Design decisions: concordance of designers and effects of the Arrow's theorem on the collective preference ranking

Design decisions: concordance of designers and effects of the Arrow's theorem on the collective preference ranking

Fiorenzo Franceschini[1] and Domenico Maisano[2]

[1] fiorenzo.franceschini@polito.it [2] domenico.maisano@polito.it
Politecnico di Torino, DIGEP (Department of Management and Production Engineering),
Corso Duca degli Abruzzi 24, 10129, Torino (Italy)

Abstract

The problem of collective decision by design teams has received considerable attention in the scientific literature of engineering design. A much debated problem is that in which multiple designers formulate their individual preference rankings of different design alternatives and these rankings should be aggregated into a collective one.

This paper focuses the attention on three basic research questions: (i) "How can the degree of concordance of designer rankings be measured?", (ii) "For a given set of designer rankings, which aggregation model provides the most coherent solution?", and (iii) "To what extent is the collective ranking influenced by the aggregation model in use?".

The aim of this paper is to present a novel approach that addresses the above questions in a relatively simple and agile way. A detailed description of the methodology is supported by a practical application to a real-life case study.

Keywords: Engineering design, collective decision-making, design teams, preference ranking, collective ranking, degree of concordance.

1. Introduction

A general problem, which may concern many design contexts, is to aggregate multiple rankings of different design alternatives into a collective (social) one. This problem may concern the early design stages, in which designers often have conflicting opinions [Fu, Cagan, Kotovsky 2010; Frey et al. 2009; Hoyle, Chen 2011; Keeney 2009; Weingart et al. 2005].

Let us consider the case in which m designers (or, more in general, *decision-making agents*: D_1 to D_m) formulate their corresponding m preference rankings among n design alternatives (or, more in general, *objects*: O_1, O_2, O_3, etc.). It is assumed that the preference rankings of designers are *complete*; i.e. all designers are able to rank all the alternatives of interest, without omitting any of them.

Each ranking can be decomposed into paired-comparison statements like $O_1 > O_2$, $O_1 \sim O_2$, $O_2 > O_1$, where symbols ">" and "~" respectively mean "strictly preferred to" and "indifferent to".

Additionally, designer rankings – which will hereafter also be referred to as "preference profiles" – may be more or less *concordant* with each other.

The collective ranking is supposed to reflect the *m* rankings as much as possible, even in the presence of diverging preferences. For this reason, it can also be defined as *social*, *consensus* or *compromise* ranking [Cook 2006; Herrera-Viedma et al., 2014; Franceschini, Maisano, Mastrogiacomo 2016].

The aggregation problem of interest may also take into account the importance of designers, which could not necessarily be equal for all of them. However, for the purpose of simplicity, in the rest of the paper designers will be assumed to be equally important.

Decision-making problems based on rankings are very diffused in the scientific literature for two reasons: (i) rankings are probably the most intuitive and effective way to represent preference judgments of design alternatives, and (ii) rankings do not require a common reference scale – neither numeric, linguistic or ordinal – to be shared by designers [Yager 2001; Chen et al., 2012].

Many approaches have been proposed in the literature to address the problem of aggregating rankings by multiple decision-making agents [Arrow 2012, Rayanaud 1986; Franssen 2005; Cook 2006; Hazelrigg 1999; Frey et al., 2010; Katsikopoulos 2009; Ladha et al. 2003; Reich 2010]. In general, different aggregation models may lead to different collective rankings of the alternatives under consideration (social choice) [Arrow and Rayanaud 1986; McComb et al., 2017]. The fact remains that it is not easy to identify the model that provides the collective ranking that best reflects the *m*-rankings. This paper will try to address this problem.

In recent years, a long and passionate debate on the applicability and effects of the Arrow's theorem on design decisions has characterized the engineering design literature [Arrow 2012, Rayanaud 1986; Reich 2010; Hazelrigg 1996, 1999, 2010; Scott and Antonsson 1999; Franssen 2005; Yeo et al. 2004; McComb et al. 2017]. In a nutshell, this theorem establishes the impossibility of a generic aggregation model to provide a solution (i.e., a collective ranking) that is always fair[1].

In line with the hypotheses of McComb et al. (2017), this paper is concerned with team decisions in the early design stages. Since design outlines are not yet well defined at this stage, the debate on the most relevant design criteria and methods to quantify/prioritize them is potentially very intense [Weingart 2005; Kaldate 2006; McComb et al. 2015; 2017]. Although there is a substantial agreement on the criteria that may condition the design process, the selection of design alternatives is generally

[1] This concept will be clarified later.

driven by the different personal experience of designers [Dwarakanath and Wallace 1995].

It is then necessary to adopt appropriate aggregation models to "transform" the decisions of individual designers into a collective design decision. In this scenario, the Theorem of Arrow finds grounds for its application, with some important practical effects [Hazelrigg 1999; Saari and Sieberg 2004; Franssen 2005; Jacobs et al. 2014]. The first effect – as the theorem itself shows – is that there is no aggregation model that can always satisfy several properties, also known as *fairness* criteria [Arrow 2012, Fishburn, 1973a, Nisan et al., 2007, Saari 2011]:

(i) *Unrestricted domain (universality).* The aggregation model is defined for problems characterized by any number of decision-making agents, any number of alternatives and any composition of preference rankings over alternatives.
(ii) *Non-dictatorship*. The aggregation model does not simply return the very same preference ranking of a specific decision-making agent (i.e., the so-called *dictator*).
(iii) *Independence of irrelevant alternatives* (IIA). The collective preference between alternatives x and y must depend solely on the relative preferences between x and y in the individual rankings.
(iv) *(Weak) monotonicity*. If any decision-making agent modifies his or her preference ranking by promoting a certain alternative, then the collective preference ranking should respond only by promoting the same alternative or remaining unchanged.
(v) *Pareto efficiency (unanimity).* If all decision-making agents prefer x to y, then the collective ranking must also prefer x to y.

