IMPROVING THE SKATING SYSTEM - II: METHODS AND PARADOXES FROM A BROADER PERSPECTIVE

by Xavier Mora*
Professor of Mathematics, Universitat Autònoma de Barcelona (Spain)
Dance Sport Scrutineer, Certified by the British Dance Council
Dance Sport Judge, Asociación Española de Baile Deportivo y de Competición

30th November 2004

> Frederick Mosteller, 1996:
> "My first paper on statistics in sports dealt with the World Series of major-league baseball. At a cocktail party ... someone asked: What is the chance that the better team in the series wins? Some people did not understand the concept that there might be a 'best' or 'better' team, possibly different from the winner."

Though *essentially self-contained*, this article is motivated by [**2**], where certain proposals were put forward for improving the system used in the scrutineering of dancesport competitions. In this article the subject is discussed from a broader perspective, with the aim of better gauging those proposals in comparison with other possibilities. In this connection, it will be demonstrated that one of the systems proposed in [**2**] is definitely superior to the traditional one. On the other hand, we shall present an alternative —the LCO System— that significatively improves upon both of them. In fact, we shall see that this alternative achieves about the best possible combination of desirable properties (including pen and paper computation). All of these statements will be substantiated by means of both rational arguments and thorough simulations. Another significant contribution of this article is a consistent method for translating the final ranking into a finely tuned rating of the contestants. Besides dancesport, we shall also have in mind the case of figure skating, whose latest developments will be analyzed in some detail. On the other hand, the heart of the matter is not specific of dancesport nor figure skating, and some parts of this article may be interesting for other kinds of judging or voting processes.

1. Over two hundred years of experience!

The scrutineering system used in dancesport is traditionally called "skating system". This is due to the fact that it was originally borrowed from figure skating. Nowadays, dancesport and figure skating use somewhat different scrutineering systems (see § 13), but the problem is still essentially the same: There are several couples (or individuals), several dances (or sections), and several judges; for every dance or section, each judge assesses the performances of the different couples and expresses his particular preference opinion about

URL: `http://puffinet.com/escrutini/iss2en.pdf` (without technical appendices),
`http://mat.uab.cat/~xmora/articles/iss2Aen.pdf` (with technical appendices).

This work benefited from the opportunity of posing the problem in the workshop *Study Groups of Mathematics and Technology*, organized by the *Universitat Politècnica de Catalunya* and held in Barcelona on July, 8th–10th, 2002.

them in the form of a ranking (or a more quantitative rating); the problem consists in combining these particular opinions into a global result.

Problems of this kind are not exclusive to dancesport or figure skating, but they occur in voting processes of all kinds, i. e. whenever a common decision must be taken on the basis of multiple individual preferences that concern a certain set of options. For instance, in political elections there are (hopefully) several candidates to choose from, and the individual preferences of the electors, possibly numbering a lot, must be summarized into a unique global result.

At first sight, these problems seem rather elementary. However, as soon as more than two options come into play, serious unexpected difficulties appear. These difficulties take often the form of paradoxes, i. e. situations where apparently reasonable methods are discovered to contradict apparently reasonable expectations. As a result, the matter becomes much less elementary than initially expected. This, together with the importance of some of its applications, has made it the subject of a whole branch of social science, which is known as **social choice theory**.

This discipline was founded essentially in the times of the French Revolution, when democratic methods were being paid a special attention. Its foundations were laid down mainly by the French mathematician Marie-Jean-Antoine-Nicolas de Caritat, marquis de Condorcet, who will be mentioned repeatedly throughout this article. Since that time, the subject has been analyzed further by many other people, and currently it is still a very active academic field. Most of the methods, difficulties and paradoxes that we are meeting in dancesport scrutineering have already been met and discussed in social choice theory. Certainly, this calls for trying to take advantage of these two hundred years of specialized experience, and this is indeed the aim of this article.

For more details about social choice theory, the reader is referred to the bibliography, where we have included some plain general overviews of the subject [**3, 4**], some detailed historical accounts [**5, 6, 7**], some specialized expositions [**9, 11, 12**], and some web sites devoted to the subject [**15, 16, 17**].

2. Some generalities.

In general terms, we are dealing with a set of different **items** which are the matter of certain **valuations**; in fact, we are given several **particular** valuations and the problem consists in combining them into a **global** one. In the case of dancesport, the items are couples, and the particular valuations correspond to different judges and possibly different dances.

By a **valuation**, or opinion, or set of preferences, about a set of items we mean any collection of data that provides information about how do they compare with each other.

Most often, this will be done by means of a **ranking**, i. e. an ordering of the items in question. A ranking consists thus in a simple enumeration of the items in their order of preference. Equivalently, it can be specified by giving the placing, or rank, of each item in that ordering, i. e. 1 for the best item, 2 for the next one, and so on. Occasionally, one is led to allow for the possibility of ties between two or more consecutive items. We shall refer to such a valuation as a **weak ranking**, and the case without ties will then be stressed by calling it a **proper ranking**.

Sometimes, valuations are expressed by means of more quantitative marks, or rates, in which case we shall be talking about a **rating**. A rating contains thus more information than a simple ranking. For the purpose of comparing methods with each other, in this article rates will often be rescaled, and possibly reversed, so that the best possible rate is 1.000 and the worst possible one equals the number of items.

The different valuations will often express the opinion of several **judges**, or voters. Another possibility is that they do not correspond to different people, but to several **fields** of valuation. Finally, it is also possible that they combine both several judges and several fields. This is usually the case of dancesport, where the fields of valuation are different dances, and the global result is supposed to reflect *at the same time* the general opinion of the judges and the all-round efficiency in the whole set of dances included.

Some important issues arise only in this last compound case. In spite of that, most of this paper will be devoted to the case of either several judges or several fields, but not both. As we shall see, the fundamental difficulties start here, and obviously a good understanding of these simpler cases is highly desirable before considering the compound one. In particular, we shall see that combining different judges has a different character than combining different fields.

Unless we say otherwise, by *Skating System* we shall always mean the scrutineering system used in dancesport (not figure skating). Besides the version currently in use, which is described in [**1**] and shall be referred to as *Traditional Skating System* (TSS), we shall consider also the two alternative versions of it presented in [**2**], namely the *Revised Skating System* (RSS) and the *Double Revised Skating System* (DRSS).

For illustrative purposes, we shall often use the example G given below, where there are seven items, labeled 61 to 67, and five different rankings, labeled A to E. As it is usual in dancesport, each ranking is specified by means of the placings given to the different items.

Example G

	Data				
Id	A	B	C	D	E
61	1	1	7	7	7
62	4	2	3	3	1
63	2	6	2	4	2
64	3	3	1	2	5
65	6	4	5	6	4
66	7	5	6	1	3
67	5	7	4	5	6

3. The plurality method.

Let us begin by noticing that in example G, couple 61 has more first placings than any other. Of course, in view of the other marks, we are not easily convinced of its deserving a global first place. Let us assume, however, that the particular valuations correspond to different judges and each of them had been asked to tell only which couple is his most preferred one, and nothing more. In other words, we do not know the whole table, but only which of its cells contain a first-place mark. In a large variety of voting contexts, one is indeed supposed to take a decision on such grounds only. In that case, certainly the only reasonable decision is to allocate a global first place to couple 61. Such a procedure is often called the *plurality* method (as opposed to a *majority* method).

When each judge expresses not only his most preferred option but his whole preference ranking, then things can look very different. In fact, in Example G we have a *majority* of judges, namely C, D and E, that agree on putting couple 61 in the last position. Therefore, a plurality winner can be at the same time the less desired option! This remark is referred to as **Borda's paradox**, after Jean-Charles de Borda, who in 1770–1784 pointed out such a possibility in connection with election procedures [**6**: chap. 5]. As a solution, he proposed the method that will be described in the next section.

Certainly, the possibility of situations like the one above makes it desirable to avoid the plurality method when we are combining different judges.

4. Addition; Borda's method.

4.1. As a solution to the preceding criticism of the plurality method, Borda proposed that each voter should express not only his most preferred item, but a complete ordering of all of them, just like it is done in dancesport final rounds; after this information has been collected, he proposed to add up the marks obtained by each item and to rate the items by the resulting values S. The results are shown in next table, where, for comparison with other methods, we have displayed not only the sum S, but also the average A, i. e. S divided by the number of particular rankings, and the resulting global ranking R. As one can see, the winner by this system is couple 62.

EXAMPLE G. ADDITION METHOD.

	Data							
Id	A	B	C	D	E	S	A	R
61	1	1	7	7	7	23	**4.600**	*5*
62	4	2	3	3	1	13	**2.600**	*1*
63	2	6	2	4	2	16	**3.200**	*3*
64	3	3	1	2	5	14	**2.800**	*2*
65	6	4	5	6	4	25	**5.000**	*6*
66	7	5	6	1	3	22	**4.400**	*4*
67	5	7	4	5	6	27	**5.400**	*7*

4.2. Borda's method can be criticized for two reasons. In [**2**: § 2.2] we already pointed out a first criticism, which we shall call the **strong flip-flop paradox**. It consists in the remark that a deletion of the worse rated items may alter the global ordering of the better rated ones. In example G this happens, for instance, if we delete all couples with A greater than 4. If the ranks of the remaining couples are changed accordingly, and one applies again the addition method, one obtains the following result:

EXAMPLE G′. ADDITION METHOD.

	Data							
Id	A	B	C	D	E	S	A	R
62	3	1	3	2	1	10	**2.000**	*2*
63	1	3	2	3	2	11	**2.200**	*3*
64	2	2	1	1	3	9	**1.800**	*1*

So, the winner is now couple 64. This is certainly most undesirable: even assuming the same performances, and the same preferences of the judges, the winner may depend on the presence or absence of other couples of a lower merit!

The preceding criticism of Borda's method was already formulated in 1785–88 by Condorcet [**6** : chap. 1, § 5.5], and in 1803 it was expressed again by Pierre-Claude-François Daunou, a French statesman and historian [**6** : chap. 11, p. 245–249].

Certainly, one would prefer a method with the property that *a deletion of the worse rated items should not alter the global ordering of the better rated ones.* Together with this condition, it makes sense to require also the analogous one where the items being deleted are the best rated ones. In the following, these two conditions will be viewed as two sides of the same property, which will be called **consistency with respect to losers and winners**.

Instead of this condition, one could require the stronger version that a deletion of any subset of items (not necessarily the worse rated or best rated ones) should not alter the global ordering of the remaining ones. In the specialized literature, this stronger condition is usually referred to as *independence of irrelevant alternatives*, and the weaker version introduced above is often called *local independence of irrelevant alternatives.* Conditions of this kind were considered in 1950 more or less independently by Kenneth J. Arrow and John Forbes Nash Jr, and later on they were discussed also by Amartya Sen, all of them future winners of the Nobel prize for Economics (respectively in 1972, 1994 and 1998). A most celebrated result of K. J. Arrow consists in the mathematical proof of the *impossibility* (!) of a system satisfying at the same time several apparently reasonable conditions together with the condition of (general) independence of irrelevant alternatives. Later on (§ 6.3) we shall see that this condition is really asking too much. By the way, this means that one cannot avoid the *weak flip-flop paradox*: sometimes a deletion of some items may alter the global ordering of the remaining ones.

4.3. A second criticism against Borda's addition method is that it is too easily affected by insincere voting. Or, to put it in more objective terms, it is *too sensitive to eccentric marks.* For instance, let us consider the example shown next at the left-hand side, where there are seven couples, of which we are interested only in two of them, and five judges:

Example H

Id	*Judges* A	B	C	D	E	*S*	*R*
71	1	1	1	1	7	**11**	*2*
72	2	2	2	2	1	**9**	*1*

Example H′

Id	*Judges* A	B	C	D	E	*S*	*R*
71	5.9	5.9	5.9	5.9	5.5	**29.1**	*2*
72	5.8	5.8	5.8	5.8	6.0	**29.2**	*1*

According to the addition method, the winner would be couple 72, in spite of the fact that a *majority* of judges, in fact all of them but one, agree that the winner should be couple 71. Notice that the problem does not lie in the ordinal character of the marks. In fact, the "quantitative" marks (the higher the better) of the right-hand side do exactly the same job. The problem is that the eccentric marking of judge E is having too much influence.

In this connection, Borda himself is reputed to have recognized that

> "My scheme is only intended for honest men".

In 1795, Pierre-Simon, marquis de Laplace, a celebrated mathematician, had tried to provide a mathematical justification of Borda's addition method. Later on, he recognized that

> "This mode of election would no doubt be the best, if it were not true that considerations alien to merit often influence the electors, even the most honest ones, and move them to put in the last places the candidates that are most threatening to their favourite, which gives a great advantage to candidates of mediocre merit. Besides, experience has led establishments that once adopted this mode of election to stop using it".

We can also quote Sir Francis Galton, a naturalist and statistician who was a cousin of Charles Darwin. In 1907 he published several articles about voting procedures, where he argued as follows:

> "How can the right conclusion be reached, considering that there may be as many different estimates as there are members? That conclusion is clearly *not* the *average* of all the estimates, which would give a voting power to 'cranks' in proportion to their crankiness".

This criticism, which will be easily shared by most dancesport people, calls for avoiding Borda's addition method whenever combining the opinion of different judges.

4.4. When combining different fields the preceding criticism does not apply, and addition fits in very well with the idea of all-round quality. However, to suit that purpose *the marks being added up should be more sensitive than simple ordinal numbers*.

Since long ago, the Skating System has been avoiding the use of the addition method for combining different judges, and it uses it only for combining different dances. The Traditional Skating System does that in Rule 9, which is indeed concerned with combining dances, but it incurs the fault of adding up ordinal numbers. The Revised Skating System corrects this fault to a certain extent, since its Steps 2.1 and 2.2 add up more information than just ordinal numbers. Finally, the Double Revised Skating System is again guilty of adding up simple ordinal numbers; this happens in Step 3 (combining dances for every judge), where there is no other information available than ordinal numbers.

In the following sections (§ 5–10) we shall specialize in the case where the particular valuations that we are combining correspond to several judges (not dances), which shall allow us to use a somewhat simpler terminology.

5. Median.

5.1. The criticism of the addition method because of its sensitivity to eccentric marking (§ 4.3) has something in common with the previous criticism of the plurality method (§ 3). Indeed, both arguments coincide in paying a fundamental attention to the opinion of a *majority* of jugdes, i. e. more than one half of them. In this connection, everybody will agree that *if a majority of judges agree on allocating the first position to the same couple, then this couple should win*. In the following we shall refer to this fundamental condition as the **majority principle**.

Much often, however, there is no couple with a majority of first places. Such a situation happens for instance in example G above. Which couple should then be declared the winner? On the other hand, how should we allocate the remaining positions? Is there a natural answer to these questions on the basis of the opinion of a majority of judges? As we shall see, there are two such natural answers, one of which will be dealt with in the remainder of this section, and the other in § 6.

5.2. The first approach is essentially the one followed by the Skating System in its first part, i. e. when combining judges within each dance. That system entered the ballroom dancing scene in 1937/38. In 1937 it was tried out at the Blackpool Festival, and in the following year it was adopted by the *Star Championships*, where

> "it was definitely announced that the judge's markings would be calculated on the 'skating system,' thus removing all danger of the strange results that had occurred in this competition in the past" (Philip J. S. Richardson, *A History of English Ballroom Dancing*, London, 1945).

In fact, figure skating had been applying the majority principle since 1895. On the other hand, even in 1937/38, this principle superseded the addition rule only in the special case where a couple had been placed first by a majority of judges; in any other case the winner was determined by means of the addition method. The form of the Skating System used today in dancesport is the result of a substantial evolution from those origins. In particular, it was not until 1948 that the most significative rules in connection with the majority criterion took its present form:

> Rule 5. "The winner of a particular dance is the couple who is placed first by a majority of the judges; second, the couple who is placed second or higher by a majority. The remaining positions are allocated in a similar way".
>
> Rule 8. "If no couple receives a majority of 'firsts' then the winner is the couple who are placed 'second and higher' by a majority of judges. If no couple receives a majority of 'first' and 'second' places, then the 'third' places (and if necessary, lower places) must be included (subject to Rules 6 and 7). The 'second' and other positions should be calculated in a similar way".

Certainly, these rules can result in ties, in which case the Traditional Skating System applies certain less essential rules, Rules 6 and 7, whose statement will be omitted.

In the case of example G the winner according to Rules 5 and 8 is couple 63, the only one that has been placed second or first by a majority of judges.

The method defined by Rules 5 and 8 above is much older than the Skating System itself. In fact, it goes back to 1792/93, when it was introduced by Condorcet in a draft for a new French Constitution [**5** : chap. 10–12, see especially p. 249 **;** **6** : chap. 8]. As we shall see in next section, Condorcet had previously introduced a different more fundamental approach. But that approach was not practical for big elections, and as a practical alternative he proposed a procedure of the kind that we are discussing, which he advocated also as "least susceptible to factions and intrigue".

Condorcet's practical method was never implemented in the French Constitution, and indirectly it cost Condorcet his life: he died because of a pamphlet where he criticized Robespierre's alternative! [**5** : § 1.6 **;** **6** : p. 37]. Shortly afterwards, however, Condorcet's practical method was being used in the city-state of Geneva, and later on a variation of it known as Bucklin's system has been used in several states of North America.

5.3. In the dancesport usual case of an odd number of judges, Rules 5 and 8 can be reformulated in a much simpler way in terms of a standard statistical parameter called the **median**. Let us consider a set of marks coming from an odd number of different judges. By definition, the median of this set is the value that lies at the central position when these marks are arranged by order of magnitude. For example, the median of $(1, 1, 7, 7, 7)$ is 7, while the median of $(2, 6, 2, 4, 2)$ is the most central value in $(2, 2, \mathbf{2}, 4, 6)$, i. e. 2. In other words, the median is the smallest value with the property that a majority of the marks under consideration are less than or equal to that value. The reader will easily convince himself that Rules 5 and 8 above are exactly equivalent to ranking the couples by the median of the marks that they have obtained from the different judges.

By the way, this method was in fact the one that Sir Francis Galton recommended in 1907 instead of the addition method:

> "The estimate to which least objection can be raised is the *middlemost* estimate, the number of votes that is too high being exactly balanced by the number of votes that it is too low".

The fact that Rules 5 and 8 of the Skating System are equivalent to a median ranking rule has been remarked by Gilbert W. Bassett Jr and Joseph Persky [**24**], who have also examined the performance of this method from a statistical point of view (see also [**26**]).

In the case of an even number of judges, we can still arrange the marks by order of magnitude, but instead of a single most central value, we have to consider now the two most central ones, which may certainly be different from each other. This introduces some indefiniteness when it comes to extending the notion of median to the even case. In order to remain in agreement with Rules 5 and 8, one would have to take the largest of those two most central values. However, the standard definition of the median in the even case is not the largest but the average (arithmetic mean) of those two most central values. With this definition, which we shall adopt from now on, the median method does not agree completely with Rules 5 and 8, but in fact it is more equitable [**2** : § 6.6] .

5.4. When two or more couples have the same median, then the natural way to break the tie is to look at progressively extended central sums (or averages), thus keeping away eccentric marks until they cannot be avoided anymore.

These *central averages* are not different from the *trimmed averages* often considered in judge-rated sports. The difference in terminology reflects the fact that here we are looking at them as progressively extended medians rather than progressively trimmed averages.

Equivalently, instead of these progressively extended central sums we can use the *adjacent sums*, which are defined as follows: After the marks have been arranged by order of magnitude, the *1st adjacent sum* L_1 is the sum of the values lying at both sides of the median, for an odd number of judges, or at both sides of the two most central values, for an even number of judges; the *2nd adjacent sum* L_2 is the sum of the next two values at both sides of the ones already considered, and so on. For example, the 1st adjacent sum of the numbers $(7, 5, 6, 1, 3)$ is the sum of the bold-faced values in $(1, \mathbf{3}, 5, \mathbf{6}, 7)$, namely 9, while their 2nd adjacent sum is 8.

This **extended median method** is the one-dance version of the Revised Skating System proposed in [**2**]. In the case of example G, this method results in the following global ranking (where the symbol '$\succ$' means 'better than'): $63 \succ 62 \succ 64 \succ 66 \succ 65 \succ 67 \succ 61$. The medians and adjacent sums that lead to this result are shown in next table.

Example G. Extended median method.

	Judges					*Rearranged*								
Id	A	B	C	D	E						M	L_1	L_2	R
61	1	1	7	7	7	1	1	7	7	7	7	8	8	*7*
62	4	2	3	3	1	1	2	3	3	4	3	5	5	*2*
63	2	6	2	4	2	2	2	2	4	6	2	6	8	*1*
64	3	3	1	2	5	1	2	3	3	5	3	5	6	*3*
65	6	4	5	6	4	4	4	5	6	6	5	10	10	*5*
66	7	5	6	1	3	1	3	5	6	7	5	9	8	*4*
67	5	7	4	5	6	4	5	5	6	7	5	11	11	*6*

5.5. In the case of an odd number of judges, this method has the good property that addition of marks is relegated to tie-breaking rules. In contrast, its main criterion, i. e. the median itself, hinges only on comparing marks with each other, which is especially suitable to the case where marks are simple ordinal numbers.

In spite of this nice feature, the median method can still be put in doubt ! For instance, in example G we have seen that the winner by this method is couple 63 . However, one can argue that couple 64 should be considered better than 63 because a majority of judges, namely B, C and D, agree on that opinion. Indeed, judge B gave a 3rd place to couple 64 against a 6th place for couple 63, judge C gave them respectively his 1st and 2nd places, and judge D gave them respectively his 2nd and 4th places.

Such a criticism against the median method was already pointed out in 1794 by Simon-Antoine-Jean L'Huillier, a Swiss mathematician [**6** : chap. 9, p. 151–160].

To put it in a slightly different way, *the median method does not satisfy the condition of consistency with respect to losers and winners* stated in § 4.2. In fact, if we delete all couples with M greater than 3, and the ranks of the remaining couples are changed accordingly, a new application of the median method looks as follows, where instead of couple 63 the winner is now couple 64:

Example G′. Extended median method.

Id	*Judges* A	B	C	D	E	*Rearranged*					M	L_1	L_2	R
62	3	1	3	2	1	1	1	2	3	3	2	4	4	*2*
63	1	3	2	3	2	1	2	2	3	3	2	5	4	*3*
64	2	2	1	1	3	1	1	2	2	3	2	3	4	*1*

This problem is present in any median method. This includes the Traditional Skating System, the Revised Skating System, and the Double Revised Skating System. What we are saying is that *all these methods are liable to the strong flip-flop paradox* even in the case of just one dance. By the way, this lack of consistency is not just a theoretical problem, but in practice it shows up easily and disturbingly in the event of disqualifications!

What is it happening? If we carefully analyze the situation, we realize that when we compute the median, or equivalently when we apply Rules 5 and 8, we are *comparing the marks obtained by the same couple from the different judges.* In contrast, when we argue that couple 64 is considered better than 63 by a majority of judges, then we are *comparing the marks given by the same judge to the different couples.* Of both points of view, the second one lies certainly on more solid grounds. In fact, the marks that we are dealing with have been given separately by each judge as an expression of the way that he himself compares the couples with each other. In contrast, the median compares the marks accross different judges.

6. Paired comparisons.

6.1. The preceding criticism leads to a different approach, namely to compare items with each other on the basis of how are they compared "within" each of the judges (in contrast to comparing their ranks accross different judges). This is the main approach followed by Condorcet since 1785 [**6** : chap. 1, § 5.4]. Later on, in the 1870s, the same idea was followed, it seems that independently, by Charles Lutwidge Dodgson, alias Lewis Carroll [**6** : chap. 12].

Let us go back to example G. When we compared couples 63 and 64 we saw that 3 of the 5 judges preferred couple 64 over 63. Since 3 out of 5 is a majority, we inferred that couple 64 should be considered better than 63. In fact, we can see that in this sense couple 64 is better than any other, and therefore it should be considered the winner.

Generally speaking, the approach of paired comparisons consists in comparing each item with every other so as to count how many judges prefer the former to the latter.

The results of all possible comparisons are conveniently arranged in what will be called the **matrix of paired-comparison scores**, or simply the matrix of scores. For example, next table shows the matrix of scores corresponding to example G:

Example G. Paired comparisons

	Judges					*Scores*						
Id	A	B	C	D	E	61	62	63	64	65	66	67
61	1	1	7	7	7	-	2	2	2	2	2	2
62	4	2	3	3	1	**3**	-	**3**	2	**5**	**4**	**5**
63	2	6	2	4	2	**3**	2	-	2	**4**	**3**	**5**
64	3	3	1	2	5	**3**	**3**	**3**	-	**4**	**3**	**5**
65	6	4	5	6	4	**3**	0	1	1	-	**3**	2
66	7	5	6	1	3	**3**	1	2	2	2	-	**3**
67	5	7	4	5	6	**3**	0	0	0	**3**	2	-

Each cell of this matrix compares the item indicated at the left with the one indicated at the top. In fact, it displays the number of judges who preferred the former to the latter. If this number is a *majority* (which we emphasize by printing it in bold face) then it is very reasonable to conclude that the former item is *globally preferred* to the latter. This leads to the following rule to determine the winner: *if an item is globally preferred to every other in the sense above then that item should be deemed the winner.* This rule is known as **Condorcet's principle**, and the winner according to this rule is called the **Condorcet winner**.

As we have already remarked, in example G the Condorcet winner is couple 64. In fact, in the matrix of scores all entries of the corresponding row are absolute majorities.

From the point of view of the matrix of scores, the situation is not so different from a *tournament* where each item had played a match against every other. We could say, for instance, that the result of the match between couples 64 and 65 was 4 to 1. By this analogy, instead of saying that a certain item is globally preferred to another, we will often say simply that the former *beats* the latter. In this terminology, Condorcet's principle says that if an item beats every other, then it should be considered the winner.

The paired comparisons approach is certainly a very natural one. In fact, it goes back as far as the thirteenth century, when the Majorcan philosopher Ramon Llull already proposed some methods of election systematically based on this approach [**8; 6**: chap. 3].

6.2. But we are interested not only in finding out the winner. Besides that, we would like to have a *procedure for ranking the whole set of items* under consideration. How can we extend Condorcet's principle to that effect?

The natural answer to this question is fairly obvious: Once we have determined the global winner, we restrict our attention to the set of remaining items, as if they had been the only ones in play, and we look for the winner within this restricted set. In terms of the matrix of scores, this amounts to deleting both the row *and* the column corresponding to the item already classified, and applying Condorcet's principle to the smaller matrix thus obtained. These steps should be repeated until we have ranked all items. However, we are taking for granted that the Condorcet winner always exists, which is not so clear ...

Let us see what happens when we apply this procedure to example G. The 1st position is allocated to the global Condorcet winner, i. e. couple 64. The 2nd position goes to couple

62 , since it beats every other but 64 . Similarly, the 3rd position goes to couple 63 , which beats all the remaining ones. However, in next step we encounter a difficulty: we are left with four couples, namely 61 , 65 , 66 , 67 , but none of them fulfils the condition of beating the other three. In other words, this restricted set of items does not have a Condorcet winner.

6.3. Of course, nothing prevents such a situation from happening at the very beginning. In fact, it suffices to imagine that couples 61 , 65 , 66 , 67 had been the only ones to take part in the competition. So we realize that *sometimes there is not a Condorcet winner.*

By inspecting the entries of the matrix of scores, we see that the problem lies in couples 65 , 66 , 67 and it has to do with the fact that they all beat the same number of opponents. In particular, each of these couples beats just one of the other two, namely 65 beats 66 , 66 beats 67 , and 67 beats 65 . The problem is that these beating relations form a vicious circle! As a consequence, none of the couples involved is a clear winner over the others. This phenomenon is called **Condorcet's paradox**, and a vicious circle like the one described is called a **Condorcet cycle**. This paradox is a crucial point of social choice theory, and somehow it lies behind most of the other paradoxes.

In particular, Condorcet cycles show that the condition of (general) independence of irrelevant alternatives mentioned in § 4.2 is really asking too much.

In mathematical terms, Condorcet's paradox means that the global preference relation defined above —one item is globally preferred to another if the former is preferred to the latter by a majority of judges— is not necessarily transitive, even though each judge expresses a transitive set of preferences.

By the way, most often the preferences of a judge are transitive just by artificial constraint. In particular, dancesport rules require the judges of a final to always express a complete ranking of all couples involved. However, it is not clear at all that the natural preferences of a judge are really transitive. After all, he is combining multiple criteria, and therefore Condorcet's paradox may well be present already in his internal decision process.

6.4. Condorcet cycles can be interpreted as ties. When the number of judges is even we can have *simple ties* between two couples, like in the case of the marks $(1, 1, 2, 2)$ against $(2, 2, 1, 1)$, or, for that matter, $(1, 1, 7, 7)$ against $(2, 2, 1, 1)$. When the number of judges is odd such a possibility is avoided (unless ties are already allowed in the preferences expressed by a judge), but *cyclic ties* between three or more couples are always possible because of Condorcet cycles. On the other hand, simple ties between two couples can be considered as Condorcet cycles of length two.

Sometimes ties are really unbreakable. For instance, let us take a more detailed look at cyclic ties in the simplest case of three items. In example G, the Condorcet cycle formed by couples 65 , 66 , 67 is homogeneous in the sense that each item beats the "following" one with exactly the same strength. In such a case the paired comparisons approach restricted to those couples does not provide any basis for singling out one of them as the best one.

In other cases, Condorcet cycles may lend themselves to a most reasonable cycle-breaking rule. In fact, if one of the beatings in the cycle is weaker than all the others, then it is very reasonable to reject it, i. e. to break the cycle at that point, and to adopt the resulting ranking. In the following, this rule for breaking cyclic ties will be called the *rule of rejection of the weakest beating.*

But this rule works fine only when the cycle contains just one weakest point. *When there are several such weakest points, then we get a multiplicity of solutions.* Let us consider, for instance, the following Condorcet cycle (which arises in example G if we reverse the order of couples 65 and 66 in ranking E):

Example G″

Id	*Scores* 65	66	67	R_1	R_2	*AR*
65	-	**4**	2	*1*	*2*	*1.5*
66	1	-	**3**	*2*	*3*	*2.5*
67	**3**	2	-	*3*	*1*	*2*

In this case 65 beats 66 by 4 to 1, 66 beats 67 by 3 to 2, and 67 beats 65 also by 3 to 2. So we have two weakest points. If we break the cycle between 67 and 65 we obtain the ranking R_1: $65 \succ 66 \succ 67$, but if we break it between 66 and 67 we obtain R_2: $67 \succ 65 \succ 66$. In particular, these two rankings have different winners, namely couples 65 and 67.

