Psychological Test and Assessment Modeling, Volume 55, 2013 (2), 207-225

Multiattributive decision theory: who are the best?[1]

Hartmann H. Scheiblechner† [2,3]

Abstract

A rank ordering of preference of multiattributive choice alternatives is suggested. The choice alternatives are characterised by several attributes (dimensions) which themselves are assumed to be given by strict partial orders (rank orders with possibly ties, or "rating scales"). A dominance relation is defined on the alternatives: an alternative dominates another if it is at least as good as the other in all dimensions and strictly superior in at least one dimension. The result is a multidimensional partial order. The problem is to choose a single best or a given number of best alternatives from the choice set. The solution must not involve comparisons of ranks of different dimensions, if the decision maker is a single individual, or of rank orders of different individuals, if the decision maker is a social group (social choice function). The (modified) percentile rank score (Scheiblechner, 2002, 2003) is suggested as scaling function. The performance of the (modified) percentile score is illustrated by the examples of the results of the competitors of the decathlon of Olympic Games at Beijing 2007 and World Championships at Berlin 2009.

Key words: Decision theory; ranking; scoring; d-dimensional isotonic probabilistic model

---

[1] This paper was submitted in February 2010; four excellent reviewers, highly qualified in psychometrics and measurement theory, demanded some minor – mainly of the same kind – revisions. However, the author was not able to follow their suggestions due to serious illness – he died shortly after in December 2010. Nevertheless, the editor in chief decided to publish the paper in its submitted version with just a few editorial changes. The editor would have suggested some revision himself; first, a conclusion in the final chapter as the reader now has to make this for him/herself. Second, the author could have included an introduction to his d-ISOP model (published in 2007 in Psychometrika) and further explanation for the reader, who is not acquainted with it, has to read it up on it by him/herself. Third, the Arrow's proof is not explained as well, and again the reader has to inform oneself of this proof (though the author gives a reference to Fishburn, 1973); however, a simple source like Wikipedia has proved to be sufficient. Fourth, the paper is not organized in chapters as usually practised. To summarize, though there are some shortcomings in structuring the paper, the reviewers and the editor in chief appraise the paper of being of high scientific value, the latter even reckons it will cause a stir in the field of sport competition scoring: At least the given example proves the unfairness of pertinent scoring rules.

[2] Philipps-University Marburg, Germany

[3] *Correspondence concerning this article should be addressed to:* Klaus D. Kubinger, Ph.D., c/o Division of Psychological Assessment and Applied Psychometrics, University of Vienna, Faculty of Psychology, Germany; email: klaus.kubinger@univie.ac.at

Examples

Decathlon/grant for gifted pupils:

- 36 athletes competed for the gold medal for the most complete athlete at the Olympic Games 2008 at Beijing (26 completed the contest). Ten athletic disciplines were requested from all competitors. What is the correct (fair) ranking?
- The applicants for a grant have grades in several qualifications. A given number of them can get a grant. Which applicants have the best qualifications?

Vacation plans, careful pater familias/the benevolent dictator:

- A family of 5 members (including the parents) has the choice between 10 possible plans for annual vacations. Each member establishes a rank order of all 10 alternatives. The 5 rank orders are coded as 5 attributes of the 10 options. Each alternative is evaluated by 5 values on rating scales. The best decision (social decision function) is the one which provides the greatest satisfaction for the largest number of group members.
- The benevolent dictator knows exactly the preferences (utilities, measured on individual ordinal scales or on strict partial orders) of all group members on a set of attributes of a set of choice alternatives. (If the utilities of the dictator are affected by the decision he considers his own utilities in exactly the same manner as the utilities of all other members of the group). The dictator decides on the best choice and can enforce his decision. There is not necessarily a unique best decision and the social value of all alternatives needs not to be defined (then the group has to decide on a new choice procedure, new attributes and/or choice function).

The problem

Let A = {*a*, *b*, *c*… *s*, *t*} be the set of all feasible alternatives (acts, decisions). Let $X_i(a)$, $i=1, 2 \ldots n$, the value (utility) of alternative *a* on the i^{th} attribute (dimensions on which the alternatives are evaluated). The set of attributes should be complete, operational, decomposable, non redundant and minimal (pp. 50, Keeney & Raiffa, 1976). Each X_i is monotone increasing (a rank order with possibly ties, or a strict partial order or a rating scale) and should be preferentially independent from the complement of the set of attributes. If the attribute is a rating scale it is not implied that alternatives or options being rated with the same grade are indifferent and that indifference is transitive; it only means that within the given level of precision there is no difference of a whole grade between tied options. The order of grades is transitive. A rating scale is a strict partial order.

The choice alternatives are represented by vectors of *n* scalar numerical evaluators or utilities (ordinal measures to which degrees *n* objectives are achieved by the alternatives; fixed values or mean values over probability distributions if the values are random variables).

We search for a scalar function υ

$$\upsilon(x_1, x_2, \ldots, x_n) = f[x_1, x_2, \ldots, x_n]$$

representing the total value of choice alternatives (the group value or social choice function if the decision maker is a group or a benevolent dictator). The chosen alternatives are the maximal values (suprema) of υ over A (if more alternatives have the same maximal value than are needed, then they are chosen randomly among equally valued alternatives).

The function must not involve *comparisons of utilities of different attributes or interindividual comparisons of utility functions*. The total value υ should be a *strict partial order* (rank order with possible ties or rating scale):

$a \succ b \Rightarrow \upsilon(a) > \upsilon(b)$ and

$a \approx b \Leftrightarrow \upsilon(a) = \upsilon(b)$ (Fishburn, 1973, p. 77).

That is, strict preferences $\succ$ are transitive and indifferences $\approx$ (alternatives have the same rank or rating, all group members attribute equal ranks or ratings to two alternatives) may be intransitive. In contrast to a *weak order* different numerical values (scale values) in the numerical representation of a partial order do not imply a strict preference (but a preference/indifference relation $\overset{\succ}{\approx}$). A scale value may be larger than another value without the decision maker being able to state a preference (subliminal difference). In a weak order the numerical values represent strictly ordered indifference classes.

