weighted centroid preserves all ratios of differences, respects the affine structure, and remains invariant under affine transformations of the original scales. This guarantees that the resulting ranking is **single, consistent, and unique**, providing a rigorous and preference-theoretically valid decision outcome. The weighted least squares distance (WLSD) is used solely as a mathematical device to derive this linear ranking structure within the LPS; the squared differences and the minimization criterion themselves do not carry preference meaning.

In contrast, many commonly used MCDM methods that rely directly on distances or pairwise comparisons—such as TOPSIS, VIKOR, Euclidean variants of AHP, and the Best–Worst Method (BWM) (see Kumar and Chan, 2021)—often violate the PFM axioms and do not guarantee linear aggregation of alternatives, leaving rankings potentially inconsistent. The WLSD-based weighted centroid within a properly constructed LPS avoids these pitfalls and provides a fully consistent framework for preference-based decision-making.

Finally, many group decision-making approaches in multi-objective optimisation—particularly those based on Pareto-optimal sets, distance-to-ideal compromises, or interactive procedures (.e.g. NIMBUS)—avoid explicit PFM-based preference aggregation altogether, as confirmed by the recent systematic review of Pajasmaa et al., 2025. This explains why, despite decades of methodological development, such methods cannot serve as true decision-making tools and merely function as mathematical constructs. While they can guide exploration of the design and decision space and/or identify non-dominated fronts, they fail to produce a single, uniquely determined best-fit alternative across all decision-makers' objectives, thus leaving the opportunity for pure design and decision-making unexploited.

APPENDIX

In this Appendix, we give the reader the opportunity to engage more deeply with the PFM theory and get better understanding through a series of clarifying, stand-alone examples. The first example, E1, provides a set of questions to offer methodological insight into Preference Function Modelling (PFM), whereas Examples E2–E4 are designed to illustrate specific conceptual pitfalls in preference aggregation. We begin with a decision analyst who wanted to review PFM before starting his new job. To do so, he first worked through the exercise in Example #E1.

Example #E1 — Why Preferences Only Have Relative Meaning (Differences)

1. A family had a daughter born on 1 January 2020 (at that moment, she was 0 days old). How many times older is their daughter 10 January 2020? And how many times older was she on 6 January 2020 compared to 2 January 2020? They also got a son, born on 2 April 2021. How many times older was he on 8 April 2021 compared to 5 April 2021?

Answer. (a) The first question is unsolvable: any ratio involving a time point measured from an arbitrary zero (0 days) is undefined, e.g., $k = (10-1)/(1-1) = 9/0$, which does not exist. (b) Only differences make sense: the interval from 1 January to 6 January is 5 days, and from 1 January to 2 January is 1 day, so $k = (6-1)/(2-1) = 5$, meaning she was 5 times as old on 6 January as she was on 2 January. (c) Same principle for the son: the interval from 2 April to 8 April is 6 days, and from 2 April to 5 April is 3 days, so $k = (8-2)/(5-2) = 2$, meaning he was 2 times as old on 8 April as he was on 5 April. One could equivalently use $k = (A-C)/(B-C)$ with C as the absolute zero-reference, i.e., date of birth = 0 living-days, giving daughter $k = 5/1 = 5$ and son $k = 6/3 = 2$.

2. Today the temperature is 0°C outside, which feels quite cold. A man decides to stay indoors. Tomorrow it will be twice as warm. Could he possibly go outside tomorrow?

Answer. Multiplying 0°C by two is meaningless because 0°C is not an absolute zero; converting to Kelvin gives $2 \times (0 - (-273))$ K = 546 K = 273°C, which is absurdly hot, showing that ratios on a non-absolute temperature scale are undefined.

3. The oldest son was born in 2002, and the oldest daughter in 2004. In 2004, the family therefore had 1 + 1 = 2 children. But what does the sum 2002 + 2004 represent? In 2006 and 2009, a third and fourth child were born respectively. Over how many years did the man become father to four children?

Answer. The sum $2002 + 2004 = 4006$ is mathematically correct but meaningless: addition is not defined on calendar years. Differences *are* meaningful: 2009 - 2002 = 7 years.

4. Given that t_1 is 14u = 2 p.m. and t_2 is 15u = 3 p.m. Is the ratio t_2/t_1 equal to $3/2 = 1.5$ or to $15/14 = 1.0714\ldots$?

Answer. Neither. Time points belong to an affine space where division is not defined. Only time *intervals* can be compared.

5. A man travels by train from Delft to Eindhoven, departing at 14:00 from Delft and arriving at 16:00 in Eindhoven. During the journey, he changes trains once in Breda. Can one calculate how many times longer the travel time from Delft to Breda is compared to Breda to Eindhoven and what the average train travel time of these subtrips is? If yes, how many times and what is this average? If no, what additional information is needed?

