Journal of Intelligent Manufacturing (2020) 31:387–402
https://doi.org/10.1007/s10845-018-1452-5

Adapting Thurstone's *Law of Comparative Judgment* to fuse preference orderings in manufacturing applications

F. Franceschini[1] · D. Maisano[1]

Received: 26 January 2018 / Accepted: 13 October 2018 / Published online: 22 October 2018

Abstract

A rather common problem in the manufacturing field includes: (i) a collection of *objects* to be compared on the basis of the degree of some *attribute*, (ii) a set of *judges* that individually express their *subjective* judgments on these objects, and (iii) a single *collective judgment*, which is obtained by fusing the previous subjective judgments. The goal of this contribution is to develop a new technique that combines the Thurstone's *Law of Comparative Judgment* with an ad hoc response mode based on preference orderings. Apart from being relatively practical and user-friendly, this technique allows to express the collective judgment of objects on a *ratio* scale and is applicable to a variety of practical contexts in the field of manufacturing. The description of the proposed technique is integrated with the application to a practical case study.

Keywords Manufacturing · Quality engineering/management · Decision making · Preference ordering · Paired comparison · Law of comparative judgment · Scaling · Ratio scale

Introduction

Subjective measures of product/process attributes (e.g., early estimation of product manufacturability, evaluation of qualitative product features, evaluation of customer satisfaction and perceptions, assessment of operator skills and knowledge, etc.) are crucial in several practical contexts such as manufacturing, quality engineering/management, product design, marketing, etc. (Krynicki 2006; Maier and Fadel 2007; Tao et al. 2016; Zheng et al. 2016; Lin and Cheng 2017).

A common problem is that in which a set of *judges* express their individual (subjective) judgments on a specific *attribute* of some *objects* and these judgments have to be fused into a collective one (Keeney and Raiffa 1993). Focusing on the manufacturing and quality engineering/management fields, possible examples concern: (i) the fusion of customer expectations on a set of product requirements, (ii) the fusion of judgments by reliability and maintenance engineers on the severity of a set of potential process failures, and (iii) the fusion of the opinions of designers and marketing experts on the brand image of several competing products. Addressing these issues correctly is crucial to guide development strategies and improvement actions of companies/organizations (Den Ouden et al. 2006; Franceschini et al. 2019; Lin and Cheng 2017).

Returning to the problem of interest, it can be structured as follows (see Fig. 1):

- A set of *objects* should be compared on the basis of the degree of some *attribute*. By attribute, we mean a specific feature of the objects, which is relevant for their comparison. It is assumed that the attribute of interest was previously determined through appropriate methods (Van Kleef et al. 2005).
- A set of *judges* individually express their *subjective* judgments (i.e., problem *input*) on these objects.
- Subjective judgments should be fused into a single *collective judgment* (i.e., problem *output*), usually expressed in the form of a *scaling*, i.e., assignment of numbers to the objects, according to a conventional rule/method.

The scientific literature encompasses a plurality of fusion techniques, which differ from each other for (at least) three features (DeVellis 2016):

✉ F. Franceschini
fiorenzo.franceschini@polito.it

1 Department of Management and Production Engineering (DIGEP), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy

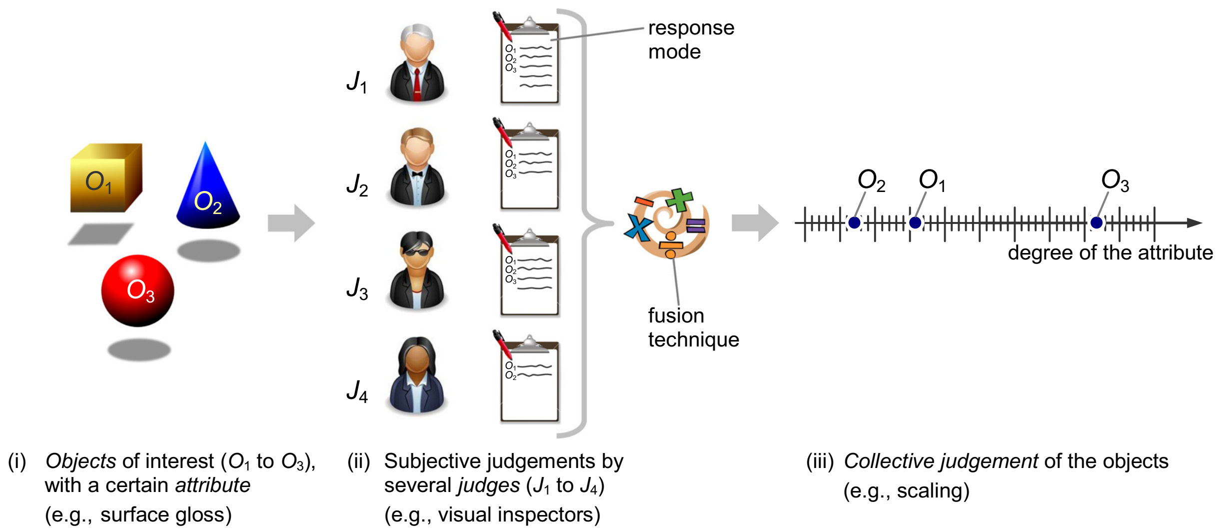


Fig. 1 Representation scheme of the problem of interest

1. the *response mode* for collecting the (input) judgments, e.g., expressed in the form of preference orderings, paired-comparison relationships, ratings, rankings, etc.;
2. the underlying *rationale* of the fusion technique, e.g., heuristic, mathematical/statistical, or fuzzy models (Çakır 2018);
3. the type of (output) *collective judgment*, e.g., expressed in the form of rankings or ordinal/interval/ratio scale values.

The simplest technique is probably that in which judges evaluate the objects using an ordinal response scale with a predetermined number of levels—e.g., five: *very low*, *low*, *intermediate*, *high*, and *very high* degree of the attribute. Then, for each object, the resulting scale levels are converted into conventional scores (e.g., 1, 2, 3, 4, 5 or − 2, − 1, 0, 1, 2, for five-level scales) and aggregated through central-tendency indicators.

In the field of *voting theory*, we recall the Borda's Count (Fishburn 1973), in which each judge ranks the objects and then their (Borda's) scores are calculated by summing the relevant rank positions. A collective ranking of the objects is finally constructed using these scores.

In the field of *multicriteria decision making*, we recall the Analytic Hierarchy Process (AHP) (Saaty 2008; Hosseini and Al Khaled 2016), in which judgments are expressed in the form of paired-comparison (ratio) relationships (e.g., "$O_i = 3 \cdot O_j$", which means that object O_i is three times more preferred than object O_j, regarding the attribute of interest) and, through a procedure based on the calculation of eigenvalues and eigenvectors, these judgments are aggregated into a vector of weights, associated with objects. Zeshui (2012) discusses other popular techniques concerned with qualitative criteria and linguistic variables.

For a thorough discussion of the more relevant existing techniques in the various scientific fields, we refer the reader to the vast literature and extensive reviews (DeVellis 2016).

Regardless of the peculiarities of the individual fusion techniques, a key element for their success is the simplicity of the response mode. Generally speaking, response modes that are relatively simple and easy to understand are more likely to be accepted and the relevant data collection process is more likely to be accurate and reliable (Franceschini et al. 2019). For example, various studies show that comparative judgments of objects (e.g., "O_i is more preferred than O_j") are generally easier than judgments in absolute terms (e.g., "the degree of the attribute of O_i is low/high") (Alwin and Krosnick 1985; Harzing et al. 2009). This requirement applies to any decision-making problem, including those in the manufacturing field, in which judges (e.g., engineers, technicians, potential customers, etc.) are generally not familiar with questionnaires characterized by sophisticated response modes.

As to the typology of collective judgments, we note that they are often treated as if they were defined on a *ratio* scale, even when they actually are not; e.g., rankings or ordinal-scale values of the objects are improperly "promoted" to ratio-scale values, in the moment in which they are combined with other indicators through weighted sums, geometric averages, or—more generally—statistics permissible to ratio-scale values (Stevens 1946; Roberts 1979). These promotions are potentially dangerous, as they can lead to significant distortions (Franceschini et al. 2019).

This article focuses on the Thurstone's *Law of Comparative Judgment* (LCJ) (Thurstone 1927), i.e., a consolidated and scientifically rigorous model that, starting from judgments expressed in the form of paired-comparison (ordinal) relationships of a set of objects, allows to construct an *interval* scaling of these objects. The LCJ is a milestone of psychometry and has stimulated the construction of numerous other models, such as that by Rasch (Andrich 1978; De Battisti et al. 2010). Despite its elegance and computational simplicity, the LCJ is not very popular in the field of manufacturing and quality engineering/management, probably because of two major limitations:

1. The response mode based on paired comparisons is inevitably tedious, especially when the number of objects tends to be high;
2. The output is defined on an *interval* scale, which is not as powerful as a *ratio* scale.