To choose the best alternatives (one or more) among a set of choice alternatives is formally equivalent to ranking a set of competitors who are evaluated on a set of attributes. In order to have a realistic and complex example of a choice set we choose the performances of sportive competitors in the decathlon of the Olympic Games at Beijing 2008.

Empirical example: decathlon Olympic Games 08

The data are the performances of 36 competitors in 10 athletic disciplines (http://results.bejing2008.cn). 10 athletes did not finish the competition (and were excluded from the final ranking). Only the 26 athletes who finished all disciplines are retained. The performances are ranked within each discipline in the present approach. The ranks range from 1 to 36 depending on how many competitors still participated in the discipline. Therefore there are tied ranks and different gaps in the rankings of the 26 finishers in different disciplines. The gaps do not alter the results.

In the example of the family annual vacations the alternative plans are the ‘competitors’ and the rankings by the 5 family members are the 5 ‘attributes’ of each plan. In the grant example the vectors of the grades in the prescribed qualifications are the ‘competitors’, each is counted with its frequency in the group of applicants. The grades are the (values of the) ‘attributes’. In the benevolent dictator example the utilities of each alternative for the group members are the attributes.

Let us return to the Olympic Games. The data are analysed by the d-ISOP model (d-dimensional isotonic probabilistic model) by Scheiblechner (2007; alternative parametric models are given in van der Linden & Hambleton, 1997). In this case the number of aspects $d = 1$ because each attribute is measured by a single numerical value and $n = 10$ for the 10 "items" or disciplines. The achievement of one athlete is ranked higher (= lower rank number, because the competitors conventionally are ranked from 1 to N, from best to last) than the achievement of another competitor, if his performances are at least equally good in all disciplines and better in at least one discipline. This gives a partial order of all 26 athletes. Some pairs of athletes can not be ordered because one is better in discipline i and the other is better in discipline $j \neq i$. The number of unordered pairs is larger the more disciplines we consider. We therefore consider also subsets of the 10 disciplines. For all partial orders the *(modified) percentile scale scores* can be computed for the competitors. Let

$\mathbf{x}_s = (x_{s1}, \ldots, x_{sn}) \quad s = 1, 2, \ldots, N$

be the attribute vector of competitor s and let competitor s dominate competitor t

$\mathbf{x}_s \succ \mathbf{x}_t$ iff $x_{sj} \geq x_{tj}$ for all $j = 1, 2, \ldots ,n$ and $x_{sj} > x_{tj}$ for at least one j.

We now can compute the percentile scale scores P of the competitors. The percentile score is a maximum likelihood estimate of the rank of an athlete (Scheiblechner, 2002, 2003) and is computed by the number of competitors an athlete has outdone minus the number of rivals by whom he was outdone himself divided by the sum of the two of them:

$P_s = (n^+(\mathbf{X}_s) - n^-(\mathbf{X}_s)) / (n^+(\mathbf{X}_s) + n^-(\mathbf{X}_s))$ where

$n^+(\mathbf{X}_s) = |\{t: \mathbf{X}_s \succ \mathbf{X}_t\}|$ the number of alternatives s dominates and

$n^-(\mathbf{X}_s) = |\{t: \mathbf{X}_s \prec \mathbf{X}_t\}|$ the number of alternatives by which s is dominated.

The percentile scale score is +1 for all alternatives for which $n^+ > 0$ and $n^- = 0$, and -1 for all alternatives for which $n^+ = 0$ and $n^- > 0$. The percentile scale score is undetermined if $n^+ = n^- = 0$. In the latter case the alternative can not be ordered relative to any other alternative.

For the decathlon, the question of dimensionality is a subtle one. For a specialized athlete, for example for a sprinter, we would require that all competitions, by which his performance is measured, be 'one-dimensional', i.e. measure the same specialized ability. By contrast, an athlete competing in the decathlon just should not be restricted to a single discipline. On the other hand we would not like to mix incompatible sports together e.g. sprint and weight lifting. All disciplines should be track and field. Therefore we do not require that all included disciplines should correlate highly positive with each other, but we also would reject essential negative correlations.

The Goodman-Kruskal measures of association (Goodman & Kruskal, 1954, 1957, 1963, 1972) between the disciplines range from .448 (100m sprint and 110m hurdles) to -.212 (1500 m and discus throw). 11 of the 45 correlations are small negative but none of them is significant (Type-I-risk 5%). 12 of the 34 positive correlations are, however, significant. The disciplines can be grouped into 'running' (R, 100m, 400m, 110m hurdles; all correlating significantly positive), 'jumping' (J, long jump, high jump, pole vault; all positively correlated), 'throwing' (T, shot put, discus throw, javelin throw; the first two significantly positively correlated) and ‘endurance’ (E, 1500m). The long jump correlates positively with the first two running disciplines, 400m and 1500m correlate positively, and javelin throw correlates positively with high jump and pole vault. This grouping of disciplines is further supported by grouping the athletes based on hierarchical clustering of the vectors of group percentile scores by single link (see below).

Further indicators of dimensionality are the indices W1 and W2, the degrees to which the axioms W1 and W2 of ISOP are satisfied, and the item-rest predictabilities (= correlations; Scheiblechner, 2003; Scheiblechner & Lutz, 2009). All indices should not be negative but not very high positive, too (see associations between disciplines). The indix W1, the amount to which the ranking of competitors in one item agrees with the rankings in all other items, ranges from .037 (pole vault) to .225 (high jump) for the disciplines and are non-negative and not significant. The item-rest predictability, the amount to which the rankings in single disciplines predict the rankings in the rest of the disciplines, range from .182 (1500m) to 1.0 (110m hurdles, long jump, pole vault). The agreement of the rank orders of the difficulty of the 10 disciplines over the competitors, W2, is .0387, indicating that the athletes tend to have different centres of gravity (areas of specialisation) of their performance.