Answer. This calculation is only possible if the intermediate times are known; otherwise, one could only say, "I'm later in Eindhoven than in Delft." Extra information is needed for a timestamp in Breda. Suppose the traveller arrives in Breda at 14:45 and departs at 15:00, then the subtrip durations are Delft–Breda = 45 min and Breda–Eindhoven = 60 min, so the ratio is $k = \frac{\text{Eindhoven–Breda}}{\text{Delft–Breda}} = \frac{60}{45} = \frac{4}{3}$. Now also the average travel time across these two subtrips can be computed as $0.5 \cdot 45 + 0.5 \cdot 60 = 52.5$ min.

6. What is the potential energy of a 1 kg ball placed on top of a 15-storey building, assuming each storey is 4 meters high? What is its potential energy when located on the first floor? What is the difference in potential energy between these two positions?

Answer. Potential energy requires choosing a zero reference level; if ground level is the zero, then $PE_{15} = 1 \times 10 \times (15 \times 4) = 600$ Nm and $PE_1 = 1 \times 10 \times (1 \times 4) = 40$ Nm, so the meaningful quantity is the difference $\Delta PE = 560$ Nm.

7. A dredger pump is installed in a supply pipe. On one side of the pump, the pressure is X, and on the other side, the pressure is Y. Under what condition will fluid flow occur?

Answer. Flow occurs only when $X \neq Y$. If $X > Y$, fluid flows from X to Y. Only pressure differences matter.

8. In a design problem, Alternative A has a value X and Alternative B has a value Y. When can a designer make a definitive choice between these alternatives? Can the designer state how many times "better" one alternative is compared to the other if the X and Y values represent (a) noise levels or (b) beauty? Now suppose $[X, Y, Z]$ are (1) $[70, 50, 10]$ and (2) $[60, 40, 0]$, where Alternative C has a value Z. Please calculate how much more "better/worse" A is compared to B, and if necessary relative to C, for cases (a)-noise and (b)-beauty, using (1) and (2) as the respective preference values.

Answer. A choice is possible only if $X \neq Y$. To say that "A is k times better than B" one needs an absolute zero reference point or a third alternative with absolute zero meaning, C, such that $k = \frac{A-C}{B-C} = \frac{A}{B}$. This only makes sense for quantities like noise levels, which have a meaningful absolute zero. For subjective criteria like beauty, only relative differences are meaningful. Thus, the only meaningful calculations are either the k-factor $k = \frac{A-C}{B-C}$, expressing that A is k times more beautiful compared with C, or applying $k = k_A/k_B$, where $k_A = \frac{A-\min}{\max-\min}$ and $k_B = \frac{B-\min}{\max-\min}$, again stating k expresses that A is k times more beautiful compared with C.
NUMERICAL RESULTS: The raw preference values satisfy $[70, 50, 10] \leftrightarrow [60, 40, 0] \leftrightarrow$

$[100, 66.67, 0]$, showing that both triples are related by the same affine (min–max) transformation. (a) Noise levels with an absolute zero reference: $k = \frac{A-C}{B-C} = 1.5$, meaning A is 1.5 times "worse" than B (A is 1.5 times louder). (b) Beauty values without a meaningful zero: using the relative reference $C = 0$, $k = \frac{A-C}{B-C} = \frac{60-0}{40-0} = 1.5$, meaning A is 1.5 times more beautiful than B *only relative* to C. Alternatively, using the min–max k-ratio with min = 0 and max = 100, $k_A = 1$, $k_B = 0.667$, and $k = k_A/k_B = 1.5$. Since beauty is subjective, this 1.5 ratio only expresses a relative A/B difference compared with C (even if C seems to be 0, it is merely an arbitrary relative reference, so no "A is k times better than B").

9. A person must choose between two jobs, A and B, based on the criteria 'growth opportunities' and 'salary', weighted at 0.6 and 0.4 respectively. Growth opportunities are 15 and 20 for positions A and B; salaries are €50,000 and €45,000 per year respectively. Which job does the person choose if he uses the weighted average mean? However, upon checking the contract, the salary is not in euros (€) but in kilo-dollars (k$). The exchange rate is 1 € = 1.1 $, so the salaries become 55 k$ and 49.5 k$ per year respectively. After recalculating, the person chooses differently. Why can't he not make a final decision?

Answer. 1) He 'unconsciously' uses a weighted average: overall weighted score yields $A = 20009 \succ B = 18012$. So he chooses A; 2) Overall weighted score yields $A = 31 \prec B = 31.8$. So he chooses B. Now he can't decide. Why? Because the weighted average is not defined in a one-dimensional affine space. Moreover, physical units are not part of the definition of the weighted average. The weighted average operator gives an infinite number of non-equivalent outcomes and is therefore unusable for decision-making!