The goal of this contribution is to develop a new fusion technique, which can be easily applied to typical problems in the field of manufacturing and quality engineering/management. The proposed technique will be based on the combination of the canonical LCJ model with an ad hoc response mode based on preference orderings. Apart from the regular objects (O_1, O_2, …, O_n), these orderings will also include two *anchor* objects, to univocally identify the zero point and a conventional unit of the output scale. In this way, the output scale will (reasonably) be considered as a *ratio* one.

The remainder of this article is organized into six sections. "Introduction of the case study" section introduces a case study (concerning the design of an automatic pallet stretch-wrapping machine) which will accompany the theoretical description of the new fusion technique. "Thurstone's LCJ" section recalls the basic principles of the LCJ. "Description of the new technique" section illustrates the response mode based on preference orderings and the procedure (*ZM*-technique) to obtain a ratio scaling of the objects; the description is integrated with a practical application to the afore-introduced case study. "Generalizing the response mode" section deals with a possible extension of the proposed technique for problems in which (i) preference orderings may include omissions and/or incomparabilities between some objects, and (ii) judges are not necessarily equally important. "Conclusions" section summarizes the original contributions of this paper and its practical implications, limitations and suggestions for future research. Further information is contained in the appendix.

Introduction of the case study

The conceptual description of the new fusion technique will be accompanied by a practical application to a real-life case study, which is illustrated below. Suppose a company designs and manufactures automatic machines for stretch-wrapping pallets (see Fig. 2).

Fig. 2 Example of automatic machine for stretch-wrapping pallets

Four design concepts (O_1 to O_4, i.e., *objects*) of automatic machines have been generated by a team of designers during the conceptual design phase (see also the short description in Fig. 3):

(O_1) conveyorized turntable;
(O_2) non-conveyorized turntable;
(O_3) rotary ring;
(O_4) conveyorized rotary arm.

The objective is to evaluate the aforementioned design concepts in terms of *user friendliness*, i.e., a measure of the ease of use of a machine, which generally implies a certain level of automation and a good user interface. In addition, this *attribute* is very important to reduce training time and operator errors (Önüt et al. 2008). Some of the factors that can positively influence user friendliness are: (i) the ability of a machine to adapt to loads with different mass, stability or "profile", (ii) the ability to be integrated within various production lines, or (iii) the reduced set-up operations for the operator.

A collective judgment should be obtained by merging the individual (subjective) evaluations of five process engineers (J_1 to J_5, i.e., *judges*). The case study will be analysed after the description of the proposed fusion technique.

Thurstone's LCJ

This section recalls the LCJ, with special attention to the so-called *case V*, i.e., the most popular variant of this model.

Fig. 3 Schematic representation and short description of four design concepts of automatic stretch-wrapping machines

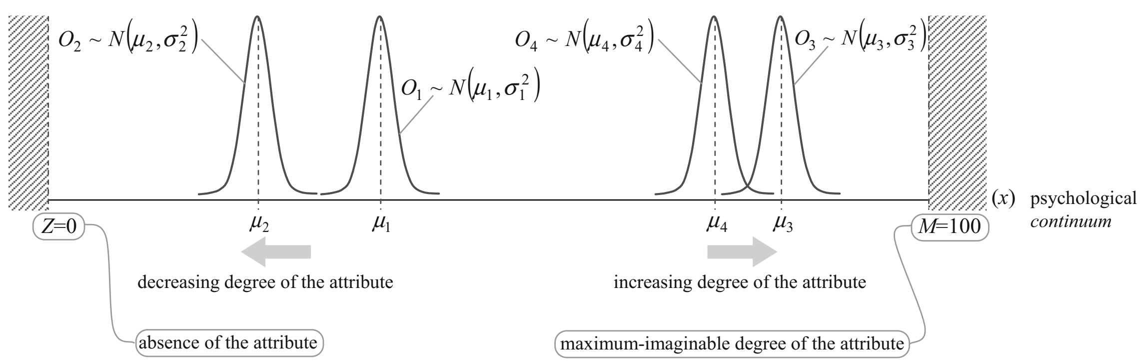


Fig. 4 Representation of the *psychological continuum* and the (unknown) positioning of four fictitious objects (O_1 to O_4), depending on the degree of the attribute of interest. For simplicity, the psychological continuum is assumed to be *single-ended*, i.e., the objects' attribute progresses in one direction only and is included in the range [*Z*, *M*]

Thurstone (1927) postulated the existence of a *psychological continuum*, that is to say an abstract and unknown unidimensional scale, in which objects are positioned depending on the degree of a certain *attribute*—i.e., a specific feature of the objects, which evokes a subjective response in each judge. The position of one object is directly proportional to the degree of the attribute, i.e., increasing to the right and decreasing to the left of the scale. For example, considering the case study introduced in "Introduction of the case study" section, Fig. 4 depicts the (supposed) positioning of the four design concepts of stretch-wrapping machines.

According to Thurstone, the position of a generic *i*-th object (O_i), will be distributed normally: $O_i \sim N(\mu_i, \sigma_i^2)$, where μ_i and σ_i^2 are the unknown mean value and variance of that object's attribute. This distribution has been postulated to reflect the intrinsic variation between judges in positioning (indirectly, as further explained) the objects in the continuum. In other words, different realizations of the same distribution reflect judgments by different judges.

Let us now make a short digression that will help to grasp the rationale of the proposed technique (in "Description of the new technique" section). The fact that the object positioning is proportional to the degree of the attribute implies that this scale has an (unknown) *absolute-zero* point (Z), that is, a point corresponding to the *absence* of the attribute; similarly, it can be assumed that the scale has another (unknown) point, corresponding to the *maximum-imaginable* degree (M) of the attribute.[1] Z and M would therefore delimit the range [Z, M]

[1] For simplicity, we consider *single-ended* psychological continua, in which the objects' attribute progresses in one direction only, starting from the absolute-zero point; this is reasonable when the attribute has a positive connotation exclusively (e.g., the importance of a set of product requirements) (Torgerson 1958).

that includes the objects to be evaluated[2] (see Fig. 4). Assigning conventional numerical values, such as $Z = 0$ and $M = 100$, one could obtain a *ratio* scale (y) with an absolute-zero point (Z) and a conventional unit ($M - Z$)/100. Of course, the scale is actually unknown, since the positions of Z and M in the psychological continuum are unknown too.

Returning to the LCJ, Thurstone and other authors assert that envisaging the psychological continuum and assigning the position of the objects directly/reliably would be very difficult for judges (Thurstone 1927; Harzing et al. 2009); on the other hand, judges would certainly find it easier to formulate comparative judgments of the objects. Let us consider an example that clarifies this concept: although a judge may find it difficult to guess the *surface gloss* of an object, based on him/her subjective physical perceptions, he/she will probably not encounter problems in comparing the surface glosses of two objects and formulate an ordinal relationship like "O_i is brighter/duller than O_j".

The LCJ (*case V*) includes the following additional postulates/assumptions (Thurstone 1927; Edwards 1957): (i) the objects' variances, which reflect the judge-to-judge variability, are all equal ($\sigma_i^2 = \sigma_j^2 = \cdots = \sigma^2$), and (ii) the inter-correlations (in the form of Pearson coefficients ρ_{ij}) between pairs of objects (O_i, O_j) are all equal too ($\rho_{ij} = \rho$, $\forall i, j$).

The application of the LCJ is based on five steps:

1. A set of (m) *judges* ($J_1, J_2, \ldots, J_m$) express their preferences[3] for each object (O_i) versus every other object (O_j), considering all possible $C_2^n = n \cdot (n-1)/2$ pairs, n being the total number of objects. Preferences are expressed through relations of *strict preference* (e.g., "$O_1 > O_2$" or "$O_1 < O_2$") or *indifference* (e.g., "$O_1 \sim O_2$") (Fishburn 1973). Results may then be aggregated into a frequency matrix ($\boldsymbol{F}$), whose general element f_{ij} represents the number of times that O_i was preferred to O_j (i.e., absolute frequency of the preference "$O_i > O_j$"). Precisely, for each judge who prefers O_i to O_j, the indicator $f_{ij} \in [0, m]$ is incremented by one unit (m being the total number of judges); if two objects are considered indifferent (i.e., "$O_i \sim O_j$"), f_{ij} and f_{ji} are both conventionally incremented by 0.5. In mathematical terms:

$$f_{ij} = |A| + 0.5 \cdot |B|, \tag{1}$$

where "| |" is the *cardinality* operator that corresponds to the number of elements of a set, and the sets $A \subseteq \{J_k$: "$O_i > O_j$"$\}$ and $B \subseteq \{J_k$: "$O_i \sim O_j$"$\}$. The complementarity relationship $f_{ij} = m - f_{ji}$ holds.
2. Next, the f_{ij} values are transformed into p_{ij} values, through the relationship:

