How Best to Rank Wines: Majority Judgment

Rida Laraki, Michel Balinski

ECOLE POLYTECHNIQUE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

HOW BEST TO RANK WINES: MAJORITY JUDGMENT

Rida LARAKI
Michel BALINSKI

June 2012

Cahier n° 2012-26

DEPARTEMENT D'ECONOMIE
Route de Saclay
91128 PALAISEAU CEDEX
(33) 1 69333033
http://www.economie.polytechnique.edu/
mailto:chantal.poujouly@polytechnique.edu

How Best to Rank Wines: Majority Judgment

Michel Balinski and Rida Laraki
CNRS and Ecole Polytechnique

Classifying and ranking wines has been a favorite activity of men and women since time immemorial. Gaius Plinius Secundus (Pliny the Elder) who died in the year 79 AD wrote in his treatise *The Natural History* [13]:

> "Who can entertain a doubt that some kinds of wine are more agreeable to the palate than others, or that even out of the very same vat there are occasionally produced wines that are by no means of equal goodness, the one being much superior to the other, whether it is that it is owing to the cask, or to some other fortuitous circumstance… The late Emperor Augustus preferred the Setinum to all others, and nearly all the emperors that have succeeded him have followed his example … The second rank belonged to the wine of the Falernian territory, of which the Faustianum was the most choice variety; the result of the care and skill employed upon its cultivation … To the third rank belonged the various wines of Alba …I am by no means unaware that most of my readers will be of opinion that I have omitted a vast number of wines, seeing that every one has his own peculiar choice … Indeed I have no wish to deny that there may be other wines deserving of a very high reputation, but those which I have already enumerated are the varieties upon the excellence of which the world is at present agreed."

And yet, although today there seems to be a large consensus about the physical conditions that should attend a serious wine tasting whose objective is to classify and to rank, how to express and then how to amalgamate the various opinions of individual judges into a collective assessment of all continues to defy description and bedevil decisions.

The intent of this article is to explain how and why the traditional methods for amalgamating the grades fail and how and why a new approach – *majority judgment* – does the job best.

1. Traditional Amalgamation Schemes and the "Judgment of Paris"

The famous – or infamous, depending perhaps on which side of the Atlantic your heart resides – wine tasting organized by the well-known English wine-expert Steven Spurrier in Paris on May 22, 1976 pitted vintage Cabernet Sauvignon wines of Bordeaux against those of California. Eleven judges participated: Spurrier, an American lady, Patricia Gallagher, and nine respected French connoisseurs. The tasting was blind, the judges graded each wine on a 0-20 scale, and the wines' average scores determined the order-of-finish. *Time* magazine wrote on June 7, 1976, "The unthinkable happened: California defeated Gaul."

We believe the facts show that is patently false: California did not defeat Gaul. California only defeated Gaul because of the method used to amalgamate the judges' opinions. Ranking the wines in accordance with their average grades is a very bad idea, as will soon be apparent.

There were 10 wines, 4 French and 6 Californian:

A. Stag's Leap 1973 (Californian)
B. Château Mouton Rothschild 1970 (French)
C. Château Montrose 1970 (French)
D. Château Haut-Brion 1970 (French)

E. Ridge Monte Bello 1971 (Californian)
F. Château Léoville-Las Cases 1971 (French)
G. Heitz Martha's Vineyard 1971 5Californian)
H. Clos du Val 1972 (Californian)
I. Mayacamas 1971 (Californian)
J. Freemark Abbey 1969 (Californian)

Judges were asked to grade the wines on a scale going from 0 (worst) to 20 (best), the traditional scale used in French schools and universities. No absolute meaning was given the grades, leaving to individual judges to use their own criteria. The tasting was blind: judges only knew they tasted all of the ten designated wines; they had no information concerning which specific ones.

The judges' grades[1] are given in table 1 where, as subsequently, a * denotes French wines.

	A	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
P. Brejoux	14.0	16.0	12.0	17.0	13.0	10.0	12.0	14.0	5.0	7.0
A. de Villaine	15.0	14.0	16.0	15.0	9.0	10.0	7.0	5.0	12.0	7.0
M. Dovaz	10.0	15.0	11.0	12.0	12.0	10.0	11.5	11.0	8.0	15.0
P. Gallagher	14.0	15.0	14.0	12.0	16.0	14.0	17.0	13.0	9.0	15.0
O. Kahn	15.0	12.0	12.0	12.0	7.0	12.0	2.0	2.0	13.0	5.0
C. Dubois-Millot	16.0	16.0	17.0	13.5	7.0	11.0	8.0	9.0	9.5	9.0
R. Olivier	14.0	12.0	14.0	10.0	12.0	12.0	10.0	10.0	14.0	8.0
S. Spurrier	14.0	14.0	14.0	8.0	14.0	12.0	13.0	11.0	9.0	13.0
P. Tari	13.0	11.0	14.0	14.0	17.0	12.0	15.0	13.0	12.0	14.0
C. Vanneque	16.5	16.0	11.0	17.0	15.5	8.0	10.0	16.5	3.0	6.0
J.-C. Vrinat	14.0	14.0	15.0	15.0	11.0	12.0	9.0	7.0	13.0	7.0

Table 1. Judges' grades.

How is one to deduce the collective opinion of the jury? An infinite number of methods exist, though the attention has typically been confined to very few.

Point-summing

The "obvious" method is to add the points given each wine – or equivalently, compute their average – then rank them in accord with their sums or averages. This is what was done in the Judgment of Paris, giving the result

Point-summing	*A*	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
Sum	155.5	155.0	150.0	145.5	133.5	123.0	114.5	111.5	107.5	106.0
Average	14.14	14.09	13.64	13.23	12.14	11.18	10.41	10.14	9.77	9.64
Rank	1	2	3	4	5	6	7	8	9	10

By this method "the golden state" defeated Gaul in that one of its wines finished first, though the four wines ranked last were all Californian whereas the four French wines were all in the first six places.