The second effect of the Arrow's theorem is that one cannot establish *ex ante* the aggregation model that most consistently reflects the preference rankings by decision-making agents. Several quantitative attempts of measuring this *coherence* (or consistency) are discussed in [Chiclana 2002, Franceschini and Maisano 2015a; 2017]. We also recall McComb et al. (2017), who hypothesized a relationship between the concept of aggregation-model *fairness* and that of *implicit agreement*. The most coherent aggregation model can generally be identified *ex-post*, since it may depend on the preference profile.

The third effect of the Arrow's theorem is that it is very difficult to measure the impact of the aggregation model on the collective ranking, for a problem characterized by a certain preference profile [Saari, Decision and Elections, 2001 page.13]: "…*the winner of an election may more accurately reflect the choice of a decision procedure rather than the view and preferences of voters*".

The scientific literature includes a wide number of aggregation models, which have been analyzed extensively from the perspective of different axioms and properties [Arrow, 2012; Fishburn, 1973;

Cook, 2006; Saary 2011; Nurmi 2012]. Some researches demonstrate the effectiveness of specific aggregation models, even though they do not satisfy some of the Arrow's fairness criteria. For instance, Dym, Wood and Scott (2002) showed that, although the Borda aggregation model may not satisfy the IIA condition, this rarely affects the most preferred alternatives. In other words, Arrow's theorem poses a considerable theoretical problem, but the practical implications are often not so worrisome.

Another research by See and Lewis (2006) proposes a structured approach to avoid severe theoretical conflicts. The paper by Katsikopoulos (2009) expressed the need for greater clarity in the discussion of design decision models; a distinction between the concepts of coherence and correspondence is necessary to structure design activities. *Coherence* is used to characterize the internal consistency of a model, while *correspondence* refers to the representation of the external performance. Additionally, Katsikopoulos (2009) suggested that the arguments by Scott and Antonsson (1999) are rooted in the framework of correspondence, while the arguments introduced by Franssen (2005) in that of coherence. Jacobs, van de Poel and Osseweijer (2014) recognized several additional issues on the concepts of uncertainty, comparability and measurability concerned with aggregation models.

Considering the above implications of the Arrow's theorem, the choice of the aggregation model depends on the specific objective(s) of the decision-making team [Dong et al. 2004; Cagan and Vogel 2012]. Selecting an adequate aggregation model is not an easy task, since this choice can importantly affect the collective ranking [Li et al. 2007; Paulus et al. 2011; Franceschini et al. 2019]. For the purpose of example, possible selection strategies could be:

(i) designers select only the most preferred alternative;
(ii) designers select the two most preferred alternatives;
(iii) designers reject the least preferred alternative(s);
(iv) …

In this paper, we try to face three basic research questions:

- “How can the degree of concordance of designer rankings be measured?”;
- “For a given set of designer rankings, which aggregation model provides the most coherent solution[2]?”;
- “To what extent, is the collective ranking influenced by the aggregation model in use?

The remainder of this article is organized into six sections. Sect. 2 introduces a case study concerning

[2] I.e., the solution that best reflects designer rankings.

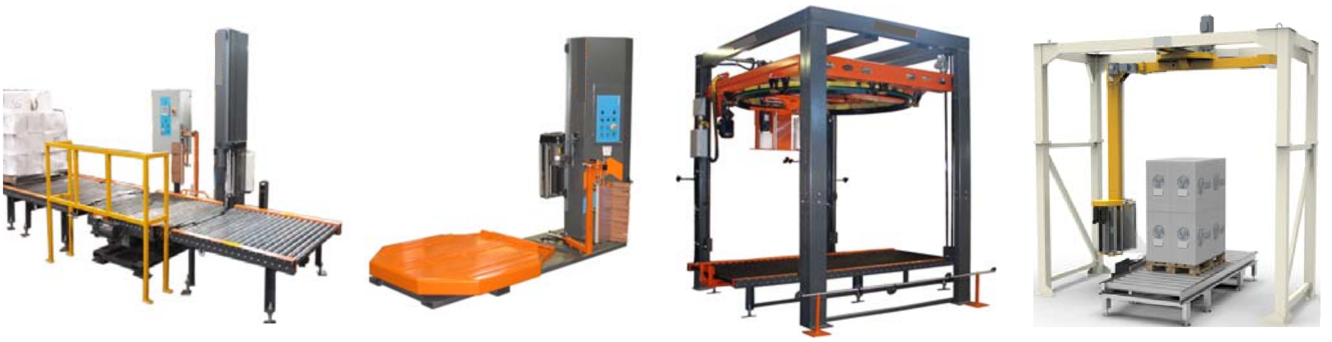
the design of an automatic pallet stretch-wrapping machine, which will accompany the theoretical description of the proposed methodology. Sect. 3 focuses on the estimation of the concordance of designer rankings, presenting some popular aggregation models. Sects. 4 proposes a novel method for testing the coherence of the collective ranking with designer rankings, trying to quantify the influence of the aggregation model in use. Sect. 5 summarizes the original contributions of this paper, practical implications, limitations and suggestions for future research.

2. Real-life case study

The conceptual description of the proposed approach 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. Four design concepts (O_1 to O_4, i.e., the alternatives or objects of the problem) of automatic machines have been generated by a team of designers during the conceptual design phase (see the short description in Fig.1):

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

(O_1) Conveyorized turntable | (O_2) Non-conveyorized turntable | (O_3) Rotary ring | (O_4) Conveyorized rotary arm



(O_1) These machines offer a relatively flexible and economical stretch-wrapping solution (i.e., relatively high load rate and reasonable weight capacity). Featuring a compact design, these machines are generally best when faced with tight fitting applications.

(O_2) Operators just have to place the pallet on the stretch wrapper and press the button to begin the automatic cycle. These machines eliminate the need for operators to get off the forklift to stretch wrap pallets and are available for both "low profile" and "high profile" wrapping operations.