$$p_{ij} = \frac{f_{ij}}{m}, \tag{2}$$

where p_{ij} represents the observed proportion of times that O_i was chosen over O_j. The p_{ij} values are aggregated into a proportion matrix ($\boldsymbol{P}$); the relationship of complementarity $p_{ij} = 1 - p_{ji}$ holds.
3. Next, p_{ij} values are transformed into z_{ij} values, through the relationship:

$$z_{ij} = \Phi^{-1}(1 - p_{ij}), \tag{3}$$

$\Phi()$ being the cumulative distribution function of the standard normal distribution. The element z_{ij} represents a unit normal deviate, which will be positive for all values of $(1 - p_{ij})$ over 0.50 and negative for all values of $(1 - p_{ij})$ under 0.50.
In general, objects are judged differently by judges. However, if all judges express the same preference for each outcome, the model is no more viable (p_{ij} values of 1.00 and 0.00 would correspond to z_{ij} values of $\pm\infty$). A simplified approach for tackling this problem is to associate values of $p_{ij} \le 0.023$ with $z_{ij} = \Phi^{-1}(1 - 0.023) = 1.995$ and values of $p_{ij} \ge 0.977$ with $z_{ij} = \Phi^{-1}(1 - 0.977) = -1.995$. More sophisticated solutions to deal with this issue have been proposed (Edwards 1957).
4. Next, the z_{ij} values related to the possible paired comparisons are reported into a matrix $\mathbf{Z}$. The element z_{ij} is reported in the i-th row and j-th column. The relationship $z_{ji} = -z_{ij}$ holds, being unit normal deviates related to complementary cumulative probabilities, i.e., $p_{ji} = (1 - p_{ij})$.
5. Next, the *scaling* can be performed by (i) summing the values into each column of the matrix $\mathbf{Z}$ and (ii) dividing these sums by n. It can be demonstrated that the result obtained for each column corresponds to the unknown average value (μ_j) of the object's attribute, up to a positive scale factor and an additive constant (Franceschini and Maisano 2015; Thurstone and Jones 1957; Edwards 1957). In other words, the LCJ results into an *interval* scaling, i.e., objects are defined on a scale with arbitrary zero point and meaningful unit (Stevens 1946; Roberts 1979). A practical application of the LCJ will be presented later on in "Application to the case study" section.

[2] This assumption is quite common for psychometric studies on subjective perceptions (Torgerson 1958; Lim 2011).

[3] According to the terminology introduced in Franceschini et al. (2007), the term "preference"—defined as *subjective* and *non-empirical* (i.e., which does not necessarily stem from a direct observation of reality) assignment of numbers/symbols to properties of objects—should be replaced with the term "evaluation"—defined as *subjective* and *empirical* (i.e., which stems from a direct observation of reality) assignment. Despite this, for the sake of simplicity the term "preference" will be hereafter used.

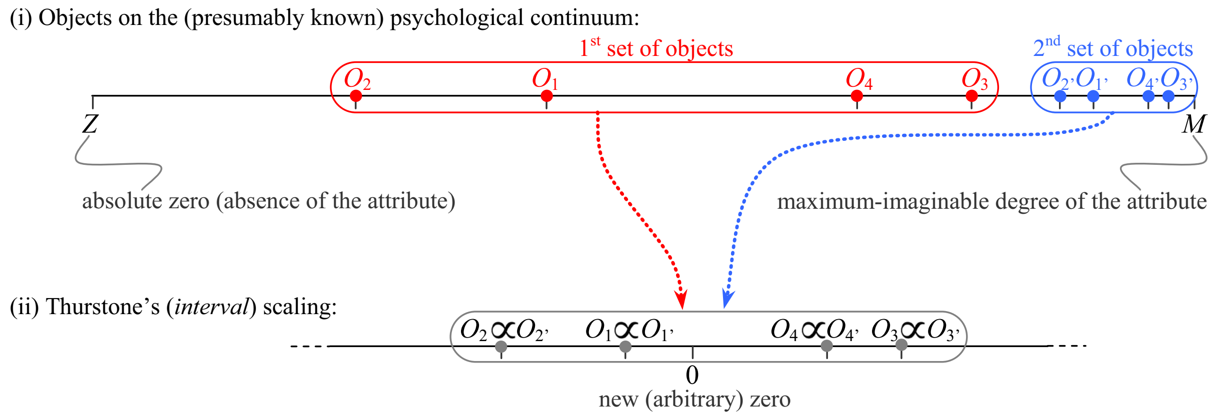


Fig. 5 Example of analogous Thurstone-scaling processes (up to a certain scale factor), obtained for two different sets of objects (O_1, to O_4). Regarding the first set, objects are spread over a (presumably known) psychological continuum, denoting wide variations in the degree of the attribute, while regarding the second set, objects are concentrated into a small portion of the psychological continuum, with generally large degrees of the attribute

A limitation of the LCJ is that, due to the arbitrary zero point and unit, the resulting (interval) scale is not "anchored" with respect to the (unknown) psychological continuum, in which objects are positioned. This limitation makes the results of different scaling processes incomparable. For the purpose of example, Fig. 5 shows an (apparently) paradoxical situation in which two considerably different displacements of four objects, in a (presumably known) psychological continuum, result in analogous scaling processes, up to a scale factor.

Another significant limitation of the LCJ is that the response mode can be laborious and tedious for judges, especially when the number of objects is large. These limitations will be (at least partly) overcome by the technique illustrated in the next section.

Description of the new technique

Introduction of preference orderings

A significant drawback of the response mode based on paired-comparison (ordinal) judgments is that it can be tedious and complex to manage, since much repetitious information is required from judges. E.g., since the number of objects in the case study is four, each judge would be required to make $C_2^4 = 6$ judgments. Although this quantity of judgments may seem practically sustainable, it tends to "explode" when increasing the number of objects.

Paired-comparison judgments can be obtained indirectly, through more practical response modes (Vasquez-Espinosa and Conners 1982; Franceschini and Maisano 2015). For example, judges can directly formulate preference orderings that are then turned into paired-comparison data (see the example in Fig. 6).

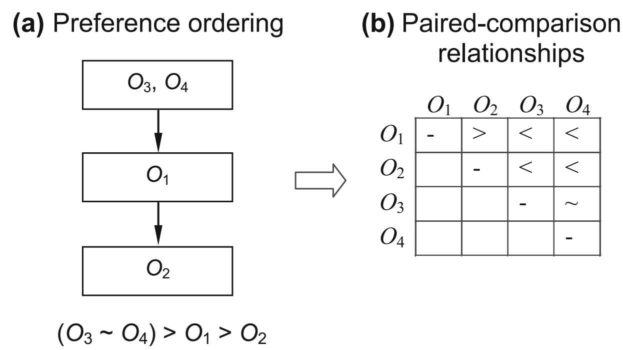


Fig. 6 Example of (indirect) determination of paired-comparison relationships from judgments expressed using a preference ordering

A practical way to do this is asking each judge to position some tags [even *immaterial* ones, through some software interface (Tarricone and Newhouse 2016)] in order of preference. The more preferred ones should be positioned in the top positions while the less preferred ones in the lower ones; those positioned at the same hierarchical level are considered indifferent (see the example Fig. 7).

In this way, each judge can construct a *linear* preference ordering, i.e., a hierarchical sequence of the objects of interest, linked by arrows depicting the *strict preference* (">") relationship. Two or more objects in the same hierarchical levels are linked by the *indifference* ("~") relationship. It can be seen than two generic objects are always comparable, since there exist a path from the first to the second one (or vice versa) that is directed downwards. The resulting number of hierarchical levels may change depending on the number of objects and their mutual relationships (Nederpelt and Kamareddine 2004). We also note that this response mode

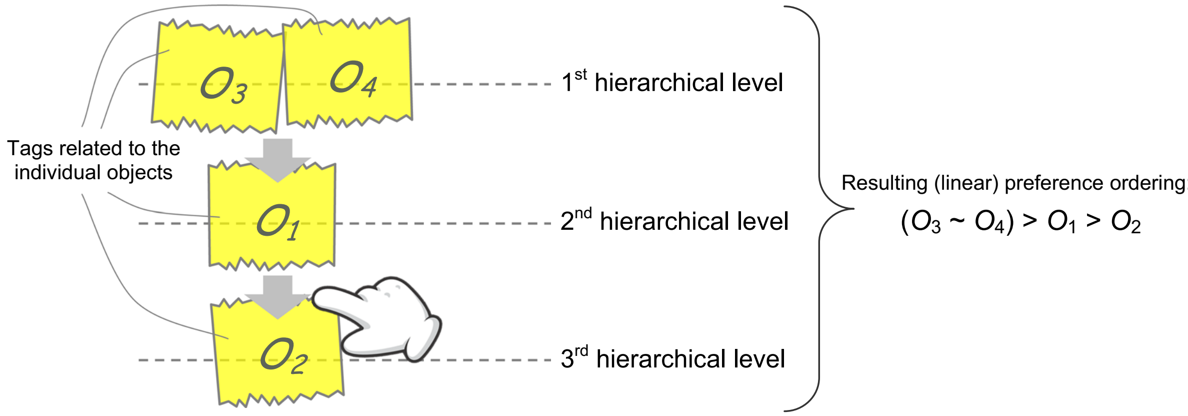


Fig. 7 Practical technique for supporting the construction of preference orderings, using tags

forces judges to be *transitive* (e.g., if "$O_1 > O_2$" and "$O_2 > O_3$", then "$O_1 > O_3$").