In the official ranking the performances in the disciplines are transformed into ‘points’ by more or less arbitrary nonlinear transformations for which information on the disciplines beyond the performances of the actual competitors of the decathlon (for example the current world record) are needed and the points then are summed up. The winner is the competitor with the largest sum of points. The official end position is used as label of the competitors in Tab. 1.

The percentile scores are an alternative evaluation of the achievements of the athletes. One advantage is that no outside information is needed and the actual data of the group of competitors are sufficient. If available the distributions of the achievements of the same group of competitors over repeated contests or of an acknowledged population of relevant athletes in the last year(s) can be used to compute the percentile scale scores in the disciplines. The disciplines can also be grouped (see below) and the scores can be computed for each group of disciplines separately. The group scores can themselves be used again to compute an overall score over all groups of disciplines (e.g. eliminating E because it only contains 1500m, Rank9 of Tab.1 would then be the official end result). By the scores of grouped disciplines we can define special types of competitors, e.g. ‘runners’ or ‘runners and jumpers’ if their score on some disciplines is much higher than on the rest of the disciplines or some antitypes of competitors if they are much weaker in one group than in the rest of the disciplines (e.g. competitor 1, the gold medallist, is an anti E type, rank 23 of 26 in 1500m).

The sum of points after 9 disciplines differs from the 10 disciplines end result by maximally 5 positions (competitor 11 and 13). The percentile score computed on the partial order of the athletes in all 10 disciplines, the ‘Score10’, differs from the official final result by maximally 6 positions (competitor 6). The ‘ScoreRJTE’, computed on the order

Table 1:
The ranks of the competitors and the percentile scores. The competitors are numbered from 1 to 26, which at the same time is their official final position. Rank9 is their position after 9 disciplines (without 1500m). The lower the rank and the more negative the score the better the performance. Score10 is their percentile score in the partial order based on the 10 separate disciplines (rank in brackets). ScoreRJTE is the percentile score based on 4 groups of disciplines (with E consisting of 1500m only; rank in brackets). ScoreR, ScoreJ, ScoreT are the percentile scores in the grouped disciplines R, J and T. (Maxima and minima in *italics*).

competitor	Rank9	Score10	ScoreRJTE	ScoreR	ScoreJ	ScoreT
1	1	*-0.638* (1)	-0.520 (3)	*-0.760*	-0.706	*-0.920*
2	2	-0.467 (2)	-0.560 (2)	-0.595	*-0.857*	0.173
3	4	-0.443 (3)	*-0.600* (1)	-0.627	-0.362	-0.227
4	3	-0.271 (5)	-0.180 (8)	0.253	-0.694	-0.707
5	7	-0.182 (9)	-0.400 (5)	0.081	-0.118	-0.333
6	5	-0.095 (12)	-0.080 (12)	0.425	-0.743	-0.173
7	10	-0.336 (4)	-0.455 (4)	-0.676	0.000	-0.147
8	9	-0.135 (10)	-0.180 (8)	0.014	-0.667	0.333
9	12	-0.262 (6)	-0.360 (6)	-0.680	0.217	-0.200
10	14	-0.207 (7)	-0.313 (7)	-0.270	-0.529	0.413
11	6	-0.062 (14)	-0.100 (11)	0.351	-0.275	-0.387
12	11	-0.128 (11)	-0.180 (10)	0.547	-0.493	-0.467
13	8	-0.190 (8)	-0.082 (13)	-0.547	-0.164	-0.040
14	15	0.099 (16)	0.071 (15)	0.493	-0.045	-0.040
15	13	-0.083 (13)	0.061 (14)	-0.547	0.194	-0.093
16	16	0.176 (19)	0.152 (16)	0.467	0.343	-0.147
17	17	0.276 (11)	0.300 (18)	0.760	0.471	-0.360
18	19	0.143 (17)	0.180 (17)	0.108	0.352	0.067
19	20	0.159 (18)	0.260 (20)	-0.333	0.429	0.440
20	18	0.045 (15)	0.232 (19)	-0.573	0.194	0.227
21	22	0.240 (20)	0.280 (22)	-0.227	0.200	0.787
22	21	0.320 (22)	0.360 (24)	0.200	0.889	-0.280
23	23	0.412 (24)	0.340 (23)	0.622	-0.014	*0.867*
24	24	*0.664* (26)	*0.780* (26)	*0.787*	0.565	0.680
25	25	0.363 (23)	0.253 (21)	0.054	*0.946*	0.413
26	26	0.577 (25)	0.740 (25)	0.680	0.803	0.120

on running R (‘ScoreR’), jumping J (‘ScoreJ’), throwing T (‘ScoreT’) and endurance E also differs from the official result by 6 positions for competitor 6 (a “jumper”, J; he is 2^{nd} in ScoreJ and 7 worst in ScoreR). The first 3 places (the medals) are reversed by the ScoreRJTE. The ‘ScoreE’ only consists of a single discipline (1500m) and is relatively overvalued in the present analysis (due to the available data). The 1500m either could be eliminated from the decathlon (corresponding to Score 9) or supplemented by two other disciplines, e.g. swimming 1000m and cycling 40km (compare triathlon). The Score10 and the ScoreRJTE differ by maximally 5 positions (competitor 13). The absolute range of dispersion for ScoreRJTE (1.380) is larger than for Score10 (1.302). The range of dispersion of the percentile scores gives an impression of how well the data are represented by a single scale value. The ranges of dispersion for the grouped disciplines are ScoreJ (1.803), ScoreT (1.787) and ScoreR (1.547). The ranges for grouped disciplines are larger than for Score 10 because the disciplines within a group can be better represented by a single score than over all disciplines and are more reliable because they rely on more defined pair comparisons between competitors.