If we are forced to choose between these two couples, we will notice that 65 is 1st in R_1 and 2nd in R_2, while 67 is 1st in R_2 but 3rd in R_1, which differentiates these couples in favour of 65. This is equivalent to adopting the result given by the average AR of R_1 and R_2. However, this result agrees neither with R_1 nor with R_2, and, more importantly, it is not consistent with respect to losers and winners: if we delete couple 66 then the winner is clearly couple 67. In contrast, the rankings R_1 and R_2 are both of them consistent with respect to losers and winners.

The extreme case of several weakest points is that of a homogeneous Condorcet cycle, where all beatings have exactly the same strength. In that case, we get as many ranking solutions as the number of items involved in the cycle. In particular, any item appears as a possible winner (but not every ranking is a solution).

6.5. In the absence of Condorcet cycles, the Condorcet winner is always defined, and the iterative procedure described in § 6.2 does always produce a ranking in a consistent way.

So it remains to see how to deal with the general case where Condorcet cycles may be present. Of course, we are looking for a method that harmonizes with the preceding ideas. In particular, its winner should coincide with the Condorcet winner whenever the latter exists. In the specialized literature such methods are known as **Condorcet completion methods**. In the following sections we shall describe the ones that seem to be more appropriate for our purposes. For other methods, see [**4, 9, 10, 11, 12, 15, 18**].

7. Copeland's method.

7.1. A very natural way to proceed consists simply in associating each item with the number of those other items that it beats. If simple ties between two items are not discarded, for instance because of an even number of judges, then they will be appropriately counted as a contribution of one half for each of the items involved.

Bearing in mind the tournament analogy pointed out in § 6.1, it will be realized that this is essentially the rating system used in soccer and other sports leagues (except possibly for the way of handling ties and for the home-away asymmetry). In social choice theory, this method is usually named after A. H. Copeland, who analyzed it with detail in 1951.

For the purpose of comparison with other methods, instead of the tally obtained by adding up the number of wins plus one half the number of ties we shall consider its complement to the total number of items, which gives an equivalent position-like number that we shall denote by C. In the case of example G, the resulting values of C are as shown in next table.

Example G. Copeland's method, with tie-breaking

	Judges					*Scores*									
Id	A	B	C	D	E	61	62	63	64	65	66	67	C	T	R
61	1	1	7	7	7	-	2	2	2	2	2	2	7	12	*7*
62	4	2	3	3	1	**3**	-	**3**	2	**5**	**4**	**5**	2	22	*2*
63	2	6	2	4	2	**3**	2	-	2	**4**	**3**	**5**	3	19	*3*
64	3	3	1	2	5	**3**	**3**	**3**	-	**4**	**3**	**5**	1	21	*1*
65	6	4	5	6	4	**3**	0	1	1	-	**3**	2	5	10	*5*
66	7	5	6	1	3	**3**	1	2	2	2	-	**3**	5	13	*4*
67	5	7	4	5	6	**3**	0	0	0	**3**	2	-	5	8	*6*

7.2. As we saw in § 6.3, example G has a Condorcet cycle involving couples 65 , 66 , 67 . In Copeland's method this cycle shows up in the form of a tie between these couples, all of which obtain $C = 5$.

However, in the matrix of scores the rows corresponding to these couples look different. In principle, rows with larger entries indicate better couples, which suggests a tie-breaking rule based on the sum of the entries of each row, which we shall denote by T. According to this criterion, the tie would break into the following ranking: $66 \succ 65 \succ 67$.

This result agrees with the both the addition method and the extended median method. However, in the case of the addition method the agreement is not a simple coincidence. In fact, the parameter T that we have just introduced is connected to the S of the addition method through the formula $S = NJ - T$, where N and J represent respectively the number of items and the number of judges.

The method just outlined is therefore equivalent to supplementing Copeland's method with a tie-breaking rule according to the addition method. Such a procedure can be traced back to 1787, when it was considered by Condorcet as a reasonable possibility for defining the winner when none of the items beats all the others [**6** : p. 35–36] .

Let us remark also that, from the tournament point of view this method is analogous to the goal difference method used in soccer and other sports leagues.

7.3. Unfortunately, *Copeland's method is not consistent with respect to losers and winners*, and therefore any tie-breaking modification has the same problem. In the following example the reader will easily check that deleting all couples with C greater than 3, i. e. couples 84 , 85 , 86 , changes the winner from couple 81 to couple 82 .

Example I. Copeland's method, with tie-breaking

	Judges			*Scores*								
Id	A	B	C	81	82	83	84	85	86	C	T	R
81	2	1	4	-	1	**3**	**2**	**3**	**2**	2	11	*1*
82	1	6	3	**2**	-	**2**	1	**2**	1	3	8	*2.5*
83	3	2	5	0	1	-	**2**	**3**	**2**	3	8	*2.5*
84	5	4	1	1	**2**	1	-	1	**3**	4	8	*4*
85	4	3	6	0	1	0	**2**	-	**2**	4	5	*5*
86	6	5	2	1	**2**	1	0	1	-	5	5	*6*

8. Minimum disagreement.

8.1. Let us assume that the global result is required to be a complete ranking of all couples involved. The problem can then be viewed as follows: we are given several rankings, one per judge, and we must produce another ranking, the result, that summarizes all the given ones in the best possible way. Of course, unless all of the judges give exactly the same ranking, any ranking that one may propose as result will contain some disagreement with some of the judges. In the words of Sir Maurice George Kendall, a British statistician who introduced these ideas in 1955, we may say that it *violates* some judges' preferences. However, we can always *look for the ranking whose disagreement with the judges is as small as possible.* For that purpose, we need a method of measuring the total amount of disagreement of a candidate ranking with respect to those that have been given by the judges. Reasonably enough, this quantity should be the sum total of the disagreements of that ranking with respect to each of the judges.

8.2. The heart of the matter consists thus in *measuring the disagreement between two different rankings.* But this is not so difficult: it suffices to count the number of **inversions**, i. e. how many pairs of items are reversed in order when going from one ranking to the other.

For example, the number of disagreements between the ranking $12 \succ 14 \succ 11 \succ 13$ and the ranking $12 \succ 11 \succ 14 \succ 13$ is just one, because there is only one pair of items in reverse order, namely 11 and 14 . Similarly, the number of disagreements between the ranking $12 \succ 14 \succ 11 \succ 13$ and the ranking $11 \succ 12 \succ 13 \succ 14$ is three, since there are exactly three pairs of items in reverse order, namely 11 and 12 , 11 and 14 , and 13 and 14 . When rankings are specified as lists of ranks, like in the columns of our tables, counting disagreements has a different feeling, but the idea is exactly the same.

By the way, these ideas could be applied to measuring the agreement between the different rankings that we are given, or between each of them and the global result. In fact, statistics provides standard methods for such purposes, some of them based exactly on the ideas discussed here. Some years ago, Julie Malcolm and Steve Nikleva suggested using such methods for evaluating the judges of dancesport competitions (*Dance News*, 1365, 25th August 1994, p. 3; *ibidem*, 1451, 18th April 1996). In this connection, however, the writer fully agrees with the views expressed on that occasion by Bryan Allen: "I would also resent the suggestion that any judge who differs from the majority is wrong! That judge might be the only one, in my opinion, who was right!" (*loc cit*). If judges were foolishly evaluated according to their agreement with the majority, in practice this could easily lead them towards not expressing their own sincere expert judgement, but merely reproducing the more or less established public opinion.

8.3. Given two different rankings we can thus produce a number that measures the disagreement between them. In particular, for any ranking that we may consider as a candidate for a global result we can calculate its disagreement with each of the judges, and, by adding up these numbers, we obtain a number that measures its total amount of disagreement with them.

Most remarkably, for any candidate ranking *its total amount of disagreement with the judges can be worked out from the matrix of scores*: As we saw in § 6.1, each off-diagonal entry of this matrix tells us how many judges supported the view that "the item indicated at the left is better than the one indicated at the top". In particular, each entry corresponds to a proposition of this kind. Now, this proposition will either agree or disagree with the candidate ranking under consideration. Obviously, in the second case that entry should be counted as a contribution to the total amount of disagreement, whereas in the first case it should be counted as a contribution to the total amount of agreement.

In other words, for any candidate ranking, we can always classify the off-diagonal entries of the matrix of scores into two classes according to whether the corresponding propositions agree or disagree with that ranking. By adding up the entries of each class we shall obtain respectively the total amount of agreement or disagreement of that ranking with the judges. In fact, the total amount of disagreement computed in this way coincides exactly with its initial definition, since we have done nothing else than counting the same inversions after having grouped them in a different way. On the other hand, one can easily check that the total amount of agreement and the total amount of disagreement always add up to $JN(N-1)/2$.

For instance, for example I and the ranking $81 \succ 82 \succ 83 \succ 84 \succ 85 \succ 86$, the entries in agreement with this ranking are the ones above the diagonal, whereas the entries in disagreement are the ones below. Therefore, the total amount of disagreement of this ranking with the judges is equal to the sum of the entries below the diagonal, i. e. 15. Similarly, the amount of agreement is 30, and both quantities do indeed add up to $JN(N-1)/2 = 3 \cdot 6 \cdot 5/2 = 45$.

8.4. Once these notions have been formulated, the idea consists simply in looking for a ranking whose total amount of disagreement with the judges in the sense above is as small as possible. Alternatively, since the amounts of agreement and disagreement defined above add up to a constant, the condition of minimum disagreement can be equivalently replaced by a condition of maximum agreement.

These ideas fit in very well with the theoretical approach that Condorcet used in 1785 [**6** : p. 31] . Having said that, when it comes to putting it as a recipe, Condorcet's wording is not precise enough, and in fact it can be interpreted also in terms of other different methods. The precise formulation given above dates from the 1950s, when it was considered by several authors, of which we have already mentioned Sir Maurice George Kendall. Today it is often named after another of these authors, John George Kemény, a distinguished mathematician and the co-inventor of the BASIC computer language.

Instead of naming it after any particular author, we have preferred to call it the method, or criterion, of **minimum total disagreement** (MTD).

8.5. But we still have to say how does one actually find the ranking that achieves the minimal disagreement with the judges. For each candidate ranking we certainly know how to compute its amount of disagreement with the judges. But how shall we find the best possibility? In this connection, one must admit that efficient general procedures are not so easy to come by. However, for a reduced number of items one can simply try all possible rankings and see which one is the best. For 6 or 7 items, which is the dancesport typical case, the number of possible rankings is respectively 720 or 5040. Certainly, trying all of them by pen an paper would take a long time, but for today's personal computers this is done in less than a second.

For a larger number of items, checking all possible rankings takes too much time, even for a computer, but more elaborated computer programs are able to cope easily with up to about 25 to 35 items.

8.6. Next table shows the results of this method when applied to example G. In this case there are three different results, R_1, R_2, R_3. In other words, there are three different rankings with the property of achieving a minimal total amount of disagreement with the judges (namely, 64 inversions).

These three rankings coincide with each other for all couples but 65 , 66 , 67 . This is not surprising, since these three couples form a Condorcet cycle where each couple beats the "following" one by exactly the same margin. Somehow, the present method does not distinguish between these couples, and they are tied for the 5th place. In fact, this is exactly the result that we obtain if we take the average R of the three rankings produced by the criterion of minimal total disagreement. It will be noticed also that this conclusion fully agrees with the result of Copeland's method. On the other hand, it must be clear that this last averaging operation does not properly belong to the MTD method.

Example G. MTD Method

Id	*Judges* A	B	C	D	E	*Scores* 61	62	63	64	65	66	67	R_1	R_2	R_3	R
61	1	1	7	7	7	-	2	2	2	2	2	2	*7*	*7*	*7*	*7*
62	4	2	3	3	1	**3**	-	**3**	2	**5**	**4**	**5**	*2*	*2*	*2*	*2*
63	2	6	2	4	2	**3**	2	-	2	**4**	**3**	**5**	*3*	*3*	*3*	*3*
64	3	3	1	2	5	**3**	**3**	**3**	-	**4**	**3**	**5**	*1*	*1*	*1*	*1*
65	6	4	5	6	4	**3**	0	1	1	-	**3**	2	*4*	*5*	*6*	*5*
66	7	5	6	1	3	**3**	1	2	2	2	-	**3**	*5*	*6*	*4*	*5*
67	5	7	4	5	6	**3**	0	0	0	**3**	2	-	*6*	*4*	*5*	*5*

In the case of example I, the MTD method produces a unique result (whose disagreement with the judges is equal to 13 inversions). In this case, however, the result does not agree with Copeland's method. In particular, both methods give a different winner.

Example I. MTD Method

Id	*Judges* A	B	C	*Scores* 81	82	83	84	85	86	R
81	2	1	4	-	1	**3**	**2**	**3**	**2**	*2*
82	1	6	3	**2**	-	**2**	1	**2**	1	*1*
83	3	2	5	0	1	-	**2**	**3**	**2**	*3*
84	5	4	1	1	**2**	1	-	1	**3**	*5*
85	4	3	6	0	1	0	**2**	-	**2**	*4*
86	6	5	2	1	**2**	1	0	1	-	*6*

8.7. The MTD method is especially significative in connection with the condition of *consistency with respect to losers and winners*. Most remarkably, we have at last a method that satisfies that condition! In fact, it is not difficult to see that if a ranking minimizes the disagreement with the judges and we delete all items whose rank number is above, or below, a certain value, the restriction of that ranking to the remaining items does still minimize the disagreement with the judges.

Although we will not go into details, it is interesting to remark also that the MTD method is not so alien to the the median concept of § 5. In fact, a ranking that minimizes the disagreement with the judges can be interpreted as a sort of median of the input rankings without disassembling them into disconnected marks. Owing to this median-like character, the MTD method shares with the median method a certain property of robustness against eccentric marking.

8.8. In spite of these good properties, the MTD method can still be a matter of criticism. In particular, one can raise several objections in the spirit of the majority principle.

Of course, for two items, there is no problem at all: in that case the MTD criterion coincides exactly with the majority principle. For more than two items, we already know that the quest for the majority principle can be frustrated by the possibility of Condorcet cycles. In fact, the presence of such a cycle means that there is no ranking with the property that each of its paired-comparison preferences agrees with a majority of judges. In other words —looking from the other side—, in that situation the pattern of majorities in the matrix of scores, i. e. the information that we are conveying by means of the distribution of bold-face type, does not correspond to any ranking. As we discussed in § 6.4, in the simplest cases it is most reasonable to break Condorcet cycles at their weakest beatings. As one can easily see, in the case of three items the MTD criterion coincides exactly with this rule. However, for more than three items, *the MTD criterion may go against the rule of rejection of the weakest beating.*

For instance, let us consider the following example with four items. There are fifteen judges, but their rankings reduce to three possibilities: more specifically, six of the judges coincide in giving the ranking marked as A, five of them give the ranking B, and the remaining four give the ranking C.

Example J. MTD Method

	6	5	4	*Scores*				
Id	A	B	C	91	92	93	94	*R*
91	1	4	3	-	**10**	6	6	*4*
92	2	1	4	5	-	**11**	**11**	*1*
93	3	2	1	**9**	4	-	**15**	*2*
94	4	3	2	**9**	4	0	-	*3*

By inspecting the matrix of scores, one easily discovers several Condorcet cycles. Specifically, we have two cycles of length three, namely α: $91 \succ 92 \succ 93 \succ 91$ and β: $91 \succ 92 \succ 94 \succ 91$, and one cycle of length four, namely γ: $91 \succ 92 \succ 93 \succ 94 \succ 91$ (here we are extending the meaning of the symbol '$\succ$' to denote paired-comparison beatings). Like any other ranking, in such a situation, the MTD ranking is forced to reject some of these paired-comparison majorities. Specifically, it happens to reject the paired-comparison majority $91 \succ 92$, which simultaneously breaks the three cycles and results in the ranking $R =$ B : $92 \succ 93 \succ 94 \succ 91$. This ranking minimizes the total disagreement with the judges to 30 inversions. However, it does not conform to the rule of breaking cycles at the weakest beating. In fact, the three cycles have been broken by rejecting $91 \succ^{10} 92$, where the superscript indicates a score of 10, but all of them had a weaker beating, namely $93 \succ^{9} 91$ for cycle α, and $94 \succ^{9} 91$ for cycles β and γ.

By breaking cycles at beatings which are not the weakest, the MTD criterion becomes exposed to rather undesirable situations: For instance, in the preceding example the rejection of the paired-comparison majority $91 \succ^{10} 92$ is most questionable: certainly, the opposite preference follows from the chain of paired-comparison majorities $92 \succ^{11} 93 \succ^{9} 91$, and also from $92 \succ^{11} 93 \succ^{15} 94 \succ^{9} 91$; however, in both cases one of the majorities involved is weaker than the one that is being rejected. So, in terms of majority sizes there is a case for claiming that item 91 should be globally preferred to 92. In the words of Stephen Eppley [**19**], we may say that *the MTD criterion is not immune to majority complaints.*

Another objection against the MTD criterion in the spirit of the majority principle could be formulated as follows: As soon as there are more than two items, two rankings may be in different degrees of agreement or disagreement with each other. As a consequence, when a particular ranking is considered as a candidate for a global result, its total amount of disagreement with the judges may be more concentrated on some judges than others. In particular, it may happen that a few judges concentrate a large amount of disagreement, in which case that candidate for a global result could be discarded because of a large disagreement with a *minority* of judges!

Let us mention also that instead of minimizing the total amount of disagreement, one could go for minimizing the number of rejected majorities, which sometimes leads to a different result (see for instance [**4**], where this alternative is called the method of minimum violations). However, this alternative criterion does not solve the problem of immunity to majority complaints.

9. Immunity to majority complaints and ranked pairs.

9.1. Let us try to give a general definition of **immunity to majority complaints** (IMC). By drawing from the preceding particular example, and abstracting the underlying idea, we can say that a ranking being immune to majority complaints means the following: every preference adopted by that ranking must beat or match its opposite either directly or indirectly, i. e. through a chain of other preferences also adopted by that ranking.

Let us explain it with more precision. We are given a matrix of paired-comparison scores and we are considering a certain ranking R as a candidate for a global result. We view this ranking as a special way of expressing which item is preferred out of every pair. In principle, we would like that these preferences agree with the majorities of the matrix of scores, i. e. we would like the following condition to be satisfied:

(MAJ) The ranking of an item a above another one b should mean that the score of a against b is greater than the score of b against a .

However, we know now that sometimes this condition is asking too much. In its place, the condition of immunity to majority complaints asks for something weaker:

(IMC) The ranking of an item a above another one b should mean that the *indirect* score of a against b is *greater than or equal to* the (direct) score of b against a .

Here, the **indirect score** of a against b means the maximum score that one can obtain through a chain of items which descends along the ranking under consideration. For instance, in example J the direct score of 92 against 91 is 5, but the indirect score of 92 against 91 through the ranking $R = 92 \succ 93 \succ 94 \succ 91$ is 9, because of the chains $92 \succ^{11} 93 \succ^{9} 91$ and $92 \succ^{11} 93 \succ^{15} 94 \succ^{9} 91$. As it is illustrated by this example, in the definition above, the "score through a chain" should be understood as the minimum score of its intermediate propositions.

9.2. Let us consider the possibility of satisfying the condition of immunity to majority complaints. As we shall see, this leads to a new class of methods with interesting properties.

Most naturally, one might begin by wondering whether we are not facing another impossible dream. But that is not the case. In fact, one can show that *the following simple procedure does always produce a ranking immune to majority complaints*:

Procedure RP. Consider the paired-comparison scores by decreasing order of magnitude; at each step, adopt the corresponding proposition unless it contradicts the already adopted ones, i. e. unless it would close a cycle; continue until a complete ranking is obtained.

For instance, in the particular case of example J this procedure leads to successively adopting the following propositions: (1) $93 \succ^{15} 94$; (2) $92 \succ^{11} 93$, after which we know that $92 \succ 93 \succ 94$; (3) $92 \succ^{11} 94$, which certainly matches the preceding information; and (4) $91 \succ^{10} 92$, which completes the ranking $S = \mathsf{A} : 91 \succ 92 \succ 93 \succ 94$. It is interesting to remark that this ranking rejects the paired-comparison majorities $93 \succ^{9} 91$ and $94 \succ^{9} 91$, completely in accordance with the rule of breaking the Condorcet cycles at the weakest beating. Let us remark also that in this case the rejected majorities were farther down the list, but in other cases they occur before the complete ranking has been established.

When there are several scores of the same magnitude we shall say that the corresponding propositions are tied. In that case the result may depend on the order in which these propositions are considered (the situation is essentially the same that we already met in § 6.4 in connection with the rule of rejection of the weakest beating). Most remarkably, it turns out that *every ranking immune to majority complaints can be obtained through the preceding procedure if the tied propositions are suitably ordered.*

Once again, the main idea of this method can be traced back to the work of Condorcet in 1785 [**5** : p. 129] . The precise formulation given above dates from 1986/87 when it was worked out concurrently by Thomas M. Zavist and T. Nicolaus Tideman [**13**], and the latter called it the rule of **ranked pairs** (RP). Later on, this rule and its variations has become increasingly popular in specialized media [**15–20**].

The ranked-pairs procedure is related to certain standard algorithms that arise in a variety of technological contexts, including communication networks and computer programming. More specifically, the ranked-pairs procedure is remarkably akin to a standard algorithm that was proposed in 1956 by Joseph B. Kruskal Jr in connection with certain ubiquitous problems which are known as the shortest-spanning-tree problem and the travelling salesman problem. In terms of algorithm theory, the ranked-pairs procedure can be viewed as a 'greedy' algorithm that replaces the MTD criterion of § 8 by a local version of it.

9.3. As we have seen, if we allow for all possible orderings of the tied propositions, the rule of ranked pairs is equivalent to looking for all rankings that satisfy the condition of immunity to majority complaints. In this sense, *the method of ranked pairs is equivalent to the criterion of immunity to majority complaints.* Now, this criterion is very suitable to our purposes not only in the spirit of the majority principle, but also by other reasons.

First of all, it is not difficult to see that *the condition of immunity to majority complaints is consistent with respect to losers and winners.* That is, if we take a ranking immune to majority complaints, say $a \succ b \succ c \succ d \succ e \succ f$, and we delete some items from the top and some items from the bottom, say we keep only the segment $c \succ d \succ e$, the resulting restricted ranking will always keep the condition of immunity to majority complaints.

This is an immediate consequence of the definition given in § 9.1. In fact, the indirect score between two items is concerned only with chains that *descend along the ranking under consideration.* Therefore, it depends only on the ranking segment limited by the two items in consideration, which is not affected by deleting losers or winners.

The method of ranked pairs is interesting also in connection with the so-called **consistency with respect to clones** (or independency of clones). In fact, this property was the aim that led Zavist and Tideman into the method of ranked pairs. To cut a long story short, let us say that this property forbids the occurrence of a certain kind of (weak) flip-flops which are not controlled by the condition of consistency with respect to losers and winners (a more detailed definition is given in the inset below). In the way that we have defined it above, the method of ranked pairs is ensured to be consistent with respect to clones as long as it produces a unique result. When the result is not unique the property of consistency with respect to clones may cease to hold for the method of ranked pairs as defined above, but it will be satisfied by certain variants to be described below.

Two or more items are said to be *clones* of each other when all of the judges rank them consecutively (independently of possible differences in the internal ordering of those items and in the ranking of the others). For a ranking method, the condition of *consistency with respect to clones* imposes the two following restrictions: (a) clones should be ranked consecutively; and (b) when clones are added or deleted the results should remain the same except for the internal ordering of the clones (and of course their addition or deletion). These restrictions are especially desirable in the case of political elections and parliamentary votations in order to disallow any manipulations based upon artificially introducing additional options which are similar to some of the already existing ones.

9.4. As we have seen, in the case of example J the method of ranked pairs gives a different result than the method of minimum total disagreement. In fact, MTD gives the ranking $R = \mathsf{B} : 92 \succ 93 \succ 94 \succ 91$, whereas RP gives the ranking $S = \mathsf{A} : 91 \succ 92 \succ 93 \succ 94$. Let us compare these two results. In § 8.8 we saw that the ranking R rejects only 1 paired-comparison majority and its total amount of disagreement with the judges is 30 inversions. In contrast, the ranking S is rejecting 2 paired-comparison majorities and its total amount of disagreement with the judges is 31 inversions. Certainly, these values for S are worse than those for R. However, the ranking S does a better job than R at minimizing a different parameter. In order to define this parameter, let us decompose the total amount of disagreement with the judges as R: $30 = \mathbf{10} + 6 + 6 + 4 + 4 + 0$ and S: $31 = \mathbf{9} + \mathbf{9} + 5 + 4 + 4 + 0$, where each term corresponds to a different pair of items and these terms are arranged in decreasing order. It is not difficult to see that the rankings obtained by the method of ranked pairs have always the property of minimizing the leading term of such a decomposition. In other words, *the method of ranked pairs has the virtue of minimizing the maximum number of judges that are disregarded in connection with every separate pair of items.*

Independently of its being satisfied by the method of ranked pairs, in 1986 Kenneth J. Arrow and Hervé Raynaud [**9**] referred also to this condition of minimum leading disagreement as a desirable property for a result to avoid being "vulnerable to 'legitimate' criticism". In their terminology, the rankings that satisfy that condition are said to be *prudent*.

Concerning the comparison between R and S, it is interesting to notice also that in this example the RP result $S = \mathsf{A}$ is fully supported by more judges than the MTD one $R = \mathsf{B}$.

9.5. When there are tied propositions, the method of ranked pairs can produce multiple results. In big elections, exactly equal scores are very improbable, but when the number of voters or judges is small, like in most dancesport competitions, tied propositions will be the rule and multiple results will occur with a certain frequency.

For instance, in the case of example G, the method of ranked pairs produces exactly the same three rankings that were obtained in § 8.6 by the MTD criterion. In RP one obtains one or another of these three rankings depending on the order in which one considers the tied propositions concerning items 65 , 66 , 67 (which form an homogeneous Condorcet cycle). In such cases, multiple results are essentially unavoidable: if we delete all items but 65 , 66 , 67 , symmetry makes clear that none of the three possible ranking results is more reasonable than the others; on the other hand, consistency somehow entails keeping this situation when the other items are added.

In other cases, the multiplicity of results is not that acceptable. For instance, this is the case of example I. Here we have only three judges, which forces the matrix of scores to contain many tied propositions. In this case, RP produces no less than eight different rankings! Certainly, this is far too much: we cannot accept a result consisting in eight rankings when we started with only three of them!

So, we still should make an effort to reduce the set of results as much as possible. As a matter of fact, there exist several refinements of the method of ranked pairs that go in this direction. For the sake of brevity, in the following we shall omit a detailed discussion of all the existing variants and alternatives. Instead, we shall concentrate on describing the two variants that in our opinion better serve our purposes.

9.6. As we have been seeing, if we want to produce a unique result in the form of a complete ranking, sometimes this will mean to single out one of several equally entitled possibilities. In such a case, if we are forced to make a choice we are led to recourses such as drawing lots or using a casting vote. In principle, such "less rational" and more practical possibilities were out of our scope. However, it happens that some of them are more consistent than others. So we better take a look into it.

Specifically, we shall consider a method based upon a casting vote. Like the other votes, this casting vote will be assumed to give a complete ranking of the items under consideration. This ranking will be used as a **tie-breaker** for the ranked-pairs procedure according to the following rule:

Rule NTB. Assume that we have a tie between several propositions. In order to decide which one to consider first, we apply successively the following criteria: 1. The propositions that agree with the tie-breaker ranking are given priority before those that disagree with it. 2. When the preceding criterion does not decide, we identify the preferred item and the unpreferred item of each proposition, and we apply successively the two following criteria: 2.1. That the preferred item be ranked better in the tie-breaker ranking. 2.2. That the unpreferred item be ranked worse in the tie-breaker ranking.

For instance, consider example G and assume that the tie-breaker coincides with judge A. In that case, rule NTB selects in first place the proposition $63 \succ 67$, and from all propositions with a score of 3, it selects the proposition $63 \succ 66$.

In particular, this rule includes the case of simple ties between two items (in the case of an even number of judges), in which case it agrees with giving priority to the item which is ranked better in the tie-breaker ranking.

The tie-breaker ranking could be provided by a special judge previously appointed to that effect. On the other hand, nothing prevents this special judge to coincide with one of the ordinary judges. In this case, his ranking will still be used also in the same way as those of the other judges. By default, in our examples we shall assume the tie-breaker ranking to be that of judge A.

Since we are using the RP procedure, the result is ensured to be immune to majority complaints, and therefore consistent with respect to losers and winners. But now we have an additional property: in fact *this method is always consistent with respect to clones.*

The tie-breaking rule that we have called rule NTB has been proposed recently by Markus Schulze [**18b; 15**: 20 May 2004] as a more natural alternative to other similar rules that had been introduced previously by several authors, starting from Zavist and Tideman in 1989 [**14**]. In the following, we shall refer to the resulting method as **ranked pairs with natural tie-breaking** (RPN).

In the case of example G, the result of this method is the ranking R_2 of § 8.6, i. e. the only one where couples 65 , 66 , 67 are ordered in the same way as in the tie-breaker ranking A. If the tie-breaker had been ranking E then that compatibility is not possible and the resulting ranking happens to be R_1. In the case of example I, the result coincides completely with the tie-breaker ranking A. In general, this happens whenever the tie-breaker ranking is itself immune to majority complaints, as in this case.

9.7. An alternative approach for always producing a unique result consists in relaxing the notion of 'result': Instead of asking for a proper ranking, where we are forced to choose one preferred item out of every pair, here we shall allow for the possibility of ties between two or more consecutive items. In the terminology of § 2, we shall allow the result to be a weak ranking. For instance, in example G it seems natural a result of the form $64 \succ 62 \succ 63 \succ 65 \sim 66 \sim 67 \succ 61$ (where the symbol '$\sim$' means 'tied with').