(O_3) Unlike traditional turntables or rotary arms, these machines rotate only the carriage itself. Featuring a compact design, these machines allow to reduce film costs dramatically, since they can start and stop the cycle anywhere on the load.

(O_4) These machines eliminate the need to rotate the pallet being stretch wrapped. Heavy or unstable loads can quickly and efficiently be wrapped with these machines. These machines are most commonly used in-line with the rest of the packaging line.

Fig. 1. Schematic representation and short description of four design concepts of automatic stretch-wrapping machines.

The objective is to evaluate the afore-mentioned 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 [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, (ii) the ability to be integrated within various production lines, or (iii) the reduced set-up operations.

A collective judgment should be obtained by merging the individual (subjective) evaluations of ten design engineers (i.e., the decision-making agents of the problem: D_1 to D_{10}). The case study actually involved a team of designers in an important company based in north-western Italy. For reasons of confidentiality, the company is kept anonymous.

The preference profile provided for the four design concepts is shown in Tab. 1.

Designer	Preference ranking
D_1	$O_1 > O_2 > O_3 > O_4$
D_2	$O_1 > O_2 > O_3 > O_4$
D_3	$O_1 > O_3 > O_4 > O_2$
D_4	$O_1 > O_4 > O_3 > O_2$
D_5	$O_1 > O_4 > O_3 > O_2$
D_6	$O_3 > O_2 > O_4 > O_1$
D_7	$O_3 > O_2 > O_4 > O_1$
D_8	$O_4 > O_2 > O_3 > O_1$
D_9	$O_4 > O_2 > O_3 > O_1$
D_{10}	$O_4 > O_2 > O_3 > O_1$

Tab. 1. Preference profile provided by ten design engineers (D_1 to D_{10}) for the four design concepts (O_1 to O_4).

3. Concordance of designers

3.1. *The W coefficient*

The problem of evaluating the degree of concordance/agreement of a set of preference rankings (preference profile variability) can be traced back to the so-called problem of m-rankings. In this context, Kendall and Smith (1939) proposed the variance of the sum of the rank positions as a measure of this concordance. Precisely, a coefficient of concordance W can be defined as [Kendall, Smith 1939; Kendall, 1962, Fishburn, 1973b; Legendre 2005; 2010]:

$$W = \frac{S}{\frac{m^2(n^3-n)}{12}}. \quad (1)$$

The term S in Eq. (1) corresponds to the observed sum of squares of the deviations of the sums of the rank positions with respect to the mean value:

$$S = \sum_{i=1}^{n}(R_i - \bar{R})^2. \quad (2)$$

The term $m^2(n^3 - n)/12$ is the maximum possible value of S, which occurs in the case of complete unanimity in the rankings ($W = 1$).

The first term R_i in the brackets of eq.(2) is the sum of the rank positions for the *i*-th object:

$$R_i = \sum_{j=1}^{m} r_{ij}, \tag{3}$$

in which terms r_{ij} represent the rank of object O_i, according to the *j*-th designer, *n* being the total number of objects and *m* the total number of rankings. The second term in the brackets in Eq. 2 is the mean value of the R_i values:

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i. \tag{4}$$

W is defined in the dominion [0, 1]; $W = 0$ indicates the absence of concordance, while $W = 1$ indicates the complete concordance (unanimity). By replacing Eqs. (2), (3) and (4) within Eq. (1), the following expression can be obtained:

$$W = \frac{12\sum_{i=1}^{n} R_i^2 - 3m^2 n(n+1)^2}{m^2 n(n^2-1)}, \tag{5}$$

The coefficient *W* may be slightly modified if, apart from strict-preference relationships, preference rankings also include tied values, characterized by *indifference* relationships (e.g., $O_1 \sim O_2$) (Kendall, 1962; Legendre, 2010).

The effect of ties is to reduce the value of *W*; however, this effect is small unless the number of ties is large. To correct for ties, we introduce the following correction factor:

$$T_j = \sum_{i=1}^{g_j} (t_i^3 - t_i); \quad \forall j = 1, \dots, m, \tag{6}$$

where t_i is the number of tied ranks in the *i*-th group of tied ranks (where a group is a set of values having the same (tied) rank), and g_j is the number of groups of ties in the set of ranks (ranging from 1 to *n*) for the *j*-th designer. Thus, T_j is the correction factor required for the set of ranks related to the *j*-th designer. Note that if there are no tied ranks for the *j*-th designer, then $T_j = 0$.

With the correction for ties, the formula of *W* (Eq. (5)) becomes that of W_T:

$$W_T = \frac{12\sum_{i=1}^{n} R_i^2 - 3m^2 n(n+1)^2}{m^2 n(n^2-1) - m\sum_{j=1}^{m} T_j}, \tag{7}$$

where $\sum_{j=1}^{m} T_j$ is the sum of the values of T_j. W_T is also defined in the dominion [0, 1]

Returning to the case study concerning the four design concepts of automatic stretch-wrapping machines, Tab. 2 shows the calculation of their ranks.

Since there are no tie-break conditions, we may immediately calculate the coefficient of concordance for the *m*-rankings using Eq. (5):

$$W^{(m)} = \frac{12(25^2+26^2+25^2+24^2) - 3\cdot 10^2 \cdot 4(4+1)^2}{10^2 4(4^2-1)} = 0.004 = 0.4\%. \tag{8}$$

Designer	Rank positions			
	O_1	O_2	O_3	O_4
D_1	1	2	3	4
D_2	1	2	3	4
D_3	1	4	2	3
D_4	1	4	3	2
D_5	1	4	3	2
D_6	4	2	1	3
D_7	4	2	1	3
D_8	4	2	3	1
D_9	4	2	3	1
D_{10}	4	2	3	1
$\boldsymbol{R_{O_i}}$	**25**	**26**	**25**	**24**

Tab. 2. Ranks for the four design concepts (O_1, O_2, O_3 and O_4), obtained from the designer rankings in Tab.1.