Although the transformed points of the disciplines are used to calculate the final positions of the competitors in the official valuation the P-scores (percentile scores), which only use rank information on the performances of the actual competitors, correlate similarly (and sometimes even higher) with the final position than the transformed points (ScoreJ 0.848, sum of points of jumping disciplines -0.791). The P-scores of the grouped disciplines R, J and T correlate lower with each other than the corresponding sums of points (thus favouring the grouping and the scaling). The final position and the Rank9 correlate with 0.964. The correlation of the final position with the Score10 and the ScoreRJTE is 0.934 and 0.928, respectively. That is, the ranking by the P-scores is a possible alternative to the present official ranking.

Table 2 gives the dominance relations between competitors as defined above based on the vector of the 3 grouped disciplines R, J and T. The percentile scores are computed first for the 3 groups of disciplines separately and then the final percentile score ScoreRJT is calculated from the vector (Perc.R, Perc.J, Perc.T).

Figure 1 depicts the minimal generator of the dominance relations of Table 2. The dominance relations implied by transitivity from the generating relations are not depicted. The generating relations $(x \to y)$ and $(y \to z)$ imply $(x \to z)$ which is not displayed.

The partial order in Figure 1 has 3 suprema (competitors 1, 2 and 6). If a unique winner is desired, we have to introduce secondary criteria, e.g. the majority of victories for competitor 1. Or the minimax criterion, the maximum of the minimal attribute (again in favour of competitor 1, if grouped disciplines are used). Similarly the dominance graph in Figure 1 has 6 infima (competitors 21 to 26, who could be arranged e.g. inversely to the order of number of losses 24, 26, 23, 25, 21, 22). Secondary criteria preferably should be properties of the dominance relations or other auxiliary attributes of the alternatives.

Table 2:
Dominance relations (Perc.R, Perc.J, Perc.T) = Perc.RJT Olympic games. Bold entries: minimal generator of the dominance graph, normal entries: dominance relations implied by transitivity. Rows 21 – 26 deleted, because there are no more dominance relations. Rows 19 and 20 interchanged, because 19 surpasses 20 in 1500m. Rows 19 and 20 interchanged because 19 surpasses 20 in 1500m (Rank9 $\neq$ Rank10).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	20	19	21	22	23	24	25	26
1	X		**1**	**1**	**1**		**1**	**1**	**1**	**1**	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2		X						**1**		**1**									**1**	1	1		1	1	1	
3			X										**1**	1	**1**	1		1	**1**	1	1		1	1	1	1
4				X							**1**	**1**		1		1	1						1	1	**1**	1
5					X									**1**		**1**		**1**				**1**	1	1		1
6						X								**1**		**1**							1	1		1
7							X								**1**	**1**		1	**1**	1	1			1	1	1
8								X															**1**	**1**	**1**	
9									X							**1**		**1**		**1**				1	**1**	1
10										X											**1**		**1**	**1**	**1**	
11											X			**1**		**1**	**1**						1	1		1
12												X					**1**						**1**	1		**1**
13													X	**1**				**1**		**1**	**1**		1	1	**1**	1
14														X									**1**	**1**		**1**
15															X			**1**		**1**	**1**			1	**1**	1
16																X								**1**		**1**
17																	X							**1**		
18																		X						**1**		**1**
20																			X	**1**	**1**			1	**1**	
19																				X				**1**		

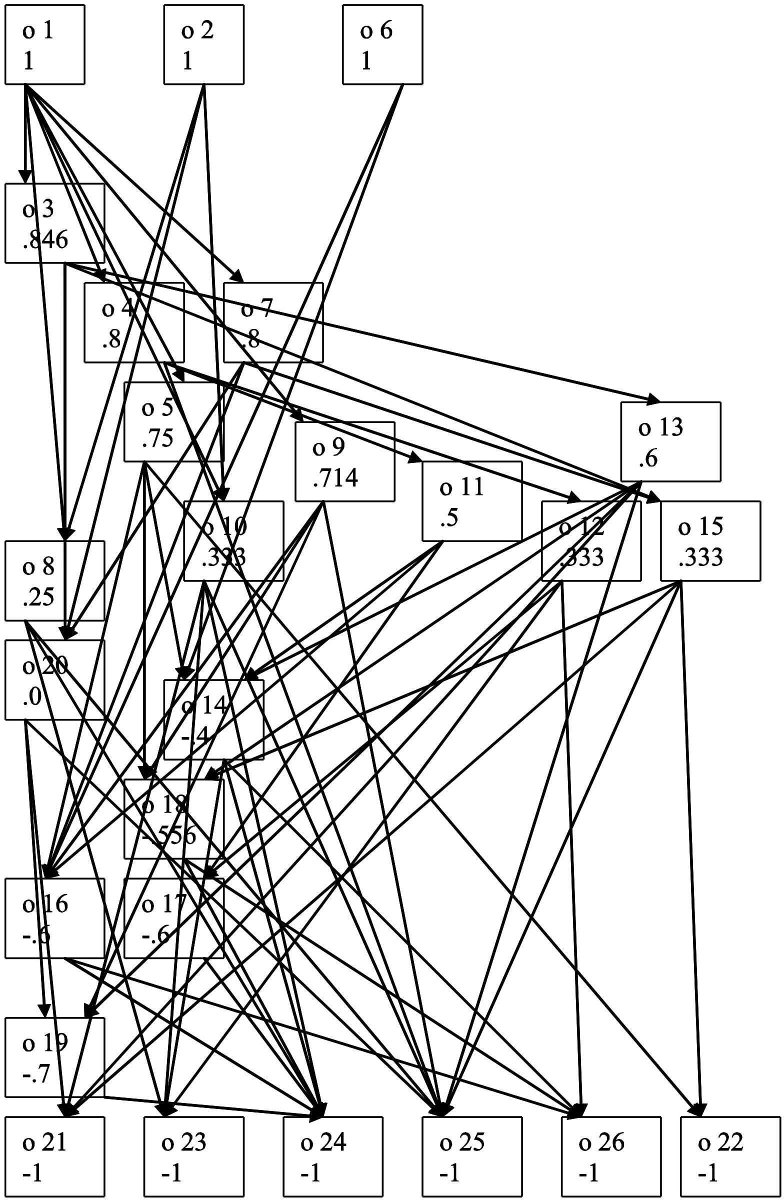

Figure 1:
Minimal generator of dominance relations Olympic Games 2008 (o 1 – o 26 the competitors). Left option *x* dominates right option *y*: $x \to y$ .The crossings of arrows illustrate the deviations of the multidimensional order from a linear order.