It is not difficult to modify the RP procedure so as to produce a unique result in the form of a weak ranking:

Procedure WRP. Consider the paired-comparison scores by decreasing order of magnitude; at each step, adopt the corresponding proposition unless it contradicts those that had been already adopted *with a strictly higher score*, i. e. unless it would close a cycle with a single weakest point (see § 6.4); continue until a complete weak ranking is obtained.

We shall refer to the resulting method as **weak ranked pairs** (WRP). This method has been proposed quite recently also by Andrew Myers (under the name of 'CIVS ranked pairs') [**20**]. A preliminary exploration, based upon both mathematical arguments and computational evidence, seems to confirm that this method satisfies suitable variants of the properties satisfied by the RPN method.

In particular, the property of immunity to majority complaints is satisfied in the following form:
(imc′) The ranking of an item a *strictly above* another one b means that
the *indirect* score of a against b is *strictly greater than* the (direct) score of b against a.

The WRP method is particularly attractive because it produces a unique result without the need for a casting vote. However, we are paying a price, namely that the result may contain ties. This is the case of example G, where the result is indeed the expected one, namely $64 \succ 62 \succ 63 \succ 65 \sim 66 \sim 67 \succ 61$. A more dramatic tie occurs in the case of example I, where the result of the WRP method is a complete tie between all couples! Naturally, ties become less probable as the number of judges becomes larger. An idea of how often can they occur will be obtained in § 12.

9.8. Although we have presented the WRP method as an alternative to the RPN method, in fact both of them can be considered as two extreme cases of a more general method where the "tie-breaker" can be any weak ranking. In the case of a proper ranking this more general method reduces to RPN, whereas WRP corresponds to the case where the "tie-breaker" is a complete tie.

When the result of the WRP method is a proper ranking, then this ranking is guaranteed to be the only one that satisfies the condition of immunity to majority complaints. In particular, it coincides with the result of the RPN method. However, and contrarily to what might be suggested by the terminology, when the result of the WRP method is not a proper ranking, then the proper ranking produced by the RPN method is not necessarily a refinement of that weak ranking. More than that, in certain cases it may happen that not one proper ranking immune to majority complaints has the property of being a refinement of the WRP result. In that connection, it must be emphasized that in the present context the terms "tie-breaking" and "tie-breaker" refer to ties between propositions (see § 9.2), but not to ties between items. So, *the RPN method is not a refinement of the WRP method*, but they should be considered as two different variants of the ranked-pairs procedure.

Let us compare the present situation, as obtained with the two preceding variants of the ranked-pairs procedure, with the one that we had with the MTD criterion of § 8. In both cases we have consistency with respect to losers and winners. However, now we have also the property of immunity with respect to majority complaints as well as the property of consistency with respect to clones. Furthermore, each of those two variants has the virtue of producing a unique result in harmony with all of those properties.

But that is not all. Ranked pairs beats the MTD criterion also in terms of computational efficiency. In fact, as we said in § 8.5, the basic procedure for applying the MTD criterion consists in trying all possible rankings, which means a lot of work even for a small number of items. This can be alleviated by means of certain more elaborate algorithms, but in general terms the problem remains out of reach for pen and paper computation. In contrast, the RP procedure has a direct character, which means much less computations. More specifically, when both items and judges are present only in small numbers, like in the case of a dancesport final, *both the RPN method and the WRP method are suitable for pen and paper computation.*

Let us remark here that the condition of immunity to majority complaints has also good implications in connection with certain methods for converting ranks to rates that will be considered in the following section.

10. From ranks to rates.

10.1. In this section we shall deal with converting the preceding ranking results into quantitative ratings. A ranking says who finished first, second, third, and so on, but it does not give any idea at all about the distance between two consecutive items. In particular, it does not say how clear was the winner. In contrast, a rating allows for the possibility of quantifying such matters.

In addition to their intrinsic significance, quantitative ratings are especially interesting for the purpose of combining the results of multiple fields. Having said that, the truth is that our final proposal in that connection will be essentially independent of the methods discussed in this section. So, the reader who wants to get to the point can jump to § 11.

We are still assuming that each judge expresses his opinion in the form of a ranking. Certainly the information provided by each of these rankings has a purely qualitative character. However, the fact of having several judges makes it possible to derive some quantitative appraisal of the differences of merit between couples, especially when the number of judges is large enough.

Such a quantitative effect is clearly present in Borda's rank addition method of § 4. For instance, in its application to example G (§ 4.1) couples `63`, `66` and `61` obtain consecutive positions, but their rates, namely 16, 22 and 23, or equivalently 3.2, 4.4 and 4.6, show a larger gap between couples `63` and `66` than between couples `66` and `61`.

As we remarked in § 4, Borda's rank addition method is liable to certain undesirable phenomena, which pushed us into median ranks and later into the paired comparisons approach. In particular, rank addition seemed especially inappropriate because ranks do not have a quantitative character. Later on, however, we saw how Borda's rates reappeared unexpectedly in § 7.2 through a different procedure quite suitable to paired comparisons.

But the paired comparisons approach lends itself to a variety of other rating methods besides Borda's. In this connection, it is worth recalling from § 6.1 that, from the point of view of the matrix of paired-comparison scores, our problem embodies the same structure as a whole tournament or league where each item had played a match against every other. As a consequence, it turns out that most of the rating systems used in league contests, i. e. contests based on one-to-one matches, like soccer or chess, can automatically be applied also to our situation.

In fact, in § 7 we already saw that Copeland's method and its tie-breaking variation are essentially equivalent to the rating system used in soccer leagues. These methods admit of interesting elaborations, proposed in 1952–55 by T. H. Wei and Sir Maurice George Kendall, where each item is rated by means of a number which depends not only on the number of wins and ties and the corresponding scores, but also on the rates of the opponents (for instance, beating a top rated item is given more value than beating a low rated one).

Another interesting class of rating methods rely on certain probability models that connect ideal rates with real judgements.

For more information (of a technical character) about rating methods in connection with paired comparisons and ranking judgements, the reader is referred to [**21** : vol. 6, p. 555–560 ; **22**] and the references therein.

Unfortunately, none of these standard rating methods that we are referring to is systematically consistent with the ranking methods of the preceding sections. On the other hand, if we want to produce a rating that agrees with a ranking determined previously, we can always start from that ranking and try to make it into a rating by adding a quantitative part according to the information provided by the matrix of paired-comparison scores. In the remaining of this section we shall consider such a way of proceeding.

10.2. So, we are given a matrix of paired-comparison scores together with a particular ranking. Our aim is to convert this ranking into a rating according to the quantitative information contained in that matrix. In the following, the ranking that that we are given will be called the **basis ranking**. For simplicity, we shall restrict the following discussion to the case where the basis ranking is a proper ranking; however, everything can be extended to the case of a weak ranking.

Specifically, we shall look for a rating that satisfies the conditions listed below. As before, N denotes the number of items.

Condition 1 : *Rank-like character.* Each rate is a number, integer or fractional, between 1 and N. The best possible rate is 1 and the worst possible one is N. The average of all rates is the same as the average of the numbers $1, 2 \dots N$, namely $(N+1)/2$.

Condition 2 : *Compatibility.* The rates order the items in the same way as the basis ranking, except for the possibility of ties. More specifically, if an item is ranked better than another in the basis ranking, then the former will be rated better than or equal to the latter. In the special case where all paired comparisons are simple ties, then all rates are equal to $(N+1)/2$.

Condition 3 : *Scale invariance.* The rates depend only on the relative scores, i. e. their value divided by the total number of judges. So, if every judge and his ranking is replaced by a fixed number of copies, the rates remain exactly the same.

Condition 4 : *Classification.* Let us consider a splitting of the items into a top class plus a low class according to whether their basis ranks are better or worse than a certain threshold. Assume also that all of the judges have put each member of the top class in front of every member of the low class. In that case, *and only in that case*, the rates can be obtained separately for each of these two classes according to the corresponding part of the matrix of scores. In other words, in that case, and only in that case, a deletion of the low-ranked items does not alter the rating of the top-ranked ones, and viceversa. All of this holds under the only qualification that the unassembled low class rates will differ from the assembled ones by the number of top class members.

In particular, the classification condition implies that the winner will be rated exactly 1 *only* when all of the judges have put that item into first place. Similarly, the loser will be rated exactly N *only* when all of the judges have put that item into last place. Moreover, by applying it repeatedly, one sees that this condition implies also that if all of the judges agree with the basis ranking, then the rates will be exactly equal to the ranks.

The term *rating* implies also a condition that the rates should have a quantitative character. In general terms, this means that rates should be sensitive to small changes in the entries of the matrix of scores. In particular, as soon as the rankings given by the judges contain some disagreement with the basis ranking, some of the rates should deviate from the basis ranks. In some sense, a larger disagreement should result in a larger deviation, which will make the rates to become closer to each other.

10.3. In the following we briefly describe a method that seems to *fulfil the conditions above whenever the basis ranking is immune to majority complaints* (and the paired-comparison scores come from the rankings given by a panel of judges). This method will be called **reduction rating**. As a practical illustration, we shall consider its application to the particular case of Example G with the ranking R_2 (see § 8.6).

Step 0 : *Preparation.* To start with, we rearrange the data, so that the items appear in the order described by the basis ranking. In our case, this ranking is R_2: 64 ≻ 62 ≻ 63 ≻ 67 ≻ 65 ≻ 66 ≻ 61 , and the matrix of scores gets rearranged as shown next at the left-hand side. After this step, all the relevant information lies above the main diagonal; so, from now on we can forget about the other half of the matrix.

Example G. Step 0.

Id	*Scores* 64	62	63	67	65	66	61
64	-	**3**	**3**	**5**	**4**	**3**	**3**
62	2	-	**3**	**5**	**4**	**5**	**3**
63	2	2	-	**5**	**4**	**3**	**3**
67	0	0	0	-	**3**	2	**3**
65	1	0	1	2	-	**3**	**3**
66	2	1	2	**3**	2	-	**3**
61	2	2	2	2	2	2	-

Example G. Step 1.

Id	*Scores* 64	62	63	67	65	66	61
64	-	**3**	**3**	**5**	**4**	**3**	**3**
62		-	**3**	**5**	**4**	**5**	**3**
63			-	**5**	**4**	**3**	**3**
67				-	**3**	**3**	**3**
65					-	**3**	**3**
66						-	**3**
61							-

Step 1 : *Projection.* The paired comparisons information is then projected in the direction of the basis ranking. This simply means that each score is replaced by the corresponding indirect score as defined in § 9.1. Although we shall not go into details, let us remark that after step 0, the computation of the indirect scores is relatively simple. In our case the only change occurs with the pair (67 , 66), whose direct score 2 is replaced by the indirect score 3.

Step 2 : *Successive dichotomy.* The next operation, and the central one, is a successive dichotomy procedure which progressively subdivides the set of items into finer classes until all items are completely separated. This process is accompanied by a progressive adjustment of the rates. More specifically, this is done as follows:

As a starting point, all items are considered to form a single class and they share the rate $(N + 1)/2$. Like in the classification condition, we then consider a splitting into a top class plus a low class according to whether the basis ranks are better or worse than a certain threshold. This can be done in $N - 1$ different ways depending on the number of members of the top class, which we shall denote by K. From all these possibilities, we shall choose the one that better agrees with the paired comparisons of the projected matrix. More precisely, the degree of agreement is measured by the **splitting gap**, which is defined by the formula $G = 2\,W/J - 1$, where W denotes the average of the $K(N - K)$ scores of the projected matrix that compare a top-class item with a low-class one, and J is the number of judges. As it is easily checked, the splitting gap G lies always between 0 and 1; the value 1 indicates that all of the judges have put each member of the top class in front of every member of the low class, whereas the value 0 is obtained when none of such comparisons had the support of a strict majority of judges.

Together with this splitting, the old rate $X_{\text{old}} = (N+1)/2$ is replaced by

$$X_{\text{new}} = \begin{cases} X_{\text{old}} - G\,(N-K)/2, & \text{for the top class,} \\ X_{\text{old}} + G\,K/2, & \text{for the low class.} \end{cases}$$

In our example, the splitting gap is found to be largest when we let the top class to include three members. In that case, we have $W = (4+3+5+3+5+4+5+3+4+3+ +5+3)\ /\ 12 = 3.9167$, and $G = 0.5667$. Correspondingly, the top class is rated 2.8667 and the low class is rated 4.8500.

Each of these two classes is then separately applied the same process of splitting and rate adjustment, and so on until we are done.

STEP 3: *Average.* In the preceding paragraph we have assumed that there is only one splitting point, i. e. one value of K, where the gap is largest. When there are several such points, then the successive dichotomy process can take different paths, which unfortunately may lead to different ratings. In that case, we define the reduction rating as the average of the ratings obtained by all possible paths.

10.4. The following table shows the reduction rating which is obtained for example G on the basis of the ranking obtained in § 9.6 by the RPN method. In addition, we give also the result which is obtained when a suitable generalization of the preceding method is applied to the weak ranking obtained in § 9.7 by the WRP method.

EXAMPLE G. RPN AND WRP WITH REDUCTION RATING

	Judges					*Scores*										
Id	A	B	C	D	E	61	62	63	64	65	66	67	*R*	*X*	*R′*	*X′*
61	1	1	7	7	7	-	2	2	2	2	2	2	*7*	5.1500	*7*	5.1500
62	4	2	3	3	1	**3**	-	**3**	2	**5**	**4**	**5**	*2*	2.8667	*2*	2.8667
63	2	6	2	4	2	**3**	2	-	2	**4**	**3**	**5**	*3*	3.0667	*3*	3.0667
64	3	3	1	2	5	**3**	**3**	**3**	-	**4**	**3**	**5**	*1*	2.6667	*1*	2.6667
65	6	4	5	6	4	**3**	0	1	1	-	**3**	2	*5*	4.7500	*5*	4.7500
66	7	5	6	1	3	**3**	1	2	2	2	-	**3**	*6*	4.9500	*5*	4.7500
67	5	7	4	5	6	**3**	0	0	0	**3**	2	-	*4*	4.5500	*5*	4.7500

As it is illustrated by the preceding examples, the results have indeed a quantitative character. As a general rule, when the input rankings contain a substantial amount of disagreement with each other, the reduction rates tend to be closer to each other.

10.5. The preceding procedure is related to certain methods that arise in a variety of scientific contexts in order to systematically arrange and classify a given set of items according to a matrix of data which measure their dissimilarity to each other. The general study of such methods is the subject of the so-called combinatorial data analysis and its cognate the hierarchical cluster analysis (or numerical taxonomy) [**23**; **21** : vol. 2, p. 1–10 and vol. 3, p. 623–630]. In particular, the algorithm of step 2 can be traced back to 1968, when it was used by Richard B. Mc Cammon in connection with certain geological studies.

In § 10.3 we stated that the method of reduction rating seems to fulfil the conditions of § 10.2 whenever the basis ranking is immune to majority complaints and the paired-comparison scores come from the rankings given by a panel of judges. From a mathematical point of view, this property is very remarkable, it does not seem to be known, and it calls for a mathematical proof. For the moment, such a proof is still lacking and that statement is based only on an extensive computational experimentation (involving more than one million cases). On the other hand, the main proposal of this paper does not hinge on the truth of that statement.

Let us briefly mention also that the reduction rates obtained above can be translated into another kind of rates which can be interpreted as winning quotas. For instance, in the case of example G the reduction rates X obtained above can be translated into the following quotas Q:

Example G. Quotas

Id	*R*	*X*	*Q* (%)
61	*7*	5.1500	5.420
62	*2*	2.8667	24.185
63	*3*	3.0667	21.065
64	*1*	2.6667	27.915
65	*5*	4.7500	7.104
66	*6*	4.9500	6.217
67	*4*	4.5500	8.093

Such quotas are based upon certain mathematical models which are briefly discussed in appendix A (included only in the technical version of this paper).

11. Multiple fields.

11.1. Let us step, at last, into the floor of multiple fields of valuation. In the case of dancesport this means multiple dances. As we said in § 2, we would like, if that is possible, to obtain a global result that reflects *at the same time* the general opinion of the judges and the all-round quality over those multiple fields. All of the methods considered in this section will implement the idea of all-round quality by means of an operation of addition or averaging over the different fields.

As a natural and straightforward generalization, all of these methods allow for the possibility of using weighted averages where each field has a different weight. Such a generalization would be suitable to figure skating.

In order to illustrate our arguments, we shall consider the example B of [**2**], which involves five dances and five judges:

Example B

	Cha-cha-cha					Samba					Rumba					Paso Doble					Jive				
Id	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
21	1	1	1	1	1	1	1	1	1	1	1	1	3	3	2	1	1	3	3	2	1	1	3	3	2
22	6	6	3	3	3	3	3	3	6	6	2	2	2	2	6	2	2	2	2	6	2	2	2	2	6
23	4	4	2	5	4	5	5	5	3	3	5	4	5	4	4	5	5	5	5	5	5	5	6	4	5
24	3	3	6	6	5	6	6	6	2	2	6	5	6	6	1	6	6	6	4	1	6	4	5	5	1
25	5	2	4	2	6	2	2	4	4	4	3	3	1	1	3	3	3	1	1	3	3	3	1	1	3
26	2	5	5	4	2	4	4	2	5	5	4	6	4	5	5	4	4	4	6	4	4	6	4	6	4

11.2. In order to deal with such a situation, the Traditional Skating System (see [**1**]) follows a procedure with two parts. The first part deals separately with each dance. As we mentioned in § 5.2–5.3, this is done according to the median ranks together with certain tie-breaking rules. As a result one obtains a ranking for each dance (save in the case of certain unbreakable ties, but we need not bother about it here). After that, the second part proceeds to combine the different dances into an all-round result. To that effect, the Traditional Skating System uses *rank addition* together with other tie-breaking rules.

For instance, in the case of example B that system produces the following summary:

Example B. TSS Summary

	Dances						
Id	C	S	R	P	J	*S*	*R*
21	1	1	2	2	2	**8**	*2*
22	2	2	1	1	1	**7**	*1*
23	3	5	4	5	6	**23**	*5*
24	6	6	6	6	5	**29**	*6*
25	5	3	3	3	3	**17**	*3*
26	4	4	5	4	4	**21**	*4*

Here, the columns C , S , R , P , J show the rankings obtained for the different dances according to the median ranks and the subsequent tie-breaking rules of part one, the next column, labeled S, shows the sum of those ranks, and R gives the final global result.

As it is illustrated by this example, *sometimes the results of the TSS are not what one would expect* after a common-sense-guided first look at the data of § 11.1. As it was pointed out in [**2** : § 2.3], such anomalies arise because the dance ranks that are added in part two are forgetting the more detailed information obtained in part one. In other words, *some of the information is being thrown away.*

Anomalies of this sort can occur even in situations involving only two couples. For instance, in a situation like

Example B′

	Cha-cha-cha							Samba							Rumba							Paso Doble							Jive						
Id	A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G
21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	2	2	2	2	1	1	1	2	2	2	2
22	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	2	2	2	1	1	1	1	2	2	2	1	1	1	1

the TSS all-round winner is couple 22 in spite of the fact couple 21 has many more first placings! TSS supporters will say that couple 21 has won only two dances, whereas couple 22 has won the other three. This is absolutely true. However, one can say also that both couples were very close to each other in three dances, but couple 21 was clearly superior in the other two!

Dancesport people are well aware of such strange features of TSS, but they have become rather used to it. Most often, their reply would be a simple “Yes, but the rule says. . . ” On the other hand, they can argue that similar effects are also present in certain well-recognized sports, like tennis. Anyway, the main question is: Should we not try to make sure that the best couple wins? Or do we simply want to play at applying certain mystic rules?

11.3. The loss of information that we have described can lead to remarkable paradoxes. For example, let us assume six couples, five judges and five dances. In one of the dances, say Waltz, we shall consider two possibilities that differ from each other only in the ranking given by one of the judges, say judge E. The following tables show these two possibilities together with the corresponding dance results according to the TSS:

Example K. TSS

Id	Waltz A	B	C	D	E	*R*
101	1	1	1	3	**1**	*1*
102	2	4	6	2	6	*3*
103	5	2	4	1	5	*2*
104	3	3	5	6	4	*6*
105	4	5	2	5	3	*4*
106	6	6	3	4	2	*5*

Example K′. TSS

Id	Waltz A	B	C	D	E	*R*
101	1	1	1	3	**6**	*1*
102	2	4	6	2	5	*5*
103	5	2	4	1	4	*3*
104	3	3	5	6	3	*2*
105	4	5	2	5	2	*4*
106	6	6	3	4	1	*6*

As it can be appreciated, the only difference between the data of K and K′ is the position that judge E gives to couple 101: this couple changes from first to last, while the others stay in the same order (all of their ranks are shifted one unit). We shall assume that judge E has perceived a major fault in couple 101, and this has caused a sincere change of his mind from K to K′. If everything works correctly, such a change should not do any good to couple 101. In other words, we can accept that this couple gets the same result as before, but certainly not a better one. However, as it is shown by the following tables, it can happen that this couple gets a better all-round result "thanks to" that penalization!

Example K. TSS Summary

Id	*Dances* W	T	V	F	Q	*S*	*R*
101	1	1	2	3	3	**10**	*2*
102	3	2	1	1	2	**9**	*1*
103	2	3	4	2	1	**12**	*3*
104	6	5	5	5	5	**26**	*5*
105	4	4	3	4	4	**19**	*4*
106	5	6	6	6	6	**29**	*6*

Example K′. TSS Summary

Id	*Dances* W	T	V	F	Q	*S*	*R*
101	1	1	2	3	3	**10**	*1*
102	5	2	1	1	2	**11**	*2*
103	3	3	4	2	1	**13**	*3*
104	2	5	5	5	5	**22**	*5*
105	4	4	3	4	4	**19**	*4*
106	6	6	6	6	6	**30**	*6*

So, even without rules 10 and 11, *the Traditional Skating System lacks a very desirable property which is known as* **monotonicity**: a displacement of one item in one of the input rankings should never produce an opposite displacement of that item in the final result.

11.4. In [**2**] the deficiencies of the Traditional Skating System were alleviated by means of an alternative that was called Revised Skating System (RSS). Its main idea consists in using the extended median method (§ 5.4) on each dance, and then combining the different dances by separately adding up each of the parameters used by that method, namely the median ranks and the adjacent sums. Finally, the all-round result is determined by the values of these aggregates according to a rule entirely analogous to the case of one dance: the median rank aggregates are used as the main criterion and ties are broken by the successive adjacent sum aggregates. For more details we refer the reader to [**2**].

In the case of example B, the application of RSS can be summarized as shown in the following table:

Example B. RSS Summary

	Dances (M)						
Id	C	S	R	P	J	M^D	*R*
21	1	1	2	2	2	**8**	*1*
22	3	3	2	2	2	**12**	*2*
23	4	5	4	5	5	**23**	*5*
24	5	6	6	6	5	**28**	*6*
25	4	4	3	3	3	**17**	*3*
26	4	4	5	4	4	**21**	*4*

Here, the columns C , S , R , P , J show the median ranks in the different dances, the next column, labeled M^D, shows the sum of those median ranks, and R gives the final global result. In this case the median rank aggregates did not contain any ties, so the adjacent sum aggregates were not needed.

As it can be appreciated, median ranks are slightly more quantitative than the dance ranks used by TSS. In example B this suffices to do justice to couple 21 . On the other hand, the Revised Skating System is ensured to always comply with monotonicity.

11.5. But there are also other issues. Until now we have taken for granted that the right way to combine multiple judges and multiple fields, consists in first combining the judges for every field and then combining the different fields. What about going the other way round, i. e. first combining the different fields for every judge and then combining the different judges? In a simple world, one would expect both procedures to produce the same results, but our world is not so simple: in fact, these two procedures can easily result in different winners!

In dancesport such a paradox came to light in 2001, when it was detected in several major events for which the weekly magazine *Dance News* uses to publish a by-judge analysis ("How The Adjudicators Saw It") in order to find out the all-round preferences of each judge. Most surprisingly, it was observed that sometimes a majority of judges preferred the couple that the Traditional Skating System had relegated to second position.

In order to see what is going on, we better look at a simplified example, like the following one borrowed from [**2**], where there are five dances and three judges:

Example E. By-dance analysis (TSS)

	W			T			V			F			Q			*Dances*						
Id	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	W	T	V	F	Q	*S*	*R*
51	1	1	2	1	2	1	2	1	1	2	2	2	2	2	2	1	1	1	2	2	**7**	*1*
52	2	2	1	2	1	2	1	2	2	1	1	1	1	1	1	2	2	2	1	1	**8**	*2*

Example E. By-judge analysis

	A					B					C					*Judges*				
Id	W	T	V	F	Q	W	T	V	F	Q	W	T	V	F	Q	A	B	C	*M*	*R*
51	1	1	2	2	2	1	2	1	2	2	2	1	1	2	2	8	8	8	**8**	*2*
52	2	2	1	1	1	2	1	2	1	1	1	2	2	1	1	7	7	7	**7**	*1*

The first table shows the marks grouped by dances and processed according to the Traditional Skating System. The all-round winner is couple 51 with a sum of 7 against the 8 of couple 52. The second table shows the same marks grouped by judge, together with a judge summary which tries to find out the all-round preference of every judge. Like before, this all-round information is found by addition over dances; the only difference is that now this is done separately for every judge and we later combine the different judges. As it can be seen, in this example all of the judges result in the same all-round winner, namely couple 52. Of course, this common result will be also the global conclusion of the by-judge analysis (in a general case we could define the global result by means of the median M). So, the by-judge analysis has a different winner than the by-dance analysis. As it was shown in [**2**], such discrepancies occur more often than suspected.

In social choice theory, such a phenomenon was clearly identified in 1976 by Douglas W. Rae and Hans Daudt. These authors named it **Ostrogorskii paradox** after the Russian political scientist Moisei Yakovlevich Ostrogorskii, whose major work, published in 1902, was essentially related to such kind of inconsistencies in democracy and the party system because of the existence of a variety of political issues as fields of valuation.

11.6. Naturally, the question immediately arises of *which way is the right one*. Is it correct to first aim at separate dance results and then combining these into an all-round result? Or, maybe the right way is to first find out the all-round valuations of every judge and then combining them into a collective result?

In favour of the by-dance analysis, one can argue that, after all, couples compete to win dances, not to "win adjudicators". In fact, the Syllabus of the Blackpool Dance Festival reads for example as follows:

> "The British Professional Modern Ballroom Dancing Championship ... Prizes will be awarded to the First, Second, Third, Fourth, Fifth and Sixth *in each dance*. The Couple showing the best 'All-round' standard in the four dances will be declared the British Professional Modern Champions".

If we take that statement and we replace the word 'dance' by 'judge' it does not make much sense.

Another significant remark in favour of the by-dance analysis can be made as follows: Although it deviates from the current dancesport custom, it would be quite reasonable that the panel of judges could vary from one dance to another, so as to allow for specialized panels. Obviously, in that case the grouping by judges does not make any sense whatsoever!

But let us go back to assuming a common panel of judges. In support of the by-judge analysis, one could argue that the main purpose of a dancesport competition is not so much to rank the couples in each particular dance, but to find out which couples are the best all-rounders. Therefore, one should certainly pay attention to what the judges are expressing in this regard.

As an additional support to that point of view, one could adduce that political elections work in this way: Most often, there are several fields of valuation on the table, but it is every elector's job to average them into his all-round preference. On the other hand, it is not so clear that this feature be a good thing. In fact, this is precisely the main point of Ostrogorskii's criticism against today's standard form of democracy. According to Rae and Daudt, his main thesis was that

> "all manner of mischief can result when issues are mixed together in a single contest".

On the other hand, in our setting the judges are not directly assessing all-round efficiencies. In the preceding discussion we have taken for granted that their all-round preferences can be derived by adding up the ranks that they give to the same couple in different dances. But this is not so clear. For instance, let us consider judge A in example E. Certainly, he has preferred couple 52 in three dances and couple 51 in the other two. But, how can we be sure that this means that he globally prefers couple 52? Would it not be possible that his preference of couple 52 to 51 in each of the three dances V, F, Q had been extremely weak, whereas in the other two dances he had found couple 51 far superior to 52, so that his all-round preference be in favour of couple 51?

The preceding argument is essentially the same that we used in § 11.2 as a criticism against the Traditional Skating System. In both cases, the problem is that ranks are not quantitative enough when it comes to combine dances by addition. In that connection, the by-dance analysis offers some prospects for being more quantitative, for instance by using the methods of § 10. In contrast, the by-judge analysis does not have such a possibility.

On the whole, the preceding arguments lead to the general conclusion that *the by-dance analysis lies on better grounds than the by-judge analysis.* However, if we want to avoid flaws like those discussed in § 11.2, *the information to be added up or averaged over dances should be more quantitative than simple ranks.*

11.7. In that connection, one possibility consists in making use of the reduction rating method of § 10. As we saw there, that method can be applied to any ranking obtained by a ranked-pairs procedure and its output is a rating that carefully quantifies the distances between consecutive items. In order to combine this information into an all-round result it will suffice to average such ratings over the different dances. More specifically, we shall consider a method of this kind where the WRP procedure of § 9.7 is the specific variant of ranked pairs used to deal with each dance. This method will be called **reduction-averaged WRP** (RAW).

As an illustration, the following table shows the application of this method to example B:

Example B. Reduction-averaged WRP

	Dances						
Id	C	S	R	P	J	*X*	*R*
21	1.0000	1.0000	2.1333	2.1333	2.1333	1.6800	*1*
22	4.1500	4.1000	2.5333	2.5333	2.5333	3.1700	*3*
23	3.7500	4.3000	4.3167	4.7667	4.8667	4.4000	*5*
24	4.6000	4.5000	5.0167	4.9667	4.4667	4.7100	*6*
25	3.7500	3.2000	2.3333	2.3333	2.3333	2.7900	*2*
26	3.7500	3.9000	4.6667	4.2667	4.6667	4.2500	*4*

Unfortunately, this method has several shortcomings. To begin with, it does not lend itself to pen and paper computation. However, ist biggest disadvantage is that it does not keep the property of consistency with respect to losers and winners.