[1] There are some very small discrepancies in the grades assigned the wines in the available sources. We follow those given by R. E. Quandt [14]. They are slightly different than those given in our book [1]; thankfully those discrepancies are small enough that they do not appreciably change the analysis.

But is this outcome *reasonable*? Was Gaul *really* defeated in 1976? Or was the ranking simply a roll of the dice that depended on the use of this particular method for amalgamating the various individual opinions? What might other methods have proposed?

Truncated point-summing

When point-summing methods are used it is sometimes modified by first eliminating a competitor's highest and lowest grades in order to dampen the influence of exaggerated grades. For example, in figure skating competitions the top two and bottom two grades are dropped. In the Judgment of Paris dropping each wine's top and bottom grades does not change the order; however, dropping their two top and bottom grades does:

Truncated point-summing	*A*	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
Sum	99.0	100.0	97.0	93.5	86.5	79.0	73.5	71.0	72.5	65.0
Average (7 grades)	14.14	14.29	13.86	13.36	12.38	11.29	10.50	10.14	10.38	9.29
Rank	2	1	3	4	5	6	7	9	8	10

Quandt's method

"Grading wines consists of assigning 'grades' to each wine, with no restrictions on whether ties are permitted to occur. While the resulting scale is not a cardinal scale, some meaning does attach to the level of the numbers assigned to each wine. Thus, if on a 20-point scale, one judge assigns to three wines the grades 3, 4, 5, while another judge assigns the grades 18, 19, 20, and a third judge assigns 3, 12, 20, they appear to be in complete harmony concerning the ranking of wines, but have serious differences of opinion with respect to absolute quality. I am somewhat skeptical about the value of the information contained in such differences. But we always have the option of translating grades into ranks and then analyzing the ranks …" [13, p. 8].

Since a judge may give two or more wines the same grade, each is assigned the average of their possible places in the judge's ranking (e.g., three wines with the second highest grade – occupying places 2, 3 and 4 in the ranking – are each ranked 3). Thus with Quandt's method table 2 gives the relevant input data.

	A	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
P. Brejoux	3.5	2.0	6.5	1.0	5.0	8.0	6.5	3.5	10.0	9.0
A. de Villaine	2.5	4.0	1.0	2.5	7.0	6.0	8.5	10.0	5.0	8.5
M. Dovaz	8.5	1.5	6.5	3.5	3.5	8.5	5.0	6.5	10.0	1.5
P. Gallagher	6.0	3.5	6.0	9.0	2.0	6.0	1.0	8.0	10.0	3.5
O. Kahn	1.0	4.5	4.5	4.5	7.0	4.5	9.5	9.5	2.0	8.0
C. Dubois-Millot	2.5	2.5	1.0	4.0	10.0	5.0	9.0	7.5	6.0	7.5
R. Olivier	2.0	5.0	2.0	8.0	5.0	5.0	8.0	8.0	2.0	10.0
S. Spurrier	2.5	2.5	2.5	10.0	2.5	7.0	5.5	8.0	9.0	5.5
P. Tari	6.5	10.0	4.0	4.0	1.0	8.5	2.0	6.5	8.5	4.0
C. Vanneque	2.5	4.0	6.0	1.0	5.0	8.0	7.0	2.5	10.0	9.0
J.-C. Vrinat	3.5	3.5	1.5	1.5	7.0	6.0	8.0	9.5	5.0	9.5

Table 2. Judges' "ranks."

The method treats the judges' individual ranks as points and ranks the wines according to their sums (or, equivalently, their averages). This gives the final ranking:

Quandt	*A*	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
Sum	41.0	43.0	41.5	49.0	55.0	72.5	70.0	79.5	77.5	76.0
Average	3.73	3.91	3.77	4.45	5.00	6.59	6.36	7.23	7.05	6.91
Rank	1	3	2	4	5	7	6	10	9	8

or,

If comparisons are the key a host of other methods immediately suggest themselves.

Majority vote

For Condorcet [4] each judge should express his opinion by giving a comparative judgment among all the wines taken pair-by-pair. This amounts to ranking all the wines yet allowing ties among them. Condorcet's basic idea was that there should be a vote among each pair of wines. For example, what is the jury's decision between the two wines *A* and *C**? Condorcet says the majority decides: *A* is ranked above *C** by 3 judges, below *C** by 5 judges, and equal by 3 judges, so *C** should be ranked above *A* with 6.5 votes to *A*'s 4.5. His hope was that there would be a wine – the *Condorcet-winner* – that defeats all others in a direct vote.

The numbers of votes for every possible confrontation is given in table 3 (e.g., *B** wins 4.5 votes against *A* and 8.0 against *E*). 5.5 votes mean a tie (indicated by below), any number above 5.5 that the wine in its row wins by a majority of the evaluations against the wine in its column (indicated by below).

	A	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
A	-	6.5	4.5	5.5	7.5	10.0	8.0	8.5	10.5	8.0
B*	4.5	-	5.0	5.5	8.0	9.0	9.0	9.0	8.0	9.0
C*	6.5	6.0	-	6.5	5.5	10.0	7.5	8.5	9.5	8.5
D*	5.5	5.5	4.5	-	6.5	7.5	7.5	8.5	8.0	7.5
E	3.5	3.0	5.5	4.5	-	6.5	9.0	8.0	6.0	9.0
F*	1.0	2.0	1.0	3.5	4.5	-	5.0	7.0	6.5	7.0
G	3.0	2.0	3.5	3.5	2.0	6.0	-	7.0	6.0	7.0
H	2.5	2.0	2.5	2.5	3.0	4.0	4.0	-	6.0	4.0
I	0.5	3.0	1.5	3.0	5.0	4.5	5.0	5.0	-	5.0
J	3.0	2.0	2.5	3.5	2.0	4.0	4.0	7.0	6.0	-

Table 3. Majority votes pair-by-pair.