Even though the direct formulation of preference orderings may sometimes be less practical then the use of ordinal response scales (e.g., in the case of telephone or street interviews) (Alwin and Krosnick 1985), the fact remains that it is less prone to the following problems:

- Ordinal scales tend to be used subjectively, as there is no absolute reference shared by all judges. For example, let us consider the five-level ordinal scale: *very low*, *low*, *intermediate*, *high*, and *very high* degree of the attribute; "indulgent" judges will tend to assign higher levels of preference whereas "severe" judges will tend to assign lower ones. For this reason, it would be questionable to aggregate judgments by different judges through indicators of central tendency.
- The number of categories in the ordinal scale may conflict with the real discriminatory power of judges; e.g., the resolution of a five-level scale may represent a limitation for judges able to distinguish among a greater number of hierarchical levels, or it can be over-detailed for judges unable to distinguish among more than a few hierarchical levels.

Anchoring the Thurstone's Scaling: the *ZM*-technique

The scientific literature includes several techniques to anchor the LCJ's output scale, even though they inevitably complicate the response mode. For example, the technique proposed by Torgerson (1958) requires that each judge directly assigns the scale values of the objects, in a range included between two anchor points: (i) a (presumed) absolute-zero point (set to 0), corresponding to the absence of the attribute, and (ii) a point corresponding to the *maximum-imaginable* degree of the attribute, conventionally set to 5. While aware of the difficulty and potential unreliability of this direct-assignment process, Torgerson suggests to use it just for the purpose of anchoring the LCJ scale (Edwards 1957). An application example of this technique is reported in "Torgerson's anchoring" section (in the Appendix).

We have developed a new anchoring technique, denominated "*ZM*-technique", that is more consistent with the response mode based on preference orderings (in "Introduction of preference orderings" section). Our proposal is to apply the LCJ including two *dummy* or *anchor* objects among the regular ones:

(Z) a dummy/anchor object corresponding to the *absence* of the attribute of interest ("Z" stands for "zero");

(M) a dummy/anchor object corresponding to the *maximum-imaginable* degree of the attribute ("M" stands for "maximum-imaginable"), consistently with the current technological and socio-economic development.

Likewise the regular objects, Z and M are assumed to project a normal distribution on the psychological continuum, with unknown mean value and unknown variance, which is equal to that of the other objects (see "Thurstone's LCJ" section). The collection of judgments is then modified by considering the regular objects (O_1, O_2, etc.) and the dummy/anchor objects (Z and M). Each judge has to formulate a preference ordering of these objects, with two important requirements ("Questionnaire" section, in the Appendix, reports a possible questionnaire with user guidance):

1. Z should be positioned at the bottom of the preference ordering, i.e., there should not be any other object with preference lower than Z. In the case the attribute of another object is judged to be absent, that object will be considered indifferent to Z and positioned at the same hierarchical level.

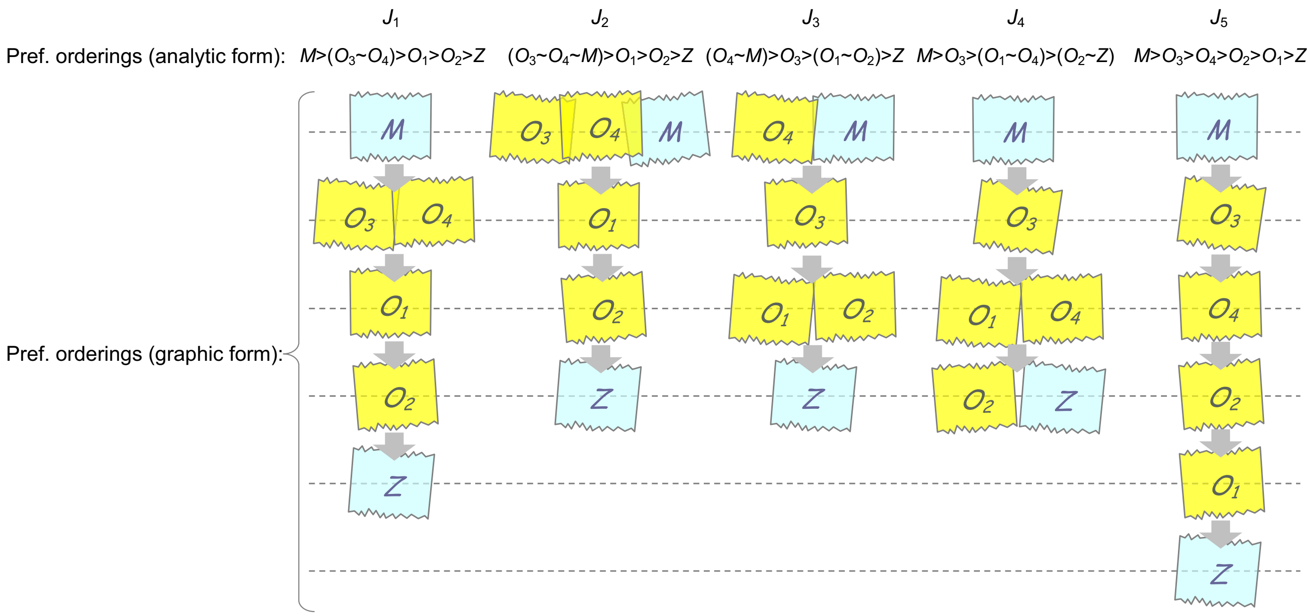


Fig. 8 Example of preference orderings formulated by five judges (i.e., J_1 to J_5), including four regular objects (O_1 to O_4) and two dummy/anchor objects (Z and M)

2. M should be positioned at the top of the preference orderings, i.e., there should not be any other object with preference higher than M. In the case the attribute of another object is judged to be the maximum-imaginable, that object will be considered indifferent to M and positioned at the same hierarchical level.

Next, the Thurstone's scaling is performed and the resulting (interval) scale is transformed into a new one, defined in the conventional range [0, 100], through the following linear transformation:

$$\frac{y-0}{100-0}=\frac{x-x_Z}{x_M-x_Z} \;\Rightarrow\; y=100\cdot\frac{x-x_Z}{x_M-x_Z}, \tag{4}$$

where x_Z and x_M are the scale values of Z and M, resulting from the LCJ; x is the scale value of a generic object, resulting from the LCJ; y is the relevant transformed scale value in the conventional range [0, 100].

It can be seen that the introduction of Z and M allows to anchor the LCJ scale into a new scale (y) with a conventional unit and a zero point (which corresponds to the absence of the attribute); it is therefore not unreasonable to consider y as a *ratio* scale. We remark that setting the value of M to 100 is a sort of normalization to make the scale unit comparable with those obtained from other LCJ processes. The adjective "comparable" means that the resulting scales have a common unit; e.g., let us assume that the LCJ is used to evaluate the courtesy of some call-center operators, according to the judgments of a sample of customers, and this evaluation is repeated annually: without proper normalization, the results obtained over several years would not be comparable.

Application to the case study

Returning to the case study in "Introduction of the case study" section, five process engineers (J_1 to J_5, i.e., judges) have to compare four design concepts (O_1 to O_4, i.e., objects) of automatic pallet stretch-wrapping machines, from the perspective of *user friendliness* (i.e., attribute).

Each judge formulates a preference ordering including the four (regular) objects and the two dummy/anchor objects: Z, i.e., a fictitious machine that is absolutely user-unfriendly (e.g., due to the difficult use by operators and high risk of error) and M, i.e., a fictitious machine that guarantees the maximum-imaginable user friendliness, in line with the current technological development. The resulting orderings are represented in Fig. 8.

It can be noticed that J_4 has positioned Z and O_2 at the bottom of the preference ordering (absence of the attribute). On the other hand, J_2 and J_3 have positioned M and other objects at the top of their orderings (maximum-imaginable degree of the attribute).

Given that the introduction of Z and M increases the information content of preference orderings, it may also cause some variation in the results. For example, the information that the degree of an attribute is zero or the maximum-imaginable one is richer than the information that it is just lower or higher than the remaining ones.