At first sight, it looks like we have come to a dead end. After all our endeavour and previous success in satisfying the condition of consistency with respect to losers and winners, at the very end this condition seems to slip through our fingers. But there is still a possibility of succeeding! As we shall see next, there exists a way to combine dances in a quantitative way which succeeds in keeping the consistency properties achieved in the preceding sections.

11.8. In some sense, the most detailed dance summary is the matrix of paired-comparison scores. Admittedly, this kind of information cannot be accepted as a final result, but only as an intermediate one. However, this does not make it less interesting for our present purposes, as long as we end up with a more understandable all-round result.

Adding up or averaging the paired-comparison scores of different dances has indeed the desired effect that inequalities are compensated in a quantitative way. For instance, in example B the scores that compare couple 51 with 52 in the different dances are respectively 5 to 0, 5 to 0, 3 to 2, 3 to 2, and 3 to 2. By adding them up we obtain 19 to 6, which is equivalent to an average of 3.8 to 1.2. In the sports-league analogy of § 6.1, it is like we added up or averaged out the detailed scores of different matches.

So, a reasonable way to deal with multiple fields consists in adding up the paired-comparison scores of the different fields and then applying the ranked-pairs procedure of § 9. More specifically, we propose to use either the WRP method, which is described in § 9.7. As we saw there, in general this method produces a weak ranking, i. e. a ranking with ties. However, in comparison with the case of a single field, the maximum score is now larger, which has the good effect that ties will be less frequent.

We propose to call this multi-field method the **LCO System** (as a tribute to Llull, Condorcet and Ostrogorskii). The following table shows the application of LCO to example B. Together with the resulting all-round ranking we show also the corresponding reduction rating, but this elaboration does not properly belong to LCO.

Example B. LCO system

	Scores							
Id	21	22	23	24	25	26	*R*	*X*
21	-	**19**	**25**	**22**	**19**	**25**	*1*	1.6000
22	6	-	**17**	**18**	9	**16**	*3*	3.6000
23	0	8	-	**15**	5	10	*5*	4.3200
24	3	7	10	-	7	10	*6*	4.5200
25	6	**16**	**20**	**18**	-	**22**	*2*	2.8400
26	0	9	**15**	**15**	3	-	*4*	4.1200

Since the LCO System is nothing but a especial application of the WRP method, the former is automatically ensured to share all of the properties of the latter, namely: consistency with respect to losers and winners, immunity to majority complaints, and consistency with respect to clones. Furthermore, for a small number of items, like in the case of a dance-sport final, then we also have the practical advantage that the procedure lends itself to pen and paper computation.

As it is immediately appreciated, the LCO System does not need a previous computation of individual dance results. Of course, if required, such results can always be obtained separately. However, it must be clear that the all-round reduction rating corresponding to the LCO result does not have to coincide with the average of the one-dance analogous ratings; in fact, this last average can even involve a different ranking.

11.9. Let us look a bit more into the structure of the LCO System. As we have seen, it applies the ranked-pairs procedure to certain all-round paired-comparison scores. These all-round scores were introduced in the preceding subsection as the result of an addition over fields. However, it is clear that they can also be viewed as the result of putting together all input rankings, without taking into account any grouping, neither by fields nor by judges (the reader conversant with the Traditional Skating System will recognize this situation as analogous to Rule 11 of that system). This does not sound quite right. In fact, until now we were led to thinking that the tasks of combining judges and combining fields required different methods.

However, the truth is that *the addition of paired-comparison scores is a method suitable both for combining judges and for combining fields*: On the one hand, at the beginning of § 11.8 we saw how such an addition does really suit the idea of all-round quality over different fields. On the other hand, for a single field assessed by several judges, each of the paired-comparison scores can also be interpreted as the result of an addition over the different judges, where the scores that we are adding up are either 1 to 0 or 0 to 1 depending on which item of the pair is preferred by the judge under consideration.

At the same time, when it comes to deal with the resulting matrix of paired-comparison scores, *the method of ranked pairs is also suitable both for combining judges and for combining fields*: On the one hand, the task of combining different judges calls for complying with the majority principle (§ 5.1) as well as Condorcet's generalization (§ 6.1) and the condition of immunity to majority complaints (§ 9.1, 9.7). As we have seen, the method of ranked pairs and its variants have the virtue of satisfying all of these conditions. On the other hand, the ranked-pairs idea of giving priority to the paired-comparison propositions with higher scores is also most reasonable for the purpose of combining different fields (notice that we are not using the word 'majority'). In contrast, other methods more directly concerned with majorities, like rank medians (§ 5) and Copeland's method (§ 7), are not so suitable to the idea of all-round merit over different fields.

The ambivalence that we are talking about, i. e. being suitable both for combining judges and for combining fields, is not a contradiction at all. As a matter of fact, in the case of two items all reasonable methods have such an ambivalence: in particular, Borda's rank *addition* is always equivalent to choosing which of the two items is preferred by a *majority*. We can say that *the method of ranked pairs manages to extend such an ambivalence to the case of more than two items.*

The fact that the LCO System makes sense of treating judges and fields in the same way has an interesting consequence: As we have seen, the all-round paired-comparison scores can be thought of as the result of an addition over both judges and fields. More specifically, for every field we add up the elementary scores coming from the different judges, and then we add up the results of the different fields. However, everybody knows that the result of an addition does not depend on the order or grouping of the numbers being added up. Therefore, if all fields have the same panel of judges, the preceding scheme is exactly equivalent to the following one: for every judge we add up the elementary scores corresponding to the different fields, and then we add up the results obtained for the different judges. Because of this equivalence, we can say that *the LCO System completely avoids the Ostrogorskii paradox.*

12. Performances compared.

12.1. According to the preceding analyses, the LCO System has many good properties. For example, we know that it completely avoids any strong flip-flops, and that it avoids even a certain kind of weak flip-flops. However, in order to better compare the different alternatives that are available to us, it is very natural to ask for more details, like for instance: How often does the Traditional Skating System lead to strong flip-flops? How often does the LCO System produce ties? Or, how do they compare with each other in terms of robustness against biased judging? In order to answer such kind of questions, the only possible way consists in trying out the different methods on many particular cases and counting how often does one encounter those situations.

Now, in order for the conclusions of such a trial to be reliable enough, the number of tried cases must be suitably large. By means of standard statistical methods, one can see that a reasonable reliability requires so many cases that it becomes unfeasible to be based entirely on real data. Fortunately, this can be solved by means of a computer, for which it is not a problem to automatically generate, and process, the needed number of cases.

Besides the Traditional Skating System, the Revised Skating System, and the LCO System, we have taken the opportunity to test several other possibilities. One of them is the method which in § 11.7 was called reduction-averaged WRP. Another one is the so-called OBO System, which has been used in figure skating from 1998 to 2002 (§ 13.4). Unlike TSS and RSS, the OBO System uses paired comparisons. More specifically, each dance is dealt with by means of Copeland's method with Borda's method as a tie-breaker (§ 7.2). However, the all-round result is still derived by averaging the ranks obtained in the different fields, like in the TSS. We have tried also a method that we call sorted rank-averaged MTD, whereby each dance is dealt with by applying the criterion of minimum total disagreement, and rank averaging is used first to combine the possible multiple results in every dance, and second to combine all fields into an all-round result. On the other hand, we have tried several methods which imitate the LCO System in that the all-round results are not obtained by combining any intermediate dance ranks or rates, but they are obtained by proceeding as if we were dealing with a single dance; such methods have been termed "unsorted". In particular, we have included an unsorted RPN method, i. e. the RPN method of § 9.6 applied to the all-round paired-comparison scores. That method completely avoids any ties at the price of sometimes making use of a tie-breaker ranking. Here we have systematically adopted as such the ranking given by the first judge in the first dance. Finally, for reference purposes we have included also two variants of Borda's rank addition method, of which one is a sorted method and the other is unsorted. The unsorted version consists simply in adding up the ranks over both dances and judges. In the sorted version, each dance is rated by adding up the ranks given by the different judges, from these results one derives the corresponding dance ranks, and finally these *ranks* are added up over dances. The complete list of methods that we have put to trial is the following:

- SB : Sorted rank-averaged Borda method.
- TSS : Traditional Skating System (§ 11.2).
- RSS : Revised Skating System (§ 11.4).
- OBO : OBO System of figure skating.
- SMTD : Sorted rank-averaged MTD.
- RAW : Reduction-averaged WRP (§ 11.7).

UB : Unsorted Borda method.

UTSS : Unsorted TSS.

URSS : Unsorted RSS.

UOBO : Unsorted OBO.

UMTD : Unsorted rank-averaged MTD.

URPN : Unsorted RPN.

LCO : LCO System, i. e. unsorted WRP (§ 11.8).

In the forthcoming simulations, these methods will be compared by examining how often do they satisfy the properties listed below. For the sake of brevity, the accompanying explanations will omit certain technicalities that arise in connection with ties.

CP : Compliance with Condorcet's principle (§ 6.1). We shall count how often the all-round winner by the method under consideration coincides with the Condorcet winner according to the all-round paired-comparison scores. The result will be expressed by means of a percentage relative to the number of cases where the Condorcet winner did exist.

IMC : Immunity to majority complaints (§ 9.1, 9.7). Like the preceding one, this property will also be examined in relation to the all-round paired-comparison scores.

LW : Consistency with respect to losers and winners, i. e. absence of strong flip-flops (§ 4.2). This property will be tested by splitting the items into two classes according to whether their all-round rank is better or worse than the middle rank, and counting how often the application of the same method to each of these subsets is free from flip-flops.

AS : Consistency with respect to arbitrary subsets, i. e. absence of weak flip-flops (§ 4.2). This property will be tested analogously as above, but here the two subsets will be determined according to whether the all-round ranks are odd or even.

B1 : Robustness against biased judging. In order to test this property, we have looked at the effect of changing the marks of a single judge in the direction of favouring a particular couple. More specifically, we have chosen at random one of the judges and one of the couples and we have modified that judge's marks by raising that couple two places higher in every dance (or the corresponding maximum when that couple is already ranked first or second). As a measure of robustness, we have counted how often this modification does not have any effect in the final all-round ranking.

B2 : Robustness against severely biased judging. This test is analogous to the preceding one but, instead of two places, the couple in question is raised up to the first position in every dance.

M : Monotonicity. In order to test this property, we have proceeded as in the case of the two preceding properties except that the displacement has been reduced to a single place, and such a modification has been made in only one of the input rankings (chosen at random). As a measure of monotonicity, we have counted how often the final ranking did not show a displacement of the same couple in the opposite direction.

OP : Consistency in connection with Ostrogorskii paradox. Here we count how often the final ranking coincides with the result of the corresponding by-judge analysis (§ 11.5). For all of the sorted methods under consideration, the by-judge analysis is carried out by first combining the different fields by rank addition and then combining the judges by the same method as in the by-dance analysis. By definition, all unsorted methods are 100 % consistent in this connection.

R1 : Absence of ties without making use of "last resort" tie-breaking rules. We regard as such the following rules: Rules 10 and 11 of TSS (see [**1**]), OBO's and UOBO's use of a tie-breaker dance, UMTD's rank-averaging of multiple ranking results, and URPN's use of a tie-breaker ranking. For the other methods all tie-breaking rules are considered equally sound.

R2 : Absence of ties when the preceding last-resort tie-breaking rules are included.

W1 : Absence of ties involving the winner, without making use of last resort tie-breaking rules.

W2 : Absence of ties involving the winner, when the last-resort tie-breaking rules are included.

On the other hand, we have examined also how often does each method coincide with TSS. In this connection, we have considered the two following possibilities:

KR : Coincidence with TSS in producing the same all-round ranking.

KW : Coincidence with TSS in producing the same all-round winner.

All of the simulations below will assume 6 items, 5 fields and 9 judges. In other simulations we have allowed the number of judges to take other values, from 3 to 21. In general terms, when the number of judges increases all performances improve, but the comparisons between methods remain essentially the same.

12.2. The simulations reported in this subsection have been carried out under the assumption that all rankings are equally probable. This amounts to say that all items are very much alike from the point of view of the judges. We shall refer to this scenario as that of an "extremely close contest". Admittedly, such a situation is *not* representative of reality, where most often there will be some differences of merit easily recognizable by the judges. However, the case of an extremely close contest is most appropriate for the purpose of discriminating between different methods: In a more realistic scenario, the rankings given by the judges will tend to be more in agreement with each other; as a consequence, the results of different methods will also tend to be more coincident with each other, and most of the percentages of occurrence considered below will be sensibly higher (an idea of their magnitude will be obtained in the following section). However, as these percentages become higher, the differences between them become more difficult to ascertain. In contrast, by focusing on the case of an extremely close contest such differences become enlarged and easier to gauge.

The results of such a simulation are shown in the following table, where all of the figures are the higher the better and heoretically exact values are indicated by means of bold face. Each of the percentages obtained by simulation is based on 250 000 cases. This allows to have a statistical confidence of 99 % that if we tried all possible cases the corresponding percentages would differ from the given ones in less than 1 %.

Simulation 1. Extremely close contest

	Aspects													
Id	CP	IMC	LW	AS	B1	B2	M	OP	R1	R2	W1	W2	KR	KW
SB	71.3	14.6	35.0	54.7	39.4	25.9	**100**	14.6	58.7	–	92.5	–	*7.6*	*64.2*
TSS	67.5	6.7	13.9	29.5	30.1	20.2	99.5	4.7	38.1	99.97	87.0	99.999	–	–
RSS	72.5	9.5	17.5	38.3	33.4	22.5	**100**	6.5	99.9	–	99.98	–	*15.5*	*73.8*
OBO	71.5	9.1	21.5	44.0	28.3	17.7	99.6	6.5	53.4	98.7	90.8	99.8	*9.1*	*65.1*
SMTD	67.6	12.5	44.7	61.0	45.9	29.7	99.998	10.1	57.0	–	91.7	–	*5.4*	*60.2*
RAW	77.5	13.1	27.2	53.1	25.6	15.8	99.7	23.5	96.9	–	99.6	–	*5.5*	*58.9*
UB	83.3	24.9	45.6	71.5	38.1	25.3	**100**	**100**	70.7	–	95.1	–	*6.9*	*63.3*
UTSS	72.3	9.9	17.3	34.0	35.3	25.1	**100**	**100**	99.7	–	99.99	–	*9.1*	*66.5*
URSS	74.5	10.0	20.7	39.1	34.9	21.4	**100**	**100**	99.9	–	99.99	–	*9.5*	*66.8*
UOBO	**100**	50.3	57.7	64.3	37.9	24.6	**100**	**100**	34.1	89.3	77.0	98.4	*6.3*	*61.6*
UMTD	**100**	84.2	91.3	68.6	45.4	30.0	99.998	**100**	81.7	96.5	94.3	99.1	*5.6*	*59.1*
URPN	**100**	**100**	**100**	51.8	48.9	32.3	100	**100**	69.5	**100**	88.6	**100**	*5.5*	*59.2*
LCO	**100**	**100**	**100**	72.8	61.5	42.8	100	**100**	69.5	–	90.8	–	*4.5*	*56.1*

12.3. The foregoing table is complemented by the following one, whose figures are estimates of the real frequency of occurrence of the properties that we are considering. In general terms, these figures are less precise than those above, which is why less digits are given.

Simulation 2. A realistic scenario

	Aspects													
Id	CP	IMC	LW	AS	B1	B2	M	OP	R1	R2	W1	W2	KR	KW
SB	96	68	84	98	86	78	**100**	76	87	–	99	–	*62*	*96*
TSS	97	66	80	97	82	78	99.96	64	85	99.99	99	100	–	–
RSS	98	71	81	97	83	79	**100**	72	99.99	–	100	–	*74*	*97*
OBO	98	71	90	98	84	81	99.98	63	86	99.9	99	100	*72*	*98*
SMTD	97	76	95	99	89	86	100	72	87	–	99	–	*66*	*97*
RAW	99	80	88	99	83	79	99.98	69	99	–	99.9	–	*63*	*97*
UB	97	72	83	99	84	75	**100**	**100**	95	–	99.6	–	*62*	*97*
UTSS	97	73	82	97	84	81	**100**	**100**	99.9	–	100	–	*70*	*97*
URSS	98	74	84	98	83	79	**100**	**100**	99.99	–	100	–	*69*	*97*
UOBO	**100**	99	99	99	86	82	**100**	**100**	98	99.7	99.8	99.99	*66*	*97*
UMTD	**100**	99	99.5	99	87	82	100	**100**	99	99.9	99.9	99.99	*66*	*97*
URPN	**100**	**100**	**100**	98	87	83	100	**100**	99	**100**	99.9	**100**	*66*	*97*
LCO	**100**	**100**	**100**	99	87	83	100	**100**	99	–	99.9	–	*65*	*97*

These figures have been obtained as follows: In order to generate a large number of realistic cases, we have taken a real competition with many judges and we have extracted 9 judges in many different ways. More specifically, this procedure has been applied to the four major competitions of the last editions of the *Elsa Wells International Championships*, the results of which are published every year in *Dance News*. Most of these competitions used 19 judges, in which case a selection of 9 judges can be made in 92 378 different ways. The idea consists in going through these numerous possibilities by means of a computer. However, it must be taken into account that the possibilities that arise from the same real competition are not mutually independent. So, instead of exhausting a single real competition, we have taken several of them and each case has been used less intensively. More specifically, we have taken 24 real competitions with 16–21 judges and each of them has been used to generate 10 000 possibilities at random. As a result, we have simulated $24 \times 10\,000 = 240\,000$ realistic competitions with 9 judges.

12.4. As we have been remarking, the values obtained in the second simulation are more representative of a general competition, but their precision is rather limited. In contrast, the values obtained in the first simulation are not so representative of a general competition, but they are better at discriminating between different methods. With this in mind, the preceding tables show that: 1. RSS performs better than TSS in all aspects. 2. In general terms, the unsorted methods perform better than their sorted counterparts. 3. The best performances are achieved by the LCO System. The only aspect where this method is not the best one is the resolution of ties. 4. However, in practice, for the LCO System with 9 judges and 5 dances, ties involving the winner occur in less than 0.1 % of the cases. For 5 judges and 3 dances (not shown in the preceding tables) this percentage stays below 0.5 %.

Besides the properties considered above, another aspect which is highly desirable is the suitability for pen and paper computation. This condition is satisfied by all of the methods that we have considered, except SMTD, RAW, and UMTD.

Altogether, there is no doubt that *the best overall performance is achieved by the LCO System.*

13. The figure skating experience.

As it was mentioned in § 5.2, the Skating System used nowadays in dancesport was borrowed from figure skating in 1937/38. Since then, the judging and scrutineering systems of dancesport and figure skating have evolved more or less by their own, and they have developed certain differences. As it will be seen shortly, in spite of these differences the scrutineering problem is still essentially the same, with similar methods and paradoxes. In fact, the core of the scrutineering system —scoring system in figure skating terminology— was very much the same until 1997/98, when the occurrence of certain paradoxical phenomena led figure skating to adopt a new system —the so-called OBO system— based on the paired comparisons approach. Seemingly, this system fell short of the expectations, for another drastic revision has taken place in 2002–04!

All of this calls for taking a closer look at the figure skating experience, to see what can we learn out of it. More precisely, we shall look at the evolution of the judging and scoring systems used by the *International Skating Union (ISU)* in the last ten years (for olympic eligible competitions). With slight variations, these systems apply not only to figure skating proper, but also to ice dance and synchronized skating (which we consider included in a broad sense use of the term 'figure skating').

13.1. A figure skating competition consists of two or maybe three sections or "programs": Short Program, Free Skating, and sometimes also a Qualifying Free Skating program. As we shall see, for scrutineering effects these different sections play an analogous role to the different dances of a dancesport competition (in fact, in ice dance they are called "dances" instead of "programs").

In contrast to dancesport competitions, the competitors do not perform simultaneusly, but one by one, which has led to a different judging procedure: Instead of directly ranking all performances, the judges rate each performance separately, using a certain numerical scale, traditionally from 0.0 to 6.0 (the higher the better). Furthermore, they are required to decompose their assessment into certain different aspects or criteria, each of which is the matter of a different rating. Traditionally, there are two such aspects, which correspond more or less to the generic notions of "technique" and "presentation" (though they go by various names depending on the discipline and section under consideration).

Anyway, we have several different ratings coming from different sections, different aspects, and different judges. Of course, the problem consists in suitably combining these multiple ratings into a global result.

13.2. Until 1997/98 this was done by means of the so-called **Ordinal System**. For the sake of comparison with other methods, we shall divide it into four successive parts:

Part A: *Aggregation of the different judging aspects.* This is done by averaging the corresponding rates (with possibly different weights). As a result, we are left with one rating per judge and section.

Part B: *Conversion of ratings into rankings.* Each of the ratings resulting from A is converted into a simple ranking. In the case of ties, preference is given to some particular judging aspect. As a result, we obtain ranking per judge and section. Sometimes, this ranking may contain some unbreakable ties.

Part C: *Combining the different judges.* This is done by a method almost identical to the one used in dancesport (which we are calling Traditional Skating System, TSS). Besides the possibility that ties be already present in the given rankings, the only difference lies in the last tie-breaking rules (Rule 7a of the TSS is replaced by Borda's rank addition rule). As in the TSS, the result is one ranking per section (possibly with ties).

Part D: *Aggregation of the different sections.* As in dancesport, this is done by averaging the results of C. A slight difference is that here the different sections may be assigned different weights. For example, in two-section figure skating competitions, Free Skating is weighted twice as much as the Short Program (which is exactly equivalent to having three sections where two of them repeat the same result). Another minor difference is the way of dealing with ties: instead of the intricate Rules 10 and 11 of the TSS, here ties are broken by giving preference to some particular section (Free Skating).

As it is apparent from the preceding description, the Ordinal System of figure skating and the TSS of dancesport cannot deny being close relatives of each other. The main difference lies in the fact that the Ordinal System does not start from one ranking per judge and section: here, each of these rankings is replaced by several ratings. However, the fact is that these ratings are used for nothing else than deriving a ranking. Probably, they originated mainly as a means to deal with several performances in succession (instead of their being simultaneous, like in dancesport).

Being so similar to each other, most of the problems that can arise with the TSS apply also to the Ordinal System of figure skating. In particular, as we saw in § 5.5, they are liable to the strong flip-flop paradox. That is, deleting or adding a competitor may alter the relative placings of the competitors in front of him.

13.3. This is precisely what happened in the 1995 World Figure Skating Championships (Ladies), where the presence of the new-comer Michelle Kwan, who finished fourth, meant a swapping of the second and third places. A similar situation happened again in January of 1997 at the European Figure Skating Championships (Men), where the skater who finished sixth caused also the swapping of the second and third places. In fact, the problem had already been pointed out in [**25**] in connection with the Free Skating section of the 1994 European Figure Skating Championships (Ice Dance).

In figure skating, the flip-flop paradox is specially conspicuous and striking because the format of several performances in succession has led to the practice of presenting interim results throughout the event. Certainly, the fact that two previous competitors may swap places as a result of the later performance of a third one must be difficult to accept for people who does not know much about the way that results are obtained. This is specially true when the third competitor stands clearly behind both the previous ones (strong flip-flop paradox).

The incidents mentioned above added to an already existing distrust of the scoring system used by figure skating. The general lack of general knowledge about it (and the failure of the skating administrators to correct this situation) had led into the impression, among spectators, TV, and even some skating officials, that the scoring system was allowing the judges too much freedom to manipulate the results. Nothing farther from the truth, but that was the state of affairs. Most significatively, at the 1997 World Figure Skating Championships, held on March of that year in Lausanne, the IOC President expressed his view that the scoring system should be more understandable to the general public.

13.4. This situation led into a hurried search for a new system that fulfiled the requirements of being more understandable to the general people and avoiding the flip-flop paradox. As we have been seeing in this paper, this is by no means not an easy task. However, the *ISU* was very quick. On June of the same year 1997, a new system was announced. It was called the **OBO System**. The acronym stands for “one-by-one”, which is supposed to allude to the paired comparisons approach. In fact, the new system modified only the part C of § 12.2 and the modification consisted in replacing the traditional median-based method by the simplest of the paired-comparison methods, namely Copeland’s method with Borda’s tie-breaker, i. e. the method described in § 7.2. The OBO System was tried for the first time in August 1997, and it was adopted as definitive in June 1998 [**27**].

This was a mistake. In fact, in § 7.3 we saw that the method used by the OBO System is still liable to the strong flip-flop paradox ! Furthermore, one can hardly say that this method is more understandable to the general public than the traditional one. And finally, as we have seen in § 12, the new system is more easily influenced by eccentric marking than the traditional one.

What is more difficult to understand, all of these remarks had been made known by several figure skating experts well before the OBO System was definitely adopted ! [**28, 29**].

As we have seen, the flip-flop paradox cannot be completely avoided if we are to respect the majority principle. The most that one can get is consistency with respect to losers and winners, i. e. avoiding the strong flip-flop paradox, and the methods that satisfy this condition are rather scarce. In fact, in October 1998, Michel Truchon, a Canadian economist, published a paper about figure skating where he advocated a method with this property, namely MTD [**30**]. But by that time the *ISU* had already “definitively” adopted the OBO System !

Copeland’s method with Borda’s tie-breaker is present also in the system currently used by the *United Country Western Dance Council (UCWDC)*. In this case, however, that method is used only in the dance-combining stage, where it acts only as a tie-breaker (instead of rule 10 of the TSS).

13.5. In the 2002 Olympic Winter Games (Salt Lake City, February, 2002) figure skating was the matter of a big scandal: one of the judges admitted to having voted against her own will because of certain instructions that she had previously received from the president of her national federation. Naturally, this incident did not do any favour to the image of figure skating. In fact, for many people this was just the tip of the iceberg. Besides treating that particular case with suitable disciplinary measures, certainly the situation called for a general action to clean up figure skating judging. In answer to this requirement, the *ISU* has embarked on a major revision of the judging system. As a byproduct, this revision of the judging system carries with it a major reformulation of the scoring system, too.

To begin with, in June 2002 the *ISU* adopted a rule that tries to make judging as anonymous as possible so as to "protect judges from external influence". According to this rule, from the whole panel of officiating judges a "secret and sealed computer" randomly selects a smaller panel, whose marks are the only ones which are used to calculate the final result. Furthermore, the marks of the officiating judges are scrambled before they are displayed so that "it is impossible to tell which marks pertain to which judges". The identity of the judges really used, and the complete information about which marks pertain to which judges is saved in a disk which is put into a sealed envelope and sent to the *ISU* Secretariat.

Very clever ... but hardly appropriate: 1. Above all, hiding something is far from being the same as cleaning it up! 2. By itself, randomly discarding some of the judges is an absurdity. What about using it as a voting procedure, for example in the *ISU* Congress? Since some members may be subject to certain influences, the scrutiny should begin by randomly throwing away a certain fraction of the votes! Anyway, from a statistical point of view, such a procedure clearly makes the results less accurate, and also less robust against biased judging than a median procedure. 3. This procedure is critically dependant on the computer. In fact, there is no possibility for independently verifying that the software is working correctly and that no human mistakes have been made. For more detailed discussions along these lines, the reader is referred to [**28, 29, 32**].

But these criticisms have not prevented the above-mentioned anonymous judging procedure from being put into practice. In fact, since September 2002 it has been used in all major *ISU* events. For about one year, the subsequent treatment of the selected data was still using the OBO System. As a whole, the resulting system was termed **Interim Judging System**. This name distinguishes it from the complete **New Judging System**, which incorporates another major innovation that will be described in the next subsection. This New Judging System was defined mainly in April 2003 [**31**], after which it has been oficially used in several *ISU* events since September 2003, and, save for some minor changes, it has been adopted as definitive in June 2004.

13.6. The second major innovation contained in the New Judging System seeks to make judging more quantitative and more based on an absolute point scale. To this effect, the technical merit is not rated from 0.0 to 6.0 on a comparative basis, but it is rated without any predetermined upper bound by adding up certain scores which are obtained in every performed "technical element". More specifically, the score obtained in a particular technical element depends on two things: (a) the value, or grade of difficulty, of that particular element, which is worked into certain numerical tables established by the *ISU*; and (b) the grade of execution of that particular realization, which is assessed by the judges by means of a seven grade scale going from -3 to $+3$. On the other hand, the presentation is still rated on a comparative basis, but in order to make it more quantitative, this aspect is decomposed into as many as five different sub-aspects or "program components" (which go by the names of "Skating Skills", "Transitions", "Performance/Execution", "Choreography", and "Interpretation"), and each of these components is rated on scale going from 0.0 to 10.0.

Together with this big change in the way that judges express their assessments, the *ISU* proposes an entirely new system for working out the global result. This new system operates as follows:

Part 0 : *Random selection of the "real" judges.* As described in § 13.5.

Part 1 : *Combining the different judges.* In contrast to the procedure described in § 13.2, this is done separately not only for every section, but also for every presentation component and even every single technical element, before any operation for putting them together. On the other hand, the method used for combining the rates coming from different judges is trimmed averaging: before taking the average one discards both the best mark and the worst one.

Part 2 : *Combining the different elements, components and sections.* This is done by adding up the results of part 1, possibly after multiplying them by certain weighting coefficients: First of all, the scores for the different technical elements are added up to form the "total technical score". On the other hand, the scores for the different presentation components are also added up, possibly after multiplying them by certain coefficients, to form the "program component score". For each section, the two preceding scores are added up to form the so-called "total segment score", which gives the result of that section. Finally, the global result is obtained by adding up the the scores of the different sections.

Certainly, this system is very different from the one described in § 13.2. Apart from the random selection, which we have already discussed, the main differences are the following:

(a) Rates are not converted into ranks.

(b) The method for combining different judges is based upon trimmed averages of those rates (instead of median ranks or paired comparisons).