The famous *paradox of Condorcet* is that majority vote may lead to no winner and no ranking. The Judgment of Paris is a striking example of its occurrence:

There is no Condorcet-winner – no first in the ranking – but there is a *Condorcet-loser* – a wine that is last in the ranking. Each of the five wines in the first group (*A, B*, C*, D*, E*) defeats all the others; for the others a wine to the left defeats a wine to the right. The first five

wines are in a *Condorcet-cycle*. Or, ignoring a good deal of information, it might be said that there is a five-way tie for first place, the others occupying second through sixth places.

Condorcet	*A*	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
Rank	1	1	1	1	1	3	2	5	6	4

Three French wines are in first place, only two California wines. Three California wines are in the last places.

Borda's method

Borda [3] agreed with Condorcet: every judge should evaluate the merits of each wine compared successively to the merits of each of its competitors. But he advocated amalgamating the opinions by summing each wine's votes against all of the others (sums of the numbers in their rows in table 3). The idea is in spirit very similar to Quandt's yet different. It happens to give the same result in this case.

Borda	*A*	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
Sum	69	67	68.5	62	55	37.5	40	30.5	32.5	34
Average	6.9	6.7	6.85	6.2	5.5	3.75	4.0	3.05	3.25	3.4
Rank	1	3	2	4	5	7	6	10	9	8

This method should actually be known under another name since it was recently discovered that Cusanus [9] had proposed it in 1433.

Black's method

Duncan Black [2] suggested that the Condorcet majority-criterion should be used and where it fails the wines in a Condorcet-cycle should be ranked according to Borda's method. This gives the ranking

Llull's method

In 1299 Ramon Llull [8] advanced what seems to be the first formal presentation of a rule of voting. It is a generalization of the idea of a Condorcet-winner. He proposed that the wines be ranked according to its sum of wins against all competitors (a tie counting as a win), instead of its sum of votes against all competitors, as does Borda. This means that if there is a Condorcet-winner, then that wine must necessarily be the winner. This method is known in the modern literature as Copeland's method [5]. It yields:

Llull	*A*	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
Wins	8	7	9	8	6	3	4	1	0	2
Rank	2	3	1	2	4	6	5	8	9	7

or

In first place a French wine, in the first four places three French wines.

Dasgupta-Maskin's method

"If no [wine] obtains a majority against all [others], then among those [which] defeat the most opponents in head-to-head comparisons, select as winner the one with the highest [Borda score]" [6, p. 97]. So if there are ties in the number of wins they are resolved with Borda's method.

According to Llull wines *A* and *D** are tied. *A* has the Borda-score 69 and *D** 62, so this method gives the ranking:

or

Dasgupta-Maskin	*A*	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
Rank	2	4	1	3	5	7	6	10	9	8

Eight methods give seven (that magical number!) different rank-orderings of the wines.

2. Why the Traditional Amalgamation Schemes Fail

Meaningfulness

There is no cardinal measure with which to rate wines (as Quandt quite correctly observes). He goes on to state, "Two scales for rating are in common use: (1) the well-known ordinal rank-scale by which wines are assigned ranks …, and (2) a 'grade'-scale, such as the well-publicized ratings by Robert Parker based on 100 points. The grade scale has some of the aspects of a cardinal scale, in that intervals are interpreted to have meaning, but is not an [interval measure]" [14, p. 2]. These statements raise the important question of "meaningfulness" and require clarification and elaboration.

How to construct a scale for measuring something is a science in itself (see e.g., [10]). The types of scales have been classified in various ways (of which one follows [15]). When numbers, names or labels indicate categories (blood type, bus number, telephone code), the scale is a *nominal measure*. When they indicate order (pain, mineral hardness, destructive power of earthquakes), it is an *ordinal measure*. Pain, for example, is usually measured on a scale from 0 to 10. The Mankowski scale defines a 2 by "Minor annoyance – occasional strong twinges. No medication needed;" a 3 by "Annoying enough to be distracting. Mild painkillers take care of it;" a 9 by "Unable to speak. Crying out or moaning uncontrollably – near delirium;" and a 10 by "Unconscious. Pain makes you pass out." [16] When in addition to order, equal intervals have the same significance (days of calendars, degrees Celsius and Fahrenheit), it is an *interval measure*. Finally, when in addition to qualifying as an interval measure, zero has an absolute meaning (dollars, grams, degrees Kelvin), it is a *ratio measure.*
Two key problems present themselves. How to assign scale values to empirical observations is the "representation problem." What analyses of observations are valid as a function of the type of scale is the "meaningfulness problem." It is obviously meaningless to add numbers that are nominal measures: the sum or average of two telephone codes bears no earthly meaning. It is also meaningless to add numbers that are ordinal measures: an increase in pain from 2 to 3 cannot be compared with an increase from 9 to 10 let alone have the same significance. For sums or averages to be meaningful the numbers must come from interval measures such as calendars – an additional day has the same significance whenever it is added in the Gregorian, Hebrew or Moslem calendars – or such as weight, distance or money – one

more carries the same meaning to whatever it is added (the last three examples are ratio measures so multiplication makes sense too).

The official results of the Judgment of Paris depend on adding or averaging numbers ranging from 0 to 20. For them to make any sense at all those numbers must be chosen from an interval measure. But that is obviously false.

First, no common definitions were given to the numbers on the scale, so each judge gave his own interpretation of their meanings. Nonetheless, it is reasonable to assume that each judge had his own benchmark wines acquired in years of experience and some absolute sense of what it means to give a 7 or a 17 to a wine; moreover, it is not unreasonable to assume that these benchmarks were fairly similar.