The subscript "$^{(m)}$" indicates that this indicator is determined by considering the m-rankings of the designers. The low value of $W^{(m)}$ denotes a very low concordance of the m-designer rankings in the evaluation of the four design concepts. Section A.1 (in the Appendix) analyses the W indicator from a statistical point of view and provides a practical method for testing its significance.

3.2. *Choice of the aggregation model*

The choice of the aggregation model usually reflects the strategy of a team of designers. The aggregation models are inspired by the need to take into account all information provided by designers, ensuring fairness criteria in the treatment of information [McComb et al. 2017]. With reference to the case study (see Sect. 2), we now consider four different aggregation models, inspired by four corresponding strategies of the design team.

(i) **Best of the best** model (BoB or *standard plurality vote*)

For each designer, the most preferred design concept is selected. With reference to Tab. 1, O_1 gets 5 preferences, O_2 gets 0 preferences, O_3 gets 2 preferences and O_4 gets 3 preferences. The collective ranking is therefore $O_1 > O_4 > O_3 > O_2$ and the preferred design concept is O_1.

(ii) **Best two** model (BTW or *vote for two*)

For each designer, the two most preferred design concepts are selected. With reference to Tab. 1, O_1 gets 5 preferences, O_2 gets 7 preferences, O_3 gets 3 preferences and O_4 gets 5 preferences.

The collective ranking is therefore $O_2 > O_1 \sim O_4 > O_3$ and the preferred design concept is O_2. In this case we observe an equivalence condition between O_1 and O_4.

(iii) **Best three** model (BTH *or vote for three*)

For each designer, the three most preferred design concepts are selected (i.e., the worst solution of each designer is not considered). In this case O_1 gets 5 preferences, O_2 gets 7 preferences, O_3 gets 10 preferences and O_4 gets 8 preferences. The collective ranking is therefore $O_3 > O_4 > O_2 > O_1$ and the preferred design concept is therefore O_3.

(iv) **Borda count** model (BC)

For each rank position a score is assigned [Borda, 1781]: respectively 3 points to the first one, 2 points to the second, 1 point to the third, and 0 points to the fourth one. With reference to Tab.1, we obtain:

$$\mathrm{BC}(O_1) = 5 \cdot 3 + 0 \cdot 2 + 0 \cdot 1 + 5 \cdot 0 = 15; \qquad \mathrm{BC}(O_2) = 14 \qquad (9)$$
$$\mathrm{BC}(O_3) = 15; \qquad \mathrm{BC}(O_4) = 16$$

$\mathrm{BC}(O_1)$, $\mathrm{BC}(O_2)$, $\mathrm{BC}(O_3)$ and $\mathrm{BC}(O_4)$ being respectively the Borda scores calculated for the four design concepts. The collective ranking is therefore $O_4 > O_1 \sim O_3 > O_2$ and the preferred design concept is O_4. Also in this case we observe an equivalence condition between O_1 and O_3.

The four aggregation models reflect different selection strategies, leading to four different collective rankings and four corresponding preferred design concepts. Which collective ranking best reflects the *m*-designer rankings?

From a conceptual point of view, all these models deserve equal dignity, since they are all potentially plausible and justifiable. However, the fact remains that they may importantly affect the solution. The following section will try to answer the above question.

Finally, we point out that the four aggregation models considered above can be classified as "*positional models*" and allow to determine a collective ranking in a single round. Alternatively, we could have used so-called "*multi-round*" aggregation models, in which the alternatives can be phased out in different rounds [Saari 2011; Bormann and Golder, 2013;].

4. Coherence of the collective ranking

4.1. *The $W^{(m+1)}$ technique*

This approach follows the conceptual scheme seen in section 3.1. The idea is to analyse the level of coherence of the *m*-designer rankings and the collective ranking resulting from the application of the aggregation model, constructing the new coefficient of concordance $W^{(m+1)}$. As an example, Table 3 shows the construction of $W^{(m+1)}$ for the BoB aggregation model.

Designer	**Preference ranking**	**Rank positions**			
		O_1	O_2	O_3	O_4
D_1	$O_1 > O_2 > O_3 > O_4$	1	2	3	4
D_2	$O_1 > O_2 > O_3 > O_4$	1	2	3	4
D_3	$O_1 > O_3 > O_4 > O_2$	1	4	2	3
D_4	$O_1 > O_4 > O_3 > O_2$	1	4	3	2
D_5	$O_1 > O_4 > O_3 > O_2$	1	4	3	2
D_6	$O_3 > O_2 > O_4 > O_1$	4	2	1	3
D_7	$O_3 > O_2 > O_4 > O_1$	4	2	1	3
D_8	$O_4 > O_2 > O_3 > O_1$	4	2	3	1
D_9	$O_4 > O_2 > O_3 > O_1$	4	2	3	1
D_{10}	$O_4 > O_2 > O_3 > O_1$	4	2	3	1
BoB model	$O_1 > O_4 > O_3 > O_2$	**1**	**4**	**3**	**2**
$\boldsymbol{R_{O_i}}$		**26**	**30**	**28**	**26**

Tab. 3. *m*-rankings and corresponding rank positions of (i) the four design concepts (O_1 to O_4) formulated by ten designers (D_1 to D_{10}) (see Table 1), and (ii) the collective ranking obtained through the BoB model.

The overall coefficient of concordance $W_{BoB}^{(m+1)}$ for the BoB model can be obtained by applying Eq. (8) to the eleven rankings in Tab. 3:

$$W_{BoB}^{(m+1)} = \frac{12(26^2+30^2+28^2+26^2)-3\cdot(10+1)^2\cdot 4(4+1)^2}{(10+1)^2 4(4^2-1)} = 0.0182 = 1.82\% \tag{10}$$

We observe that $W_{BoB}^{(m+1)} > W^{(m)}$, i.e., the overall degree of concordance between the (10 + 1) rankings has increased. This output appears reasonable since the collective ranking resulting from the BoB model is a “composition” of the 10 previous rankings and is therefore supposed to reflect them all.