From a formal point of view, these changes have the following consequences (see [**28, 29, 32**]): 1. As a consequence of (a), the resulting system is completely free from any flip-flops. So, it looks like the *ISU* finally discovered how to get rid of flip-flops ... but there is a price to pay (which is why such methods were discarded many years ago)**:** 2. Trimmed averaging does not comply with the majority principle. For example, in the following table the winner is couple 71 in spite of the fact that couple 72 was the preferred one for a majority consisting in 5 judges out of 7:

Example H″

	Judges								
Id	A	B	C	D	E	F	G	*TA*	*R*
71	7.50	7.50	7.50	7.50	7.50	8.75	8.75	**7.75**	*1*
72	7.75	7.75	7.75	7.75	7.75	7.50	7.50	**7.70**	*2*

3. Trimmed averaging is less robust to biased judging than the median rank used in the Ordinal System.

13.7. But the main point against features (a) and (b) of the new scoring system has a more general character.

Clearly, these features go in the direction of treating the marks of the judges as if they were measurements of objective magnitudes and all of them were using the same absolute scale; furthermore, the using of trimmed averages instead of medians implies an assumption that deviations from the "true" value will not be so important.

Certainly, the changes in the judging procedures contain a significative effort to achieve a certain unification of scales.

However, in the best of cases such scales make sense only for "technique", but not at all for "presentation". In fact, concerning the latter aspect, the new judging system is essentially the same as before, with the only change that instead of one scale from 0.0 to 6.0 we have now five scales from 0.0 to 10.0. But unifying a scale is surely something more than merely choosing two extreme values. In that connection, the *ISU* provides a description of each of the five presentation components in terms of certain "criteria". However, most of these criteria, if not all of them, are quite subjective.

The unavoidable fact is that "presentation" is a subjective matter, and trying to deal with it by means of quantitative rates and their averages is so out of place as trying to use such methods for measuring the artistic qualities of a painting. Truly, the best way to deal with such matters consists in using ranks coming from different judges, the more judges the better, and processing this information by the methods that we have been discussing in this article.

Of course, if one decides to use rating methods for technique but ranking methods for presentation, then the problem arises of how to combine both parts. For that purpose, one could probably devise some procedure based upon the method for converting ranks into rates that is described in § 10 of this article (which requires the methods of § 9). However, the simplest, and probably also the safest, possibility is still to use ranking methods in both cases (without anonymous judging, of course).

Not without reason, the *ISU* reform described above is meeting with a strong disapproval coming from many figure skating circles. But that reform is only a little part of the story. For the whole of it, the interested reader is referred to [**33**].

14. Concluding discussion.

14.1. Let us begin this concluding section by recognizing that the question that we have been considering, i. e. how to combine several ranking judgements into a global result, is more intricate than it seemed at first sight. Even if this article has served only to convince the reader that the question is a technical one, it has still accomplished a great deal. As we have seen, underestimating the complexity of the matter easily leads to harmful mistakes. As it is put by Iain Mc Lean and Arnold B. Urken [**6** : p. 63] :

> "We still need to design choice rules according to principles from books. Otherwise, we might find ourselves having to believe six impossible things before breakfast".

The complexity of the matter does not lie in the calculations. In fact, most of the time we are simply comparing numbers and counting occurrences. The problem rather has to do with what we expect of the results. Certainly, we expect them to be "fair". But what does that mean? In this connection, it looks like most people let themselves be guided by the simplest case of two items and they take for granted that the generalization to three or more items will be straightforward. Now, for two items, all of the methods that we have considered are exactly equivalent to each other. However, this equivalence ceases to hold as soon as a third item comes into play. As we have been seeing, for more than two items the concept of correctness of a method unfolds into a variety of principles and desirable properties. The problem is that there is no method that satisfies all of these criteria at the same time. On the other hand, not all of these desirable properties are completely out of reach: some of them are still fully ensured by certain methods, and, even among the methods that do not completely ensure a given property, some methods may be found to fail less often than others. So, it still makes sense to choose a method in accordance with such considerations (as well as the particular features of dancesport).

In that connection, one of the main purposes of this article was to compare the Traditional Skating System (TSS) with the proposals that were put forward in [**1**], as well as any alternative that could arise from a more comprehensive analysis of the matter. Throughout this article we have been seeing that the TSS has two main contenders, namely the Revised Skating System (RSS) proposed in [**1**], and the LCO System formulated in this article. In general terms, they compare as follows: RSS does better than the TSS, and LCO does better than both TSS and RSS. In the following we summarize the arguments that substantiate such a statement. After that, we shall finish with a concise specification of the LCO procedure and a few final remarks.

The Double Revised Skating System (DRSS) of [**1**] can be described as an artificial compromise between the two conflicting points of view that produce the Ostrogorskii paradox (§ 11.5). However, in § 11.6 we found reasons to stick to the traditional point of view of combining first judges and then dances, which is the one adopted by both the TSS and the RSS.

14.2. For a single dance, TSS and RSS are not so different from each other. In fact, both of them use as primary criterion the median rank (§ 5.3). As a consequence, they comply with a very desirable property, which we have called the *majority principle* : if a majority of judges agree on allocating the first position to the same couple, then this couple should win. Another good property of the median rank is a remarkable *robustness against eccentric marks*; in fact, eccentric marks do not come into play unless it becomes necessary for breaking ties. For our purposes, these two features make median ranks far superior to rank addition, i. e. Borda's method, which ballroom dancing had been using before TSS. For a single dance, the only differences between TSS and RSS lie in the interpretation of the median in the case of an even number of judges, and in the criteria used for breaking ties. These differences make RSS slightly more robust than TSS against eccentric marking.

14.3. The main problems with TSS occur in the case of several dances. In that case, TSS summarizes each dance by means of the resulting ranking and then arrives at an all-round result by adding up these *dance ranks*. However, dance ranks are not able to differentiate between different degrees of closeness between two couples. As a consequence, TSS is affected by the following problems:

Problem 1, with TSS : *Lack of compensation.* When one couple gets ranked in front of another in some dances but behind it in the others, and they are in different degrees of closeness to each other depending on the dance, then the all-round result may be not equitable because of spurious effects. This is the case of examples B and B′ of § 11.1–11.2. A real case nearly as dramatic as example B took place recently in a certain domestic competition in Spain. On the other hand, a situation of the kind of example B′ occurred for instance in the *1994 German Open Professional Standard Championship*.

Problem 2, with TSS : *Lack of monotonicity.* As it has been illustrated by example K–K′ of § 11.3, a couple can get a better all-round result "thanks" to a worse mark by some of the judges. Similarly, it can get a worse all-round result "thanks" to a better mark by some of the judges. Certainly, this is in glaring contradiction with the essential principles of what a collective decision procedure is about.

Solution to problems 1 and 2, by means of RSS: In order to solve these problems, RSS does not summarize each dance by means of the resulting ranking, but it keeps a more accurate and detailed information about each dance. This allows for the proper compensating effects to take place when this information is added up into an all-round result.

14.4. As we already acknowledged in [**1**], RSS leaves the two following problems unsolved:

Problem 3, with TSS and RSS: *Ostrogorskii paradox.* When the all-round preferences of the judges are analyzed (for instance by rank addition) sometimes it is found that a majority of judges prefer a couple different from the winner according to TSS or RSS. The essentials of this paradox are illustrated by example E of § 11.5. In dancesport this paradox came to light in 2001, when it was detected in several major events; for more details the reader is referred to [**1**]. According to § 12.3, this paradox is estimated to happen with an approximate frequency of 36 % in the case of TSS and of 28 % in the case of RSS.

Problem 4, with TSS and RSS: *Strong flip-flops.* Let us consider a partial set of couples with *consecutive* final placements, and let us imagine that they had been the only ones to take part in the competition. The preferences of the judges about these couples are assumed to remain unchanged. In spite of that, it is quite possible that the same method produces now a different ranking. In other words, even assuming the same performances and the same preferences by the judges, the result may depend on the presence or absence of other couples of lower merit!

As an illustration, let us look at the *2003 United Kingdom Open Amateur Standard Rising Star*. According to *Dance News*, 1809 (March 6th, 2003), this event went as follows:

Example L. Strong TSS flip-flops in a real event

	Waltz									Tango									Foxtrot								
Id	A	B	C	D	E	F	G	H	I	A	B	C	D	E	F	G	H	I	A	B	C	D	E	F	G	H	I
7	2	3	1	1	2	1	2	4	6	4	2	4	2	3	3	3	5	6	2	3	1	1	3	1	3	4	6
17	4	4	6	4	5	4	1	6	5	2	3	6	3	4	1	1	3	5	3	4	6	4	4	4	2	6	4
91	6	6	5	5	6	3	6	1	3	5	6	3	5	6	2	6	2	3	5	5	3	6	6	2	6	2	3
104	3	2	4	6	4	5	5	5	2	3	1	2	4	5	6	4	4	2	4	1	2	5	5	5	4	5	2
134	5	5	2	3	1	6	4	2	4	6	5	1	1	1	4	2	1	4	6	6	4	2	1	6	1	1	5
168	1	1	3	2	3	2	3	3	1	1	4	5	6	2	5	5	6	1	1	2	5	3	2	3	5	3	1

	Quickstep									Viennese Waltz									*Dances*						
Id	A	B	C	D	E	F	G	H	I	A	B	C	D	E	F	G	H	I	W	T	F	Q	V	*S*	*R*
7	3	3	1	2	3	2	3	6	6	3	4	1	2	3	2	2	3	5	1	3	1	2	3	**10**	*1*
17	4	4	6	3	4	1	2	4	4	2	3	5	4	5	5	3	5	4	5	2	3	4	4	**18**	*4*
91	5	6	5	5	6	4	6	2	5	6	6	4	6	6	3	6	4	6	6	6	6	6	6	**30**	*6*
104	2	1	4	4	5	3	4	5	2	4	1	6	5	4	6	5	6	2	4	4	5	5	5	**23**	*5*
134	6	5	2	1	1	6	1	1	3	5	5	2	1	2	4	1	1	3	3	1	4	1	2	**11**	*2*
168	1	2	3	6	2	5	5	3	1	1	2	3	3	1	1	4	2	1	2	5	2	3	1	**13**	*3*

Let us consider now the three top couples, namely 7, 134 and 168, and let us imagine that they had been the only ones to take part in the competition. In that case, the same preferences of the judges would have produced the following marks and results:

Example L′. Strong TSS flip-flops in a real event

Id	Waltz A	B	C	D	E	F	G	H	I	Tango A	B	C	D	E	F	G	H	I	Foxtrot A	B	C	D	E	F	G	H	I
7	2	2	1	1	2	1	1	3	3	2	1	2	2	3	1	2	2	3	2	2	1	1	3	1	2	3	3
134	3	3	2	3	1	3	3	1	2	3	3	1	1	1	2	1	1	2	3	3	2	2	1	3	1	1	2
168	1	1	3	2	3	2	2	2	1	1	2	3	3	2	3	3	3	1	1	1	3	3	2	2	3	2	1

Id	Quickstep A	B	C	D	E	F	G	H	I	Viennese Waltz A	B	C	D	E	F	G	H	I	*Dances* W	T	F	Q	V	*S*	*R*
7	2	2	1	2	3	1	2	3	3	2	2	1	2	3	2	2	3	3	1	2	2	3	3	**11**	*3*
134	3	3	2	1	1	3	1	1	2	3	3	2	1	2	3	1	1	2	3	1	2	1	2	**9**	*1*
168	1	1	3	3	2	2	3	2	1	1	1	3	3	1	1	3	2	1	2	3	2	2	1	**10**	*2*

As it can be appreciated, these results are different from the previous ones: with six couples the winner was couple 7, whereas now the winner is couple 134. By the way, if we consider the five top couples, then the winner is neither 7 nor 134, but 168.

As it was mentioned in [**1**], in the European Standard Professional Championships of 1953 and 1954 such a situation was foreseen to happen, which motivated a previous agreement (!) that only four couples would be recalled to the final round, instead of the usual six. Definitely, the possibility that the winner depends on the number of finalists is most undesirable.

As we saw in § 13.3, such inconsistencies are especially conspicuous in figure skating, where the format of several performances in succession has led to the practice of presenting interim results throughout the event. In such circumstances, the inconsistency manifests itself in that two previous competitors may swap places as a result of the later performance of a third one. As we explained in § 13, until 1997/98 figure skating was using a system very similar to the TSS, but the occurrence of such phenomena in certain major events motivated the adoption of another system. (which was liable to the same phenomena!).

Fortunately, the format used in dancesport does not allow to present such partial results. Even so, we have already mentioned how such inconsistencies conditioned certain major events of the past. On the other hand, they show up easily and disturbingly in the event of disqualifications (which are not infrequent in some countries). Anyway, according to § 12.3 such inconsistencies are not rare at all, but they are lurking behind many real competitions.

Solution to problems 3 and 4, by means of LCO: The flip-flop inconsistencies of TSS and RSS have to do with the fact that these methods compare the marks obtained by *each couple* from *different judges*; the problem is that this does not conform to the true meaning of these marks: in fact, they have been given separately by *each judge* as an expression of the way that he himself compares the *different couples* with each other.

This remark leads to the point of view of *paired comparisons*, where two couples are compared according to the number of judges who prefer one to the other, independently of the absolute magnitude of the corresponding marks. In that context, the majority principle admits of the following natural extension, which is known as *Condorcet's principle*: if a couple beats every other in such paired comparisons, then that couple should win.

Unfortunately, it can still happen that no couple satisfies such a condition. So, in order to solve any possible case a further extension of the majority principle is still required. This final extension admits of several approaches, from which we have chosen a particular one which has especially good properties and determines not only the winner but also a whole ranking. This approach is known by the name of *ranked pairs* and it can be characterized by the so-called *criterion of immunity to majority complaints.*

In the context of paired comparisons, the proper way to combine different dances is by *addition of the paired-comparison scores*. By combining this principle with a particular version of the ranked-pairs procedure we obtain what has been called the LCO System. The LCO System avoids all of the preceding problems.[1] Of course, it complies with the majority principle, as well as all of its extensions mentioned above. On the other hand, it keeps the condition of being amenable to pen an paper computation. Finally, the simulations of § 12 show that it even improves upon TSS and RSS in terms of robustness against eccentric marking.

14.5. We stated that any method is bound to leave out some desirable property, and the LCO System is no exception. In the following we shall point out two desirable properties that are not always satisfied by LCO. In fact, they cannot be completely fulfiled by any method, at least if we want to keep the previous achievements. Accordingly, we shall refer to these issues as "unsolvable problems". Even so, we shall see that the performance of LCO in connection with these issues can still be considered better than TSS and RSS.

Unsolvable problem 5, with TSS, RSS and LCO : *Weak flip-flops.* The only difference with respect to strong flip-flops (problem 4) is that the latter refer to several couples with *consecutive* final placements, whereas weak flip-flops do not include such a condition. So, in a weak flip-flop two couples swap places because of a third one that gets in between (whereas in a strong flip-flop the third couple gets placed either above of below both of the place-swapping ones). Though still disturbing, weak flip-flops are somehow less offending than strong ones. Anyway, they are unavoidable (due to Condorcet cycles). On the other hand, the simulations of § 12 show that LCO produces less weak flip-flops than both TSS and RSS.

Unsolvable problem 6, with TSS, RSS and LCO : *Ties.* Even if the judges are constrained to give proper rankings, as it is the case in dancesport finals, and even if they are odd in number, one cannot rule out certain symmetrical situations where several proper rankings are equally entitled to be deemed as final results. Such cases lead to allow for results in the form of a weak ranking, i. e. a ranking with ties.

The TSS rarely leads to ties because of its rules 10 and 11; however, these rules lie on very poor grounds (see [**1**]). The LCO System is not so keen at avoiding ties, but the frequency of ties is still very low. Anyway, the LCO System produces ties far less often than TSS makes use of its poorly grounded rules 10 and 11 (for 9 judges and 5 dances the respective frequencies are roughly estimated at 1 % and 15 %; in the case of ties involving the winner these percentages reduce respectively to 0.06 % and 1.1 %).

[1] Unfortunately, we have not been able to give a general proof that it satisfies the monotonicity property as stated in § 11.3. However, this statement is supported by an extensive computational experimentation. On the other hand, one can prove the following restricted version of it: if the winner is promoted to a better position in some of the input rankings then it continues being the winner. Notice that the example of § 11.3 shows that the TSS lacks even such a restricted version.

14.6. In brief, the LCO System can be described as follows:

Starting point. For every dance of the final round, each of the judges ranks all of the couples without any ties.

Step 1: *Combining judges for each dance.* For each dance, one writes down the corresponding matrix, or table, of paired-comparison scores. The rows and columns of this table are both of them headed by the identifiers of the different couples. Each off-diagonal cell is filled in with the number of judges whose ranking supports the proposition that the couple indicated at the left is better than the one indicated at the top.

Step 2: *Addition over dances.* The paired-comparison scores of the different dances are added up so as to obtain the matrix of all-round paired-comparison scores.

Step 3: *WRP procedure.* The all-round ranking is obtained from this matrix by means of the WRP procedure (§ 9.7), that is: The different paired-comparison propositions are taken into consideration in the order determined by the magnitude of their scores, the larger the earlier. Every time that a proposition is taken into consideration, that proposition is adopted unless it contradicts those that had been already adopted with a strictly higher score, The process is continued until the adopted propositions form a complete ranking.

Step 4: *Reduction rating* (optional). If required, the reduction rating can be computed by applying the procedure described in § 10.3.

Step 5: *Separate dance results* (optional). If separate dance results are required, they are obtained by applying the preceding methods separately to the paired-comparison scores of each dance.

14.7. Although the reduction rating is not an integral part of the LCO System, it is very interesting as a supplement because it provides a finely tuned rating of the constestants. In particular, it quantifies how clear was the winner.

As an illustration, the following tables show the results of applying the LCO System and the reduction rating procedure to two recent major events. For completeness, we have included also the reduction ratings that are obtained for each separate dance. However, it must be clear that the all-round rating can differ from the average of the separate-dance ones; in particular, this average could even involve a different ranking.

Example M. LCO System and reduction rating
2003 Elsa Wells International Championships · Professional Standard.

		Dances						
Id		W	T	F	Q	V	*R*	*X*
4	Timothy Howson & Joanne Bolton	1.9737	2.5614	1.6579	2.7281	2.0263	*2*	**2.2500**
38	Jonathan Wilkins & Katusha Demidova	3.5395	3.5789	3.6447	3.4912	3.2368	*4*	**3.3342**
59	William Pino & Alessandra Bucciarelli	4.3114	2.1667	4.5307	2.6754	3.5000	*3*	**3.3026**
70	Alan & Donna Shingler	4.7588	5.0000	4.6623	5.5263	5.1579	*5*	**5.0789**
83	Jonathan Crossley & Lyn Marriner	4.8114	5.3684	4.7939	5.0526	5.2105	*6*	**5.0895**
118	Christoffer Hawkins & Hazel Newberry	1.6053	2.3246	1.7105	1.5263	1.8684	*1*	**1.9447**

Example N. LCO System and reduction rating
2003 IDSF World Latin Championship.

Id		*Dances* S	C	R	P	J	*R*	*X*
12	Ricardo Cocchi & Joanne Wilkinson	2.0000	2.0000	2.0556	1.6667	1.7222	*2*	**1.8889**
19	Franco Formica & Oksana Nikiforova	1.2222	1.2222	1.2778	1.5556	1.3889	*1*	**1.3333**
33	Eugene Katsevman & Maria Manusova	4.3889	3.8704	3.8056	3.9259	3.9537	*3*	**3.9963**
36	Klaus Kongsdal & Viktoria Franova	3.7222	3.9815	3.6944	4.0370	4.5648	*4*	**4.0185**
70	Peter Stokkebroe & Kristina Juel	5.1667	5.5000	5.2500	5.5000	5.5278	*6*	**5.3889**
78	Maurizio Vescovo & Melinda Törökgyörgy	4.5000	4.4259	4.9167	4.3148	3.8426	*5*	**4.3741**

As it can be appreciated, reduction ratings are much more precise than the corresponding rankings. For instance, in example M couples 70 and 83 are ranked respectively 5th and 6th, but their rates differ only in about one hundredth of a placement!

14.8. Most often, the LCO System will produce the same all-round ranking as the Traditional Skating System. However, the percentage of cases where these two systems lead to different results is not negligible at all. According to § 12.3, it is estimated that in real competitions with five dances they produce *different all-round rankings with a frequency of about 35 %*. For instance, in the case of example N the all-round ranking produced by TSS differs from the LCO one given above in that couples 33 and 36 exchange the 3rd and 4th placements. On the other hand, nothing prevents such discrepancies from affecting the winner. According to § 12.3, LCO and TSS produce *different all-round winners with a frequency of about 3 %*. Roughly speaking, the latter percentage amounts to something like one or two major events per year.

14.9. Summing up, the main conclusions of this article are the following:

Conclusion 1. The rightness and fairness of a method for combining several ranking judgements into an all-round result can be analyzed in terms of whether and how much does that method satisfy certain desirable properties. In many cases, this kind of analysis allows to conclude that a certain method is better than another.

Conclusion 2. The Revised Skating System is better than the Traditional Skating System. On the other hand, both of these systems are based on the same main criterion.

Conclusion 3. The LCO System is significatively better than both the Traditional Skating System and the Revised Skating System. In particular, by using the paired-comparison approach, the LCO System is more consistent with the true meaning of the judges' marks.

Conclusion 4. The LCO System can be consistently supplemented with a finely tuned rating of the contestants, namely the reduction rating.

To make it easier to test it, the LCO System with reduction rating will be implemented in a forthcoming version of *MiniSkate* [**34**].

Appendix A. From reduction rates to quotas.

This appendix explains the method used for converting reduction rates into quotas, or winning probability estimates, as mentioned in the note at the end of § 10.5. This method is based upon certain probability models which represent the process of producing a ranking judgement. These models associate each item with a theoretical parameter that represents its quality. As we shall see, the values of these parameters can be put in correspondence with the winning probabilities of the different items, and also with their mean ranks. Our method consists simply in interpreting the reduction rates of § 10 as estimates of the mean ranks, and using the preceding correspondences to convert them into estimates of the winning probabilities.

A.1. The approach that we are referring to was introduced in the 1920s independently by the American psychologist Louis Leon Thurstone and the German mathematician Ernst Zermelo, whose works were published respectively in 1927 and 1929. Thurstone was motivated by psychophysics, a branch of psychology that tried to quantify psychological sensations, and he formulated a general theory about the very problem of obtaining a quantitative rating out of qualitative paired-comparison data; most remarkably, he pointed out that his method could be used for the accurate rating of psychological judgments of a highly subjective nature, such as aesthetic judgments. Zermelo's motivation was much more specific: he was looking for a method to rate chess players according to the outcomes of their matches.

At first sight, the methods introduced by these two authors look very different from each other, but a suitable analysis shows that Zermelo's method can be reformulated in such a way that it fits in the main framework of Thurstone's theory. Following other authors, we shall refer to this general framework as **Thurstone's theory of comparative judgement**. On the other hand, Zermelo's original formulation is very meaningful by itself. In fact, it has been reinvented on several occasions, and today it is often associated with the names of Ralph Allan Bradley and Milton E. Terry, who arrived at essentially the same formulation in 1952. This alternative point of view was developed in more depth by the American mathematician Robert Duncan Luce who in 1959 made it into what is now called **Luce's theory of choice**. For general information concerning these theories, the reader is referred to [**21** : vol. 5, p. 167–170 ; vol. 9, p. 237–241] and the references therein.

This conceptual framework will be very useful for our purposes, but for a proper application to our situation we need a certain elaboration. In fact, both Thurstone and Zermelo assumed that each paired comparison is independent of the others, i. e. its result is not influenced by those of the other paired comparisons. This assumption is most suitable to sports leagues, and it may be reasonable for certain judging settings. However, it does not apply to our situation, where judges are required to express their opinion in the form of a ranking. For instance, if all of these rankings agree on preferring item a to b and b to c, then the comparison of a with c is certainly forced to be that all rankings prefer a to c. Fortunately, the ideas of Thurstone and Zermelo admit of a suitable *extension to the ranking situation*. Such an extension appears already in the reports of the British statisticians P. A. P. Moran and H. E. Daniels in a crucial *Symposium on Ranking Methods* that was held in 1950 before the *Royal Statistical Society*. From now on, the above-mentioned theories shall be regarded as including this extension to the ranking situation.

A.2. Each of these theories is based upon a *probability model* that relates rates to preferences. The word *model* is especially appropriate here because of certain hypotheses that shall be made about the mechanism of judgement, i. e. the way that judges arrive at their decisions. Certainly, what happens in the judges' minds is not clear at all, and it may well be different from judge to judge. However, the models that we are about to introduce do not intend to describe the actual mechanism of judgement but only a virtual equivalent of it. In other words, even though the actual mechanism may be very different, we shall assume that the outcome of a judgement is the same *as if* the judge had proceeded according to the imaginary model in question.

Having said that, it will be seen that the mechanism of judgement postulated by Luce's theory is not so different from the actual procedure used by many dancesport judges. In contrast, the mechanism postulated by Thurstone's theory is more related to the rating procedures used in figure skating.

On the other hand, the fact that we are using the term *probability* means that we are admitting a certain unpredictable variability in the outcome of different judgements concerning the same items. This variability may be ascribed either to the judges or to the conditions of observation. Even so, we assume that each possible outcome has a certain probability of occurring. For instance, we might say that the probability that a is preferred to b is 85 %. This probability can be viewed as the proportion of times that the outcome in question would occur in an hypothetical endless series of judgements, i. e. if we asked an endless number of judges and each of them decided anew countlessly many times.

A.3. The basic hypotheses of Thurstone's theory of comparative judgement can be stated as follows:

Hypothesis T1. An elementary judgement consists in rating an item along a certain absolute scale of merit. The result will be called the **judged rate** of that item. A ranking is obtained by separately rating all items and then ordering them according to their judged rates.

Hypothesis T2. The judged rates are concentrated around a particular value, which represents the **ideal rate** of that item. More specifically, this happens according to a certain **probability law**.

A probability law is a mathematical formula that, based on the ideal rates of the alternatives in consideration, allows to compute the probabilities of different kinds of events. In particular, one can compute the probability that a given item is rated better than a certain value, or the probability that it is rated better than another, of the probability that several items are rated in agreement with a certain specified ranking.

Therefore, the model is not completely set until one adopts a specific probability law, for which there are several possibilities. By its standard status in the mathematical theory of probability and statistics, a most natural option is the so-called normal probability law. This was indeed the option adopted by Thurstone. Generally speaking, however, the theory allows for other probability laws.

A.4. As it was mentioned earlier on, Luce's theory of choice is exactly equivalent to a particular case of Thurstone's theory of comparative judgement, i. e. it corresponds to a certain particular choice of the probability law of hypothesis T2. However, this particular case admits of an alternative formulation which is specially meaningful. In fact, Luce's theory was originally formulated in this alternative way and its compatibility with Thurstone's framework was not revealed until some years later.

In the following we shall adopt the particular hypotheses and point of view of Luce's theory of choice. The reasons for doing so are the following: (a) Luce's conceptual framework is especially suitable to our purposes; (b) Changing to other cases of Thurstone's theory is known to have little effect on rating results; (c) Computations are much easier.

A.5. The basic hypotheses of Luce's theory of choice can be stated as follows:

Hypothesis L1. An elementary judgement consists in **choosing** the best of several items. Any other judgement is obtained as a combination of such elementary judgements. In particular, a ranking is obtained by first choosing the winner, i. e. the best of the whole set, then choosing the best of the remainder, and so on. Each of these choices is made with independence of all the others.

Hypothesis L2 : *Choice Axiom.* For every item, its probability of being chosen depends on which other items are under consideration. However, for any pair of items, their probabilities of being chosen are always in the same ratio, independently of which other items are under consideration.

In particular, this ratio does not change when the set of alternatives under consideration is enlarged or diminished. Therefore, hypothesis L2 amounts to a quantitative form of the principle of independence of irrelevant alternatives. However, it is important to notice that this principle appears here as a property of the model, but not as a property of any method for estimating the ideal probabilities from experimental results.

According to the preceding hypothesis, the probabilities of different choice events are not completely independent of each other. As a matter of fact, one can see that hypothesis L2 implies that every item can be associated a number, traditionally called its **strength**, so that these numbers determine the probability of any choice event. Specifically, the probability of any choice event is determined by the following rule: *for any set of candidate items, each of its members is chosen with a probability equal to the ratio between its strength and the total strength of all candidates.* By a repeated use of this rule, one can work out also the probability of any particular ranking. In fact, it suffices to multiply the probabilities of the successive choices that correspond to that ranking according to hypothesis L1.

For example, for three items $\mathsf{a}, \mathsf{b}, \mathsf{c}$ with strengths 0.60, 0.20, 0.20, the probability of choosing a as the winner is $0.60/(0.60 + 0.20 + 0.20) = 0.60$, the probability of choosing a from the set $\{\mathsf{a}, \mathsf{b}\}$ is $0.60/(0.60+0.20) = 0.75$, and the probability of choosing b from the set $\{\mathsf{b}, \mathsf{c}\}$ is $0.20/(0.20+0.20) = 0.50$. As a consequence, the ranking $\mathsf{a} \succ \mathsf{b} \succ \mathsf{c}$ has a probability of $0.60 \times 0.50 = 0.30$, whereas the ranking $\mathsf{c} \succ \mathsf{a} \succ \mathsf{b}$ has a probability of $0.20 \times 0.75 = 0.15$.

When there are several items with strength exactly equal to 0, then the computation of certain probabilities leads to indeterminacy. Such situations require a special treatment which essentially consists in considering those items separately. For simplicity, in the following we shall keep away from such especial situations.

According to the rule above, choice probabilities are given by strength ratios. As a consequence, multiplying all strengths by the same number has no effect on the result. So, in analogy with many physical magnitudes, the numerical value of a strength depends on which unit is being used. In a specific context, say a particular dancesport final, it is most natural to take as unit the total strength of all items under consideration. In that case, the numerical value of a strength coincides with the probability that the corresponding item be chosen as the best one.

Unless we state it otherwise, in the following we shall always **normalize** strengths in that way. Consequently, they will then be represented by numbers from 0 to 1 with their sum total equal to 1.

In a parliamentary election based on party lists, the strengths would give the ideal proportions, or quotas, for the assignment of seats to the parties.