Second, as is so often true when a numerical scale is used, the higher the grade the more difficult it is to raise it: there seems to be a human reluctance to give very high grades. That is certainly the case here although these were a very fine set of wines: of 110 grades none were above 17 and only four reached that level. The scale was not an interval measure, because increasing the grade of a wine from (say) 17 to 18 was much more difficult for a judge to do than increasing a grade from an 11 to a 12. But this immediately implies that the sums and averages of such grades are meaningless, so point-summing fails.

When judges compare wines and the comparisons are amalgamated, as is done in all the methods described above except point-summing, the same difficulties arise. Every one of those methods involves using Borda's basic idea that treats a place in the order as though it were an interval measure. But it is certainly not an interval measure since these methods treat the very different judges' inputs of (3, 4, 5), (18,19, 20) and (3, 12, 20) as exactly the same, which they are not: the first believes the wines are all bad and differ little, the second that all are excellent and differ little, the last that there is a huge difference among the quality of all three. Indeed, one cannot but question whether statistical analyses based on "places in the order" are meaningful.

Manipulability

There is also a second major reason that point-summing methods fail in competitions. They are among the most highly manipulable of all methods. As Sir Francis Galton so aptly remarked: "Each voter [of the jury] has equal authority with each of his colleagues. How can the right [collective] conclusion be reached, considering that there may be as many different [grades] as there are members? That conclusion is clearly not the average of all the [grades], which would give a voting power to 'cranks' in proportion to their crankiness. One absurdly large or small [grade] would leave a greater impress on the result that one of reasonable amount, and the more a [grade] diverges from the bulk of the rest, the more influence would it exert" [7]. A "cranky'' wine judge may be one who errs from lack of judgment or finesse; she may also be a judge who "cheats'' willfully either because she has some predefined agenda (e.g., a high place for friends' wines and a low one for enemies' wines that she believes she can identify) or simply because she wishes to impose her superior will on the collective jury decision. The director of a wine competition complained that in Australia – where typically juries consist of three judges who assign points from a 0 to 20 scale that are averaged to determine decisions - two judges may both give gold medal scores of 18.5 but a cranky third judge can completely thwart the majority opinion by giving a low score.

A glance at the grades given cannot but make one wonder whether the 2's, 3's and 5's were cranky grades. But put that question aside and assume the grades were "honest" evaluations. How might a judge have manipulated and what success might he have had?

Since no judge gave either a 0 (the minimum) or a 20 (the maximum) to any wine, *every* judge can manipulate the final score of *every* wine either up or down. Take, for example, C. Dubois-Millot. He could achieve the final order he prefers among *A*, *B**, *C**, *D**, and *E* as well as the order he prefers among *F**, *G*, *H*, *I*, and *J* by changing the grades he assigns as follows:

Point-summing	*A*	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
Dubois-Millot (actual)	16.0	16.0	17.0	13.5	7.0	11.0	8.0	9.0	9.5	9.0
Sum (actual)	155.5	155.0	150.0	145.5	133.5	123.0	114.5	111.5	107.5	106.0
Dubois-Millot (new)	13.0	13.5	20.0	13.5	7.0	0.0	4.0	8.5	13.5	14.0
Sum (new)	152.5	152.5	153.0	145.5	133.5	112.0	110.5	111.0	111.5	111.0
Rank (new)	2	2	1	3	4	5	8	7	6	7

Manipulation enables him to obtain *exactly* the order he wishes except that *E* is not last.

C. Vanneque's preferred order-of-finish is very different than that given by point-summing:

Manipulation enables him to obtain the order he wishes except for *H* and a near miss in the order of *C** and *E*.

Point-summing	*A*	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
Vanneque (actual)	16.5	16.0	11.0	17.0	15.5	8.0	10.0	16.5	3.0	6.0
Sum (actual)	155.5	155.0	150.0	145.5	133.5	123.0	114.5	111.5	107.5	106.0
Vanneque (new)	9.0	8.5	0.0	20.0	20.0	0	11.0	20.0	3.0	8.0
Sum (new)	148.0	147.5	139.0	148.5	138.0	115.0	115.5	115.0	107.5	108.0
Rank (new)	2	3	4	1	5	7	6	7	9	8

When "places in the order" are inputs there are again wide-open opportunities for judges to manipulate. If, for example, a judge is intent on wishing one wine V to lead over another W it suffices to place V first (or high) and W last (or low). Dubois-Millot could again achieve the order among *A*, *B**, *C**, *D**, and *E* he prefers (as well as *A*, *B**, *C**, *D**, and *F** in that order above *H*, *I* and *J*) as follows:

Quandt	*A*	*B**	*C**	*D**	*E*	*F**	*G*	*H*	*I*	*J*
Dubois-Millot (actual)	2.5	2.5	1.0	4.0	10.0	5.0	9.0	7.5	6.0	7.5
Sum (actual)	41.0	43.0	41.5	49.0	55.0	72.5	70.0	79.5	77.5	76.0
Dubois-Millot (new)	4.0	2.0	1.0	3.0	7.0	5.0	10.0	8.0	6.0	9.0
Sum (new)	42.5	42.5	41.5	48.0	52.0	72.5	71.0	80.0	77.5	77.5
Rank (new)	2	2	1	3	4	6	5	8	7	7

There is ample experimental evidence showing Borda's method is highly manipulable, as are point-summing methods (truncated or not) [1]. They are also, as was seen, meaningless. Point-summing methods have one redeeming property that the others do not: they are "coherent" in that the final order between any two wines does not depend on whether other wines are competing (they avoid "Arrow's paradox").

Coherence

A method of amalgamation is *coherent* if it ranks one wine X above another Y whatever other wines participate or do not participate in the competition. Theorem: *Every method based on comparisons is incoherent*.