(a) Judgements, paired-comparison relationships and f_{ij}, p_{ij}, z_{ij} indicators

Paired comparison	J_1	J_2	J_3	J_4	J_5	f_{ij}	p_{ij}	z_{ij}
(O_1, O_2)	>	>	~	>	<	3.5	0.70	-0.524
(O_1, O_3)	<	<	<	<	<	0	0.00*	1.995
(O_1, O_4)	<	<	<	~	<	0.5	0.10	1.282
(O_2, O_3)	<	<	<	<	<	0	0.00*	1.995
(O_2, O_4)	<	<	<	<	<	0	0.00*	1.995
(O_3, O_4)	~	~	<	>	>	3	0.60	-0.253
(O_1, Z)	>	>	>	>	>	5	1.00*	-1.995
(O_2, Z)	>	>	>	~	>	4.5	0.90	-1.282
(O_3, Z)	>	>	>	>	>	5	1.00*	-1.995
(O_4, Z)	>	>	>	>	>	5	1.00*	-1.995
(O_1, M)	<	<	<	<	<	0	0.00*	1.995
(O_2, M)	<	<	<	<	<	0	0.00*	1.995
(O_3, M)	<	~	<	<	<	0.5	0.10	1.282
(O_4, M)	<	~	~	<	<	1	1.00*	-1.995

(b) matrix ***F***

	O_1	O_2	O_3	O_4	Z	M
O_1	2.5	3.5	0.0	0.5	5.0	0.0
O_2	1.5	2.5	0.0	0.0	4.5	0.0
O_3	5.0	5.0	2.5	3.0	5.0	0.5
O_4	4.5	5.0	2.0	2.5	5.0	1.0
Z	0.0	0.5	0.0	0.0	2.5	0.0
M	5	5	4.5	4	5	2.5

(c) matrix ***P***

	O_1	O_2	O_3	O_4	Z	M
O_1	0.50	0.70	0.00*	0.10	1.00*	0.00*
O_2	0.30	0.50	0.00*	0.00*	0.90	0.00*
O_3	1.00*	1.00*	0.50	0.60	1.00*	0.10
O_4	0.90	1.00*	0.40	0.50	1.00*	0.20
Z	0.00*	0.10	0.00*	0.00*	0.50	0.00*
M	1.00*	1.00*	0.90	0.80	1.00*	0.50

(d) matrix ***Z***

	O_1	O_2	O_3	O_4	Z	M
O_1	0.00	-0.52	2.00	1.28	-2.00	2.00
O_2	0.52	0.00	2.00	2.00	-1.28	2.00
O_3	-2.00	-2.00	0.00	-0.25	-2.00	1.28
O_4	-1.28	-2.00	0.25	0.00	-2.00	0.84
Z	2.00	1.28	2.00	2.00	0.00	2.00
M	-2.00	-2.00	-1.28	-0.84	-2.00	0.00
Σ_j	-2.75	-5.23	4.96	4.18	-9.26	8.11
$\mu_j' = \Sigma_j / n$	-0.46	-0.87	0.83	0.70	-1.54	1.35
μ_j'' [0,100]	**37.5**	**23.2**	**81.9**	**77.4**	**0**	**100**

Notes:
Z is a dummy/anchor object denoting the zero preference level;
M is a dummy/anchor object denoting the maximum-imaginable preference level;
n=6 is the total number of objects, including Z and M;
(*)values of $p_{ij} \leq 0.023$ and ≥ 0.977 have been conventionally associated with z_{ij} = 1.995 and -1.995 respectively;
f_{ij} denotes the number of times that O_i is preferred to O_j;
p_{ij} denotes the proportion of times that O_i is preferred to O_j;
$z_{ij} = \Phi^{-1}(1 - p_{ij})$;
μ_j' is the (interval) scale value of the j-th object, resulting from the LCJ;
μ_j'' is the μ_j' value transformed in the conventional range [0, 100] (transformation in Eq. 4).

Fig. 9 Example of LCJ application to the preference orderings in Fig. 8: **a** paired-comparison relationships, **b** matrix ***F***, **c** matrix ***P***, **d** matrix ***Z*** and resulting scaling

The price to pay for this information enrichment is the increased effort of judges, who should formulate slightly more complicated judgments: apart from considering the regular objects, they should also envisage the two dummy/anchor objects and their "absolute" meaning. We are aware that the concepts of *zero* (regarding Z) or *maximum-imaginable* degree of an attribute (regarding M) are inevitably blurred, as they may depend on the judges' experience, expectations and techno-economic context; for example, several engineers (judges) may have different expectations on the degree of user friendliness (attribute) guaranteed by "exemplary" machines (objects). This problem is common to many scale-anchoring techniques (Torgerson 1958; Paruolo et al. 2013).

The preference orderings are then translated into paired-comparison relationships and the LCJ is applied (see Fig. 9a). We have verified that the new anchoring technique provides results in line with those obtained from other existing techniques. For example, it can be seen that the results in Fig. 9 are strongly correlated with those obtained through the Torgerson's technique (see "Torgerson's anchoring" section, in the Appendix). Additionally, it was empirically observed that this correlation tends to increase for problems with a larger number of objects and/or judges.

Comparison with other techniques

To reveal the advantages of the proposed technique, it has been compared with two other techniques: namely the *Borda's Count* and the *Method of Single Stimuli*.

Borda's Count consists of two basic steps (Fishburn 1973): (i) each (i-th) object in the preference ordering of each (j-th judge) is associated with a score (k_{ij}) that corresponds to its rank position, and (ii) the scores obtained by each object are synthesized into the so-called Borda's score (B_i):

$$B_i = \sum_{j=1}^{m} k_{ij}, \tag{5}$$

Table 1 Application of the Borda's Count to the preference orderings in Fig. 8

Object	Rank position in the preference orderings					B_i	Resulting rank position
	J_1	J_2	J_3	J_4	J_5		
O_1	4	4	4	5	3	20	4th
O_2	5	5	4	4	5	23	5th
O_3	2	1	3	2	2	10	2nd
O_4	2	1	1	3	3	10	2nd
Z	6	6	6	6	5	29	6th
M	1	1	1	1	1	5	1st

The resulting rank position of objects is obtained by ordering the objects decreasingly with respect to their B_i values

m being the total number of judges. Of course, the more important objects are those with lower B_i values. The application of the Borda's Count to the scores in Fig. 8 leads to the results in Table 1.

The Method of Single Stimuli requires that each judge directly assigns the objects' scale values, with respect to two anchors: (1) a (presumed) absolute zero ($Z=0$), corresponding to the absence of the attribute, and (2) the maximum-imaginable degree of the attribute, conventionally set to 5 ($M=5$) (Torgerson 1958). Subsequently, judge assignments are aggregated—object by object—through a central tendency indicator, such as the mean value (g_j). For more information on this method, see "Torgerson's anchoring" section (in the Appendix).

In this specific case, the five judges are also asked to make the scale-value assignments for the six objects of interest (i.e., the four "regular" objects plus the two dummy ones). Table 3 (in the Appendix) shows these assignments and the results of the aggregation.

Table 2 contains the results of the comparison of the three techniques in terms of rank positions of the objects of interest.

Comparing the results of the three techniques, we notice that they are very close to each other, in many cases even coincident. This empirical example shows that the proposed technique is relatively robust, consistently with the notion of *convergent validity* (Trochim et al. 2016). Unfortunately, the absence of a "gold standard"—i.e. a "true" reference result related to a specific problem—does not allow to establish the superiority of one technique over the others. Nevertheless, we believe that the proposed technique has two conceptual advantages:

- The response mode of the proposed technique (i.e., preference orderings) is certainly simpler than that of the Method of Single Stimuli, which requires the direct assignment of the objects on a (presumed) *ratio* scale. This means that input data are likely to be more accurate and reliable (Torgerson 1958; Franceschini et al. 2019).
- The proposed technique results into a *ratio* scaling, without any undue scale "promotion" of input data (i.e., preference orderings, which exclusively contemplate relations of indifference and strict preference among objects). Borda's Count only provides a rank ordering of the set of objects, without any interval or ratio scaling.

Generalizing the response mode

To fit a relatively large amount of practical contexts, the previously proposed response mode can be adapted to more

Table 2 Results of the comparison of the proposed techniques with two other techniques (i.e., Borda's Count and method of single stimuli)

Object	(a) Proposed technique		(b) Borda's Count		(c) Method of single stimuli	
	μ_j''	Rank posit.	B_j	Rank posit.	g_j	Rank posit.
O_1	37.5	4th	20	4th	2.2	4th
O_2	23.2	5th	23	5th	1.2	5th
O_3	81.9	2nd	10	2nd	4	2nd
O_4	77.4	3rd	10	2nd	3.8	3rd
Z	0	6th	29	6th	0	6th
M	100	1st	5	1st	5	1st

For each (j-th) object, μ_j'' is the resulting ratio-scale value, according to the proposed technique, B_j is the so-called Borda's score, while g_j is the average value of direct assignments by judges

Table 3 Direct assignments of the scale values for six objects (i.e., O_1 to O_4, Z and M), by five judges (J_1 to J_5)

Object	Rank position in the preference orderings					g_j (mean)	Resulting rank position
	J_1	J_2	J_3	J_4	J_5		
O_1	2	4	2	2	1	2.2	4th
O_2	1	1	2	0	2	1.2	5th
O_3	4	5	3	4	4	4	2nd
O_4	4	5	5	2	3	3.8	3rd
Z	0	0	0	0	0	0	6th
M	5	5	5	5	5	5	1st

The rating scale in use is included between 0 (absence of the attribute) and 5 (maximum-imaginable degree) and has a unitary resolution

general cases, in which (i) preference orderings may include *omissions* and/or *incomparabilities* between some objects, and/or (ii) judges are not necessarily *equi-important*. These two cases will be treated separately in the following two subsections.