Repeating the construction for the other aggregation models (BTW, BTH and BC), results in Table 4 are obtained. The same table compares the $W_i^{(m+1)}$ values with the respective $W^{(m)}$ value (the same for all aggregation models) for the four aggregation models of interest.

It can be noticed that the Borda Count model obtains the highest value of $W_i^{(m+1)}$; it is therefore the aggregation model with the collective ranking that best fits with the *m* rankings.

Aggregation Model	$W_i^{(m+1)}$	$W^{(m)}$
Best of the Best (BoB)	1.82%	0.4%
Best two (BTW)	0.58%	0.4%
Best three (BTH)	1.48%	0.4%
Borda Count (BC)	2.00%	0.4%

Tab. 4. $W_i^{(m+1)}$ values for the four aggregation models (*i*: BoB, BTW, BTW, BC) compared to the relevant $W^{(m)}$ values.

4.2. *Further considerations on the $W^{(m+1)}$ technique*

If designer rankings are very *polarized* (i.e., they are characterized by a relatively low variability), we expect that the output of different aggregation models converge towards a common collective ranking. E.g., in the extreme case in which all the m-designer rankings coincide, the four aggregation models would converge into that very specific (collective) ranking.

As a pedagogical example, let us consider a new problem characterized by the set of m-rankings in Tab. 6. With reference to Tab. 1, we imagine that the three designers who indicated the ranking $O_4 > O_2 > O_3 > O_1$ (i.e., D_8, D_9 and D_{10}) converted their preference into the ranking $O_1 > O_2 > O_3 > O_4$; see the corresponding rank positions on the right hand side of Table 5.

Designer	Preference ranking	Rank positions			
		O_1	O_2	O_3	O_4
D_1	$O_1 > O_2 > O_3 > O_4$	1	2	3	4
D_2	$O_1 > O_2 > O_3 > O_4$	1	2	3	4
D_3	$O_1 > O_2 > O_3 > O_4$	1	4	2	3
D_4	$O_1 > O_2 > O_3 > O_4$	1	4	3	2
D_5	$O_1 > O_2 > O_3 > O_4$	1	4	3	2
D_6	$O_1 > O_3 > O_4 > O_2$	4	2	1	3
D_7	$O_1 > O_4 > O_3 > O_2$	4	2	1	3
D_8	$O_1 > O_4 > O_3 > O_2$	4	2	3	1
D_9	$O_3 > O_2 > O_4 > O_1$	4	2	3	1
D_{10}	$O_3 > O_2 > O_4 > O_1$	4	2	3	1
R_{O_i}		**26**	**30**	**28**	**26**

Tab. 5. New *m*-rankings and corresponding rank positions of the four design concepts (O_1 to O_4) formulated by ten designers (D_1 to D_{10}).

In this case, the $W^{(m)}$ is supposed to increase with respect to that of Tab.1, due to the reinforced polarization of preferences. Applying Eq. (5), we obtain:

$$W^{(m)} = \frac{12(16^2+26^2+25^2+33^2)-3\cdot 10^2\cdot 4(4+1)^2}{10^2 4(4^2-1)} = 0.29 = 29\%. \tag{11}$$

Not surprisingly, this value is conspicuously higher than that in Eq. (8).

Tab. 6 summarizes the collective rankings resulting from the m-rankings in Tab.5, for each (i-th) aggregation model and corresponding $W_i^{(m+1)}$ and $W^{(m)}$ values.

Aggregation Model	**Collective ranking**	$\boldsymbol{W_i^{(m+1)}}$	$\boldsymbol{W^{(m)}}$
Best of the Best (BoB)	$O_1 > O_3 > O_2 \sim O_4$	32.6%	29%
Best two (BTW)	$O_1 > O_2 > O_3 > O_4$	33.2%	29%
Best three (BTH)	$O_3 > O_1 > O_2 > O_4$	30.6%	29%
Borda Count (BC)	$O_1 > O_3 > O_2 > O_4$	33.5%	29%

Tab.6. Collective rankings provided by four different aggregation models (BoB, BTW, BTW, BC) for the (m) designer rankings reported in Tab.5, and corresponding $W_i^{(m+1)}$ and $W^{(m)}$ values.

Even in this second scenario, the Borda Count model shows the highest coefficient of concordance. This is not a general rule, since – when changing the preference profile – the model that best reflects it may change.

In addition, we note that the condition $W_i^{(m+1)} \geq W^{(m)}$ holds for each of the four aggregation models, when considering both the preference profiles in Tab. 2 and Tab. 5. However, this is not a general rule, as shown by the counterexample in Tab. 7.

Designer	**Preference ranking**	**Rank positions**			
		O_1	O_2	O_3	O_4
D_1	$O_1 > O_2 > O_3 > O_4$	1	2	3	4
D_2	$O_1 > O_2 > O_3 > O_4$	1	2	3	4
Collective ranking by the H-model	$O_4 > O_3 > O_2 > O_1$	4	3	2	1

Tab. 7. m-rankings and corresponding rank positions of (i) four design concepts (O_1 to O_4) formulated by two designers (D_1, D_2), and (ii) the collective ranking obtained through a hypothetical aggregation model (H-model).

Suppose we use a hypothetical aggregation model (H-Model) that provides the following collective ranking $O_4 > O_3 > O_2 > O_1$, which is the opposite to those by the two designers of interest. In this case, it can be verified that $W^{(m)} = 1$, while:

$$W_H^{(m+1)} = \frac{12\left(6^2+7^2+8^2+9^2\right)-3\cdot(2+1)^2\cdot 4(4+1)^2}{(2+1)^2 4(4^2-1)} = 0.111 = 11.1\%. \quad (12)$$

Although this exercise may seem somewhat implausible, it does show that – from a formal point of view – the condition $W_H^{(m+1)} \geq W^{(m)}$ does not necessarily hold.

The $W^{(m+1)}$-technique considers the coefficient of concordance W^m as the fulcrum for checking the concordance of the m-rankings and the coherence with the collective ranking. The same technique can also be applied with other potential coefficients of concordance, such as, for example, the mean value of the Spearman coefficient (*rho*) between the $\binom{m}{2}$ possible pairs of designers [Kendall, 1962].