10. A professor receives an overall teaching evaluation score of 7, based on the weighted average of three equally weighted subscores: 'clarity', 'comprehensibility', and 'explainability'. At home, he tells his children about this 'good' result, upon which one child says: "That result means nothing to me, Dad!" Is the child right? If no, explain why not. If yes, what is needed to assess a 'good' result? Is the weighted average a reliable basis for this assessment?

Answer. The child is correct. A '7' is meaningless without knowing the scale and the reference group. The three subscores cannot be meaningfully added, and the weighted average is undefined on an affine scale. A claim of "good" must be based on normalized scores and comparison relative to a reference point (e.g., group mean or best performer). Only normalized differences (e.g., weighted least-squares in affine preference space) yield meaningful results.

11. Consider a five-dimensional conceptual space: the first three dimensions (x, y, z) represent *space*, the fourth dimension t represents *time*, and the fifth dimension p represents *preference*. Assume that motion in space requires a height difference Δh, motion in time requires a temporal difference Δt, and choice requires a preference difference Δp. What is the fundamental difference between space, time, and preference with respect to (a) their ontological status (object, object–subject, subject), (b) the existence of a meaningful zero reference, and (c) the condition under which motion or action occurs?

Answer. Space is an *objective physical domain*, belonging to the object *in itself* and directly observable. A zero reference in space (zero length or height) is physically meaningful and does not depend on human choice. Motion in space requires a difference in height: $\Delta h \neq 0 \Rightarrow$ motion occurs; if $\Delta h = 0$, the system remains at rest. Time has no physical zero inherent in the object; a zero point can only be defined by a subject through human convention (e.g., a clock start or calendar origin). Time is therefore a *conceptual coordinate system* imposed by the subject upon objective processes. Only time intervals are meaningful: $\Delta t \neq 0 \Rightarrow$ change can be ordered; $\Delta t = 0 \Rightarrow$ no progression occurs. Preference is entirely *subjective* and has no absolute zero; it arises from the synthesis of objective properties of the object and the subject's experience, valuation, and free will. Action or decision requires a preference difference: $\Delta p \neq 0 \Rightarrow$ a choice is made; $\Delta p = 0 \Rightarrow$ indifference.

The fundamental distinction can be summarized in the following compact table:

Dimension	**Ontological Status**	**Zero Reference**	**Difference Causes**
Space (x, y, z)	Object	Physically meaningful	Motion via Δh (energy)
Time (t)	Object–Subject	Conventionally defined	Progression via Δt
Preference (p)	Subject	No absolute zero	Decision via Δp

Space and time provide the *conditions of possibility*, while preference provides the *conditions of desirability*, which in synthesis form the basis for *design into action*. Thus, physical and social dynamics are structurally analogous—potential differences cause motion, while preference differences cause decisions. Where no difference is experienced, indifference prevails, and no movement, choice, or change occurs. Design and decision-making therefore unfold within an action space defined by *five dimensions*: spatial extension (x, y, z), temporal development (t), and preference (p), through which freedom becomes effective in the world.

BONUS: Time and preference cannot exist without the subject and are inherently part of the affine space. However, they develop in opposite directions: time moves forward along a single arrow, and in the absence of the subject, allows the object to decay or drift toward disorder (high entropy). Preference, in contrast, can be understood as a "future-oriented time" approaching the present, in which the subject actively brings an object into existence (reducing disorder, low entropy). Preference can branch into multiple sub-preferences—many "arrows" of preference—reflecting the layered, complex nature of human free will and decision-making. In this sense, preference functions as a form of "counter-time". By logical extension, one could reasonably posit the existence of a corresponding "counter-space," though its exploration lies beyond the scope of this Question 11.

Example #E2 — Why the Weighted Mean Fails in MCDM

A decision analyst has to evaluate two job offers, **Job A** and **Job B**, based on two criteria: **growth opportunities (C1)** and **salary (C2)**. The criteria are weighted as $w_1 = 0.6$ for growth and $w_2 = 0.4$ for salary (see also Example #E1-Q9). The raw values for growth and salary per job are given in the table below.

	C_1(growth)	C_2(salary-€)
A_1 (Job A)	15	50,000
A_2 (Job B)	20	45,000
$w_{1,2}$	0.6	0.4

Using the values from the table above, the analyst computes the weighted average mean. With the original €-salaries, the result is $A \succ B$. After discovering that the salaries were actually given in k$, and converting them accordingly, the same weighted average mean now yields $B \succ A$, even though the underlying decision problem has not changed (see the exact values in Example #E1-Q9).
Thus, the weighted average produces an infinite number of non-equivalent outcomes and therefore fails in decision-making. The analyst then decides to apply Preference Function Modelling (PFM) to determine whether he can make a unique and sound choice.