Empirical example: decathlon world championship 09

The decathlon world championship 09 was held in Berlin. 34 athletes completed the contest. Table 3 gives the results of the competitors.

The subjects are identified by their final position according to the sum of points (1= best). Rank of P9 gives the rank of the sum of points after 9 competitions (without 1500m). The largest difference between points 10 and points 9 is 8 positions for subject 26. Points 9 and points 10 have 9 out of 10 components in common or 90%. For 90% of common variance we would expect a correlation of 0.95. The actual correlation is 0.943 (Tab. 4, which shows again that 1500m does not fit well with the rest of the disciplines). Points 9 and the percentile score 9 correlate at 0.98. The largest difference in positions is 5 for subject 25 (weakness in T, throwing). Percentile 9 is the percentile score for 9 competitions (attributes). Rank of Perc.9 and rank of Perc.RJT correlate at 0.98. Rank of Perc.RJT is the rank of the percentile scores after computing the percentile scores of R (running, 100m, 400m, 110m hurdles), of J (jumping, long, high, pole vault) and of T (throwing, shot put, discus, javelin) separately. The largest difference between Perc.9 and Perc.RJT is 4 positions for subjects 12 and 21. The advantage of Perc.RJT is that you can give separately the ranks of the percentile scores of R, J and T. The largest difference in positions in R, J, and T is 31 positions for subject 18 (strong in J, weak in T). A further advantage of Perc.RJT is that varying achievements between disciplines are compensated within groups of more similar disciplines first.

The group percentile scores Perc.R, Perc.J and Perc.T have an advantage e.g. versus Perc.10 because they are based on a greater number of dominance relations between competitors because dominance relations on a smaller number of correlated disciplines are more frequent within groups than between groups. They therefore are more precise and more stable. The group percentile scores also give a reasonable clustering of competitors (see Fig. 2).

Figure 2 gives the dendrogram of the 34 competitors based on the rank correlations between their vectors of group percentile scores. The competitors can be clustered by hierarchical cluster analysis with the method of nearest neighbour (SPSS 15.0 for Windows). The competitors with the most similar rank orders of group percentile scores are grouped closest together. There result 3 clusters or types of athletes.

The largest cluster with 19 athletes is cluster I with the members (in the order of recruitment, more similar members in brackets):

(3,19,2,17,11), (5,34), (8,24,1), (12,26), (21,31,7,6,33,20,28).

Cluster I comprises 14 athletes who have their greatest strength in T (throwing). They perhaps are characterised by muscular explosivity and coordination (between arms and legs).

The second cluster is cluster II with the members:

(18,30,4,16,13), (14,27,15,25,9).

(8 from 10 members of cluster II have their greatest strength in R (running). Perhaps rhythmic expenditure of force of legs over several seconds is their advantage.

Table 3:
The ranks of the competitors of the decathlon world championship 09. The competitors are numbered by their official final position (1 = best). Points9 = official points without 1500m running. Perc. = rank of percentile score. 10, 9, R, J, T, RJT…ten, nine disciplines, running, jumping, throwing, combined R, J, T.

Competitor Points10	Points9	Perc.10	Perc.9	Perc.R	Perc.J	Perc.T	Perc.RJT
1	1	1	1	1	4	2	1
2	3	3	3	12	6	4	4
3	2	4	2	16	1	3	2,5
4	6	2	4	3	5	12	2,5
5	11	15	13	20	16	11	15
6	5	5	6	7	19	1	6
7	10	8	11	15	21	7	12
8	4	6	5	2	17	6	5
9	7	7	7	4	10	14,5	7
10	9	9	9	13,5	2	21,5	10
11	8	13	8	24	7	5	9
12	13	14	15	21	23	8,5	19
13	20	12	16	8	9	29	13
14	14	11	12	5,5	18	19	11
15	12	10	10	5,5	14	16	8
16	16	16	17	11	15	23	17
17	15	18	18,5	27	13	8,5	16
18	19	19	18,5	9	3	34	14
19	18	22	21	32	12	13	21
20	21	20	20	18,5	25	18	23
21	17	17	14	17	24	10	18
22	27	26	30	31	20	31	30
23	22	21	22	26	11	21,5	22
24	23	23	23	10	29	14,5	20
25	30	24	25	13,5	27	33	27
26	34	28	31	29	30	25	31
27	26	25	26,5	18,5	28	29	29
28	25	27	28	22	31	20	26
29	33	29,5	32	30	26	32	33
30	24	29,5	24	23	22	27	25
31	29	32	29	25	25	17	28
32	28	31	26,5	15	8	29	24
33	32	33	33	28	33	26	32
34	31	34	34	34	32	24	34