A.6. Clearly, stronger items have more probability of being preferred over others. Therefore, strengths provide a way of quantitatively rating the items under consideration. Such ratings differ from rank-like ratings in several respects. To begin with, a larger strength means a better item (contrarily to rank-like rates). On the other hand, the proper way to compare strength rates is by their ratio (rather than their difference). However, as we shall see next, there is a natural way to translate strengths into rank-like rates, and viceversa.

In fact, we have seen how the strengths allow us to calculate the probability of every single ranking. Now, by combining these probabilities we can work out the probability that a particular item achieves a certain rank. Finally, once we know the probability of each rank for a particular item, it is a simple matter to calculate the corresponding **mean rank**. According to the elementary theory of probability, this parameter can be viewed as the average of the ranks obtained by that item in an endless series of ranking judgements.

For example, consider again the case of three items `a` , `b` , `c` with strengths 0.60, 0.20, 0.20. In that case, we obtain that item `a` will be ranked 1st, 2nd, 3rd with respective probabilities 0.60, 0.30, 0.10 (the first of which we already knew), and therefore the mean rank of `a` is $0.60\times1+0.30\times2+0.10\times3 = 1.5$. Similarly, for items `b` and `c` we obtain that both of them will be ranked 1st, 2nd, 3rd with respective probabilities 0.20, 0.35, 0.45, which gives a mean rank equal to 2.25. So, the strengths 0.60, 0.20, 0.20 translate into the mean ranks 1.5, 2.25, 2.25.

When the number of items begins to be of some importance, the preceding procedure requires many calculations. However, the mean ranks can be computed also by means of an alternative formula that involves far less computations, namely:

$$X_i = N - \sum_{j\neq i} \frac{Q_i}{Q_i + Q_j},$$

where X_i, Q_i and N represent respectively the mean rank of item i, its strength, and the number of items.

This formula is analogous to the formula $S/J = N - T/J$ of § 7.2, where instead of theoretical means and probabilities we were dealing with the corresponding practical averages and frequencies.

By means of the preceding formula it is very easy to convert strengths into mean ranks. Going in the opposite direction, as it is the case in § 10.5, is not so easy: Unfortunately, the formula above cannot be rearranged so as to convert mean ranks into strengths. However, it can still be used to devise certain iterative algorithms that do this job and are easily implemented with the help of a computer.

Appendix B. Mathematical justification of the method of ranked pairs with natural tie-breaking. (Version 2, 15th May 2006, improved thanks to the kind remarks of Rosa Camps, Jaume Moncasi and Laia Saumell).

In this appendix we give mathematical proofs of the crucial properties of the method of ranked pairs (RP) and its variation the method of ranked pairs with natural tie-breaking (RPN). These proofs provide a guarantee that these properties will be satisfied at absolutely all times. Many of the ideas of these proofs can be found in [**13, 14, 15, 16, 19**]. However, in some aspects these sources do not cover exactly our situation, and in other aspects they are not clear enough. So, we felt it worthwhile giving these proofs here. A similar study of the method of weak ranked pairs (WRP) is in progress.

B.1. Main tools.

Besides elementary logic, most of the time these proofs will use nothing else than elementary set theory (with a strong flavour of graph theory and graph algorithms). To begin with, we must distinguish the **set of items** A. The number of items is assumed to be finite, and it will be denoted by N. We shall be particularly concerned with **relations** on A. Stating that two items a and b are in a certain relation ρ is equivalent to saying that the (ordered) **pair** formed by these two items is a member of a certain set ρ.

The pair formed by a and b will be denoted as (a, b) or simply as ab. Instead of writing $ab \in \rho$, sometimes we shall use the alternative notation $a\,\rho\,b$; in that case, $ab \notin \rho$ will be expressed as $a\,\bar{\rho}\,b$. Sometimes, pairs will be denoted without making reference to its components; for instance, we can represent a pair by π. In that case, π' will denote the pair opposite to π, i.e. if $\pi = ab$ then $\pi' = ba$.

The pairs that consist of two copies of the same item, i.e. those of the form aa, are not relevant for our purposes. So, it will be convenient to restrict our attention to proper pairs, i.e. those pairs of the form ab with $a \neq b$. The set formed by all such pairs of items from A will be denoted by $\Pi(A)$, or simply by Π. So $\pi \in \Pi$ implies $\pi' \neq \pi$. From now on we shall restrict our attention to relations contained in Π; in other words, our relations will systematically exclude any pairs of the form aa (such relations are sometimes called 'irreflexive'). In particular, the relation that includes the whole of Π will be called **complete tie**. For every relation $\rho \subset \Pi$, we shall denote by ρ' the relation that consists of all pairs of the form π' where $\pi \in \rho$; ρ' will be called the **converse** of ρ. On the other hand, we shall denote by $\bar{\rho}$ the relation that consists of all pairs π for which $\pi \notin \rho$; $\bar{\rho}$ will be called the **complement** of ρ.

A relation $\rho \subset \Pi$ will be called:

- **total**, or complete, when at least one of $ab \in \rho$ and $ba \in \rho$ holds for every pair ab.
- **antisymmetric** when $ab \in \rho$ and $ba \in \rho$ cannot occur simultaneously.
- **transitive** when the simultaneous occurrence of $ab \in \rho$ and $bc \in \rho$ implies $ac \in \rho$.
- a **weak ranking** when it is at the same time transitive and total.
- a (strong) **ranking** when it is at the same time transitive, total and antisymmetric.

Here we are deviating from the standard terminology of elementary set theory, where the terms 'order' or 'ordering' are used instead of 'ranking'.

Besides pairs, we shall be concerned also with longer sequences $a_0 a_1 \ldots a_n$. They will be often referred to as **paths**, and in the case $a_n = a_0$ we shall call them **cycles**. When $a_i a_{i+1} \in \rho$ for every i, we will say that the path $a_0 a_1 \ldots a_n$ is **contained in** ρ. In that case we will say that the pair a_0a_n is **supported by a path in** ρ, and also that a_0 and a_n are **indirectly related through** ρ. When ρ is transitive, the condition "a is indirectly related to b through ρ" implies $ab \in \rho$. In general, however, it defines a new relation, which will be called the **transitive closure** of ρ, and will be denoted by ρ^*. The transitive-closure operator is easily seen to have the following properties: $\alpha^* \subset \beta^*$ whenever $\alpha \subset \beta$; $(\alpha \cap \beta)^* \subset (\alpha^*) \cap (\beta^*)$; $(\alpha^*) \cup (\beta^*) \subset (\alpha \cup \beta)^*$; $(\alpha^*)^* = \alpha^*$; $\{\pi\}^* = \{\pi\}$. On the other hand, one can easily check that α^* is antisymmetric if and only if α contains no cycle; more specifically, $ab, ba \in \alpha^*$ if and only if α contains a cycle that includes both a and b.

A subset of items $C \subset A$ is called a **segment**, or an interval, of a relation ρ when the simultaneous occurrence of $ax \in \rho$ and $xb \in \rho$ with $a, b \in C$ implies $x \in C$. When C is a segment of ρ, it will be useful to consider a new set of items $\widetilde{A}$ and a new relation $\widetilde{\rho}$ defined in the following way: $\widetilde{A}$ is obtained from A by replacing the set C by a single item $\widetilde{c}$, i.e. $\widetilde{A} = (A \setminus C) \cup \{\widetilde{c}\}$; for $a, b \in A \setminus C$, $ab \in \widetilde{\rho}$ if and only if $ab \in \rho$, $a\widetilde{c} \in \widetilde{\rho}$ if and only if there exists $c \in C$ such that $ac \in \rho$, and $\widetilde{c}b \in \widetilde{\rho}$ if and only if there exists $c \in C$ such that $cb \in \rho$. We shall refer to this operation as the **contraction** along the segment C. If ρ is a ranking or a weak ranking on A, then $\widetilde{\rho}$ is respectively a ranking or weak ranking on $\widetilde{A}$.

When ρ is a ranking, every item a can be associated a different number $r_\rho(a)$ from 1 to N so that $ab \in \rho$ is equivalent to saying that $r_\rho(a) < r_\rho(b)$. This number is called the **rank** of a in ρ. Sometimes the notation $r(a, \rho)$ will be used instead of $r_\rho(a)$. Notice that a higher rank in ordinary language means a lower value of $r(a)$. An **inversion** is an operation that transforms a ranking ρ into another one σ by exchanging the places of two particular consecutive items, without any other change. In other words, there is a pair ab such that $r(a, \sigma) = r(a, \rho) - 1 = r(b, \rho) = r(b, \sigma) - 1$, but $r(x, \sigma) = r(x, \rho)$ for any x other than a and b.

B.2. Aggregation of preferences by the method of ranked pairs.

These tools will be applied to the problem of aggregating several individual preferences into a global one. We assume that the **individual preferences** are specified by certain rankings ρ_j with possibly different weights w_j. In general, they may be accompanied by a **tie-breaker** ρ_{tb}, i.e. a ranking which will be used for breaking certain ties. The problem consists in summarizing all of this information into a **global ranking** R in accordance with certain desirable properties.

Most of the proofs below are still valid or can easily be adapted to a more general situation where the individual preferences need not be rankings, but they can be more general relations. In contrast, the assumption that the tie-breaker is a ranking is not so easy to do without.

We start by converting the data into a set of paired-comparison **scores** $s(\pi)$. To this effect, we translate each relation ρ_j into a set of elementary scores $s_j(\pi)$ according to the following rule:

$$s_j(\pi) = \begin{cases} 1, & \text{if } \pi \in \rho_j \text{ and } \pi' \notin \rho_j; \\ 1/2, & \text{if } \pi \in \rho_j \text{ and } \pi' \in \rho_j; \\ 0, & \text{if } \pi \notin \rho_j. \end{cases}$$

These numbers are then aggregated into the global scores $s(\pi)$ by means of a weighted average:

$$s(\pi) = \Big(\sum\nolimits_j w_j\, s_j(\pi)\Big) \Big/ \Big(\sum\nolimits_j w_j\Big).$$

Clearly, these numbers satisfy $0 \leq s(\pi) \leq 1$.

Two natural candidates for giving the global preference are the **majority** relation μ, which consists of all pairs π for which $s(\pi) > 1/2$, and the **weak majority** relation ν, which consists of all pairs π for which $s(\pi) \geq 1/2$. When the the ρ_j are total, in particular when they are rankings, the scores $s(\pi)$ satisfy the equality $s(\pi)+s(\pi')=1$. As a consequence, μ is always antisymmetric and ν is always total. In the absence of simple ties, i.e. when $\mu=\nu$, one can easily check that this relation is transitive, and therefore a ranking, if and only if it contains no cycles. However this need not always be the case.

In order to deal with the general case, we shall use the method of **ranked pairs**. This method uses in a crucial way a ranking H defined on Π. Let us emphasize that this ranking is defined not on A but on Π. So, the elements of H are "pairs of pairs". In the following we shall refer to H as an **hyperranking**, and in order to express that the pair $\pi = ab$ precedes the pair $\omega = cd$ in H we shall use the notation $ab\,H\,cd$. The hyperranking H is connected with the paired-comparison scores $s(\pi)$ by the following fundamental assumption:

$$\pi\,H\,\omega \quad \text{whenever} \quad s(\pi) > s(\omega). \tag{HR}$$

So H ranks all pairs $\pi \in \Pi$ in such a way that the corresponding scores $s(\pi)$ form a non-increasing sequence. In that connection, we shall use the following notation:

π_k denotes the pair whose rank in H is k.

τ_k denotes the relation that consists of all π_l with $l \leq k$.

σ_k denotes the relation that consists of all pairs π for which $s(\pi) \geq s(\pi_k)$.

The following statements are an immediate consequence of the definitions: $s(\pi_k) \geq s(\pi_{k+1})$; $\tau_k \supset \tau_{k-1}$; $\sigma_k \supset \sigma_{k-1}$; $\sigma_k \supset \tau_k$; $\sigma_k = \tau_k$ if and only if $s(\pi_k) > s(\pi_{k+1})$.

Once the hyperranking H is defined, the ranked-pairs procedure determines the global ranking R entirely from H (without any further reference to the initial data). More specifically, R is obtained as the final stage of a sequence of relations R_k which are defined by induction from $R_0 = \emptyset$ according to the formula

$$R_k = \begin{cases} R_{k-1}, & \text{if } R_{k-1} \text{ already contains } \pi_k \text{ or } \pi'_k; \\ (R_{k-1} \cup \{\pi_k\})^*, & \text{otherwise.} \end{cases} \tag{RP}$$

Clearly, the relations R_k are transitive and satisfy $R_k \supset R_{k-1}$. Another immediate consequence of (RP) is the following: the only way to have $\pi_k \notin R_k$ is that $\pi'_k \in R_{k-1}$. On account of the inclusion $R_{k-1} \subset R_k$, this ensures that every R_k contains either π_k or π'_k. Since the list π_k $(k = 1, 2, \ldots)$ includes all pairs, it follows that a value of k will be reached for which R_k is total. Since then on, R_k will remain the same, and this final relation defines R. On the other hand, Lemma B.3.2 below shows that each of the relations R_k is antisymmetric. Summing up, the final relation R is transitive, total and antisymmetric, i.e. it is a ranking.

In order to emphasize the dependence of R on the hyperranking H, we shall write $R = \mathrm{Rp}(H)$, or $R = \mathrm{Rp}(A, H)$, and we shall refer to Rp as the **ranked-pairs operator**.

If all the scores $s(\pi)$ are different, then condition (HR) uniquely determines H. Otherwise, there are several possibilities for H which may lead to different global rankings R. The method that we have called **ranked pairs with natural tie-breaking** uses the tie-breaker ρ_{tb} to make a choice according to rule NTB of § 9.6. With the present notation, this rule can be formulated as follows:

$uv\,H\,xy$ if and only if one of the following alternatives holds:

$$\begin{array}{ll}
0. & s(uv) > s(xy), \\
1. & s(uv) = s(xy) \text{ and } u\ \rho_{\mathrm{tb}}\ v \text{ and } y\ \rho_{\mathrm{tb}}\ x, \\
2.11 & s(uv) = s(xy) \text{ and } u\ \rho_{\mathrm{tb}}\ v \text{ and } x\ \rho_{\mathrm{tb}}\ y \text{ and } u\ \rho_{\mathrm{tb}}\ x, \\
2.12 & s(uv) = s(xy) \text{ and } v\ \rho_{\mathrm{tb}}\ u \text{ and } y\ \rho_{\mathrm{tb}}\ x \text{ and } u\ \rho_{\mathrm{tb}}\ x, \\
2.21 & s(uv) = s(xy) \text{ and } u\ \rho_{\mathrm{tb}}\ v \text{ and } x\ \rho_{\mathrm{tb}}\ y \text{ and } u = x \text{ and } y\ \rho_{\mathrm{tb}}\ v, \\
2.22 & s(uv) = s(xy) \text{ and } v\ \rho_{\mathrm{tb}}\ u \text{ and } y\ \rho_{\mathrm{tb}}\ x \text{ and } u = x \text{ and } y\ \rho_{\mathrm{tb}}\ v.
\end{array} \tag{RPN}$$

Remark B.2.1 (see Cretney, 2001 [**16** : 22 Feb 2001]). *The cases 2.21 and 2.22 of rule* (rpn) *can be omitted: if several pairs with the same preferred item occupy consecutive places in the hyperranking H, then the result of the ranked-pairs procedure does not change when these pairs are rearranged in any different way.*

Proof. Obviously, any rearrangement of several consecutive pairs can be decomposed into a sequence of inversions between them. So it suffices to consider the case where the hyperranking H is replaced by another one $\widetilde{H}$ which differs from H only by one inversion between two pairs with the same preferred item. The claim that such an inversion does not alter the global result will be established in Corollary B.8.3. □

Equivalently, the rule (rpn) can be formulated also in the following way:

$$\begin{array}{ll} uv\,H\,xy \quad \text{if and only if} & \text{one of the following alternatives holds:} \\ & 0.\;\; s(uv) > s(xy), \\ & \text{A.}\;\; s(uv) = s(xy) \;\text{ and }\; s^{\text{tb}}(uv) > s^{\text{tb}}(xy), \end{array} \qquad (\text{RPN}')$$

where for every pair ab we define the tie-breaking score $s^{\text{tb}}(ab)$ by the formula

$$\begin{array}{l} s^{\text{tb}}(ab) \;=\; \operatorname{sgn}(q-p) \;-\; (p/N) \;+\; (q/N^2), \\ \text{with} \;\; p = r(a,\rho_{\text{tb}}), \;\; q = r(b,\rho_{\text{tb}}), \;\text{ and }\; \operatorname{sgn}(x) = \begin{cases} +1, & \text{if } x > 0, \\ -1, & \text{if } x < 0, \end{cases} \end{array}$$

where it should be recalled that N denotes the number of items, and $r(a,\rho_{\text{tb}})$ denotes the rank of a in ρ_{tb}.

B.3. Basic lemmas.

In this subsection we put together several lemmas which will play a fundamental role in the sequel. We start by collecting the most immediate consequences of the definition (rp):

Lemma B.3.1. *The following facts hold for every $k = 1, 2, \dots$:*

(a) *R_k is transitive.*
(b) *$R_{k-1} \subset R_k$.*
(c) *$\pi_k \notin R_k$ implies $\pi'_k \in R_{k-1}$.*
(d) *R_k contains either π_k or π'_k.*

Next, we deal with the antisymmetry of R_k and some of its consequences (which include the converse of B.3.1.c):

Lemma B.3.2. *R_k is antisymmetric, i.e. $\pi \in R_k$ implies $\pi' \notin R_k$.*

Proof. This property will be obtained by induction. Obviously, it holds for $k = 0$ and going from $k-1$ to k is immediate in the case $R_k = R_{k-1}$. So, the heart of the matter consists in going from $k-1$ to k under the assumption that $R_k \neq R_{k-1}$. In that case $R_k = (R_{k-1} \cup \{\pi_k\})^*$. So, in order to ensure that R_k is antisymmetric we must prove that $R_{k-1} \cup \{\pi_k\}$ contains no cycle. This will be proved by *reductio ad absurdum.*

In fact, because of the induction hypothesis that R_{k-1} is antisymmetric, such a cycle must contain π_k among its links. Let π_k be the pair ab. So we are saying that R_k contains a cycle of the form $x \overset{(1)}{\dots} ab\,[\dots ab] \overset{(2)}{\dots} x$, where we can assume that the indicated appearances of ab are respectively the first and last of them and the middle piece between square brackets could be missing. But this means that all of the links in $x \overset{(1)}{\dots} a$ and $b \overset{(2)}{\dots} x$ belong to R_{k-1}. Therefore, the path $b \overset{(2)}{\dots} x \overset{(1)}{\dots} a$ is contained in R_{k-1}. Since R_{k-1} is transitive, this means that $\pi'_k = ba \in R_{k-1}$. But according to (rp), this contradicts our assumption that $R_k \neq R_{k-1}$. □

Corollary B.3.3. *$\pi_k \in R_k$ implies $\pi'_k \notin R_{k-1}$.*

Proof. This is an immediate consequence of the antisymmetry of R_k together with the inclusion $R_{k-1} \subset R_k$. □

Corollary B.3.4. *$\pi_k \in R$ implies $\pi_k \in R_k$.*

Proof. Assume the contrary, namely $\pi_k \notin R_k$. By B.3.1.c, this implies that $\pi'_k \in R_{k-1} \subset R$. From here, the antisymmetry of R allows to derive that $\pi_k \notin R$, which contradicts the hypothesis. □

We shall often use also the following lemma: At every stage k, a pair ab is contained in R_k if and only if it is supported by a path $a_0 a_1 \ldots a_n$ (with $a_0 = a$ and $a_n = b$) where each link $a_i a_{i+1}$ is one of the pairs that have been considered and accepted up to that stage, i. e. $a_i a_{i+1}$ coincides with π_l for some $l \leq k$ and that π_l was accepted into $R_l \subset R$. More concisely:

Lemma B.3.5. *$R_k = (R \cap \tau_k)^*$.*

Proof. We shall proceed by induction with every step divided in two cases: (i) $\pi_k \notin R_k$; (ii) $\pi_k \in R_k$.

Case (i): $\pi_k \notin R_k$. By (rp) this implies that $R_k = R_{k-1}$. On the other hand, Corollary B.3.4 ensures that $\pi_k \notin R$ and therefore $R \cap \tau_k = R \cap \tau_{k-1}$.

Case (ii): $\pi_k \in R_k$. In this case, we can write the following chain of equalities:

$$R_k \;=\; (R_{k-1} \cup \{\pi_k\})^* \;=\; ((R \cap \tau_{k-1})^* \cup \{\pi_k\})^* \;=\; ((R \cap \tau_{k-1}) \cup \{\pi_k\})^* \;=\; (R \cap \tau_k)^*,$$

where we use successively: the rule (rp); the induction hypothesis; the general identity $(\alpha^* \cup \beta^*)^* = (\alpha \cup \beta)^*$; and the present hypothesis that $\pi_k \in R_k \subset R$. □

B.4. Minimum leading disagreement.

Given a system of paired-comparison scores $s(\pi)$, we define the **leading disagreement** of a relation ρ as the largest value of $s(\pi)$ when $\pi \notin \rho$. This number will be denoted by $\lambda(\rho)$. We are interested in minimizing the leading disagreement $\lambda(P)$ under the condition that P is a ranking. In that connection, we shall use the following notation:

λ^{mld} denotes the number $\min\{\lambda(P) \mid P \text{ is a ranking}\}$.

ρ^{mld} denotes the relation that consists of all π with $s(\pi) > \lambda^{\mathrm{mld}}$.

In [**9**: p. 92–100] λ^{mld} and ρ^{mld} are denoted respectively $\underline{\beta}$ and $R_{\underline{\beta}+1}$.

Lemma B.4.1. *A ranking P minimizes $\lambda(P)$ if and only if $P \supset \rho^{\mathrm{mld}}$.*

Proof. 1. If a ranking P minimizes $\lambda(P)$ and $\pi \in \rho^{\mathrm{mld}}$, then $\pi \in P$: Otherwise, the preceding definitions of $\lambda(P), \lambda^{\mathrm{mld}}, \rho^{\mathrm{mld}}$ would allow to derive that $s(\pi) \leq \lambda(P) = \lambda^{\mathrm{mld}} < s(\pi)$, which yields an *absurdum*.

2. If $P \supset \rho^{\mathrm{mld}}$ then $\lambda(P) \leq \lambda^{\mathrm{mld}}$: By definition $\lambda(P) = \max\{s(\pi) \mid \pi \notin P\}$. But under the present hypothesis, $\pi \notin P$ entails $\pi \notin \rho^{\mathrm{mld}}$, and, by the definition of ρ^{mld}, this implies that $s(\pi) \leq \lambda^{\mathrm{mld}}$. So the maximum of such $s(\pi)$ will also be less than or equal to λ^{mld}. □

Proposition B.4.2 (see [**13**: p. 199]). *Any ranking obtained by the ranked-pairs procedure minimizes the leading disagreement.*

Proof. Let R be a ranking obtained by the ranked-pairs procedure (rp). By the preceding lemma, it suffices to show that R includes ρ^{mld}. Now, by its definition, $\rho^{\mathrm{mld}} = \tau_k$ for some k.

On the other hand, we can easily see that ρ^{mld} contains no cycles. In fact, by the preceding lemma ρ^{mld} is contained in any ranking that minimizes the leading disagreement. Now, the fact that $\tau_k = \rho^{\mathrm{mld}}$ is acyclic implies that $R_l = \tau_l^*$ for every $l \leq k$. This is easily obtained by induction: From $R_{l-1} = \tau_{l-1}^*$, the acyclic character of $\tau_l \subset \tau_k$ implies that $\pi_l' \notin \tau_{l-1}^* = R_{l-1}$, and therefore, by (RP), $R_l = (R_{l-1} \cup \{\pi_l\})^* = (\tau_{l-1}^* \cup \{\pi_l\})^* = \tau_l^*$. Consequently, $\rho^{\mathrm{mld}} = \tau_k \subset R_k \subset R$. □

B.5. Immunity to majority complaints.

In connection with a given system of paired-comparison scores $s(\pi)$ we shall adopt also the following terminology: The **score of a path** $a_0a_1 \ldots a_n$, denoted $s(a_0a_1 \ldots a_n)$, is the minimum value of $s(a_ia_{i+1})$ for $0 \leq i < n$. For any relation ρ and any pair ab, the **indirect score of** ab **through** ρ, denoted $s^\sharp(ab, \rho)$, is the maximum value of $s(a_0a_1 \ldots a_n)$ when $a_0a_1 \ldots a_n$ is a path from a to b contained in ρ. A relation ρ is said to be **immune to majority complaints** when each $ab \in \rho$ satisfies $s^\sharp(ab, \rho) \geq s(ba)$, i. e. ab is supported by a path in ρ whose score is larger than or equal to that of the pair ba.

In [**14** : p. 172] the rankings immune to majority complaints are called *stacks*.

The following lemma shows that the condition of immunity to majority complaints entails Condorcet's principle.

Lemma B.5.1. *Assume that w is a Condorcet winner, i. e. $s(wb) > 1/2$ for every $b \in A\setminus\{w\}$. If a relation ρ is immune to majority complaints, then it cannot contain any pair of the form aw. In particular, a Condorcet winner is ranked first by any ranking immune to majority complaints.*

Proof. Assume the existence of some $a \in A \setminus \{w\}$ such that $aw \in \rho$. According to the definition of immunity to majority complaints, the pair aw is suppported in ρ by a path $a_0a_1 \ldots a_n$ with the property that $s(a_ia_{i+1}) \geq s(wa)$. Now, for $i = n-1$ this says that $s(a_{n-1}w) \geq s(wa)$, and since $s(a_{n-1}w) + s(wa_{n-1}) = 1$, this is equivalent to saying that $s(wa) + s(wa_{n-1}) \leq 1$. But this is incompatible with w being a Condorcet winner. □

Theorem B.5.2 (Zavist, Tideman, 1989 [**14** : p. 172]). *Any ranking obtained by the ranked-pairs procedure is immune to majority complaints.*

Proof. The result will be obtained by showing that all of the relations R_k have that property. This will be shown by induction. Consider $ab \in R_k$. If $ab \in R_{k-1}$ then the existence of a path as required by the condition of immunity to majority complaints follows from the induction hypothesis. If $ab \in R_k \setminus R_{k-1}$ then it can be obtained in the following way: According to Lemma B.3.5, ab is supported by a path $a_0a_1 \ldots a_n$ contained in $R \cap \tau_k$, which implies that $s(a_ia_{i+1}) \geq s(\pi_k)$. Comparing this inequality with our goal, namely $s(a_ia_{i+1}) \geq s(ba)$, it is clear that it would suffice to show that $s(ba) \leq s(\pi_k)$. Now, as it is shown in the following paragraph, we can show that $ba \notin \tau_k$, which implies that inequality.

So, we claim that $ab \in R_k \setminus R_{k-1}$ implies $ba \notin \tau_k$. In order to prove it we shall use *reductio ad absurdum.* So, let us suppose that $ba \in \tau_k$. By the definition of τ_k, this means that $ba = \pi_l$ for some $l \leq k$. Now, the hypothesis that $ab \notin R_{k-1}$ implies $ab \notin R_{l-1}$. On the other hand, by Lemma B.3.2, the hypothesis that $ab \in R_k$ implies that $ba \notin R_k$ and therefore $ba \notin R_{l-1}$. So, we have obtained that $ab, ba \notin R_{l-1}$, that is $\pi_l', \pi_l \notin R_{l-1}$. But in these conditions (RP) entails that $\pi_l \in R_l$, that is $ba \in R_l$, and therefore $ba \in R_k$, which contradicts one of the preceding. □

By the way, the preceding arguments lead immediately to the following remark:

Proposition B.5.3. *Let R be a ranking obtained by the ranked-pairs procedure. For any $ab \in R$, the indirect score of ab through R is given by $s(\pi_k)$ where k is the first integer for which $ab \in R_k$.* □

Theorem B.5.4 (see Zavist, Tideman, 1989 [**14**: p. 172]). *Assume that P is a ranking immune to majority complaints and that, in addition to* (HR), *the hyperranking H satisfies also one of the two following conditions:*

$$s(uv) = s(xy),\ uv \in P,\ yx \in P \quad \Longrightarrow \quad uv\, H\, xy. \tag{HS}$$

$$s(uv) = s(xy),\ ux \in P,\ yv \in P \quad \Longrightarrow \quad uv\, H\, xy. \tag{HT}$$

Then the global ranking obtained by the ranked-pairs procedure coincides with P.

Proof. In order to prove that the ranked-pairs procedure leads to P, it suffices to see that $R_k \subset P$ for every k. This property will be obtained by induction. As before, the non-trivial part consists in going from $k-1$ to k in the case $R_k \neq R_{k-1}$.

Let us assume that $R_{k-1} \subset P$ but $R_k \not\subset P$. According to (RP), this requires $\pi_k \notin P$. Now, since P is antisymmetric, $\pi'_k \in P$. On the other hand, the hypothesis that P is immune to majority complaints ensures that π'_k is supported by a path $a_0a_1 \dots a_n$ which is contained in P and satisfies the condition $s(a_ia_{i+1}) \geq s(\pi_k)$. By the definition of σ_k, the last inequality is saying that the path $a_0a_1 \dots a_n$ is contained also in σ_k. Now, in the presence of (HS) or (HT), we can see that this path is contained in the smaller set τ_{k-1}. In order to prove this statement we have to show that $a_ia_{i+1}\, H\, a_na_0$ even in the case $s(a_ia_{i+1}) = s(a_na_0)$. In the case of (HS) this follows from $a_ia_{i+1} \in P$ and $a_0a_n \in P$. In the case of (HT), it follows from $a_ia_n \in P$ and $a_0a_{i+1} \in P$.

So, we know that π'_k is supported by a path contained in $P \cap \tau_{k-1}$. But $P \cap \tau_{k-1} \subset R_{k-1}$: by B.3.1.d, $\pi \in \tau_{k-1}$ implies that either $\pi \in R_{k-1}$ or $\pi' \in R_{k-1}$; since $R_{k-1} \subset P$ and P is antisymmetric, when we add the information that $\pi \in P$ the only possibility is $\pi \in R_{k-1}$.