Unique Decision Making Using the Aggregated P_i^*

The decision analyst considers the following three approaches to the problem:

1. Starting from the raw performance scores, either in € or in k$, the values are directly normalized into z-scores, after which the aggregated preferences P_i^* are computed.

2. Starting by converting all criterion performance scores, whether expressed in € or in k$ (which becomes irrelevant), into raw preference scores between 0 and 100 ('best'=100 and 'worst'=0), with A_1 (Job A) having scores (0, 100) and A_2 (Job B) having scores (100, 0).

Both approaches produce the same normalized z-score patterns and so the unique P_i^*:

	C_1	C_2	P_i^*
A_1	-1	1	-0.2
A_2	1	-1	0.2
$w_{1,2}$	0.6	0.4	

Even though both approaches start from different numerical representations (€, k$, or raw preference values), after z-normalization they all lead to : $\boldsymbol{B \succ A}$. This demonstrates that the PFM / z-normalized centroid is **representation-invariant**, unlike the weighted average mean, which produced contradictory rankings. The decision analyst was very pleased, as at that moment the most important life decision for him could be made in a unique and transparent way.

Undefined Weighted Arithmetic or Geometric Mean

After making the proper job decision for Job B, he started working in the Data Science and Engineering (DSE) department. His first task was to analyse the following design evaluation problem, which involved two alternatives and two criteria, with the preference scores given as follows:

	C_1(quality)	C_2(safety)
A_1	0.9	0.2
A_2	0.7	0.6
$w_{1,2}$	0.5	0.5

The decision analyst calculated the P_i^* using z-scores and concluded that both alternatives were equal (i.e., an indifferent choice: 50/50 and equal ranks). He went back to his manager and said he could not make a decision. The manager overruled him and instructed him to use the Weighted Arithmetic Mean (WAM) or the Weighted Geometric Mean (WGM) [3], as he should have been taught and gave him the following references: Kumar and Chan, 2021 and Krejčí and Stoklasa, 2018. The decision analyst returned, disappointed in his first project, but did not give up and proceeded with the following analysis. He used, as instructed, the Weighted Arithmetic Mean $\text{WAM}_i = w_1 \cdot x_{i1} + w_2 \cdot x_{i2}$ and the Weighted Geometric Mean $\text{WGM}_i = x_{i1}^{w_1} \cdot x_{i2}^{w_2}$, and also performed the $P_i^*(z)$ computation (see Equation (5)). He constructed the following four Examples
For his storyline to his manager, he began his analysis with the following 2×2 problem — two alternatives and two criteria, C1 (quality) and C2 (safety) — as presented in the preferences table:

	C_1	C_2
A_1	0.90	0.35
A_2	0.60	0.60
w_j	w_1	w_2

For Example 1, the criteria weights are set as $(w_1, w_2) = (0.6, 0.4)$. He computed the following seemingly consistent ranking:

A_i	WAM	WGM	$P_i^*(z)$	Rank
A1	0.68	0.618	0.2	1
A2	0.60	0.600	-0.2	2

However, he proceeded with the following. In Example 2/3, he considered two sets of weights for the criteria, $(w_1, w_2) = (0.54, 0.46)$ and $(w_1, w_2) = (0.48, 0.52)$, and used the same preference scores as in Example 1.
As he was able to demonstrate, both the WAM and WGM rankings failed, as they reversed the ranking order and were, at times, inconsistent with $P_i^*(z)$. He then proceeded to analyse the following Example 4:

	C_1	C_2
A_1	0.90	0.20
A_2	0.50	0.55
$w_{1,2}$	0.5	0.5

[3] In general, WAM and WGM are widely used aggregation methods.The absolute weighted arithmetic mean (WAM) is defined as $\text{WAM} = \sum_{j=1} w_j |p_j|$, and the weighted geometric mean (WGM) as $\text{WGM} = \prod_{j=1} |p_j|^{w_j}$, where p_j is the preference score on criterion j and w_j is the corresponding weight.

w_j	A_i	WAM	WGM	$P_i^*(z)$
$w_1 = 0.54$	A_1	1	2	1
$w_2 = 0.46$	A_2	2	1	2
$w_1 = 0.48$	A_1	1	2	2
$w_2 = 0.52$	A_2	2	1	1

Table 1: Aggregated rankings under different weight vectors.