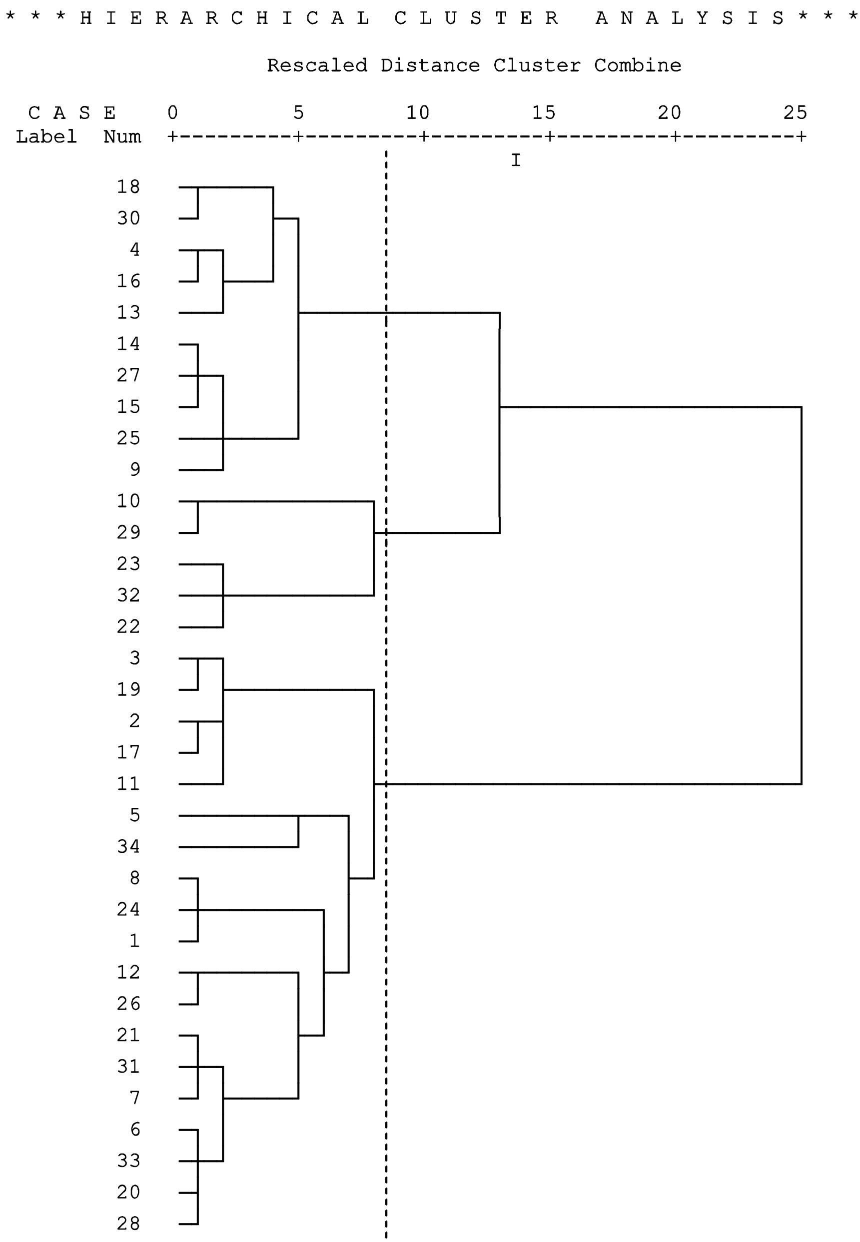

Figure 2:
Dendrogram of 34 competitors using single linkage. The broken line marks the point where further clustering causes relatively larger increase in heterogeneity (distance) within clusters than in heterogeneity between clusters.

The third cluster is cluster III with members:

(10,29), (23,32,22).

All of them have their greatest strength in J (jumping) requiring perhaps concentration on an external mark and explosive expenditure of force of legs.

How well are the 3 disciplines within the groups R, J, T represented by the group percentile score Perc.R, Perc.J, Perc.T? Since all single disciplines and the group scores basically are coded by the ranks 1 to 34 there should be exactly no difference between the difficulties of disciplines or groups of disciplines. The actual data, however, have different frequencies and positions of tied ranks and this results in small (negligible) differences larger than 0. The goodness of fit of a single group percentile score to the vector of the 3 component disciplines can not be tested strictly formally (Scheiblechner, 2003). But there are descriptive deviation statistics which can be roughly compared to likelihood ratio statistics between the data ordered nominally and the data ordered by ranks. The descriptive statistic for the deviation is twice the log likelihood ratio of the two models and the degrees of freedom are the difference of the numbers of parameters in the two models.

The descriptive statistics for the fit of the ordinal model (ISOP) to the nominal model and the degrees of freedom for the three groups are

	Deviation	Degrees of freedom
R (running)	322.4	2805
J (jumping)	389.2	2913
T (throwing)	430.5	2633

For random data the expectation of the deviation statistic is the number of degrees of freedom which here is more than 5 times the actual deviation in all groups. That means that in all cases the 3 disciplines are very well represented by their common percentile score or that the rank of the group percentile score represents a great portion of the variance of the ranks in the component disciplines.

Similarly the fit of an additive interval model (ADISOP; Scheiblechner, 1999) to the ordinal model (ISOP) can be judged by a deviation statistic and corresponding degrees of freedom:

	Deviation	Degrees of freedom
R (running)	82.2	195
J (jumping)	16.8	181
T (throwing)	30.8	173

The expectation of the deviation for random data is more than twice the actual deviation in all groups. The data could be represented acceptably by an additive interval model.

The corresponding descriptive statistics can be computed for the representation of the group scores Perc.R, Perc.J, Perc.T by the percentile score Perc.RJT

	Deviation	Degrees of freedom
Nominal versus ordinal (ISOP)	364.8	1919
Ordinal (ISOP) versus interval (ADISOP)	54.4	192

Table 4:
Rank correlations of points P10 (sum of points of 10 disciplines), points P9 (sum of points of 9 disciplines), and percentile scores Perc.10 (percentile score for 10 disciplines), Perc.9 (percentile score for 9 disciplines), Perc.R (percentile score for running), Perc.J (percentile score for jumping), Perc.T (percentile score for throwing) and Perc.RJT (percentile score for the 3 percentile scores).

Spearman -Rho	R of P10	R of P9	R of Perc.10	R of Perc.9	R of Perc.R	R of Perc.J	R of Perc.T	R of Perc.RJT
R of P10	1	.943	.966	.942	.677	.691	.718	.919
R of P9		1	.951	.979	.676	.703	.804	.960
R Perc.10			1	.973	.784	.688	.698	.959
R Perc. 9				1	.749	.720	.749	.981
R Perc. R					1	.398	.356	.775
R Perc. J						1	.302	.763
R Perc. T							1	.705
RPerc.RJT								1

The expectations of the deviations for random data are more than 5 times larger than the observed deviations and the data can be represented very well by the successively more restrictive models. The largest difference in rank positions between Rank Points9 and rank Perc.RJT is 7 for subject 13.