Take either Borda's or Quandt's method. When used to rank-order all the wines they yield

However, when used to rank-order *A*, *B**, *C**, and *D** alone they yield
when used to rank-order *A*, *B** and *C** alone they give ; and when *A* and *C** alone they conclude

Llull's and Dasgupta-Maskin's methods are also incoherent. Llull's order among all is

and since *A*'s Borda-score is 69 and *D*'s* 62, Dasgupta-Maskin's order is

But among only the wines *B**, *C**, and *D** Llull yields

Llull	*B**	*C**	*D**
Wins	1	2	1
Rank	2	1	2

or

which is incoherent with the Llull order among all. Since *B**'s Borda-score is 67 and *D*'s* 62, Dasgupta-Maskin's order among the three is

incoherent with the Dasgupta-Maskin order among all.

These are all occurrences of Arrow's paradox: depending upon the presence or absence of other wines the order between two (or more) may change. It is quite clearly an unacceptable property. *No method that depends on comparisons alone avoids it.*

3. Majority judgment

The U.I.Œ. (Union internationale des Œnologues) is an international federation of national œnological associations. Until 2009 they advocated the use of a standard tasting sheet for each wine of a competition (see table 5). Each attribute of a wine is evaluated in a language of seven grades: *Excellent, Very Good, Good, Passable, Inadequate, Mediocre, Bad*. This is an important improvement over assigning undefined points or relying on orders since these evaluations have meaning that are by and large shared by judges (who can refer to the shared benchmarks of years of experience). To each evaluation of every attribute is associated a

number of points. Their total points determine the wines' awards and their rank-order. To be awarded a "Grand Gold" a wine must have a total score of 90 or above; a "Gold" a score of at least 85 but below 90; a "Silver" at least 80 but below 85; and a "Bronze" at least 75 but below 80. This is a point-summing method where meaning is given to the points accorded, which is essential. Nevertheless, these points do not constitute an interval measure, so their sums remain meaningless.

	Excellent	*Very Good*	*Good*	*Passable*	*Inadequate*	*Mediocre*	*Bad*
Aspect							
Limpidity	6	5	4	3	2	1	0
Nuance	6	5	4	3	2	1	0
Intensity	6	5	4	3	2	1	0
Aroma							
Frankness	6	5	4	3	2	1	0
Intensity	8	7	6	5	4	2	0
Finesse	8	7	6	5	4	2	0
Harmony	8	7	6	5	4	2	0
Taste, flavor							
Frankness	6	5	4	3	2	1	0
Intensity	8	7	6	5	4	2	0
Body	8	7	6	5	4	2	0
Harmony	8	7	6	5	4	2	0
Persistence	8	7	6	5	4	2	0
After-taste	8	7	6	5	4	2	0
Global opinion	8	7	6	5	4	2	0

Table 5. U.I.Œ. "Sensorial Analysis Tasting sheet for Wine Judging Competitions," still wines, 2006. A judge circles her evaluations in each row.

All of the methods based on comparisons depart from the primitive idea that when the majority of a jury prefers wine *A* to wine *B* then that should be the jury's decision. One difficulty with this is that Condorcet's paradox may occur so there is no rank-order and no winner. But there is a deeper, much less appreciated difficulty: there is no real justification for accepting a majority preference *even between two wines alone*. For suppose – using the U.I.Œ. grades –11 judges gave two wines, *A* and *B*, the following grades:

	Excellent	*Good*	*Mediocre*	*Bad*
A:	4	3	2	2
B:	3	2	2	4

Since *A*'s high grades dominate *B*'s there is little doubt that the jury's collective opinion is But what would ordinary majority voting say? That would depend on the comparisons of the judges. If their evaluations were

	2 judges	**2 judges**	**3 judges**	**1 judge**	**1 judge**	**2 judges**
A:	*Excellent*	*Excellent*	*Good*	*Mediocre*	*Mediocre*	*Bad*
B:	*Good*	*Mediocre*	*Bad*	*Bad*	*Excellent*	*Excellent*

(each wine has the distribution of grades given above) then the majority vote would yield by a vote of 8 to 3. If, however, their evaluations were

	3 judges	4 judges	2 judges	2 judges
A:	*Good*	*Excellent*	*Mediocre*	*Bad*
B:	*Excellent*	*Bad*	*Good*	*Mediocre*

(both wines again have the same distribution of grades) then the majority vote would yield by a vote of 7 to 4. *Moral*: simple majority voting can make the wrong decision! With more information, using a generally accepted scale of grades, such anomalous results may be avoided.

The key fact is summarized in the Theorem: *A method of amalgamation is coherent and avoids the Condorcet paradox if and only if a wine's place in the ranking depends only on the grades it receives* (see [1]). Thus the only way to make coherent decisions and be certain to avoid cyclic jury decisions is precisely *to ignore each judge's implicit comparisons* and pay attention only to the grades she gives.

Point-summing methods heed this injunction: it is coherent and there is no ambiguity in the rank-orders it determines. It fails because it gives meaningless results and is highly manipulable. What then is to be done?

Majority-grade and majority-ranking

First, a common-language of grades must be defined. The U.I.Œ.'s – *Excellent, Very Good, Good, Passable, Inadequate, Mediocre, Bad* – is an excellent choice. More or fewer grades may be chosen. They must be well defined and understood by all the judges.

Let us assume that the number grades of the Judgment of Paris constituted an acceptable common language of grades. Which judge gave the grades cannot be taken into account, so each wine's grades may be arranged from highest to lowest (as in table 6). A wine's *majority-grade* is the grade supported by a majority against any other grade. Wine *C**'s majority grade is 14.0 because it obtains a majority of at least 7 to 4 against any lower grade and it obtains a majority of at least 8 to 3 against any higher grade. Using statistical jargon a wine's majority-grade is the median or middlemost of its grades. Had the grades been words or letters this definition is valid (so number grades are not necessary). The grades – in this case numbers – are never added or averaged (since sums are meaningless), their only significance is ordinal, a higher number means a higher grade or better evaluation.