This again resulted in an ambiguous ranking for both WAM and WGM, and was not consistent with $P_i^*(z)$:

A_i	WAM	WGM	$P_i^*(z)$
A1	1	2	1=2
A2	2	1	1=2

To completely impress his manager, he performed a final 4×3 Example 5, i.e., four alternatives and three criteria — C1 (technical quality), C2 (safety), and C3 (reliability) — with the following preference scores:

	C_1(quality)	C_2(safety)	C_3(reliability)
A_1	0.87	0.62	0.53
A_2	0.80	0.72	0.72
A_3	0.77	0.70	0.54
A_4	0.79	0.93	0.56
$w_{1,2,3}$	0.50	0.25	0.25

A quick conclusion yields that WAM selects A4 as the best alternative, WGM selects A2 as the best alternative, and $P_i(z)$ selects A1 as the best alternative.

A_i	WAM	WGM	$P_i^*(z)$
A1	4	3	1
A2	2	1	2
A3	3	4	4
A4	1	2	3

In this way, he demonstrated to his manager, using these examples, that the WGM and the WAM can yield infinitely many possible answers with different outcomes, non-consistent rankings, and are therefore both unsuitable as multi-criteria decision-making (MCDM) methods. Eventually, the roles were reversed: the manager was completely lost and had to abandon his incorrectly taught MCDM tools, while the decision analyst continued his work in the DSE department successfully. From that point on, they used only the PFM-based MCDM approach, as it was clear that all other methods (WAM and WGM) simply failed.

Example #E3 — Why We Need Commensurate Criteria and a Stable Zero-Reference

The decision analyst from the previous Example #E2 now wants to evaluate five alternatives on two criteria ('5x2 problem'), using the following preference scores:

dataset (0)	C_1 (cost)	C_2 (time)
A_1	100	40
A_2	0	60
A_3	20	45
A_4	85	50
A_5	60	55
$w_{1,2}$	0.5	0.5

Having learned in Example #E2 that directly applying $P_i^*(z)$ yields an immediate, unique, and consistent ranking, he was—after the euphoria of that result—perhaps slightly overconfident. He began to wonder whether there might exist an alternative approach that would avoid z-normalization altogether. In doing so, he decided first to explore the structure of the raw preference scales, hoping to deepen his understanding and, admittedly, to impress his manager even more than he already had in the previous Example.
Therefore, before applying any transformation, he examined whether the criteria scales were inherently comparable, noting that both C_1 and C_2 use 0–100 preference scores and share equal weights.

Scale Comparability

For a comparability check all criteria must at least have equal ranges and equal minimum values, i.e. $\forall j, k \in \{1, \ldots, m\} : p_{\min,j} = p_{\min,k} \ \wedge \ p_{\max,j} = p_{\max,k}$.
This condition ensures that a difference of Δp on any criterion contributes proportionally, independently of the original scale. This is exactly the essence of Barzilai's *scale validity* (see Barzilai, 2010). By checking the above condition, it became clear that these scales are not directly comparable. Intuitively, he could also see that a difference of $\Delta p = 10$ has very different implications for the two criteria: for C1 (cost), which ranges from 0 to 100, a difference of 10 represents 10% of the total spread and is therefore substantial, while for C2 (time), which only varies from 40 to 60, a difference of 10 represents 50% of the smaller spread; however, because the actual observed differences between alternatives on C2 are smaller (mostly 5–10 units), the criterion has less discriminative power in practice. Consequently, C2 contributes less to the overall ranking compared with C1. Moreover, C1 represents costs ranging from €10,000 to €1,000,000 ('lower is better'), whereas C2 represents time varying only from 80 to 120 days ('shorter is better'). Therefore, C1 and C2 cannot be directly aggregated without proper normalization, because it is their spreads—not their weights—that determine their influence on the aggregation. This criterion dominance is not what weights represent; it is purely an artefact of the mapping, not true importance. Consequently, he decided to perform an affine transformation on C2, mapping all values to the interval $[0, 100]$.

dataset (1)	C_1	C_2
A_1	100	0
A_2	0	100
A_3	20	25
A_4	85	50
A_5	60	75
$w_{1,2}$	0.5	0.5

Now it can be seen from Dataset (1) that the scales are comparable, since $p_{\min} = 0$ and $p_{\max} = 100$. After this analysis, the decision analyst then considered whether the k-ratio approach could still 'uniquely' determine a proper aggregated preference ranking. To address this question, he first compiled the following interlude.