Multiattributive decisions

Let for each attribute $X_i(a)$ a strict partial order of the elements of the decision set A be given. In a social group decision problem the attributes may be the rank orders (with possible ties) of the alternatives by the group members or the alternatives rated by the group members by the ordered categories very attractive, attractive, … , not attractive. The strict orders between different grades are transitive and the indifferences within one grade may be intransitive. In an individual decision problem the attributes are the rank orders or ratings of the alternatives with respect to single attributes. We search for a social decision function which does not rely on comparisons of preference intensities of different individuals or an individual decision function which does not necessitate the comparison of preferences among different attributes.

A scaling function which attributes an overall value to each choice alternative (combination of attributes) is a decision function because it specifies the best alternative(s) for each subset of feasible or available alternatives (and for each subset of attributes).

Lexicographical ordering:
If the attributes (or group members) are linearly ordered in importance (rank order without ties) then the lexicographical order is the desired decision function.

Lexicographical ordering with aspiration level:
The elements of the decision set that lay below the aspiration levels x_i^0 on given attributes *i* are excluded and the remaining elements are ordered lexicographically.

The (modified) percentile score:
The competitors of the decathlon could be ranked in many different ways, depending on how the performances are transformed into "points" within the disciplines and how the points are combined over disciplines. We want to have a social decision function which does neither require a differential weighting of disciplines nor a comparison of differences between achievements across different disciplines. We only require strict partial orders and no metric measurement (metric = at least interval scales) within disciplines. For example, the running disciplines are seemingly measured on the interval or rational scale of physical time, but to run 100m in 9.6 seconds instead of 9.7 is much more difficult than to run it in 12.1 instead of 12.2 seconds (the physical time scale is no appropriate scale of difficulty or level of achievement, it functions only as ordinal scale within single disciplines).

How the percentile score does satisfy these requirements was demonstrated with the decathlon example. However, an artificial example with only two subjects in school is much more transparent. An academic board has to assign grants to the most gifted (successful) 10% of 100 applicants. The achievements of the pupils in 2 subjects rated from 1 to 5 points are known (Tab. 5).

The highest percentile score 1 is given to applicants with 5 points in E and 5 points in D and also to 4 points in E and 5 points in D ((5, 5) and (4, 5) for short, Tab. 6). There are no pupils with these achievements. Therefore (5, 4) and (3, 5) get the score 1 as local suprema and there are 2+3 = 5 pupils with score 1.

We can assign 10 grants. We need 5 more applicants. The next highest are (5, 3) and (4, 4) with percentile score 0.9524. There are 3+3 = 6 such pupils. We select 5 pupils from this group at random. We needed no decision about relative weights of points in D and E and we needed no assumption about distances between neighbouring points within and between subjects. Similarly we could eliminate the 50% weakest applicants if, say, only 50% are admitted to an advanced course.

Table 5:
Frequency of points in a difficult subject D and an easy subject E.

		D					
	points	1	2	3	4	5	Sum
E	1	8	7	4	1	0	20
	2	8	5	4	1	2	20
	3	10	9	5	3	3	30
	4	7	6	4	3	0	20
	5	2	3	3	2	0	10
	Sum	35	30	20	10	5	100

Table 6:
Percentile scores of grade vectors of Tab. 2.

	1	2	3	4	5
1	-1	-0.7576	-0.3478	0.1515	0.6
2	-0.8	-0.3521	0.1034	0.48	0.8537
3	-0.5152	0.0857	0.5068	0.7714	1
4	0.0612	0.5652	0.8025	0.9524	1
5	0.6098	0.8507	0.9524	1	1

The Nobel committee can seemingly assign 0 to 3 prices in one domain (we are not sure whether a price can be transferred from one domain to another).

Consider a community which can invest an amount *a* to health care and an amount *b* to local traffic. The decision is found by voting (cf. Fig. 3).

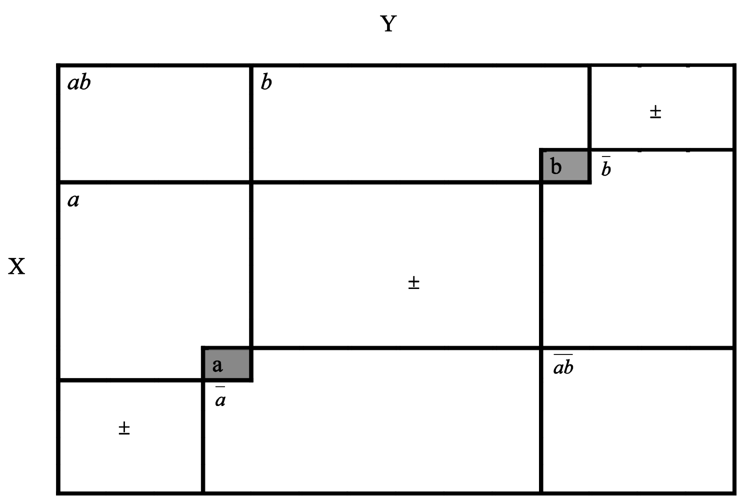

Figure 3:
A voting of option *a* versus *b*. Voters to the left and above box 'a' promote alternative 'a' because their position is smaller in X and Y and 'a' is an improvement to them. Voters to the right and below 'a' reject 'a' because 'a' is a deterioration in X and Y to them. Voters in *ab* promote both alternatives, voters in $\overline{ab}$ reject both alternatives. Voters in $\pm$ are indifferent because they improve in one attribute and loose in the other. Voters in *a* promote option 'a' and voters in $\bar{a}$ reject 'a'. If the number of voters in $(a+\bar{b})-(b+\bar{a})$ is positive then 'a' is the social choice. The other voters are indifferent and waive their right to vote or do their vote randomly.