So we have obtained that π'_k is supported by a path contained in R_{k-1}. But then $\pi'_k \in R_{k-1}$, and by (RP) this implies that $R_k = R_{k-1}$, in contradiction with one of the starting assumptions. □

Corollary B.5.5. *The method of ranked pairs with natural tie-breaking has the following property: if the tie-breaker ranking ρ_{tb} is immune to majority complaints then the global ranking coincides with ρ_{tb}.*

Proof. It suffices to check that the hyperranking defined by the rule (RPN) always satisfies (HS) for $P = \rho_{\mathrm{tb}}$. □

Corollary B.5.6. *Any ranking immune to majority complaints can be obtained by the ranked-pairs procedure by suitably ordering the pairs with equal scores.* □

B.6. Consistency with respect to losers and winners.

In this subsection and the following ones we shall look at the effect of certain changes in the data. In particular, we shall consider situations where the set of items A is replaced by a certain subset $\widetilde{A}$. In that connection we shall use the following notation:

$\widetilde{\Pi}$ denotes $\Pi(\widetilde{A})$ as a subset of $\Pi = \Pi(A)$.

$\widetilde{\rho}_j, \widetilde{\rho}_{\mathrm{tb}}$ denote the restriction of $\rho_j, \rho_{\mathrm{tb}}$ to $\widetilde{A}$, i. e. $\widetilde{\rho}_j = \rho_j \cap \widetilde{\Pi}$ and $\widetilde{\rho}_{\mathrm{tb}} = \rho_{\mathrm{tb}} \cap \widetilde{\Pi}$.

$\widetilde{R}$ denotes the global ranking on $\widetilde{A}$ obtained from $\widetilde{\rho}_j, \widetilde{\rho}_{\rm tb}$.

A method is said to be **consistent with respect to losers and winners** when it has the following property: if $\widetilde{A}$ is a segment of R, then $\widetilde{R} = R \cap \widetilde{\Pi}$.

As it was remarked in § 9.3, the definition of immunity to majority complaints immediately implies that if a ranking R is immune to majority complaints and $\widetilde{A}$ is a segment of R, then $R \cap \widetilde{\Pi}$ is immune to majority complaints as a ranking on $\widetilde{A}$. In view of Theorem B.5.2 and Corollary B.5.6, this means that the method of ranked pairs is consistent with respect to losers and winners in the following sense: If $\mathrm{Rp}(A, H) = R$, where H is a particular hyperranking on Π compatible with the scores $s(\pi)$, and $\widetilde{A}$ is a segment of R, then there exists some hyperranking $\widetilde{H}$ on $\widetilde{\Pi}$ compatible with the scores, such that $\mathrm{Rp}(\widetilde{A}, \widetilde{H}) = R \cap \widetilde{\Pi}$. In the following we shall strengthen this result by showing that one can take as $\widetilde{H}$ the restriction of H to $\widetilde{\Pi}$, i. e. $\widetilde{H} = H \cap (\widetilde{\Pi} \times \widetilde{\Pi})$.

Theorem B.6.1. *The ranked-pairs procedure is consistent with respect to losers and winners in the following sense: If* $\mathrm{Rp}(A, H) = R$ *and* $\widetilde{A} \subset A$ *is a segment of* R*, then* $\mathrm{Rp}(\widetilde{A}, H \cap (\widetilde{\Pi} \times \widetilde{\Pi})) = R \cap \widetilde{\Pi}$.

Proof. We shall make use of the notations $\widetilde{H} = H \cap (\widetilde{\Pi} \times \widetilde{\Pi})$ and $\widetilde{R} = \mathrm{Rp}(\widetilde{A}, \widetilde{H})$. Analogously, the sequence of relations that lead to $\widetilde{R}$ will be denoted by $\widetilde{R}_k$, $\widetilde{\pi}_k$ will denote the element of $\widetilde{\Pi}$ whose rank in $\widetilde{H}$ is k, and $\widetilde{\tau}_k$ will denote the set of all $\widetilde{\pi}_l$ with $l \leq k$. In addition, $\varphi(k)$ will denote the rank of $\widetilde{\pi}_k$ in H, so that $\widetilde{\pi}_k = \pi_{\varphi(k)}$ and $\widetilde{\tau}_k = \tau_{\varphi(k)} \cap \widetilde{\Pi}$. Since $\widetilde{H}$ is a restriction of H, it is clear that $\varphi(k)$ increases with k, i. e. $\varphi(k-1) < \varphi(k)$. In this connection, one should keep in mind that $\varphi(k-1) < l < \varphi(k)$ implies $\pi_l \notin \widetilde{\Pi}$. In other words, $\tau_{\varphi(k)-1} \cap \widetilde{\Pi} = \tau_{\varphi(k-1)} \cap \widetilde{\Pi} = \widetilde{\tau}_{k-1}$.

The main idea of the proof will consist in showing that $\widetilde{\pi}_k \in \widetilde{R}$ if and only if $\widetilde{\pi}_k \in R$. This property will be obtained by induction. In order to see that it implies the equality $\widetilde{R} = R \cap \widetilde{\Pi}$, and for better organizing the proof, it will be convenient to consider the following statements:

$$\widetilde{\pi}_k \in R \ \Rightarrow\ \widetilde{\pi}_k \in \widetilde{R}, \tag{A_k}$$

$$\widetilde{\pi}_k \in \widetilde{R} \ \Rightarrow\ \widetilde{\pi}_k \in R, \tag{B_k}$$

$$R \cap \widetilde{\tau}_k \ \subset\ \widetilde{R} \cap \widetilde{\tau}_k, \tag{C_k}$$

$$\widetilde{R}_k \ =\ (\widetilde{R} \cap \widetilde{\tau}_k)^* \ \subset\ R \cap \widetilde{\Pi}. \tag{D_k}$$

According to the definitions, (C_k) is equivalent to saying that (A_l) holds for $l \leq k$, and (D_k) is easily checked to be equivalent to saying that (B_l) holds for $l \leq k$ (the equality at the left of (D_k) is ensured by Lemma B.3.5). On the other hand, when k grows large, the combination of (C_k) and (D_k) implies the desired equality $\widetilde{R} = R \cap \widetilde{\Pi}$ (since $\widetilde{\tau}_k$ ends up being equal to $\widetilde{\Pi}$). For the statements (C_k) and (D_k) we shall admit the possibility that $k = 0$, in which case $\widetilde{\tau}_0 = \widetilde{R}_0 = \emptyset$. Certainly, these equalities make both (C_0) and (D_0) trivially true. So, our aim will be fulfilled if we show that (A_k) follows from (D_{k-1}) while (B_k) follows from (C_{k-1}) (for every $k \geq 1$).

1. (D_{k-1}) implies (A_k): Assume that $\widetilde{\pi}_k \in R$. Since R is antisymmetric, this implies that $\widetilde{\pi}'_k \notin R$, and therefore (D_{k-1}) ensures that $\widetilde{\pi}'_k \notin \widetilde{R}_{k-1}$. Now, according to (RP), the latter implies that $\widetilde{\pi}_k \in \widetilde{R}_k \subset \widetilde{R}$.

2. (C_{k-1}) implies (B_k): Assume that $\widetilde{\pi}_k = \pi_{\varphi(k)} \notin R$. By B.3.1.c, it follows that $\pi'_{\varphi(k)} \in R_{\varphi(k)-1}$. According to Lemma B.3.5, this is equivalent to saying that $\pi'_{\varphi(k)} \in (R \cap \tau_{\varphi(k)-1})^*$, i. e. $\pi'_{\varphi(k)}$ is supported by a path contained in $R \cap \tau_{\varphi(k)-1}$. Now, since $\pi'_{\varphi(k)} = \widetilde{\pi}'_k \in \widetilde{\Pi}$ and

$\widetilde{A}$ is a segment of R, all of the intermediate items of this path must be also members of $\widetilde{A}$. So, we have $\pi'_{\varphi(k)} \in (R \cap \tau_{\varphi(k)-1} \cap \widetilde{\Pi})^*$. On the other hand, by the remark made at the end of the first paragraph of this proof, this is equivalent to saying that $\pi'_{\varphi(k)} \in (R \cap \widetilde{\tau}_{k-1})^*$. Finally, ($\mathrm{C}_{k-1}$) allows to derive that $\pi'_{\varphi(k)} \in (\widetilde{R} \cap \widetilde{\tau}_{k-1})^*$, i. e. $\widetilde{\pi}'_k \in \widetilde{R}_{k-1}$ (by Lemma B.3.5), which implies that $\widetilde{\pi}_k \notin \widetilde{R}$ (because of the inclusion $\widetilde{R}_{k-1} \subset \widetilde{R}$ and the antisymmetry of $\widetilde{R}$). □

Corollary B.6.2. *The method of ranked pairs with natural tie-breaking is consistent with respect to losers and winners.*

Proof. It suffices to check that for $u, v, x, y \in \widetilde{A} \subset A$ the rule (RPN) for determining whether uv precedes xy in the hyperranking does not depend on whether uv and xy are considered as members of $\widetilde{\Pi} = \Pi(\widetilde{A})$ or as members of $\Pi = \Pi(A)$. □

B.7. Consistency with respect to clones.

A subset of items $C \subset A$ will be called a **cluster** (of clones) for the data $\rho_j, \rho_{\mathrm{tb}}$ when C is a segment of both each ρ_j and ρ_{tb}. In that case, the members of C will be said to be **clones** of each other. A method is said to be **consistent with respect to clones** when it has the following properties: (a) Every cluster is a segment of the global ranking R; (b) If C is a cluster and $\widetilde{\rho}_j, \widetilde{\rho}_{\mathrm{tb}}$ are the contractions of $\rho_j, \rho_{\mathrm{tb}}$ along C, then the global ranking $\widetilde{R}$ obtained from $\widetilde{\rho}_j, \widetilde{\rho}_{\mathrm{tb}}$ coincides with the contraction of R along C.

As an immediate consequence of the definition, if C is a cluster, the scores $s(\pi)$ are bound to have the following property:

$$\begin{array}{l}\text{For any } c, d \in C \text{ and } a, b \in A \setminus C, \\ s(ac) = s(ad) \quad \text{and} \quad s(cb) = s(db).\end{array} \tag{SC}$$

More than the preceding condition on the scores $s(\pi)$, in the following we shall be especially interested in the case where the hyperranking H satisfies the following one:

$$\begin{array}{l}\text{For any } c, d \in C,\ a, b \in A \setminus C, \text{ and } x, y \in A, \\ ac\,H\,xb \Leftrightarrow ad\,H\,xb \quad \text{and} \quad cb\,H\,ay \Leftrightarrow db\,H\,ay.\end{array} \tag{HC}$$

Theorem B.7.1. *If H satisfies* (HC)*, then C is a segment of $R = \mathrm{Rp}(H)$.*

Proof. This result will be obtained by showing that C is a segment of each of the relations R_k, i. e. if $cx \in R_k$ and $xd \in R_k$ with $c, d \in C$, then $x \in C$. By virtue of Lemma B.3.5, it suffices to show the impossibility that $R \cap \tau_k$ contains a path of the form $cx_1 \dots x_n d$ with $c, d \in C$, $n \geq 1$ and $x_i \notin C$ for some $1 \leq i \leq n$. By considering a suitable segment of such a path, we can say even more: it will suffice to show the impossibility that $R \cap \tau_k$ contains a path of that form with $x_i \notin C$ for *all* $1 \leq i \leq n$. This impossibility will be shown by induction; the non-trivial part consists in going from $k-1$ to k in the case where the path $cx_1 \dots x_n d$ is contained in $R \cap \tau_k$ but not in $R \cap \tau_{k-1}$, i. e. when one of its links coincides with π_k. We shall distinguish three cases: (i) $\pi_k = cx_1$; (ii) $\pi_k = x_n d$; (iii) $\pi_k = x_i x_{i+1}$ for some $1 \leq i < n$. In each of them we shall derive that $\pi'_k \in R_{k-1}$. Since $R_{k-1} \subset R$ and R is antisymmetric, this implies that $\pi_k \notin R$, which contradicts the hypothesis that the path under consideration was contained in $R \cap \tau_k$.

Case (i): $\pi_k = cx_1$. This implies that $x_n d\,H\,cx_1$, and therefore, by (HC), $x_n c\,H\,cx_1$. So, $x_n c \in \tau_{k-1}$. Now we claim that $x_n c \in R_{k-1}$. In fact, the contrary would imply that

$cx_n \in R_{k-1}$ (by B.3.1.c), and therefore the path cx_nd would be contained in R_{k-1}, which is contrary to the induction hypothesis. So we obtain that R_{k-1} already contains the path $x_1 \ldots x_n c$, which forbids the acceptance of cx_1 into R_k.

Case (ii): $\pi_k = x_n d$. This case is analogous to the preceding one. In this case we get that R_{k-1} already contains the path $dx_1 \ldots x_n$, which forbids the acceptance of $x_n d$ into R_k.

Case (iii): $\pi_k = x_i x_{i+1}$ for some $1 \le i < n$. This implies both $x_n d \, H \, x_i x_{i+1}$ and $cx_1 \, H \, x_i x_{i+1}$. Now we have two subcases: (a) $x_n d \, H \, cx_1 \, H \, x_i x_{i+1}$; (b) $cx_1 \, H \, x_n d \, H \, x_i x_{i+1}$. Let us consider the subcase (a): Like in case (i), (HC) implies that $x_n c \, H \, cx_1$, from which one can derive that $x_n c \in R_{k-1}$. As a consequence, we see that R_{k-1} already contains the path $x_{i+1} \ldots x_n c x_1 \ldots x_i$, which forbids the acceptance of $x_i x_{i+1}$ into R_k. Analogously, in the subcase (b) we get that R_{k-1} already contains the path $x_{i+1} \ldots x_n d x_1 \ldots x_i$, which forbids the acceptance of $x_i x_{i+1}$ into R_k. □

In the following we shall consider contractions along C. So we shall consider a new set of items of the form $\widetilde{A} = (A \setminus C) \cup \{\widetilde{c}\}$ and the corresponding set of pairs $\widetilde{\Pi} = \Pi(\widetilde{A})$. More specifically, here we shall identify $\widetilde{c}$ with a particular member of C. In this case, $\widetilde{A}$ and $\widetilde{\Pi}$ can be considered as subsets of A and Π.

Theorem B.7.2. *If H satisfies* (HC)*, then, for any $\widetilde{c} \in C$, the ranking on $\widetilde{A}$ given by* $\mathrm{Rp}(\widetilde{A}, H \cap (\widetilde{\Pi} \times \widetilde{\Pi}))$ *coincides with* $\mathrm{Rp}(A, H) \cap \widetilde{\Pi}$.

Proof. One can follow almost exactly the proof of Theorem B.6.1. Like there, we write $R = \mathrm{Rp}(A, H)$. The only step that requires a proof specific to the present situation is the claim made in the last paragraph of that proof that $\pi'_{\varphi(k)} \in (R \cap \tau_{\varphi(k)-1})^*$ implies $\pi'_{\varphi(k)} \in (R \cap \tau_{\varphi(k)-1} \cap \widetilde{\Pi})^*$. So we must show that, if π'_l is supported by a path contained in $R \cap \tau_{l-1}$, then it is supported also by a path which, besides being contained in $R \cap \tau_{l-1}$, has the additional property that all of its intermediate items are also members of $\widetilde{A} = (A \setminus C) \cup \{\widetilde{c}\}$. We shall distinguish three cases: (i) $\pi_l = \widetilde{c}a$ for some $a \in A \setminus C$; (ii) $\pi_l = b\widetilde{c}$ for some $b \in A \setminus C$; (iii) $\pi_l = ba$ for some $a, b \in A \setminus C$. The proof will be based on the fact that C forms a segment in R, as ensured by Theorem B.7.1, and the hypothesis that H satisfies (HC).

Case (i): $\pi_l = \widetilde{c}a$ for some $a \in A \setminus C$. Let us assume that $\pi'_l = a\widetilde{c}$ is supported by a path contained in $R \cap \tau_{l-1}$. Since C forms a segment in R, this path must have the form $ax_1 \ldots x_n c_1 \ldots c_n \widetilde{c}$ with $x_i \in A \setminus C$ and $c_i \in C$. By hypothesis, all of the links of this path precede $\pi_l = \widetilde{c}a$ in H. In particular, $x_n c_1 \, H \, \widetilde{c}a$. Now, property (HC) allows to derive that $x_n \widetilde{c} \, H \, \widetilde{c}a$. On the other hand, $x_n \widetilde{c} \in R$ since it is supported by a segment of the path under consideration. Therefore, we have obtained that $\pi'_l = a\widetilde{c}$ is supported by a path contained in $R \cap \tau_{l-1} \cap \widetilde{\Pi}$, namely $ax_1 \ldots x_n \widetilde{c}$.

Case (ii): $\pi_l = b\widetilde{c}$ for some $b \in A \setminus C$. This case is analogous to the preceding one. In this case we get that $\pi'_l = \widetilde{c}b$ is supported by a path of the form $\widetilde{c}y_1 \ldots y_n b$ with $y_i \in A \setminus C$.

Case (iii): $\pi_l = ba$ for some $a, b \in A \setminus C$. Since C is known to form a segment in R, if $\pi'_l = ab$ is supported by a path contained in $R \cap \tau_{l-1}$, this path must have the form $ax_1 \ldots x_n c_1 \ldots c_n y_1 \ldots y_n b$ with $x_i, y_i \in A \setminus C$ and $c_i \in C$. On the other hand, we are assuming that $x_n c_1 \, H \, ba$ and $c_n y_1 \, H \, ba$. From these facts, property (HC) allows to derive that $x_n \widetilde{c} \, H \, ba$ and $\widetilde{c} y_1 \, H \, ba$. On the other hand, we must have $x_n \widetilde{c} \in R$ and $\widetilde{c} y_1 \in R$, because the contrary would contradict the fact that C forms a segment in R (for instance, $x_n \widetilde{c} \notin R$ implies $\widetilde{c} x_n \in R$, while the path under consideration includes $x_n c_1 \in R$ and assumes $x_n \in A \setminus C$). Therefore, we have obtained that $\pi'_l = ab$ is supported by the following path contained in $R \cap \tau_{l-1} \cap \widetilde{\Pi}$, namely $ax_1 \ldots x_n \widetilde{c} y_1 \ldots y_n b$. □

If the only cases of equal scores are those that appear in condition (SC), then any hyperranking that satisfies (HR) is ensured to satisfy also (HC). When there are other cases of equal scores besides those implied by (SC), then (HR) is not enough to guarantee (HC). However, this condition and its consequences are always ensured if we choose H according to the rule (RPN):

Corollary B.7.3. *The method of ranked pairs with natural tie-breaking is consistent with respect to clones.*

Proof. It suffices to check that rule (RPN) causes H to satisfy (HC) whenever C is a cluster for the data $\rho_j, \rho_{\rm tb}$. This is easily checked as a consequence of (SC) and the fact that C is a segment of $\rho_{\rm tb}$. In particular, the latter ensures that, for any $c, d \in C$ and any $a, b \in A \setminus C$, $a\,\rho_{\rm tb}\,c$ if and only if $a\,\rho_{\rm tb}\,d$, $c\,\rho_{\rm tb}\,a$ if and only if $d\,\rho_{\rm tb}\,a$, and $c\,\rho_{\rm tb}\,b$ if and only if $d\,\rho_{\rm tb}\,b$. By (RPN), these equivalences guarantee that $ac\,H\,xb$ if and only if $ad\,H\,xb$, as required in the first part of (HC). The second part of (HC) is obtained in an analogous way. □

Remark B.7.4. The preceding corollary depends in a crucial way on the fact that C is a segment of $\rho_{\rm tb}$. In that connection, the tie-breaking rule proposed by Zavist and Tideman in 1989 [**14**] has the remarkable property of being consistent with respect to clones even when the definition of a cluster is weakened by requiring C to be a segment of every ρ_j but not necessarily of $\rho_{\rm tb}$. In exchange, however, the rule of Zavist and Tideman lacks the following property of rule (RPN) (Corollary B.5.5): When $\rho_{\rm tb}$ is immune to majority complaints, the global ranking coincides with $\rho_{\rm tb}$.

Bibliography

[**1**] Xavier Mora, 2001.
The Skating system (2nd edition).
Available in: `http://puffinet.com/escrutini/skating2en.pdf`.

[**2**] Xavier Mora, Alistair Braden, 2002.
Improving the Skating system - I.
Available in: `http://puffinet.com/escrutini/iss1en.pdf`.
Published in 8 parts in *Dance News*, 1769 – 1776 (May 23rd – July 11th, 2002).

[**3**] H. Peyton Young, 1988.
Condorcet's theory of voting.
American Political Science Review, 82: 1231–1244.

[**4**] Jonathan Levin, Barry Nalebuff, 1995.
An introduction to vote-counting schemes.
Journal of Economic Perspectives, 9: 3–26.

[**5**] Iain McLean, Fiona Hewitt (eds.), 1994.
Condorcet. Foundations of Social Choice and Political Theory.
Edward Elgar.

[**6**] Iain McLean, Arnold B. Urken (eds.), 1995.
Classics of Social Choice.
The University of Michigan Press, Ann Arbor.

[**7**] Duncan Black, 1998.
The Theory of Committees and Elections (revised 2nd edition).
Kluwer Academic Publishers.

[**8**] M. Drton, G. Hägele, D. Haneberg, F. Pukelsheim, W. Reif, 2001–2002.
The Augsburg Web Edition of Llull's Electoral Writings.
Available in: `http://www.math.uni-augsburg.de/stochastik/llull/`.

[**9**] Kenneth J. Arrow, Hervé Raynaud, 1986.
Social Choice and Multicriterion Decision-Making.
MIT Press.

[**10**] Douglas R. Woodall, 1996–1997.
[**a**] *Monotonicity and single-seat election rules.* *Voting Matters*, 6: 9–14 (1996).
Available in: `http://www.mcdougall.org.uk/vm/main.htm`.
[**b**] *Monotonicity of single-seat preferential election rules.* *Discrete Applied Mathematics*, 77: 81–98 (1997).

[**11**] Hannu Nurmi, 1999.
Voting Paradoxes and How to Avoid Them.
Springer Verlag.

[**12**] Donald G. Saari, 2001.
[**a**] *Decisions and Elections · Explaining the Unexpected.* Cambridge University Press.
[**b**] *Chaotic Elections! · A Mathematician Looks at Voting.* American Mathematical Society.

[**13**] T. Nicolaus Tideman, 1987.
Independence of clones as a criterion for voting rules.
Social Choice and Welfare, 4: 185–206.

[**14**] Thomas M. Zavist, T. Nicolaus Tideman, 1989.
Complete independence of clones in the ranked pairs rule.
Social Choice and Welfare, 6: 167–173.

[**15**] Election Methods Mailing List, since 1996.
Available at: `http://electorama.com/em/`.

[**16**] Ranked Pairs Mailing List, since 2000.
Archived at: `http://www.topica.com/lists/RankedPairs/`.

[**17**] Blake Cretney, 2000.
Condorcet.org.
Web site located at: `http://condorcet.org/`.

[**18**] Markus Schulze, 2003.
[**a**] *A new monotonic and clone-independent single-winner election method.*
Voting Matters, 17: 9–19.
Available in: `http://www.mcdougall.org.uk/vm/main.htm`.
[**b**] *A new monotonic, clone-independent, reversal symmetric, and Condorcet-consistent single-winner election method.* Manuscript submitted for publication.

[**19**] Stephen Eppley, 2003–2004.
The Maximize Affirmed Majorities voting procedure (MAM).
Web site located at: `http://www.alumni.caltech.edu/~seppley/`.

[**20**] Andrew Myers, 2004.
Condorcet Internet Voting Service.
Web site located at: `http://www5.cs.cornell.edu/~andru/civs/`.

[**21**] Samuel Kotz, Norman L. Johnson (eds.), 1982–1988.
Encyclopedia of Statistical Sciences.
John Wiley & Sons.

[**22**] Michael A. Fligner, Joseph S. Verducci (eds.), 1992.
Probability Models and Statistical Analyses for Ranking Data.
Springer-Verlag.

[**23**] Lawrence J. Hubert, Phipps Arabie, Jacqueline Meulman, 2001.
Combinatorial Data Analysis: Optimization by Dynamic Programming.
Society for Industrial and Applied Mathematics.

[**24**] Gilbert W. Basset Jr, Joseph Persky, 1994.
Rating Skating.
Journal of the American Statistical Association, 89: 1075–1079.

[**25**] Johan Bring, Kenneth Carling, 1994.
A paradox in the rating of figure skaters.
Chance, 7 (4): 34–37.

[**26**] Edmund L. Russell III, 1994–1997.
[**a**] *Figure skating: What is a 5.8 really?* *Proceedings of the Section on Statistics in Sports, American Statistical Association*, 1994.
[**b**] *Scoring in figure skating: What does it measure?* Ibidem, 1995.
[**c**] *Amateur figure skating: Is the ranking system out of date?* Ibidem, 1995.
[**d**] *Choosing a ranking system for figure skating.* *STATS: The Magazine for Students of Statistics*, 20: 13–17 (1997).

[**27**] *International Skating Union, ISU*, 1998.
New ISU Figure Skating Results System.
ISU Communication No. 997.
Available in: `http://ww2.isu.org/news/997.html`.

[**28**] George S. Rossano, 1996–2004.
Ice Skating International: Online.
Web site located at: `http://www.iceskatingintnl.com/index.htm`.

[**29**] Sandra Loosemore, 1997–2004.
The Figure Skating Page.
Web site located at: `http://www.frogsonice.com/skateweb/articles.shtml`.

[**30**] Michel Truchon, 1998–2002.
[**a**] *Figure Skating and the Theory of Social Choice.*
Cahier de Recherche 9814, Dépt. d'Économique, Univ. de Laval, Québec, Canada.
Available in: `http://www.ecn.ulaval.ca/w3/recherche/cahiers/1998/9814.pdf`.
[**b**] *Choix social et comités de sélection: le cas du patinage artistique.*
CIRANO Burgundy Reports, 2002RB-02.
Available in: `http://www.cirano.qc.ca/pdf/publication/2002RB-02.pdf`.

[**31**] *International Skating Union, ISU*, 2003.
New Judging System · Figure Skating / Ice Dancing.
ISU Communication No. 1207.
Available in: `http://www.isu.org/`.

[**32**] Edmund L. Russell III, 2003.
Winning and losing in figure skating.
Available in: `http://www.worldskating.org/news/russell.htm`.

[**33**] Sonia Bianchetti Garbato, 2004.
Cracked Ice : Figure Skating's Inner World.
Libreria dello Sport.
Available through: `http://www.soniabianchetti.com/`.

[**34**] Xavier Mora, 2004.
MiniSkate, version 5.
To appear in: `http://puffinet.com/escrutini/miniskate/en/index.html`.

Xavier Mora
Departament de Matemàtiques
Universitat Autònoma de Barcelona
08193 Bellaterra, Spain
`xmora@mat.uab.cat`

B.8. Monotonicity.

Assume that some of the rankings ρ_j and $\rho_{\rm tb}$ are replaced by new ones $\widetilde\rho_j$ and $\widetilde\rho_{\rm tb}$ so that a particular item a is promoted to a better rank without affecting the relative ordering of the other items, i.e. $\widetilde\rho_j$ and $\widetilde\rho_{\rm tb}$ coincide respectively with ρ_j and $\rho_{\rm tb}$ on $A \setminus \{a\}$ but $r(a,\widetilde\rho_j) \le r(a,\rho_j)$ and $r(a,\widetilde\rho_{\rm tb}) \le r(a,\rho_{\rm tb})$. We shall say that a method is **monotonous** when the resulting global ranking $\widetilde R$ compares with R in the following way: $\mathsf{P}(a,\widetilde R) \subset \mathsf{P}(a,R)$, or equivalently $\mathsf{S}(a,\widetilde R) \supset \mathsf{S}(a,R)$; here $\mathsf{P}(a,\rho)$ and $\mathsf{S}(a,\rho)$ denote respectively the sets of predecessors and successors of an item a in a relation ρ, as introduced in § B.1. In particular, the preceding inclusions imply that $r(a,\widetilde R) \le r(a,R)$.

In order to establish the monotonicity of the method of ranked pairs, we shall begin by analyzing the effect of changing the hyperranking in the following way:

$$\widetilde H \text{ differs from } H \text{ by only one inversion, namely } \; uv\,\widetilde H\,xy \; \text{ but } \; xy\,H\,uv. \tag{E}$$

Like in the preceding sections, we shall continue using a tilde to distinguish between homologous objects corresponding to H and $\widetilde H$. With such a notation, the case described by (E) implies the existence of an integer h such that

$$\begin{aligned} &\pi_h = xy, \qquad \widetilde\pi_h = uv,\\ &\pi_{h+1} = uv, \quad \widetilde\pi_{h+1} = xy,\\ &\pi_k = \widetilde\pi_k \quad \text{for every} \quad k < h \quad \text{and every} \quad k > h+1. \end{aligned}$$

Lemma B.8.1. *Consider the situation* (E) *and assume that $\widetilde R$ differs from R. Then R includes xy but not uv whereas $\widetilde R$ includes uv but not xy.*

Proof. The conclusions concerning R and $\widetilde R$ are symmetric to each other. So it suffices to prove one of them, say that R includes xy but not uv. This will be done by checking that the contrary implies $\widetilde R = R$. Now, the contrary of R including xy but not uv can be divided in three cases: (i) R includes both uv and xy; (ii) R includes uv but not xy; and (iii) R includes neither uv nor xy. We shall see that each of these hypotheses allows to derive that $\widetilde R = R$. Since $\pi_k = \widetilde\pi_k$ for every $k > h+1$, it suffices to show that $\widetilde R_{h+1} = R_{h+1}$. On the other hand, since $\pi_k = \widetilde\pi_k$ for every $k < h$, it is clear that we have at least $\widetilde R_{h-1} = R_{h-1}$.