Interlude

Following the basic k-ratio (see Equation (1)), it can be shown that for multiple alternatives: let $i = 1, \ldots, n$ index the alternatives, and multiple criteria : let $j = 1, \ldots, m$ index the criteria the k-ratio of alternative i on criterion j is defined as

$$k_{i,j} = \frac{p_{i,j} - p_{\min,j}}{p_{\max,j} - p_{\min,j}}, \tag{7}$$

where $p_{\min,j}$ and $p_{\max,j}$ denote the minimum and maximum preference scores for criterion j. Then he assumed that a subset of the linear operations were valid and since the k-scores are affine invariant, a consistent aggregation across criteria could be obtained using the following weighted centroid:

$$\Pi_i^*(k) = \sum_{j=1} w_j \, k_{i,j} \tag{8}$$

where w_j is the weight of criterion j, with $\sum_{j=1} w_j = 1$. For $p_{\min,j} = 0$ and $p_{\max,j} = 100$, we obtain $\Pi_i^*(k) = \frac{1}{100} \sum_{j=1} w_j \, k_{i,j}$.

Comparison $\Pi_i^*(k)$ and $P_i^*(z)$ Ranking

Now he wanted to test and compare his new k-score–based $\Pi_i^*(k)$ ranking with the $P_i^*(z)$ ranking (see Equation (5)). In addition to Dataset (1), he therefore constructed the following additional Dataset (2), in which the preference values were slightly adjusted to become fully symmetric ([0,100]), and in which an extra intermediate reference was also introduced:

dataset (2)	C_1	C_2
A_1	0	100
A_2	100	0
A_3	50	50
A_4	80	30
A_5	20	70
$w_{1,2}$	0.5	0.5

He then performed the computational analysis for both $P_i^*(z)$ and $\Pi_i^*(k)$, obtaining the following results:

dataset	$\boldsymbol{\Pi_i^*(k)}$	$\boldsymbol{P_i^*(z)}$
(1)	$A_4 = A_5 \succ A_1 = A_2 \succ A_3$	$A_5 \succ A_4 \succ A_2 \succ A_1 \succ A_3$
(2)	$A_4 \succ A_1 = A_2 = A_3 \succ A_5$	$A_4 \succ A_1 \succ A_3 \succ A_2 \succ A_5$

To his astonishment, he found that for both datasets he could not arrive at a unique ranking, as $\Pi_i^*(k)$ yielded different results than $P_i^*(z)$. He was particularly struck by the fact that $\Pi_i^*(k)$ contained several ties or indifferent outcomes, even for three alternatives, preventing a unique decision from being made. He then proceeded to re-examine the following 3x2-datasets to determine whether these ties were specific or indicative of a more general issue.

dataset (10)	C_1	C_2
A_1	0	0
A_2	0	100
A_3	100	60
$w_{1,2}$	0.30	0.70

dataset (15)	C_1	C_2
A_1	-50	-50
A_2	-50	50
A_3	50	10
$w_{1,2}$	0.30	0.70

dataset (20)	C_1	C_2
A_1	90	20
A_2	50	55
A_3	10	100
$w_{1,2}$	0.50	0.50

dataset (30)	C_1	C_2
A_1	95	10
A_2	60	55
A_3	10	100
$w_{1,2}$	0.51	0.49

After computing $P_i^*(z)$ and $\Pi_i^*(k)$ for the datasets (10), (15), (20), and (30), he obtained the following remarkable results:

dataset	$\boldsymbol{\Pi_i^*(k)}$	$\boldsymbol{P_i^*(z)}$
(10)	$A_3 \succ A_2 \succ A_1$	$A_2 \succ A_3 \succ A_1$
(15)	$A_3 \succ A_2 \succ A_1$	$A_2 \succ A_3 \succ A_1$
(20)	$A_1 = A_3 \succ A_2$	$A_1 \succ A_3 \succ A_2$
(30)	$A_1 \succ A_2 \succ A_3$	$A_2 \succ A_1 \succ A_3$

After careful analysis, the decision analyst concluded that $\Pi_i^*(k)$ could not yield a unique or consistent decision, whereas only $P_i^*(z)$ provided a stable ranking. The main reasons for $\Pi_i^*(k)$ violating the PFM axioms are:

- **Zero-Reference Requirement:** aggregation requires a stable, meaningful zero per criterion to ensure differences are linear and interpretable.
- **Arbitrary Zero in $\Pi_i^*(k)$:** $k_{ij} = p_{ij}/p_{\max,j}$ sets 0 arbitrarily; $p_{\max,j}$ depends on the dataset, so differences are not relative to a common zero.
- **Non-linear Differences:** k_{ij} is linear in p_{ij} but not in differences from a stable reference; shifting all p_{ij} changes k_{ij}, violating Axiom 3.
- **Incomparable Differences:** zeros vary across criteria, so differences cannot be aggregated meaningfully.
- **Non-uniqueness:** in symmetric datasets, $\Pi_i^*(k)$ may match $P_i^*(z)$ rankings, but this is incidental; true compliance requires a stable zero-reference.