A	16.5	16.0	15.0	15.0	14.0	*14.0*	14.0	14.0	14.0	13.0	10.0
*B**	16.0	16.0	16.0	15.0	15.0	*14.0*	14.0	14.0	12.0	12.0	11.0
*C**	17.0	16.0	15.0	14.0	14.0	*14.0*	14.0	12.0	12.0	11.0	11.0
*D**	17.0	17.0	15.0	15.0	14.0	*13.5*	12.0	12.0	12.0	10.0	8.0
E	17.0	16.0	15.5	14.0	13.0	*12.0*	12.0	11.0	9.0	7.0	7.0
*F**	14.0	12.0	12.0	12.0	12.0	*12.0*	11.0	10.0	10.0	10.0	8.0
G	17.0	15.0	13.0	12.0	11.5	*10.0*	10.0	9.0	8.0	7.0	2.0
H	16.5	14.0	13.0	13.0	11.0	*11.0*	10.0	9.0	7.0	5.0	2.0
I	14.0	13.0	13.0	12.0	12.0	*9.5*	9.0	9.0	8.0	5.0	3.0
J	15.0	15.0	14.0	13.0	9.0	*8.0*	7.0	7.0	7.0	6.0	5.0

Table 6. Judgment of Paris: grades given wines, their majority-grades italicized.

The majority-grades are used to obtain the *majority-ranking* of the wines, but – as here – there may be ties that need to be resolved. The rationale for resolving them is simple. Consider wines *A* and *B**. A majority decided each should have the grade 14, so a majority should again decide which should be classed ahead among the remaining grades (when that one 14 is dropped from each). Wine *A*'s second majority-grade is 14 (with at least 8 against a lower grade and at least 6 against a higher grade). Now, however, there is a difficulty because with an even number of grades a tie may occur, as happens here for *B**: 5 are for 15 or higher, 5 against.

A	16.5	16.0	15.0	15.0	14.0	——		14.0	14.0	13.0	10.0
*B**	16.0	16.0	16.0	15.0	15.0	——		14.0	12.0	12.0	11.0

When there is an even number of grades the theory dictates [1] there must be an absolute majority for the higher grade and only a relative majority for the lower grade, so *B**'s second majority-grade is 14. Again a tie, so the procedure is repeated. Wine *A*'s third majority-grade is 14, Wine *B*'s third majority-grade is 15: therefore, *B** must be ranked ahead of *A*, or

A	16.5	16.0	15.0	15.0		——	——	14.0	14.0	13.0	10.0
*B**	16.0	16.0	16.0	15.0		——	——	14.0	12.0	12.0	11.0

The majority-ranking for the wines of the Judgment of Paris is:

The majority-ranking can contain a tie *only* if two wines have exactly the same set of grades.

Majority judgment ranks the ten wines differently than any of the other methods previously considered. With what we believe is the only valid method of amalgamating judges' evaluations, the thinkable happened, Gaul defeated California: Chateau Mouton Rothschild 1970 is first, all of the four French wines are among the first six; none of the French wines are among the last four.

Why Majority Judgment

Majority judgment is meaningful. There are two scales for rating in common use, an ordinal scale where wines are assigned ranks and an interval scale where wines are assigned cardinal numbers. But there is also a middle ground that asks for more than ranks but less than an interval scale: an ordinal scale of merit. The U.I.Œ.'s word grades and the Mankowski pain scale are examples. Piano, figure skating, gymnastics, diving and many other competitions use number scales whose meanings are carefully defined and/or come to have very definite meanings much as the measurements of length and weight. So long as they are treated as ordinal and not assumed to constitute interval scales the approach is perfectly valid. The very notion of determining a consensual jury decision implies some commonality in a language of absolute grades, determined by widely shared benchmarks. As Wittgenstein so aptly said, "the meaning of a word is its use in the language." Otherwise, meaningful decisions cannot be made.

Majority judgment is the least manipulable. Both theory and experiments show that among the widely known and recommended methods of amalgamation majority judgment is the least subject to strategic manipulation.

It is *strategy-proof-in-grading* meaning that if a judge's objective is that a particular wine should be evaluated (say) *Very Good* then honesty is her optimal strategy: to assign *Very Good*. Consider, for example, wine *A* of the Judgment of Paris, with a majority-grade of 14 but which C. Dubois-Millot evaluated as 16. Disappointed, could he have upped the grade he

A	16.5	16.0	15.0	15.0	14.0	*14.0*	14.0	14.0	14.0	13.0	10.0

assigned to 20 and thereby raised *A*'s majority-grade? The answer is no because that would change nothing in the majority opinion. Symmetrically, P. Tari gave 13 to *A*, so he too was disappointed but could do nothing to lowering *A*'s majority-grade even by changing the 13 to 0 because, again, majority opinion remains the same. This is an important property for it allows the judge whose primary wish is for the jury decision to assign the grade he believes is merited to forget about strategizing and concentrate on making the correct evaluation.

A judge, however, may have a different agenda in mind: to assign his grades so as to realize the majority-ranking – the final order-of-finish – that he believes in rather than to determine the "correct" majority-grades – the final grades of the wines. One would wish for a method that is *strategy-proof-in-ranking*, meaning that each judge's optimal strategy when she has the final rankings in mind is again to give her honest evaluations of each wine. Regrettably, Theorem: *No method of amalgamation is strategy-proof-in-ranking* [1].