The Condorcet paradox

A frequently used group decision procedure is that each member establishes a linear rank order of the choice alternatives (without ties) and the group decision is the alternative with the largest sum of ranks if the ranks are weighted from m-1 > m-2 > ... > 1 > 0 or by $a_1 > a_2 > \ldots > a_m$ from best to least preferred of m positions. Condorcet showed that this 'method of marks' may result in a decision which does not have a strict simple majority in pair comparisons over all other alternatives. Consider the $N = 5$ linear rank orders of $m = 5$ alternatives (Fishburn, 1973, p 147):

Rater 1. *x y a b c*

Rater 2. *y a c b x*

Rater 3. *c x y a b*

Rater 4. *x y b c a*

Rater 5. *y b a x c*

Alternative *y* has the largest sum of marks or the majority of best positions but it does not have the simple majority if the choice set is restricted to $\{x, y\}$ because 3 members prefer *x* to *y*.

The flaw of the method of marks is that it tacitly assumes that the distances between succeeding marks are the same in all rank orders and that distances can be added or compensate (cancellation) each other thus making some sort of interval scales of the rank orders.

The correct approach to the above problem is to consider the ranks of the alternatives in the 5 rank orders as 5 attributes of the objects *a, b, c, x, y* giving:

Alternative x: ranks (1, 5, 2, 1, 4)

Alternative y: ranks (2, 1, 3, 2, 1)

Alternative a: ranks (3, 2, 4, 5, 3)

Alternative b: ranks (4, 4, 5, 3, 2)

Alternative c: ranks (5, 3, 1, 4, 5)

In this multiattributive choice set y dominates *a* and *b* and the percentile scores are as in Figure 4.

The percentile scores of *x* and *c* are undetermined showing that they can not be integrated in the partial order of the remaining objects. By the weak Condorcet condition (Fishburn, 1973, p. 146) *x* has the simple majority over *y* and the objects lying between them should play no role. Excluding dominated alternatives {*a*, *b*} from the choice set by the reduction condition (Fishburn, 1973, p. 148) gives {*x*, *y*} as most preferred alterna-

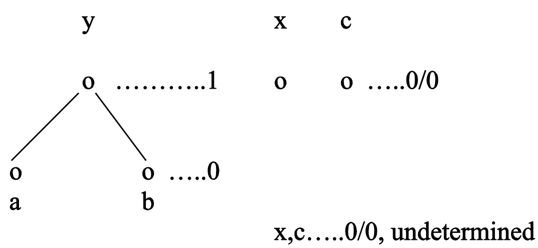

Figure 4:
Partial order of objects of Condorcet paradox – Percentile scores.The percentile scores of *x* and *c* are undetermined.

tives by the method of marks. The percentile scores are undetermined for the reduced set. The simple majority still selects *x*. There exists no logically satisfactory social choice function for the set of all 5 alternatives.

Another example where an undetermined percentile scale score indicates the non-existence of a stable solution (a saddle point) is the prisoner's dilemma game. Let the payoff matrix be

		C	
		confess	deny
R	confess	(2,2)	(4,1)
	deny	(1,4)	(3,3)

where R is the row player, C the column player; the first number in each cell is the rank of the outcome for R, the second number is the rank of the outcome for C, both from 1 best to 4 worst. The outcomes (4, 1) and (1, 4) have undetermined percentile scale scores, the cell (confess, confess) has +1 and the cell (deny, deny) has score -1. (The cell (confess, confess) is the solution for repeated games).

Theorem. If the choice set is fixed and does not contain options that can not be ordered relative to at least one other option ($n^- + n^+ > 0$ for all options) and if the minimal generator of the graph of the order (Fishburn, 1973, pp. 78) does not split into several isolated components (subgraphs; the graph is connected) then the (modified) percentile score gives the scale value (ordinal scale) of the options. The suprema of the order are standardized to the value +1, the infima to -1.

If in Arrow's proof of the 'impossibility theorem' we replace 'weak order' by 'strict partial order' in conditions C3 and C4 of the 7 conditions (cf. Fishburn, 1973, theorem 16.1, pp. 204) then we get the above 'possibility theorem'.

If the graph of the partial order is divided into two or more disconnected branches subgraphs), then the decision problem is not well defined. Some alternatives or some attributes have to be eliminated.

References

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

Goodman, L.A., &, Kruskal, W.H. (1954, 1959, 1963, 1972). Measures of association for cross classifications, Part I – IV. *Journal of the American Statistical Association, 49* (732-764), *54* (123-163), *58* (310-364), *67* (415-421).

Scheiblechner, H. (1995). Isotonic ordinal probabilistic models (ISOP). *Psychometrika, 60*, 281-304.

Scheiblechner, H. (1999). Additive conjoint isotonic probabilistic models (ADISOP). *Psychometrika, 64*, 295-316.

Scheiblechner, H. (2002). Note on nonparametric IRT: Scoring functions, subject parameter estimation and further model controls of isotonic probabilistic models (ISOP). Preliminary draught: URL:http://www.staff.uni-marburg.de/~scheible/isoscore2.pdf.

Scheiblechner, H. (2003). Nonparametric IRT: testing the bi-isotonicity of isotonic probabilistic models (ISOP). *Psychometrika, 68*, 79-96.

Scheiblechner, H. (2007). A unified nonparametric IRT model for d-dimensional psychological test data (d-ISOP). *Psychometrika, 72*, 43-67.

Scheiblechner, H., &, Lutz, R. (2009). Die Konstruktion eines optimalen eindimensionalen Tests mittels nichtparametrischer Testtheorie (NIRT) am Beispiel des MR SOC. (The construction of an optimal uni-dimensional test by means of nonparametric test theory (NIRT) by way of the MR SOC as example). *Diagnostica, 55*, 41-54.

van der Linden, W.J., &, Hambleton, R.K. (eds.), (1997). *Handbook of modern item response theory*. New York: Springer.