Overall, $\Pi_i^*(k)$ fails to satisfy PFM Axioms because it lacks a common zero-reference and does not preserve linearity in meaningful differences, whereas $P_i^*(z)$ satisfies all axioms by design, yielding a unique and consistent ranking.

NOTE: k-Ratio vs. z-Normalisation — Preferences represent **differences**, not (vector) distances. They are relative, with no intrinsic absolute zero, and even z-normalized preferences remain points in a linear preference space. The zero corresponds to the sample mean and a unit difference (standard deviation), serving only as a reference for linear operations. In a z-normalized space, an aggregated ranking $P_i^*(z)$ can be computed with as few as two alternatives. In the p-domain (raw scores), at least three alternatives are generally needed to define meaningful k-ratios ($k_i = (p_i - p_{\min})/(p_{\max} - p_{\min})$), because no natural stable zero-reference exists. The z-normalization introduces a functional zero at the mean (μ_j) for each criterion, rendering all differences affine-invariant and allowing linear aggregation with only two alternatives. The functional mean can be interpreted as a "third" reference alternative in the k-ratio framework.

Example #E4 — Why Distances Cannot Represent Preferences

In this final example, the decision analyst aims to reveal a common pitfall in preference aggregation, namely the frequent—but incorrect—treatment of preferences as distances and the use of (non-linear) distance-optimisation–based rankings. To this end, he starts by evaluating a set of three alternatives and two criteria with the following p- or z-scores.

p-scores	C_1	C_2
A	90	10
B	50	50
C	10	90
$w_{1,2}$	0.5	0.5

z-scores	C_1	C_2
A	1.0	−1.0
B	0.0	0.0
C	−1.0	1.0
$w_{1,2}$	0.5	0.5

Comparison of Distance-Based D_i and P_i^* Rankings

In any vector space, one may define a total weighted least squared difference (WLSD). In the present context, this first takes the form of a weighted Euclidean distance (squared) between an alternative i and an ideal (target) point $(1, 1, \ldots, 1)$, as commonly used in TOPSIS-style constructions (see, e.g., Kumar and Chan, 2021). The normalized Euclidean WLSD is then given by

$$D_i^E = \sum_{j=1} w_j \, (z_{ij} - 1)^2,$$

where w_j are the criterion weights and z_{ij} the normalized preference score of alternative i on criterion j. The resulting ranking is obtained by $\arg\min_i(D_i^E)$, i.e., by minimizing the distance to the ideal point over all alternatives. The intention is to compare this ranking with the one induced by maximizing the aggregated preference score P_i^* (see Equation (5)). This resulted in:

	C_1	C_2	$\boldsymbol{P_i^*}$	$\boldsymbol{D_i^E}$
A	90	10	0.0	2.0
B	50	50	0.0	1.0
C	10	90	0.0	2.0
w_j	0.5	0.5	–	–

He concluded, for the time being, that the Euclidean distance-based ranking $D_i^E : B \succ A = C$ was inadequate, as it neither corresponded to his intuition—according to which a clear tie should emerge nor to the outcome of P_i^*, which indeed indicated a full tie (as expected a-priori). As an initial reflection, however, he did not discard the distance-based ranking altogether, reasoning that the observed discrepancy might be attributable to the complete symmetry of the dataset. To gain further confidence, the analyst considered a second distance-based ranking to assess the sensitivity of the results to the choice of norm, namely the Manhattan distance. Analogous to the WLSD Euclidean-distance approach, the Manhatten-based ranking is obtained by:

$$\arg\min_i (D_i^M), \quad \text{with} \quad D_i^M = \sum_{j=1} w_j \, |z_{ij} - 1|.$$

In this second step, he applied the Manhattan-based WLSD to a set of non-symmetrical datasets for further evaluation, together with their corresponding p-scores.

dataset (1)	C_1	C_2
A	60	0
B	0	100
C	100	20
$w_{1,2}$	0.3	0.7

dataset (2)	C_1	C_2
A	90	20
B	50	55
C	10	100
$w_{1,2}$	0.5	0.5

dataset (3)	C_1	C_2
A	0	0
B	0	100
C	100	60
$w_{1,2}$	0.3	0.7

To his astonishment, he obtained the following results: all distance-based rankings—whether Euclidean or Manhattan (with "lowest is best")—differed from the aggregated P_i^* preference ranking (with "highest is best"). Moreover, the distance-based rankings could even differ among themselves. NOTE: all values reported in the tables are rounded. In essence, this demonstrates that distance-based optimisation can lead to inconsistent rankings and therefore cannot guarantee a unique ranking in MCDM. Such approaches fail to comply with the axioms of Preference Function Modelling (PFM), which are required for consistent and unique preference aggregation.