4.3. *Impact of the aggregation models*

This section deals with the problem of measuring the impact of the aggregation model in use on the collective ranking. Consistently with the contents of the previous sections, the following synthetic indicator can be considered:

$$b_i^{(m)} = \frac{W_i^{(m+1)}}{W^m}. \tag{13}$$

It can easily be proved that $b_i^{(m)} \in]0, \infty[$. For a certain set of m-rankings, if $b_i^{(m)} \geq 1$, it means that the *i*-th aggregation model provides a somehow coherent collective ranking (positive impact). On the contrary, if $b_i^{(m)} < 1$, it means that the *i*-th aggregation model provides a somehow incoherent collective ranking (negative impact).

In practice, for a specific set of m-rankings, $b_i^{(m)}$ allows to identify the most coherent aggregation model, i.e., that with:

$$b^{(m)*} = max_i \, b_i^{(m)} \tag{14}$$

By way of example, Table 8 reports the value of $b_i^{(m)}$ for the four discussed aggregation models, considering the m-rankings in Tab.6.

Aggregation Model	$W_i^{(m+1)}$	$W^{(m)}$	$b_i^{(m)}$
Best of the Best (BoB)	32.6%	29%	1.12
Best two (BTW)	33.2%	29%	1.14
Best three (BTH)	30.6%	29%	1.06
Borda Count (BC)	33.5%	29%	1.16

Tab. 8. $b_i^{(m)}$ values for the four aggregation models of interest, considering the m-rankings in Tab. 6.

In this case, the BC is the aggregation model that produces the collective ranking that best reflects the m-rankings.

To test the significance of an observed value of $b_i^{(m)}$, one should consider the relevant distribution, which arises from the composition of the two distributions of $W^{(m)}$ and $W_i^{(m+1)}$ (see also Sect A.1, in the Appendix). Precisely, one can consider the distribution of $b_i^{(m)}$ in the $(n!)^{(m+1)}$ possible sets of rank positions and use it to reject or accept the hypothesis that designer rankings are concordant or not [Kendall 1962].

As a further example, Tab. 9 shows the value of $b_i^{(m)}$ for the four aggregation models discussed in Sect. 3.2, related to the (m) designer rankings in Tab.3.

Aggregation Model	$\boldsymbol{W_i^{(m+1)}}$	$\boldsymbol{W^{(m)}}$	$\boldsymbol{b_i^{(m)}}$
Best of the Best (BoB)	1.82%	0.4%	4.55
Best two (BTW)	0.58%	0.4%	1.45
Best three (BTH)	1.48%	0.4%	3.70
Borda Count (BC)	2.00%	0.4%	5.00

Tab.9. $b_i^{(m)}$ values for the four aggregation models of interest, considering the *m*-rankings in Tab. 3.

Even in this case, the BC is the aggregation model that produces the collective ranking that best reflects the *m*-rankings.

In practice, $b_i^{(m)}$ can be used to provide a kind of metric for evaluating the *relative impact* of two compared models. For a certain preference profile, we define as *relative impact* the ratio $s_{ij}^{(m)}$ between the values $b_i^{(m)}$ of two generic models:

$$s_{ij}^{(m)} = \frac{b_i^m}{b_j^m} = \frac{W_i^m}{W_j^m} \tag{15}$$

As an example (see Tab.9), for BoB e BTW models we obtain:

$$s_{BoB,BTW}^{(m)} = \frac{b_{BoB}^m}{b_{BTW}^m} = \frac{4.55}{1.45} = \frac{W_{BoB}^m}{W_{BTW}^m} = 3.14 \tag{16}$$

The relative impact $s_{BoB,BTW}^{(m)} = 3.14$ shows how strongly the BoB model proposes the O_1 alternative as opposed to the BTW model, which supports the O_2 alternative (see section 3.2).

Let us now analyse in detail the $b_i^{(m)}$ indicator (Eq. 13). The denominator ($W^{(m)}$) depends exclusively on the *m*-rankings, while the numerator ($W_i^{(m+1)}$) depends both on the *m*-rankings and the collective ranking. For problems characterized by a relatively large number (m) of rankings, the contribution of the collective ranking will therefore tend to have a lower "weight", i.e., 1/(m+1).

Moreover, $b_i^{(m)}$ values tend to be higher for problems characterized by relatively discordant *m*-rankings (and therefore lower $W^{(m)}$ values). For example, the $b_i^{(m)}$ values in Table 9 are significantly higher than those in Tab. 8, as they result from a problem with significantly more discordant m-rankings and therefore relatively lower $W^{(m)}$ values.

5. Discussion

This paper proposed a novel method to support the decisions of teams of designers in early design stages, trying to provide a plausible answer to three different research questions: (i) how can the degree of concordance of designer rankings be measured? (ii) For a given set of designer rankings, which aggregation model provides the most coherent solution? (iii) To what extent is the collective ranking influenced by the aggregation model in use?

The answer to the first research question is represented by the Kendall's coefficient of concordance W, or by other similar multi-rater coefficients of agreement. The next two research questions were interpreted as consequences of Arrow's theorem, when deployed in the context of engineering design. In this field, designers are often in contrast on alternative design solutions and it is not easy to decide how aggregate the choices of individual designers into a single collective one. The answer to the second research question was given by a special use of the coefficient W. This coefficient of concordance is applied not only to m-rankings but also to the collective ranking produced by the aggregation method of interest. The new indicator $W^{(m+1)}$ allows a quick comparison between alternative aggregation models, if they are applied to the same specific problem, describing their "effectiveness".

For the third research question, paraphrasing Arrow's theorem, it is difficult to understand if and to what extent a certain aggregation model impact on the collective ranking. A practical answer is given by the $b_i^{(m)}$ indicator, which depicts the ability of a certain aggregation model to provide a plausible synthesis. $b_i^{(m)}$ can also be used to provide a quantitative *metric* for evaluating the *relative impact* of different models.