Partial orderings

Let us assume that the response mode admits orderings in which some objects are omitted and/or incomparable with each other. According to the Mathematics' *order theory*, this kind of ordering is classified as *partial* (Nederpelt and Kamareddine 2004) and can be diagrammed as a graph with branches, which determine different possible paths from the element(s) at the top to that one(s) at the bottom (see the example in Fig. 10). If two objects are not comparable, there exists no direct path from the first to the second one (or vice versa); e.g., in Fig. 10b, O_1 and O_5 are incomparable since they lie along two different paths.

From an operational point of view, the procedure described in "Description of the new technique" section would require a few adjustments:

- Each preference ordering is translated into paired-comparison relationships of strict preference ">", indifference "~", and incomparability "||". The latter relationship is applied in the case of paired comparisons of objects included in the ordering but on different paths, or paired comparisons where at least one of the objects is omitted from partial ordering.
- The indicator f_{ij}—which takes into account the strict-preference and indifference relationships between O_i and O_j—is associated with a new indicator m_{ij}, which counts the "usable" paired-comparison relationships, i.e., those in which the alternatives of interest are not incomparable.
- The p_{ij} indicator would therefore be calculated as:

$$p_{ij} = \frac{f_{ij}}{m_{ij}}. \quad (6)$$

Equation 2, in which judges formulate orderings with all usable paired-comparison relationships between O_i and O_j, represents a particular case of the more general case in Eq. 6. The p_{ij} values are then aggregated in the $\boldsymbol{P}$ matrix. If this matrix is *complete*—i.e., all the elements can be determined—the canonical LCJ is applied. If the matrix is *incomplete*—i.e., for some paired comparisons $m_{ij} = 0$ and therefore f_{ij}, p_{ij} and z_{ij} cannot be determined—the procedure is a bit more complicated. Scientific literature contains several techniques for applying the LCJ to incomplete $\boldsymbol{F}$, and therefore $\boldsymbol{P}$ and $\boldsymbol{Z}$, matrices; the most consolidated one is that proposed by Morrissey (Morrissey 1955) and later refined by Gulliksen (Gulliksen 1956; Westland et al. 2014).

Although the problem is technically solvable even in these cases of "incompleteness", the fact remains that the presence of many incomparabilities denotes the difficulty of judges in comparing several objects. This indication should not be ignored by questionnaire administrators.

Non equally important judges

In the canonical LCJ, judges are implicitly assumed to be equi-important, i.e., they contribute to the collective judgment equally. In the presence of an *importance hierarchy* of judges, expressed through a set of weights ($w_1, w_2, \ldots, w_m$) (Vora et al. 2014; Ngan et al. 2016), this hypothesis could be relaxed and the procedure described in "Description of the new technique" section would be changed as follows:

- Preference orderings are translated into paired-comparison relationships. Then, the f_{ij} indicators (previously defined in Eq. 1) are reformulated as:

$$f_{ij} = \sum_{k \in A} w_k + 0.5 \cdot \sum_{k \in B} w_k, \quad (7)$$

being $A \subseteq \{k$: "$O_i > O_j$", for $J_k\}$ and $B \subseteq \{k$: "$O_i \sim O_j$", for $J_k\}$.

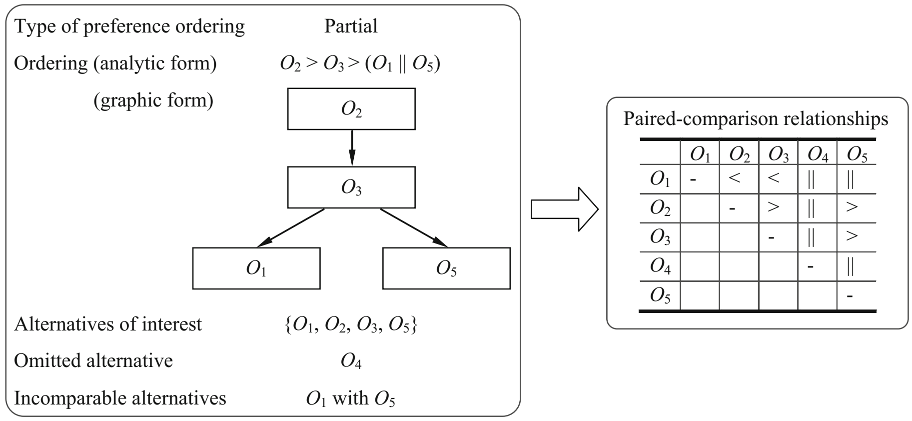


Fig. 10 Example of *partial* preference ordering and transformation into paired-comparison relationships. Symbols ">", "~" and "||" respectively mean "strictly preferred to", "indifferent to" and "incomparable to"

- The p_{ij} indicators (previously defined in Eq. 2) are reformulated as:

$$p_{ij} = \frac{f_{ij}}{\sum_{\forall k} w_k} = \frac{\sum_{k \in A} w_k + 0.5 \cdot \sum_{k \in B} w_k}{\sum_{\forall k} w_k}. \quad (8)$$

This is equivalent to adopting a weighted voting system in which judges do not necessarily have the same amount of influence. The p_{ij} value would therefore represent the portion of weighed votes for which O_i is preferred to O_j. The rest of the procedure would remain unchanged.

For the purpose of example, let us consider a variant of the case study in "Introduction of the case study" section, in which judges (i.e., the five process engineers) are weighted proportionally to their "professional experience" (e.g., number of years of activity), as this feature is supposed to influence the accuracy of the response significantly while being relatively easy to evaluate.

With further adjustments, the proposed procedure could also be adapted to problems in which the importance hierarchy of judges is expressed through a preference ordering (e.g., $J_1 > J_2 > (J_3 \sim J_4) > \cdots$) and not through a set of weights, as formalized in Franceschini et al. (2016).

Conclusions

This paper proposed a new technique to fuse multiple subjective judgments into a collective judgment that combines the canonical LCJ model with an ad hoc response mode based on preference orderings. Apart from the regular objects, these orderings will also include two dummy/anchor objects, to univocally identify the zero and the maximum-possible value of the LCJ scale. This allows to represent the objects on a conventional ratio scale, included between 0 and 100, without any conceptually-prohibited promotion. In addition, the response mode based on preference orderings is relatively practical and user-friendly, and could help diffuse the LCJ even in contexts where it is not commonly used, such as manufacturing and quality engineering/management.

The technique is largely automatable and, with a few adjustments, can be applied to problems in which (i) preference orderings may include omissions and/or incomparabilities between some objects, and (ii) judges are not necessarily equally important.

The proposed technique has some limitations:

- The problem was addressed by considering a *single-ended* psychological continuum, where the attribute's degree grows from an absolute-zero point to a point with maximum-imaginable degree. With some changes, the technique could be adapted to *double-ended* continua.
- The introduction of the two anchor objects (Z and M) into the preference orderings requires an additional effort of imagination for judges.

Regarding the future, we plan to apply the proposed technique to more structured real-life problems in the field of manufacturing and quality engineering/management [e.g. within the context of *Quality Function Deployment* (Chen et al. 2017) and/or that of *Failure Mode, Effects and Criticality Analysis* (Qazi et al. 2017)]. Besides, we plan to further simplify the response mode, assuming that judges do not have to formulate complete orderings but just (partial) orderings that include the more/less preferred objects (e.g. the top-three or bottom-three ones). In fact, it has been observed that judges tend to focus on the more/less preferred objects, providing more reliable judgments about them, to the detriment of the remaining objects (Harzing et al. 2009; Franceschini and Maisano 2018). Finally, we plan to develop a statistically-sound procedure to estimate the uncertainty related to the solution (i.e., the scaling of objects).