dataset (1)	$\boldsymbol{P_i^*}$	$\boldsymbol{D_i^E}$
A	-0.599	2.805
B	0.582	1.690
C	0.017	1.505
ranking	$B \succ C \succ A$	$C \succ B \succ A$

dataset (2)	$\boldsymbol{P_i^*}$	$\boldsymbol{D_i^M}$
A	0.027	1.197
B	-0.051	1.051
C	0.024	1.248
ranking	$A \succ C \succ B$	$B \succ A \succ C$

dataset (3)	$\boldsymbol{D_i^E}$	$\boldsymbol{D_i^M}$
A	4.571	2.121
B	0.887	0.607
C	0.542	0.710
ranking	$C \succ B \succ A$	$B \succ C \succ A$

After the previous evaluations, he realised that when it comes to a life-or-death decision (e.g., choosing an operation for person A, B, or C), it is of paramount importance that the ranking is unique and consistent. The result, however, was non-unique under both the Euclidean and Manhattan distances, demonstrating that distance-based rankings—whether

Euclidean, Manhattan, or based on other norms—are arbitrary and cannot provide a proper ranking of alternatives. Finally, he concluded that even within distance-based rankings, an infinite number of different outcomes can occur, and that **preferences are not (metric) distances**, but points in a one-dimensional affine space, with relative meaning. Overall, it is concluded that distance-based rankings are norm-dependent and unstable, whereas the weighted centroid $P_i^*(z)$ in a linear preference space is unique by design, as confirmed by step-by-step numerical examples. This can be attributed to at least the following:

1. P_i^* is additive, linear, and therefore **consistent** with the preference axioms governing a linear preference space (LPS).
2. Distance-based measures D_i construct rankings using absolute (Manhattan) or squared (Euclidean) deviations, which are strictly non-negative by definition. As a result, they cannot be congruent with the signed centroid aggregation P_i^*, representing a linear preference projection, and may **directly contradict the preference ranking**. Any coincidence in ranking is incidental rather than structural.
3. Distance-based aggregation relies on **an absolute metric zero**, which does not exist in preference scales, rather than on a meaningful preference centre. In contrast, P_i^* is defined within a centred linear preference space, where the mean provides a functional zero and preserves all relative differences.

References

[1] Pajasmaa, J., Miettinen, K., & Silvennoinen, J. (2025). Group decision making in multiobjective optimization: A systematic literature review. *Group Decision and Negotiation*, 34(2), 329–371. https://doi.org/10.1007/s10726-024-09915-8

[2] Teuber, L.G., van Heukelum, H.J., & Wolfert, A.R.M. (2025). Advancing strategic planning and dynamic control of complex projects. *arXiv*, https://doi.org/10.48550/arXiv.2408.12422.

[3] French, S. (2023). Reflections on 50 years of MCDM: Issues and future research needs. *EURO Journal on Decision Processes*, 11-100030, https://doi.org/10.1016/j.ejdp.2023.100030

[4] Wolfert, A.R.M. (2023). *Open Design Systems*. IOS Press/ TUDelft-OPEN, https://doi.org/10.3233/RIDS10 or www.odesys.nl.

[5] Barzilai, J. (2022). *Pure Economics*. FriesenPress.

[6] Kumar, U., Chan, F.T.S. (Eds.). (2021). *A Handbook on Multi-Attribute Decision-Making Methods*. Wiley.

[7] Krejčí, J., & Stoklasa, J. (2018). Aggregation in the analytic hierarchy process: Why weighted geometric mean should be used instead of weighted arithmetic mean. *Expert Systems with Applications, 114*, 97–106. https://doi.org/10.1016/j.eswa.2018.06.060

[8] Figueira, J. R., Greco, S., Ehrgott, M. (Eds.). (2016). *Multiple Criteria Decision Analysis: State of the Art Surveys* (Vol. 233 & 234). Springer.

[9] Barzilai, J. (2010). Preference function modelling: The mathematical foundations of decision theory. In M. Ehrgott, J. R. Figueira, & S. Greco (Eds.), *Trends in Multiple Criteria Decision Analysis* (pp. 57–86). Springer. https://doi.org/10.1007/978-1-4419-5904-1_3

[10] Barzilai, J. (2006). Preference modeling in engineering design. In K. E. Lewis, W. Chen, & L. C. Schmidt (Eds.), *Decision Making in Engineering Design* (pp. 43–47). ASME Press.

[11] Strang, G. (2006). *Linear algebra and its applications* (4th ed.). Cengage Learning.

[12] Barzilai, J. (2005). Measurement and preference function modelling. *International Transactions in Operational Research, 12*(2), 173–183. https://doi.org/10.1111/j.1475-3995.2005.00496.x