Having said that, let us now consider the proposed approach in its entirety. This approach is easily implementable and, with few additional adjustments, can also be applied to problems including *partial* preference rankings [Franceschini, Maisano 2018].

The main limitations of the proposed approach are:

- the method does not consider nor the (possible) uncertainty in designer rankings, neither possible incompleteness in preference rankings;
- the method allows only an *ex-post* analysis of the impact of aggregation models.

Regarding the future, we plan to develop an interactive on-line tool for supporting the activity of design teams. In addition, this tool will be used for more structured problems in the field of design for manufacturing and quality engineering/management (e.g. integrating it with the *Quality Function Deployment* and other design support tools) [Olewnik, Lewis 2008; Franceschini, Maisano, Mastrogiacomo 2015b; 2015c; Chen et al., 2017]. Finally, we plan to consider problems in which designers may have a different (ordinal or cardinal) weight, for example due to a different design experience.

Acknowledgments

This research was partially supported by the award "TESUN-83486178370409 Finanziamento dipartimenti di eccellenza CAP. 1694 TIT. 232 ART. 6", which was conferred by "Ministero dell'Istruzione, dell'Università e della Ricerca - ITALY".

Appendix

A.1 *Statistical meaning of W*

This subsection analyses the statistical significance of *W*. In general, the *W* distributions are available in tabular terms for small values of *m* and *n* [Kendall, 1962]. For higher values, the Fisher distribution can be used:

$$\mathrm{F} = \frac{(m-1)\cdot W^{(m)}}{[1-W^{(m)}]}\,, \tag{A.1}$$

with parameters ν_1 and ν_2 defined respectively as:

$$\begin{cases} \nu_1 = n - 1 - \frac{2}{m} \\ \nu_2 = (m-1)(n - 1 - \frac{2}{m}) \end{cases}. \tag{A.2}$$

When $n > 7$, $W^{(m)}$ can be described by a chi-square distribution $\chi_r^2 = m(n-1)W^{(m)}$. χ_r^2 is distributed as a χ_{n-1}^2 with $\nu = n - 1$ degrees of freedom.

For example, considering the data in Tab.6, where $m = 10$ designers and $n = 4$ design concepts, $W^{(m)} = 0.29$. Applying Eq. (8), it can be obtained:

$$\mathrm{F} = \frac{(10-1)\cdot 0.29}{(1-0.29)} = 3.67, \tag{A.3}$$

The degrees of freedom are respectively:

$$\begin{cases} \nu_1 = 4 - 1 - \frac{2}{10} = 2.8 \approx 3 \\ \nu_2 = (\mathrm{m}-1)\left(\mathrm{n} - 1 - \frac{2}{\mathrm{m}}\right) = 25.2 \approx 25 \end{cases} \tag{A.4}$$

From the tables of the Fisher distribution for a significance of 5%, it is obtained $F_{5\%;3;25} = 2.99$.

Since $F > F_{5\%;3;25}$, the significance of the coefficient of concordance for the preference profile in Tab.6 is confirmed.

References

Arrow, K.J., Rayanaud, H. (1986), *Social choice and multicriterion decision-making*, Cambridge: MIT Press.

Arrow K.J. (2012), *Social Choice and Individual Values*, 3rd edition, Yale University Press, New Haven.

Borda, J.C. (1781) *Mémoire sur les élections au scrutin, Comptes Rendus de l'Académie des Sciences*. Translated by Alfred de Grazia as Mathematical derivation of an election system, Isis, **44**:42-51.

Bormann, N.C., Golder, M. (2013). *Democratic electoral systems around the world*, 1946–2011. Electoral Studies, 32(2), 360-369.

Cagan J., Vogel C.M. (2012), *Creating Breakthrough Products: Innovation From Product Planning to Program Approval*, 2nd edn., FT Press.

Chen, S., Liu, J., Wang, H., Augusto, J.C. (2012) Ordering Based Decision Making–A Survey. *Information Fusion*, **14**(4): 521-531.

Chiclana, F., Herrera, F., Herrera-Viedma, E. (2002) A note on the internal consistency of various preference representations. *Fuzzy Sets and Systems*, **131**(1): 75-78.

Cook, W.D. (2006) Distance-based and ad hoc consensus models in ordinal preference ranking. *European Journal of Operational Research*, **172**(2): 369-385.

Dong A., Hill A.W., Agogino A.M. (2004), A document analysis method for characterizing design team performance. *Journal of Mechanical Design*, **126**(3):378–385.

Dwarakanath S., Wallace K.M. (1995), Decision-making in engineering design– observations from design experiments. *Journal of Engineering Design*, **6**(3):191–206.

Dym C.L.,Wood W.H., Scott M.J. (2002), Rank ordering engineering designs: pairwise comparison charts and Borda counts. *Research in Engineering Design*, **13**: 236–242.

Fishburn, P.C. (1973a), Voter concordance, simple majorities, and group decision methods, *Behavioral Science*, **18**:364-376.

Fishburn, P.C. (1973b) *The Theory of Social Choice*. Princeton University Press, Princeton, USA.

Franceschini, F., Maisano, D. (2015a) Checking the consistency of the solution in ordinal semi-democratic decision-making problems. *Omega*, **57**(1): 188-195.

Franceschini F., Maisano D., Mastrogiacomo L. (2015b), Customer Requirements prioritization on QFD: a new proposal based on the Generalized Yager's Algorithm. *Research in Engineering Design*, **26**(2):171-187.

Franceschini, F., Galetto, M., Maisano, D, Mastrogiacomo, L. (2015c) Prioritization of Engineering Characteristics in QFD in the case of Customer Requirements orderings. *International Journal of Production Research*. **53**(13):3975-3988.

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.

Franceschini, F., Maisano, D. (2017). Consistency analysis in quality classification problems with multiple rank-ordered agents. *Quality Engineering*, **29**(4):672-689.