(a) Judgements, paired-comparison relationships and f_{ij}, p_{ij}, z_{ij} values

Paired comparison	J_1	J_2	J_3	J_4	J_5	f_{ij}	p_{ij}	z_{ij}
(O_1, O_2)	>	>	~	>	<	3.5	0.70	-0.524
(O_1, O_3)	<	<	<	<	<	0	0.00*	1.995
(O_1, O_4)	<	<	<	~	<	0.5	0.10	1.282
(O_2, O_3)	<	<	<	<	<	0	0.00*	1.995
(O_2, O_4)	<	<	<	<	<	0	0.00*	1.995
(O_3, O_4)	~	~	<	>	>	3	0.60	-0.253

(b) matrix ***F***

	O_1	O_2	O_3	O_4
O_1	2.5	3.5	0	0.5
O_2	1.5	2.5	0	0
O_3	5	5	2.5	3
O_4	4.5	5	2	2.5

(c) matrix ***P***

	O_1	O_2	O_3	O_4
O_1	0.50	0.70	0.00*	0.10
O_2	0.30	0.50	0.00*	0.00*
O_3	1.00*	1.00*	0.50	0.60
O_4	0.90	1.00*	0.40	0.50

(d) matrix ***Z***

	O_1	O_2	O_3	O_4
O_1	0.000	-0.524	1.995	1.282
O_2	0.524	0.000	1.995	1.995
O_3	-1.995	-1.995	0.000	-0.253
O_4	-1.282	-1.995	0.253	0.000
Σ_j	-2.753	-4.515	4.244	3.024
$\mu_j' = \Sigma_j / n$	**-0.688**	**-1.129**	**1.061**	**0.756**

Notes:
f_{ij} denotes the number of times that O_i is preferred to O_j;
p_{ij} denotes the proportion of times that O_i is preferred to O_j;
$z_{ij} = \Phi^{-1}(1 - p_{ij})$;
(*) for $p_{ij} \leq 0.023$, z_{ij} is conventionally set to 1.995, while for $p_{ij} \geq 99.977$, it is set to -1.995;
μ_j' is the (interval) scale value of the j-th object, resulting from the LCJ.

Fig. 11 Example of application of the LCJ, considering the paired-comparison relationships by five judges (J_1 to J_5) on four objects (O_1 to O_4). These relationships are identical to those in the example in Fig. 9, except that those with at least one of the dummy/anchor objects are not present

Appendix

A.1 Torgerson's anchoring

This section exemplifies the anchoring technique by Torgerson (1958, p. 196), applying it to the LCJ scaling in Fig. 11. Focussing on this scaling process, it can be seen that the (input) paired-comparison relationships are identical to those in the example in Fig. 8, except that those with at least one of the dummy/anchor objects are not present. The resulting (non-anchored) scale is reported in Fig. 11d.

The rationale of the Torgerson's anchoring is that results of the LCJ are (at least roughly) correlated with those resulting from the so-called *Method of Single Stimuli*, in which each judge directly assigns the objects' scale values, with respect to two anchors: (1) a (presumed) absolute zero, corresponding to the absence of the attribute, and (2) the *maximum-imaginable* degree of the attribute, conventionally set to 5. While aware of the difficulty and potential roughness of these direct assignments, Torgerson (1958, p. 196) suggests their use just for the purpose of anchoring the LCJ scale.

Subsequently, judge assignments are aggregated—object by object—through a central tendency indicator, such as the mean or median value (g), and plot against the scale values (x) computed from the LCJ. Then, a straight line to the points is fitted and the intercept on the horizontal axis ($g=0$) is taken as estimate of the position of the absolute-zero point (Z) and that on the horizontal line ($g=5$) as estimate of the position of the point with maximum-imaginable degree (M) of the attribute.

Considering the example in Fig. 11, we hypothesize that the five judges directly assign the objects' scale values on a rating scale from 0 to 5, with unitary resolution; the zero point corresponds to the absence while the maximum value (i.e., 5) corresponds to the maximum-imaginable degree of the attribute. Table 3 collects these assignments.

Assignments are then aggregated using the *arithmetic mean*. The graph in Fig. 12 plots the resulting mean values (g) against the scale values (x) obtained through the LCJ (see Fig. 11). Then, a straight tendency line is fitted (through a linear least-squares regression) and the intersection of this line with the horizontal axis ($g=0$) determines an estimates of the absolute-zero point (Z, i.e., first anchor), while that

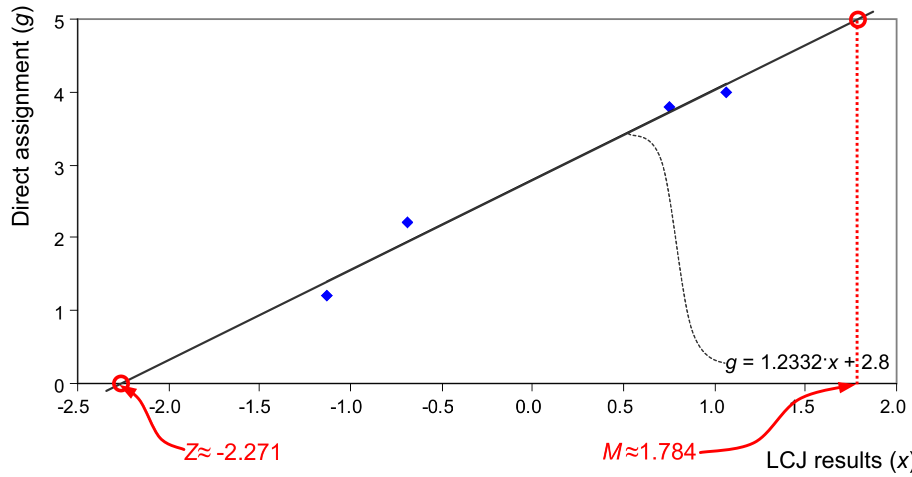
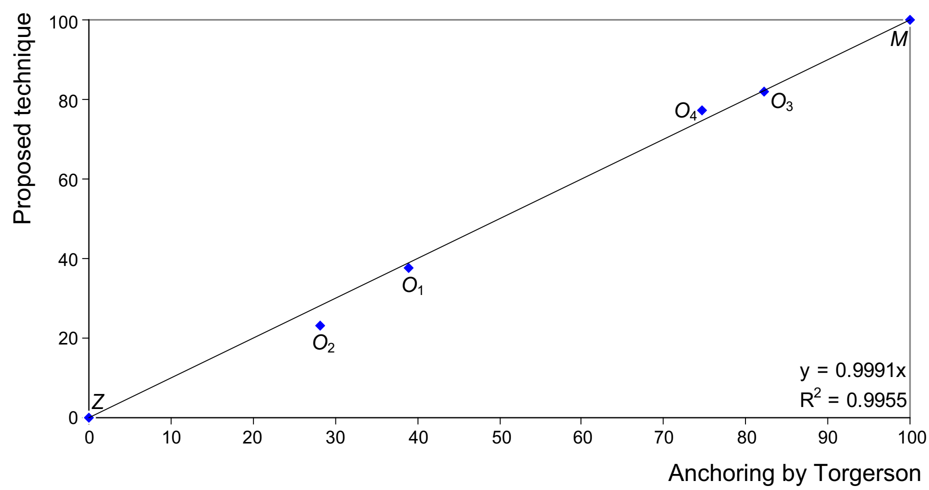


Fig. 12 Comparison of the scale values resulting from the Thurstone's LCJ and those resulting from a direct scale-value assignment (*Method of Single Stimuli*) for four objects (O_1 to O_4)

Table 4 Anchoring of the LCJ scale (in Fig. 11d), applying the technique by Torgerson

	O_1	O_2	O_3	O_4	Z	M
Results of the LCJ	− 0.688	− 1.129	1.061	0.756		
Anchor values					− 2.271	1.784
Scale values transformed into [0, 100]	39.0	28.2	82.2	74.6	0	100



	O_1	O_2	O_3	O_4	Z	M
Anchoring by Torgerson	39.0	28.2	82.2	74.6	0	100
Proposed technique	37.5	23.2	81.9	77.4	0	100

Fig. 13 Comparison between two anchoring techniques (i.e., that by Torgerson, exemplified in Table 4, and the proposed technique, exemplified in Fig. 9), with reference to the same LCJ-scaling problem

with the horizontal line $g = 5$ determines an estimate of the point (M, i.e., second anchor) of the maximum-imaginable degree of the attribute on the Thurstone's scale. Next, the LCJ scale values are normalized in the conventional range [0, 100], through the linear transformation in Eq. 4. This scale can reasonably be considered as a ratio one (see Table 4).

We have verified that the new anchoring technique (presented in "Anchoring the Thurstone's Scaling: the *ZM*-technique" section) provides results in line with those obtained from the Torgerson's technique. e.g., Fig. 13 shows that these two anchoring techniques, when applied to the same scaling problem, are strongly correlated. Also, we have empirically observed that the correlation tends to increase for problems with a larger number of objects and/or judges.

A.2 Questionnaire

Figure 14 reports an example of questionnaire to guide the construction of preference orderings.

Questionnaire

Instructions for Judges

- A ***preference ordering*** is an ordered sequence of ***objects*** (O_1, O_2, …), depending on the degree of preference of a certain ***attribute***.
- The judge has to position the objects, depending on the degree of preference of their attributes: most preferred objects at the top and least preferred at the bottom of the sequence.
- Two are the possible *relationships* between each pair of objects:
  1. *strict preference,* e.g., "O_1 is preferred to O_2", then O_1 is positioned at a higher hierarchical level than O_2;
  2. *indifference*, e.g., "O_1 has the same preference level of O_2", then the two objects are positioned at the same hierarchical level.
- The number of hierarchical levels is not fixed in advance, since it may depend on the number of objects and their mutual relationships.
- Apart from the *regular* objects (O_1, O_2, …), the judge has to include two *dummy* objects in his/her preference ordering:

  Z object with a *zero* degree of preference of the attribute;

  M, object with a *maximum-possible* degree of preference of the attribute.