Case (i): R includes both uv and xy. Since $uv = \pi_{h+1}$, the hypothesis that $uv \in R$ implies that $uv \in R_{h+1}$ (Corollary B.3.4), and this implies that $vu \notin R_h$ (by B.3.1.c). Now, since $R_{h-1} \subset R_h$, it follows that $vu \notin R_{h-1} = \widetilde R_{h-1}$, and therefore $\widetilde\pi'_h = uv$ is accepted into $\widetilde R_h \subset \widetilde R$. Now we claim that $\widetilde\pi_{h+1} = xy$ is also accepted into $\widetilde R_{h+1} \subset \widetilde R$. In fact, the contrary entails that $yx \in \widetilde R_h = (\widetilde R_{h-1} \cup \{uv\})^* = (R_{h-1} \cup \{uv\})^* \subset (R_h \cup \{uv\})^* = R_{h+1} \subset R$, which contradicts the assumption that $xy \in R$ since R is antisymmetric.

Case (ii): R includes uv but not xy. As above, we get $vu \notin R_{h-1} = \widetilde R_{h-1}$, which causes uv to be accepted into $\widetilde R_h$. On the other hand, by B.3.1.c $\pi_h = xy \notin R$ implies that $yx \in R_{h-1} = \widetilde R_{h-1} \subset \widetilde R_h$, which prevents from accepting xy into $\widetilde R_{h+1}$. Therefore, $\widetilde R_{h+1} = R_{h+1} = (R_{h-1} \cup \{uv\})^*$.

Case (iii): R includes neither uv nor xy. As in case (ii), we get $yx \in R_{h-1} = \widetilde R_{h-1} \subset \widetilde R_h$, which prevents from accepting xy into $\widetilde R_{h+1}$. On the other hand, $\pi_{h+1} = uv \notin R$ implies that $vu \in R_h$ (by B.3.1.c). But $R_h = R_{h-1}$ because $\pi_h = xy \notin R$. So we have that $vu \in R_{h-1} = \widetilde R_{h-1}$, which prevents from accepting uv into $\widetilde R_h$. Therefore, $\widetilde R_{h+1} = R_{h+1} = R_{h-1}$. □

Corollary B.8.2. *Consider the situation* (E) *and assume that either u is the winner in R or x is the loser in $\widetilde R$. In that case, $\widetilde R = R$.*

Proof. The present hypotheses preclude one of the conclusions of Lemma B.8.1, namely that R includes xy but not uv. Therefore, the hypothesis of that lemma is also precluded. □

Corollary B.8.3. *Consider the situation* (E) *with either* $u = x$ *or* $v = y$. *In that case,* $\widetilde{R} = R$.

Proof. The cases $u = x$ and $v = y$ are analogous to each other. So it suffices to consider one of them, say $u = x$. We shall proceed by *reductio ad absurdum*. So, let us assume that $\widetilde{R} \neq R$. According to the preceding lemma, in that case we have $xy \in R\backslash\widetilde{R}$ and $xv \in \widetilde{R}\backslash R$. Now, $\pi_{h+1} = xv \notin R$ implies that $vx \in R_h$ (by B.3.1.c). On the other hand, $\pi_h = xy \in R$ implies that $R_h = (R_{h-1} \cup \{xy\})^*$. By combining these two facts, we get that vx is supported by a path contained in $R_{h-1} \cup \{xy\}$. However, this path cannot contain xy, because then it would contain a cycle $xy \ldots x$, which violates the antisymmetry of R_h. So, vx is supported by a path entirely contained in R_{h-1}. Therefore, $vx \in R_{h-1} = \widetilde{R}_{h-1} \subset \widetilde{R}$. But we already obtained that $xv \in \widetilde{R}$. So, we have arrived at a contradiction with the antisymmetry of $\widetilde{R}$. □

The one-inversion situation (E) allows for the final rankings R and $\widetilde{R}$ being quite different from each other. For example, let $A = \{x, y, u, v, a, b\}$, let the hyperranking H start in the following way: $vx\,H\,yu\,H\,xy\,H\,uv\,H\,av\,H\,ua\,H\,ub\,H\,bv\,H\,ab$, and let $\widetilde{H}$ differ from H only in having $uv\,\widetilde{H}\,xy$ instead of $xy\,H\,uv$; it is easily checked that in this case R and $\widetilde{R}$ are respectively the rankings $a\,v\,x\,y\,u\,b$ and $y\,u\,a\,b\,v\,x$. In spite of such differences, one can still ensure the following general result:

Theorem B.8.4. *In the situation* (E) *the rankings* R *and* $\widetilde{R}$ *compare in the following way:*

$$\mathsf{P}(u, \widetilde{R}) \subset \mathsf{P}(u, R), \qquad \mathsf{P}(v, \widetilde{R}) \supset \mathsf{P}(v, R), \qquad \mathsf{P}(x, \widetilde{R}) \supset \mathsf{P}(x, R), \qquad \mathsf{P}(y, \widetilde{R}) \subset \mathsf{P}(y, R).$$

Proof. This is obviously true when $\widetilde{R} = R$, so from now on we assume that $\widetilde{R} \neq R$. By Lemma B.8.1, the only possible way for $\widetilde{R}$ to be different from R is that

$$\begin{array}{rl} R & \text{includes } \ \pi_h = xy \ \text{ but not } \ \pi_{h+1} = uv, \\ \text{whereas } \widetilde{R} & \text{includes } \ \widetilde{\pi}_h = uv \ \text{ but not } \ \widetilde{\pi}_{h+1} = xy. \end{array}$$

By Corollary B.3.4, in this statement one can replace R and $\widetilde{R}$ respectively by $R_h = R_{h+1}$ and $\widetilde{R}_h = \widetilde{R}_{h+1}$. Furthermore, by B.3.1.c and B.3.3, the fact that $uv \in \widetilde{R}_h \setminus R_{h+1}$ implies that $vu \in R_h \setminus \widetilde{R}_{h-1} = R_h \setminus R_{h-1}$; analogously, $yx \in \widetilde{R}_h \setminus R_{h-1} = \widetilde{R}_h \setminus \widetilde{R}_{h-1}$. Therefore, R_h and $\widetilde{R}_h$ contain respectively the following preferences:

$$\begin{array}{ccc} x & \longrightarrow & y \\ \uparrow & R_h & \downarrow \\ v & & u \end{array} \qquad\qquad \begin{array}{ccc} x & & y \\ \uparrow & \widetilde{R}_h & \downarrow \\ v & \longleftarrow & u \end{array}$$

where an horizontal arrow from a to b indicates that a is preferred to b, and a vertical arrow from a to b indicates that either a is preferred to b or $a = b$. If there are no other items than x, y, u, v, the theorem follows immediately from these facts.

Let us now go for the general case.

In the following we shall concentrate upon the statement about u. The others are proved in a similar way. The desired property, namely $\mathsf{P}(u, \widetilde{R}) \subset \mathsf{P}(u, R)$, will be obtained as the final stage of the following ones, which will be shown to hold for any k:

$$\mathsf{P}(u, \widetilde{R}_k) \subset \mathsf{P}(u, R_k), \qquad (\mathrm{F}_k)$$

$$\mathsf{S}(u, \widetilde{R}_k) \supset \mathsf{S}(u, R_k). \qquad (\mathrm{G}_k)$$

Notice that these two statements are not equivalent to each other until k is large enough so that both R_k and $\widetilde{R}_k$ have reached their respective final values R and $\widetilde{R}$ (then, and only then, $\mathsf{S}(a, R)$ is exactly the complement of $\mathsf{P}(a, R)$ in $A \setminus \{a\}$). Properties (F_k) and (G_k) will be obtained by induction. Clearly, they are true for $k < h$ since then $\widetilde{R}_k = R_k$.

They are also true for $k = h, h+1$: Clearly, the pair $\widetilde{\pi}_h = uv$ cannot appear in a path from z to u in $\widetilde{R}_h$. Therefore, $\mathsf{P}(u, \widetilde{R}_h) = \mathsf{P}(u, \widetilde{R}_{h-1}) = \mathsf{P}(u, R_{h-1}) \subset \mathsf{P}(u, R_h)$. Similarly, the pair $\pi_h = xy$ cannot appear in a path from u to z in R_h (since $xu \in R_h$. Therefore, $\mathsf{S}(u, R_h) = \mathsf{S}(u, R_{h-1}) = \mathsf{S}(u, \widetilde{R}_{h-1}) \subset \mathsf{S}(u, \widetilde{R}_h)$.

Induction, using that $\widetilde{\pi}_k = \pi_k$ ... (**no funciona!**)

Versió anterior:

Furthermore, by B.3.1.c and B.3.3, the fact that $uv \in \widetilde{R}_h \backslash R_{h+1}$ implies that $vu \in R_h \backslash \widetilde{R}_{h-1} = R_h \setminus R_{h-1}$; analogously, $yx \in \widetilde{R}_h \setminus R_{h-1} = \widetilde{R}_h \setminus \widetilde{R}_{h-1}$. Therefore, R_h and $\widetilde{R}_h$ contain respectively the following preferences:

$$\begin{array}{ccc} x & \longrightarrow & y \\ \Big\uparrow & R_h & \Big\downarrow \\ v & & u \end{array} \qquad\qquad \begin{array}{ccc} x & & y \\ \Big\uparrow & \widetilde{R}_h & \Big\downarrow \\ v & \longleftarrow & u \end{array}$$

where an horizontal arrow from a to b indicates that a is preferred to b, and a vertical arrow from a to b indicates that either a is preferred to b or $a = b$. If there are no other items than x, y, u, v, the theorem follows immediately from these facts.

In order to settle the general case, it will suffice to establish the following property:

$$ab \in R \setminus \widetilde{R} \;\Leftrightarrow\; ba \in \widetilde{R} \setminus R \;\;\Rightarrow\;\; au \in R,\; vb \in R,\; bx \in \widetilde{R},\; ya \in \widetilde{R}. \tag{F}$$

In fact, one can easily see that the theorem follows from this property by taking ab in each of the forms uz, zv, zx, yz. For instance, for $ab = uz$ one of the conclusions of (F) reads $uu \in R$. But this is impossible: since none of the π_k is equal to uu, the only way that uu could become included in R is by transitive closure; but this would amount to say that R contains a cycle of the form utu for some $t \in A$, which is not possible because of Lemma B.3.2. Therefore, one cannot have $uz \in R \setminus \widetilde{R}$. In other words, $uz \in R$ implies $uz \in \widetilde{R}$. Since z is arbitrary, this ensures that $r(u, \widetilde{R}) \leq r(u, R)$.

Property (F) will be obtained as the final stage of the following one, which will be shown to hold for any k:

$$ab \in R_k \cap \widetilde{R}'_k, \;\Leftrightarrow\; ba \in \widetilde{R}_k \cap R'_k \;\;\Rightarrow\;\; au \in R,\; vb \in R,\; bx \in \widetilde{R},\; ya \in \widetilde{R}. \tag{F$_k$}$$

Here we use the notation introduced in § B.1 according to which $ab \in \rho'$ is equivalent to say that $ba \in \rho$. Since R and $\widetilde{R}$ are total and antisymmetric, one has the equalities $R \setminus \widetilde{R} = R \cap \widetilde{R}'$ and $\widetilde{R} \setminus R = \widetilde{R} \cap R'$. Therefore, when k grows large enough (F$_k$) turns into (F).

Property (F$_k$) will be obtained by induction. In order to better organize the proof, it will be convenient to consider the following auxiliary statements:

$$\pi_k = ab \in R \setminus \widetilde{R} \;\;\Rightarrow\;\; au \in R,\; vb \in R,\; bx \in \widetilde{R},\; ya \in \widetilde{R}. \tag{G$_k$}$$

$$\widetilde{\pi}_k = ba \in \widetilde{R} \setminus R \;\;\Rightarrow\;\; au \in R,\; vb \in R,\; bx \in \widetilde{R},\; ya \in \widetilde{R}. \tag{H$_k$}$$

The properties (F$_k$), (G$_k$) and (H$_k$) are trivially true for $k < h$ because in this case the sets $R_k \cap \widetilde{R}'_k$ and $\widetilde{R}_k \cap R'_k$, are empty and the conditions $\pi_k \in R \setminus \widetilde{R}$ and $\widetilde{\pi}_k \in \widetilde{R} \setminus R$ are never satisfied. Now, for $k = h$ (where $\pi_h = xy$ and $\widetilde{\pi}_h = uv$) properties (G$_h$) and (H$_h$) reduce to what has been established in the first paragraph of this proof. Furthermore, for $k = h+1$ (where $\pi_{h+1} = uv$ and $\widetilde{\pi}_{h+1} = xy$) properties (G$_{h+1}$) and (H$_{h+1}$) are again trivially true because their hypotheses are not satisfied. In the following we shall see that: (1) (F$_k$) holds as soon as (G$_l$) and (H$_l$) hold for every $l \leq k$; (2) (G$_k$) follows from (F$_{k-1}$) whenever $k > h+1$; and (3) (H$_k$) follows from (F$_{k-1}$) whenever $k > h+1$. Altogether, this establishes (F$_k$) for any k, which eventually gives (F) and therefore proves the theorem.

1. (F$_k$) holds as soon as (G$_l$) and (H$_l$) hold for every $l \leq k$: By definition, the condition $ab \in \widetilde{R}'_k$ is equivalent to $ba \in \widetilde{R}_k$, and similarly, $ba \in R'_k$ is equivalent to $ab \in R_k$. As a result,

$ab \in R_k \cap \widetilde{R}'_k$ is certainly equivalent to $ba \in \widetilde{R}_k \cap R'_k$, as stated in ($\mathrm{F}_k$). We want to show that these conditions imply the right-hand side of (F_k). We begin by noticing that since $\widetilde{R}$ is antisymmetric, $ba \in \widetilde{R}_k \subset \widetilde{R}$ implies that $ab \notin \widetilde{R}$. Similarly, we obtain also that $ba \notin R$. Now, according to Lemma B.3.5, $ab \in R_k$ ensures that ab is supported by a path contained in $R \cap \tau_k$. Let $a_0a_1 \dots a_n$ be such a path. Since $ab \notin \widetilde{R}$, at least one of its links a_pa_{p+1} is not contained in $\widetilde{R}$. By construction, we have either $a = a_p$ or $aa_p \in R$, and similarly, either $a_{p+1} = b$ or $a_{p+1}b \in R$. But we know that $a_pa_{p+1} = \pi_l$ for some $l \leq k$. Therefore, (G_l) ensures that $a_pu \in R$ and $va_{p+1} \in R$, from which the transitive property allows to conclude that $au \in R$ and $vb \in R$. A similar argument based upon (H_l) for $l \leq k$ allows to derive the properties $bx \in \widetilde{R}$ and $ya \in \widetilde{R}$ from the information that $ba \in \widetilde{R}_k \setminus R$ (which follows from the left-hand side of (F_k)).

2. (G_k) follows from (F_{k-1}) whenever $k > h+1$: Assume the hypothesis of (G_k), namely $\pi_k = ab \in R \setminus \widetilde{R}$. Two cases: (i) $\pi_k \in R_{k-1}$; (ii) $\pi_k \in R_k \setminus R_{k-1}$. Case (i): $\pi_k \in R_{k-1}$. This certainly implies that $\pi'_k \in R'_{k-1}$. On the other hand, by B.3.1.c the hypothesis that $\pi_k \notin \widetilde{R}$ implies that $\pi'_k \in \widetilde{R}_{k-1}$ (here we are using the fact that $\widetilde{\pi}_k$ coincides with π_k for every $k > h+1$). So we have $\pi'_k = ba \in \widetilde{R}_{k-1} \cap R'_{k-1}$, from which the desired conclusion follows as an application of (F_{k-1}). Case (ii): **pendent!**.

3. (H_k) follows from (F_{k-1}) and (H_{k-1}) whenever $k > h+1$: This implication is obtained by an argument entirely analogous to the preceding one. □

A partir del teorema B.8.4:

Corollary B.8.5. *Assume that H and $\widetilde{H}$ are related in the following way for some $a \in A$:*

$$\begin{aligned} &\text{(a)}\quad uv\,\widetilde{H}\,xy \text{ if and only if } uv\,H\,xy \text{ whenever } u,v,x,y \in A \setminus \{a\};\\ &\text{(b)}\quad r(av,\widetilde{H}) \leq r(av,H), \text{ for any } v \neq a;\\ &\text{(c)}\quad r(xa,\widetilde{H}) \geq r(xa,H), \text{ for any } x \neq a. \end{aligned} \tag{HM}$$

Then $r(a,\widetilde{R}) \leq r(a,R)$.

Proof. The result follows by a repeated application of Theorem B.8.4 when H is gradually transformed into $\widetilde{H}$ by a sequence of inversions where the ranks $r(av)$ and $r(xa)$ are respectively decreased and increased one unit at a time. □

In the following we consider the dependence of the hyperranking H on the individual preferences ρ_j and the tie-breaker ranking ρ_{tb}. The dependence on the individual preferences ρ_j is assumed to be through the paired-comparison scores $s(\pi)$. In order to express this dependence we shall write $H = \mathrm{Hr}(s,\rho_{\mathrm{tb}})$.

Corollary B.8.6. *Assume that $H = \mathrm{Hr}(s,\rho_{\mathrm{tb}})$ satisfies condition* (HR) *for any paired-comparison scores $s(\pi)$ and any tie-breaker ranking ρ_{tb}. Assume also that $H = \mathrm{Hr}(s,\rho_{\mathrm{tb}})$ and $\widetilde{H} = \mathrm{Hr}(s,\widetilde{\rho}_{\mathrm{tb}})$ satisfy* (HM) *whenever the paired-comparison scores $s(\pi)$ remain fixed and $\widetilde{\rho}_{\mathrm{tb}}$ differs from ρ_{tb} by only one inversion of the form $a\,\widetilde{\rho}_{\mathrm{tb}}\,b$ but $b\,\rho_{\mathrm{tb}}\,a$. In such conditions the resulting method of ranked pairs is monotonous.*

Proof. Assume that $\widetilde{\rho}_j$ and $\widetilde{\rho}_{\mathrm{tb}}$ coincide respectively with ρ_j and ρ_{tb} in the way that they compare the elements of $A \setminus \{a\}$. Assume also that $r(a,\widetilde{\rho}_j) \leq r(a,\rho_j)$ and $r(a,\widetilde{\rho}_{\mathrm{tb}}) \leq r(a,\rho_{\mathrm{tb}})$. We want to prove that $r(a,\widetilde{R}) \leq r(a,R)$. To that effect, it suffices to solve the two following special cases:

(i) $r(a,\widetilde{\rho}_j) = r(a,\rho_j) - 1$, $r(a,\widetilde{\rho}_i) = r(a,\rho_i)$ for $i \neq j$, $r(a,\widetilde{\rho}_{\mathrm{tb}}) = r(a,\rho_{\mathrm{tb}})$.

(ii) $r(a,\widetilde{\rho}_j) = r(a,\rho_j)$ for all j, $r(a,\widetilde{\rho}_{\mathrm{tb}}) = r(a,\rho_{\mathrm{tb}}) - 1$.

In fact, any other case can be taken care of by means of several steps where each step falls in one of these two cases.

Case (i). In this case the only difference between the two sets of data lies in the existence of a single $b \in A$ such that $ab \in \widetilde{\rho}_j$ whereas $ba \in \rho_j$. As a consequence, we will have $\widetilde{s}(ab) > s(ab)$ and

$\widetilde{s}(ba) < s(ba)$, which entails that $r(ab, \widetilde{H}) \leq r(ab, H)$ and $r(ba, \widetilde{H}) \geq r(ba, H)$. Other than this, H and $\widetilde{H}$ compare the elements of $\Pi \setminus \{ab, ba\}$ in exactly the same way. Clearly, this situation falls within the hypotheses of Corollary B.8.5, which ensures our claim.

Case (ii). In this case the only difference between the two sets of data lies in the existence of a single $b \in A$ such that $ab \in \widetilde{\rho}_{\rm tb}$ whereas $ba \in \rho_{\rm tb}$. By hypothesis, in that case the resulting hyperrankings H and $\widetilde{H}$ satisfy (HM). So we fall again into the hypotheses of Corollary B.8.5, which ensures our claim. □

Corollary B.8.7. *The method of ranked pairs with natural tie-breaking is monotonous.*

Proof. It suffices to check that rule (RPN) fulfils the hypotheses of Corollary B.8.6. The only non-trivial part consists in checking that $H = \mathrm{Hr}(s, \rho_{\rm tb})$ and $\widetilde{H} = \mathrm{Hr}(s, \widetilde{\rho}_{\rm tb})$ satisfy (HM) whenever $\widetilde{\rho}_{\rm tb}$ differs from $\rho_{\rm tb}$ by only one inversion of the form $a\,\widetilde{\rho}_{\rm tb}\,b$ but $b\,\rho_{\rm tb}\,a$. In that situation, part (a) of (HM) is easily obtained as a consequence of (RPN). In contrast, parts (b) and (c) are better established through (RPN′). In fact, by the definition of $s^{\rm tb}(\pi)$, the inequalities $r(a, \widetilde{\rho}_{\rm tb}) < r(a, \rho_{\rm tb})$ and $r(v, \widetilde{\rho}_{\rm tb}) \geq r(v, \rho_{\rm tb})$ for any $v \neq a$ imply that $\tilde{s}^{\rm tb}(av) > s^{\rm tb}(av)$ and $\tilde{s}^{\rm tb}(xa) < s^{\rm tb}(xa)$ whenever $v, x \neq a$. Now, $s^{\rm tb}$ and $\tilde{s}^{\rm tb}$ take the same values but in a different order. This, together with the hypothesis that the paired comparison scores $s(\pi)$ remain the same, allows to derive that $r(av, \widetilde{H}) \leq r(av, H)$ and $r(xa, \widetilde{H}) \geq r(xa, H)$ whenever $v, x \neq a$. □

Potser valdria la pena escriure (RP) en la forma següent:

$$R_k = \begin{cases} R_{k-1}, & \text{if } R_{k-1} \text{ already contains } \pi'_k \quad (\pi_k \text{ is rejected});\\ R_{k-1} = (R_{k-1} \cup \{\pi_k\})^*, & \text{if } R_{k-1} \text{ already contains } \pi_k \quad (\pi_k \text{ is confirmed});\\ (R_{k-1} \cup \{\pi_k\})^*, & \text{if } R_{k-1} \text{ contains neither } \pi_k \text{ nor } \pi'_k \quad (\pi_k \text{ is accepted}). \end{cases}$$

[**x**] R. Greiner 1909, F. Esscher 1924, Maurice G. Kendall 1938.
(references in Kendall)

[**x**] G. Th. Guilbaud, 1952.
Les théories de l'interêt général et le problème logique de l'agrégation.
Économie Appliquée, 5 (4) 1952.
Reprinted in: Eléments de la Théorie des Jeux (Paris, Dunod, 1968).
English translation in: Readings in Mathematical Social Sciences (Chicago, Science Research Associates, 1966): 262–307.

[**x**] G. Th. Guilbaud, 1953.
Sur une difficulté de la théorie du risque.
Colloques Internationaux du Centre National de Recherche Scientifique, Econometrie, 40: 19–25.

[**x**] M. Barbut, 1959.
Quelques aspects mathématiques de la décision rationelle.
Les Temps Modernes, 15 (164): 725–745.

[**x**] G. Th. Guilbaud, 1970.
Préférences stochastiques.
Mathématiques et Sciences Humaines, 32: 45–56.

[**x**] Eric Jacquet-Lagrèze, 1969.
L'agrégration des opinions individuelles.
Informatique et Sciences Humaines, 4.
→ UCM (aquest volum manca!)

[**?**] L. Hubert, 1973.
Monotone invariant clustering procedures.
Psychometrika, 38 (1): 47–62.
→ Humanitats (aquest volum manca!)

[**x**] G. Brossier, 1980.
Représentation ordonnée des classifications hiérarchiques.
Statistique et Analyse des Données, 2: 31–44.
→ `http://www.fsagx.ac.be/si/biblio.htm` (Gembloux, Bèlgica).

[**x**] Thomas More Zavist, 1986.
Toward consensus.
Quoted by T. N. Tideman, 1987.

[**x**] Bernard Debord, 1987.
Caractérisation des matrices de préférences nettes et méthodes d'agrégation associées.
Mathématiques et Sciences Humaines, 97: 5–17
→ UAB?

[**xx**] Ian Stewart, 1989, 1991.
Game, set and math.
Basil Blackwell (Cambridge, Mass., USA) and Penguin (London, UK).
→ UGR

[**xx**] J. Bring, K. Carling, 1994.
A paradox in the ranking of figure skaters.
Chance, 7: 34–37
→ UCM

[**15c**] Edmund L. Russell III, 1995.
Scoring in figure skating: What does it measure?.
Proceedings of the Section on Statistics in Sports. American Statistical Association, 1995:

xx–xx.
→ `asainfo@amstat.org` (anterior a 1996)

[**15e**] Edmund L. Russell III, 1997.
Choosing a ranking system for figure skating.
STATS: The Magazine for Students of Statistics, 20: 13–17.
→ `asainfo@amstat.org` (sense ser-ne membre?) (demanat)

[**xx**] Douglas R. Woodall, xxxx.
Properties of single-winner preferential election rules II.
X X X.

[**xx**] Laurent Vidu, 1999 *Social Choice and Welfare*, 16: 159–167.
→ UAB

Hubert (CDA)
Hansen + Jaumard (Canada, Espaliers)
Chepoi + Fichet
Rouillon (Canada, Nouvelle Democracie)
Cretney (EM)
Schulze?
Eppley
EM
Rossano (patinatge) -
Loosemore (patinatge) -
Scrutiny3
francesos
alemany
Braden ***
Leach (Dance News) ***
Smith-Hampshire? ***
Koponen (IDSF)
V. Bain (IDSF)
Fenn (ball esportiu)
Marco ***
AEPBS
Burgauner (ball esportiu + filosofia)
McLean (Oxford)
Tideman?
Zavist?
Saari
Laslier
Truchon
Girbau -
Rosa -
Aureli -
Warren -
Enric Nart
P Puig
Muñoz
Josep Lluís
Josep Vives
Frederic -
Juanjo -
Julià
Carles -
Solà-Morales -
Bonet ***
Calsina -
Sílvia -
D Jou
Grífols?
Joan Carles Gil
Ferran Rovira -
Sebastià Sànchez -
Dave Getty
Tony Lee
Jordi Viñas -
Maydeu-Olivares
Soler
Pérez?

∗ Conjecture. For every profile, and every basis ranking immune to majority complaints, the reduction rating procedure satisfies the conditions of § 10.2.

Remarks:

0. The name *reduction rating* obeys to the fact / has been chosen because he situation considered by the classification condition is mathematically known as reducibility of the matrix of scores. For sharp scores (margins $\in \{-1, 0, 1\}$) the projection step 1 always results in a reducible matrix or a complete tie). Cf. definition of reducibility in terms of standard powers of matrices: for sharp scores, standard powers are equivalent to boolean powers, which are relevant to the projection step 1 and immunity to majority complaints).

1. Every profile: Not true for every paired-comparison matrix (example: Condorcet cycle with sharp scores → failure of classification condition). Only those that come from a profile. Characterization? For $N = 3$ ... General N: difficult (related to permutation polytopes?).

2. Every basis ranking immune to majority complaints: Essential.

3. Reduction rating procedure: Step 1 is needed. "Saari" projection is not enough (example). The matrices resulting from step 1 can be characterized by a property that generalizes anti-Robinson (see Chepoi + Fichet, 1997). Anti-Robinson is known to have good properties in connection with problems similar to the one that we are considering (refs?).

4. Reduction rating procedure: Step 3 might be unnecessary for the result to hold, but is included to avoid a multiplicity of rating results.

5. The conditions of § 10.2: Notice that the compatibility condition admits the possibility of ties. Tied rates happen when either the paired-comparison matrix already contains some kind of ties, or when the basis ranking is not suitable. (Related to the characterization of paired-comparison matrices that com from a profile?).

6. Relation to the "espaliers" of Hansen and Jaumard, 1996. The proposition above corresponds to espaliers having no diagonals?

7. For $N = 2$, where there are only two possible rankings, the reduction rates coincide exactly with the rank averages as long as the basis ranking is taken to be the "correct" one. If the basis ranking is taken in the other way, then the reduction rating is a complete tie.

8. For $N \geq 3$ the conditions of § 10.2 can be seen to be incompatible with any linear relationship between the paired-comparison scores and the rates. Accordingly, the reduction rating method involves some non-linear operations.

9. The reduction rating does not depend continuously on the entries of the matrix of scores (or the judges weights). In other words, sometimes a slight change in the scores can result in a large variation of the rates. Example: $\{\{*, a, 1\}, \{1 - a, *, b\}, \{0, 1 - b, *\}\}$ where a and b can be greater than $1/2$: there is a discontinuity for $a = b$. Step 3 → mid point, but does'nt avoid the discontinuity. Maybe the conditions of § 10.2 imply a topological obstruction to continuity (cf. Chichilnisky, 1980ss).

10. Another source of discontinuity: Changes in the basis ranking (a discrete object). This contrasts with Borda's method, but is not strange at all for a non-linear (optimization) algorithm.

11. The reduction rates tend to be closer to each other in the measure that the projected matrix approaches a complete tie. In its turn, this may be due to two causes: one possibility is a large amount of disagreement between the input rankings; alternatively, the input rankings may be in much agreement with each other, but not with the basis ranking. Let us assume that the basis ranking is not fixed, but we allow it to vary. Certainly, the projection step above is strongly related to the ideas of § 8, and more specifically to the amount of agreement or disagreement with the judges as discussed in § 8.3. In general terms, one can say that a large disagreement will cause the projection step to introduce more ties, and as a consequence the resulting rates will be closer to each other. Contrarily, a ranking with a large agreement will cause the reduction rates to be more widely distributed. This immediately suggests a way to convert the ideas of this section into a self-contained rating method, namely, to look for the maximum dispersion. Of course, the result will depend on the way of measuring dispersion, for which there are several possibilities. Certainly, it would be very interesting that this method were always in agreeement with the method of ranked pairs, but a preliminary exploration seems to discard such a possibility. Difficult: when a parameter is varied continuously and the RP ranking

changes, at the moment of change both ratings should have exactly the same dispersion. We have a multidimensional object, whose structure is known only indirectly through the matrix of dissimilarities = distances (we do not know even its dimension); we are projecting it in one dimension; we are choosing the direction in which the items become more spread.