Franceschini, F., Maisano, D. (2018). Fusion of partial orderings for decision problems in Quality Management. Proceedings of the 3rd International Conference on Quality Engineering and Management (ICQEM 2018), July 11-13, 2018, Barcelona (Spain).

Franceschini, F., Galetto, M., Maisano, D. (2019) *Designing Performance Measurement Systems: Theory and Practice of Key Performance Indicators*, Springer International Publishing. Cham, Switzerland, ISBN: 978-3-030-01191-8.

Franssen M. (2005), Arrow's theorem, multi-criteria decision problems and multi-attribute preferences in engineering design. *Research in Engineering Design*, **16**(1–2):42–56.

Frey D.D.et al. (2009), The Pugh controlled convergence method: model-based evaluation and implications for design theory. *Research in Engineering Design*, **20**(1):41–58.

Frey D.D., et al. (2010) Research in engineering design: the role of mathematical theory and empirical evidence. *Research in Engineering Design*, **21**(3):145–151.

Fu K., Cagan J., Kotovsky K. (2010), Design team convergence: the influence of example solution quality. *Journal of Mechanical Design*, **132**(11): 111005.

Hazelrigg G. A. (1996), The implications of Arrow's impossiblity theorem on approaches to optimal engineering design. *Journal of Mechanical Design*, **118**(2):161-164.

Hazelrigg G.A.(1999), An axiomatic framework for engineering design. *Journal of Mechanical Design*, **121**(3):342.

Hazelrigg G.A. (2010), The Pugh controlled convergence method: model-based evaluation and implications for design theory. *Research in Engineering Design*, **21**(3):143–144.

Herrera-Viedma, E., Cabrerizo, F.J., Kacprzyk, J., Pedrycz, W. (2014) A review of soft consensus models in a fuzzy environment. *Information Fusion*, **17**:4-13.

Hoyle C., Chen W. (2011), Understanding and modelling heterogeneity of human preferences for engineering design. *Journal of Engineering Design*, **22**(8):583-601.

Jacobs J. F., van de Poel I., Osseweijer P. (2014), Clarifying the debate on selection methods for engineering: Arrow's impossibility theorem, design performances, and information basis. *Research in Engineering Design*, **25**(1):3-10.

Kaldate A., et al. (2006), Engineering parameter selection for design optimization during preliminary design. *Journal of Engineering Design*, **17**(March 2015): 291-310.

Katsikopoulos K. (2009), Coherence and correspondence in engineering design: informing the conversation and connecting with judgment and decision-making research. *Judgment and Decision Making*, **4**(2):147-153.

Kendall M.G., Smith B.B. (1939), The problem of m-rankings, *Annals of Math. Statist*, **10**:275-287.

Kendall (1962), *Rank correlation methods*, Griffin & C., London.

Keeney R. L. (2009), The foundations of collaborative group decisions. *International Journal of Collaborative Engineering*, **1**,4.

Ladha K., Miller G., Oppenheimer, J. (2003), *Information Aggregation by Majority rule: Theory and Experiment*. http://www.gvptsites.umd.edu/oppenheimer/research/jury.pdf.

Legendre P. (2005), Species associations: the Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics*, **10**: 226.

Legendre, P. (2010). *Coefficient of concordance* in: Encyclopedia of Research Design, Vol. 1. pp. 164-169, N. J. Salkind, (editor), SAGE Publications, Inc., Los Angeles.

Li H., Bingham J. B., Umphress, E. E. (2007), Fairness from the top? Perceived procedural justice and collaborative problem solving in new product development. *Organization Science*, **18**(2):200-216.

McComb C., Goucher-Lambert K., Cagan J.(2015), *Fairness and manipulation: an empirical study of Arrow's impossibility theorem*. In International Conference on Engineering Design, Milan, Italy, pp. 267-276.

McComb C., Goucher-Lambert K., Cagan J. (2017), Impossible by design? Fairness, strategy and Arrow's impossibility theorem. *Design Science*, **3**:1-26.

Nisan N. (Ed.) (2007), *Algorithmic Game Theory*. Cambridge University Press.

Nurmi H. (2012), *On the relevance of theoretical results to voting system choice*. In Electoral Systems: Studies in Choice and Welfare (ed. D. S. Felsenthal & M.Machover), pp. 255-274, Springer.

Olewnik A. T., Lewis K. (2008), Limitations of the House of Quality to provide quantitative design information. *International Journal of Quality and Reliability Management*, **25**(2):125-146.

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.

Paulus P. B., Dzindolet M. T., Kohn N. (2011), *Collaborative creativity, group creativity and team innovation*. In Handbook of Organizational Creativity (ed. M. D.Mumford), pp. 327-357. Elsevier.

Reich Y. (2010), My method is better! *Research in Engineering Design*, **21**(3):137-142.

Saari D. G., Sieberg K. K. (2004), Are partwise comparisons reliable? *Research in Engineering Design*, **15**(1):62-71.

Saari D. G. (2011), *Geometry of Voting*. Elsevier B.V.

Saari D. G. (2011), *Decision and elections*, Cambridge University Press.

Scott M. J., Antonsson E. K. (1999) Arrow's theorem and engineering design decision making. *Research in Engineering Design*, **11**:218-228.

See T.K., Lewis K. (2006), A formal approach to handling conflicts in multiattribute group decision making. *Journal of Mechanical Design*, **128**(4):678.

Weingart L. R., et al. (2005), *Functional diversity and conflict in cross-functional product development teams: considering representational gaps and task characteristics*. In Understanding Teams (ed. L. L. Neider & C. A. Schriesheim), pp. 89-110. Information Age Publishing.

Yager, R.R. (2001) Fusion of multi-agent preference orderings. *Fuzzy Sets and Systems*, **117**(1):1-12.

Yeo S. H., Mak M.W., Balon S. A. P. (2004), Analysis of decision-making methodologies for desirability score of conceptual design. *Journal of Engineering Design*, **15**(2):195-208.