  Regular objects with zero-preference degree should be positioned at the same hierarchical level of *Z* (indifference relationship) but **never below**, while objects with maximum-possible preference degree should be positioned at the same hierarchical level of *M*, but **never above**.

Example

Two judges (J_1, J_2) construct their preference orderings on the **aesthetics** (i.e, the *attribute* of interest) of four **car models** (i.e, the *objects* of interest, O_1, O_2, O_3 and O_4).

As regards J_1, O_4 is preferred to O_3 and, in turn, to O_1 and to O_2; since O_4 reaches the maximum-possible degree of preference, it is considered indifferent to *M*.

As regards J_2, O_3 is preferred to O_1 and O_4 (tied), which are, in turn, preferred to O_2; since O_2 has a zero preference degree, it is considered indifferent to *Z*.

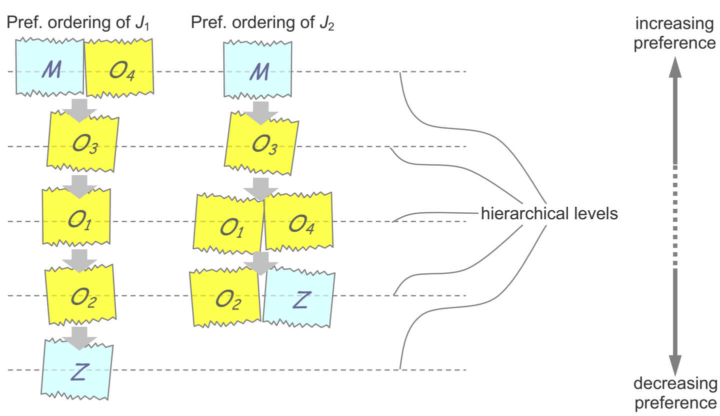


Fig. 14 Example of questionnaire with guidance for the formulation of preference orderings

References

Alwin, D. F., & Krosnick, J. A. (1985). The measurement of values in surveys: A comparison of ratings and rankings. *Public Opinion Quarterly, 49*(4), 535–552.

Andrich, D. (1978). Relationships between the Thurstone and Rasch approaches to item scaling. *Applied Psychological Measurement, 2*(3), 451–462.

Çakır, S. (2018). An integrated approach to machine selection problem using fuzzy SMART-fuzzy weighted axiomatic design. *Journal of Intelligent Manufacturing, 29*(7), 1433–1445.

Chen, L. H., Ko, W. C., & Yeh, F. T. (2017). Approach based on fuzzy goal programing and quality function deployment for new product planning. *European Journal of Operational Research, 259*(2), 654–663.

De Battisti, F., Nicolini, G., & Salini, S. (2010). The Rasch model in customer satisfaction survey data. *Quality Technology & Quantitative Management, 7*(1), 15–34.

Den Ouden, E., Yuan, L., Sonnemans, P. J., & Brombacher, A. C. (2006). Quality and reliability problems from a consumer's perspective: An increasing problem overlooked by businesses? *Quality and Reliability Engineering International, 22*(7), 821–838.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (4th ed.). London: Sage.

Edwards, A. L. (1957). *Techniques of attitude scale construction*. New York: Irvington Publishers.

Fishburn, P. C. (1973). *The theory of social choice*. Princeton: Princeton University Press.

Franceschini, F., Galetto, M., & Maisano, D. (2007). *Management by measurement: Designing key indicators and performance measurement systems*. Berlin: Springer.

Franceschini, F., Galetto, M., & Maisano, D. (2019). *Designing performance measurement systems: Theory and practice of key performance indicators*., Management for professionals Berlin: Springer.

Franceschini, F., & Maisano, D. (2015). Prioritization of QFD customer requirements based on the law of comparative judgments. *Quality Engineering, 27*(4), 437–449.

Franceschini, F., & Maisano, D. (2018). Fusion of partial orderings for decision problems in quality management. In *Proceedings of the 3rd international conference on quality engineering and management (ICQEM 2018)*, July 11–13, 2018, Barcelona (Spain).

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016). A new proposal for fusing individual preference orderings by rank-ordered agents: A generalization of the Yager's algorithm. *European Journal of Operational Research, 249*(1), 209–223.

Gulliksen, H. (1956). A least squares solution for paired comparisons with incomplete data. *Psychometrika, 21,* 125–134.

Harzing, A. W., Baldueza, J., Barner-Rasmussen, W., Barzantny, C., Canabal, A., Davila, A., et al. (2009). Rating versus ranking: What is the best way to reduce response and language bias in cross-national research? *International Business Review, 18*(4), 417–432.

Hosseini, S., & Al Khaled, A. (2016). A hybrid ensemble and AHP approach for resilient supplier selection. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-016-1241-y.

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value trade-offs*. Cambridge: Cambridge University Press.

Krynicki, J. C. (2006). Introduction to soft metrology. In *XVIII IMEKO World Congress*, 17–22 September, Rio de Janeiro (Brazil).

Lim, J. (2011). Hedonic scaling: A review of methods and theory. *Food Quality and Preference, 22*(8), 733–747.

Lin, C. J., & Cheng, L. Y. (2017). Product attributes and user experience design: How to convey product information through user-centered service. *Journal of Intelligent Manufacturing, 28*(7), 1743–1754.

Maier, J. R. A., & Fadel, G. M. (2007). A taxonomy and decision support for the design and manufacture of types of product families. *Journal of Intelligent Manufacturing, 18*(1), 31–45.

Morrissey, J. H. (1955). New method for the assignment of psychometric scale values from incomplete paired comparisons. *JOSA, 45*(5), 373–378.

Nederpelt, R., & Kamareddine, F. (2004). *Logical reasoning: A first course*. London: King's College Publications.

Ngan, T. T., Tuan, T. M., Son, L. H., Minh, N. H., & Dey, N. (2016). Decision making based on fuzzy aggregation operators for medical diagnosis from dental X-ray images. *Journal of Medical Systems, 40*(12), 280.

Önüt, S., Kara, S. S., & Efendigil, T. (2008). A hybrid fuzzy MCDM approach to machine tool selection. *Journal of Intelligent Manufacturing, 19*(4), 443–453.

Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: Voodoo or science? *Journal of the Royal Statistical Society: Series A (Statistics in Society), 176*(3), 609–634.

Qazi, A., Quigley, J., Dickson, A., & Ekici, Ş. Ö. (2017). Exploring dependency based probabilistic supply chain risk measures for prioritising interdependent risks and strategies. *European Journal of Operational Research, 259*(1), 189–204.

Roberts, F. S. (1979). *Measurement theory: With applications to decisionmaking, utility, and the social sciences* (Vol. 7)., Encyclopedia of mathematics and its applications Reading, MA: Addison-Wesley.

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences, 1*(1), 83–98.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103,* 2684.

Tao, F., Zhang, L., & Laili, Y. (2016). *Configurable intelligent optimization algorithm*. Berlin: Springer. ISBN 9783319088396.

Tarricone, P., & Newhouse, C. P. (2016). Using comparative judgement and online technologies in the assessment and measurement of creative performance and capability. *International Journal of Educational Technology in Higher Education, 13*(1), 16.

Thurstone, L. L. (1927). A law of comparative judgments. *Psychological Review, 34*(4), 273.

Thurstone, L. L., & Jones, L. V. (1957). The rational origin for measuring subjective values. *Journal of the American Statistical Association, 52*(280), 458–471.

Torgerson, W. S. (1958). *Theory and methods of scaling*. Oxford: Wiley.

Trochim, W., Donnelly, J. P., & Arora, K. (2016). *Research methods: The essential knowledge base* (2nd ed.). Boston: Cengage. ISBN 9781133954774.

Van Kleef, E., Van Trijp, H. C., & Luning, P. (2005). Consumer research in the early stages of new product development: A critical review of methods and techniques. *Food Quality and Preference, 16*(3), 181–201.

Vasquez-Espinosa, R. E., & Conners, R. W. (1982). *The law of comparative judgment: Theory and implementation* (No. RSIP/TR-403.82). Louisiana State, University of Baton Rouge (Remote Sensing and Image Processing Lab).

Vora, A., Paunwala, C. N., & Paunwala, M. (2014). Improved weight assignment approach for multimodal fusion. In *Proceedings of the 2014 international conference on circuits, systems, communication and information technology applications (CSCITA 2014)* (pp. 70–74).

Westland, S., Li, Y., & Cheung, V. (2014). Monte Carlo analysis of incomplete paired-comparison experiments. *Journal of Imaging Science and Technology, 58*(5), 50506.1–50506.6.

Zeshui, X. (2012). *Linguistic decision making: Theory and methods*. Berlin: Springer. ISBN 9783642294402.

Zheng, P., Xu, X., & Xie, S. Q. (2016). A weighted interval rough number based method to determine relative importance ratings of customer requirements in QFD product planning. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-016-1224-z.