Since perfection cannot be achieved one must accept the best possible. Suppose the final-order placed *A* above *B* and some judge preferred *B* to *A*, so wished to reverse that ranking. That judge would be tempted to increase *A*'s grade and decrease *B*'s. With majority judgment if the judge is able to increase *A*'s grade she cannot decrease *B*'s, and if she can decrease *B*'s she cannot increase *A*'s – the method is *partially strategy-proof-in-ranking*. In fact, Theorem: *Majority judgment is the only method that is partially strategy-proof-in-ranking* [1].

To see what all of this means in practice consider, first, the judge Dubois-Millot (who could so easily have manipulated the order-of-finish were either point-summing or Quandt's method used). Contrast the majority-ranking with Dubois-Millot's preferences,

Majority-ranking:

Dubois-Millot's preferences:

Dubois-Millot cannot raise *C** in the majority-ranking. Lowering *B**'s grade is not sufficient to place *B** below *C** but puts *B** below *A*, not his intention. He cannot lower *E* in the majority-ranking, nor can he raise *F**. He can raise *I* (by changing the 9.5 to 12) above H, but cannot raise *J* nor lower *G*. All he is able to do is move *I* up above *H*.

C. Vanneque's wishes diverge more from the majority-ranking:

Majority-ranking:

Vanneque's preferences:

He can do nothing to push *D** above *B*, A,* or *C**; nor can he push either *B*, A,* or *C** below *D**. He cannot push *A* above *B**, but he can push *B** below *A* (by lowering its grade from 16 to 12). He cannot place either *H* or *G* above *F**, nor can he place *J* above *I*. Thus Vanneque, who can easily manipulate the outcome with point-summing (putting *D** first, *A* second and *B** third), can at most change the majority-ranking to

Majority judgment is coherent. Of crucial importance, majority judgment is necessarily coherent because wines are assigned individual grades by judges instead of being compared (the root of all the evils!).

To finish the case for "why" it should be said that all of the assertions made in the above rest on formal proof. Indeed, majority judgment is characterized mathematically as the only method for amalgamating judges' opinions that is coherent and meaningful, does the best in combating manipulation, and heeds the majority opinion (in particular, cranky judges cannot counter the majority will).

4. Majority judgment in use

"To give a global grade to a wine, one should not immediately think in terms of a numerical value, but rather classify its quality; the grade, in the chosen scale, will follow automatically." So said the great œnologist Émile Peynaud in his classic treatise [11, p. 104]. Indeed, insiders say that professional judges often work backward: they first decide on a wine's quality and then they assign numerical grades to the various attributes so that their sums give the desired outcome. Grading a wine strictly on the basis of the quality of its individual characteristics may miss the point for it "has difficulty in detecting exceptional wines by overly favoring wines that are 'taste-wise correct'"[12, p. 109]. A wine that is truly outstanding in some one attribute yet has clear flaws in others may well classify as sublime, well above all the other competing wines, yet lag the others when the evaluations are based on attributes.

The previous discussion has shown that the use of numbers without specific meanings gives meaningless results; moreover, the very use of numbers suggests that they will be summed to determine final decisions (though summing is meaningless since they do not constitute interval measures), so they induce strategic behavior, the attribution of points that may not be honest. For these reasons, a set of grades such as the seven given by the U.I.Œ., should be used, though six may be sufficient (in most competitions *Bad* is not used at all).

Experts of great taste and experience will integrate for themselves the importance of the various attributes, so should simply be asked to assign one of the six or seven grades. However, juries composed of judges of limited experience and tasting ability should be asked to assign one of the six or seven grades to each of the attributes. They may not be able to integrate the relative importance of the attributes, they may overlook some of them in their evaluations, and they are undoubtedly more at ease when faced with the more specific task of addressing specific qualities. This means that majority judgment must be extended to ranking wines that are given a grade for each of several attributes.

The U.I.Œ. revised its scoring sheets in 2009 (Table 7). The two lowest grades were dropped probably because they were little used. Regrettably the word descriptions were dropped except for *Excellent* and *Inadequate*, thereby emphasizing the comparative aspects rather than the absolute nature of the evaluations. However, an additional row was adjoined below the score sheet, "Eliminated due to a major defect": a wine with two such mentions cannot be awarded a medal. The former scoring sheet (Table 5) contained 14 attributes to be evaluated individually, all of approximately the same numerical weight or importance, the new one (Table 7) contains 10 attributes, but the total importance of *Visual* (= *Aspect*), *Nose* (= *Aroma*), *Taste* and *Overall* (= *Global*) is about the same as before.

	Excellent				*Inadequate*
Visual					
Limpidity (VL)	5	4	3	2	1
Aspect other than limpidity (VA)	10	8	6	4	2
Nose					
Genuineness (NG)	6	5	4	3	2
Positive intensity (NI)	8	7	6	4	2
Quality (NQ)	16	14	12	10	8
Taste					
Genuineness (TG)	6	5	4	3	2
Positive intensity (TI)	8	7	6	4	2
Harmonious persistence (TP)	8	7	6	5	4
Quality (TQ)	22	19	16	13	10
Harmony – Overall judgment (HO)	11	10	9	8	7

Table 7. U.I.Œ. "Score Sheet," still wines, 2009.
A judge circles her evaluations in each row.

We advocate restoring words as grades, eliminating numbers altogether, and using an even number of grades – for example, *Excellent, Very Good, Good, Passable, Inadequate, Bad* – so as to eliminate the possibility of "opting for the middle."

Les Citadelles du Vin

The Citadelles du Vin is an annual wine competition held in the Bordeaux area every June that is organized by the well-known œnologist Jacques Blouin. In 2006 some sixty judges organized into twelve juries of five judges classified 1,247 wines. Two methods were used to amalgamate judges' opinions, the official U.I.Œ. point-summing method and majority judgment with a single global criterion and five grades, *Excellent, Very Good, Good, Fair*, and *Mediocre*. These words appeared on the scoring sheets (*Very Good, Good*, and *Fair* replacing in that order the three 's). [2]

The responses of one jury on three white wines, A, B, and C are given in table 8. The numbers that correspond to the word grades are given in keeping with practice. Thus, for example, Judge 1 evaluated Wine *A*'s "Visual limpidity (VL)" as 5 or *Excellent* for the U.I.Œ. method, and *Very Good* (at the right) for majority judgment.

[2] We are indebted to Jacques Blouin for giving us access to the data of the results in 2006.

Attribute	VL	VA	NG	NI	NQ	TG	TI	TP	TQ	HO	Maj. Jdg. Grade
Judge	Wine *A:*										
1	5	8	7	5	14	7	5	7	19	10	*Very Good*
2	5	10	6	5	12	6	5	7	16	9	*Good*
3	5	10	8	5	14	7	5	8	16	10	*Very Good*
4	5	10	8	6	16	8	6	8	22	11	*Excellent*
5	5	8	7	5	14	6	4	6	16	9	*Fair*
	Wine *B:*										
1	5	8	7	5	14	7	5	7	16	9	*Very Good*
2	5	10	6	5	12	6	5	7	16	9	*Good*
3	5	10	7	5	14	7	4	7	16	9	*Good*
4	5	10	7	5	12	7	5	6	19	10	*Very Good*
5	5	10	7	5	12	6	4	6	13	8	*Fair*
	Wine *C:*										
1	5	10	7	5	14	7	4	7	16	10	*Very Good*
2	5	8	7	5	14	7	5	6	16	9	*Good*
3	5	10	7	4	12	7	5	7	16	10	*Good*
4	5	10	7	4	12	7	4	6	19	10	*Very Good*
5	5	8	7	5	14	7	5	7	16	10	*Good*

Table 8. A jury's grades for three white wines, les Citadelles du Vin, 2006.

Majority judgment for expert juries

We believe (as has already been said) that truly expert juries should give their global evaluations in a single scale preferably of six grades (instead of the five used here).

The *majority judgment procedure* is simple. List the grades of each from highest to lowest:

Wine *A*:	*Excellent*	*Very Good*	*Very Good*	*Good*	*Fair*
Wine *B*:	*Very Good*	*Very Good*	*Good*	*Good*	*Fair*
Wine *C*:	*Very Good*	*Very Good*	*Good*	*Good*	*Good*

The *majority-grade* is the middlemost grade. *A*'s majority-grade is *Very Good*, *B*'s and *C*'s *Good*. Thus *A* is judged the best of the three. To decide on the order between *B* and *C*, the remaining grades must decide, so drop the "first" (equal) majority-grades (see below). When there is an even number of grades the *majority-grade* is the lower middlemost grade.

Wine *B*:	*Very Good*	*Very Good*	~~*Good*~~$_1$	*Good*$_2$	*Fair*
Wine *C*:	*Very Good*	*Very Good*	~~*Good*~~$_1$	*Good*$_2$	*Good*

The second majority-grades of *B* and *C* are *Good,* again equal. So drop them:

Wine *B*:	*Very Good*	*Very Good*$_3$	~~*Good*~~$_1$	~~*Good*~~$_2$	*Fair*
Wine *C*:	*Very Good*	*Very Good*$_3$	~~*Good*~~$_1$	~~*Good*~~$_2$	*Good*

The third majority grades are again the same, *Very Good*. So repeat:

Wine *B*:	*Very Good*	~~*Very Good*~~$_3$	~~*Good*~~$_1$	~~*Good*~~$_2$	*Fair*$_4$
Wine *C*:	*Very Good*	~~*Very Good*~~$_3$	~~*Good*~~$_1$	~~*Good*~~$_2$	*Good*$_4$

B's fourth majority-grade is *Fair*, *C*'s is *Good*, so *C* is judged better than *B*, giving the majority-ranking:

This is the general procedure, though in this case it is immediately evident that *C* is judged better than *B* since they only differ in one grade.

This order happens to agree with the usual U.I.Œ. point-summing method determined by the respective wines' averages of the total points over all attributes given by the judges, *A* obtaining an average of 87.2, *B* of 82.0 and *C* of 83.6.

Multi-criteria majority judgment for ordinary juries

Evaluating several attributes independently has the advantages and disadvantages already discussed. The several – in this case10 – independent evaluations of a wine by 3 judges may be thought of as 30 independent judges evaluating the merits of different aspects of the wine. If the different aspects bear the same importance then majority judgment may be applied as though there were one global evaluation of each wine by 30 judges. We believe that when possible it is best for the attributes to be defined so that they do bear the same importance (as is suggested by the points accorded in the 2006 version of the U.I.Œ.'s scoring sheet).

In the 2009 version, however, the attributes have differing importance or weights. One approximation of the relative weights of importance of the attributes is to simply add the points used for each criterion (given in table 8).

Attribute	VL	VA	NG	NI	NQ	TG	TI	TP	TQ	HO
Weight	15	30	27	20	60	27	20	30	80	45
	Wine *A:*									
Judge 1	*Exc*	*VG*	*VG*	*VG*	*VG*	*VG*	*VG*	*VG*	*VG*	*VG*
Judge 2	*Exc*	*Exc*	*G*	*VG*	*G*	*G*	*VG*	*VG*	*G*	*G*
Judge 3	*Exc*	*Exc*	*Exc*	*VG*	*VG*	*VG*	*VG*	*Exc*	*G*	*VG*
Judge 4	*Exc*	*Exc*	*Exc*	*Exc*	*Exc*	*Exc*	*Exc*	*Exc*	*Exc*	*Exc*
Judge 5	*Exc*	*VG*	*VG*	*VG*	*VG*	*G*	*G*	*G*	*G*	*G*
	Wine *A:*									
Majority-grade	*Exc*	*Exc*	*VG*	*VG*	*VG*	*VG*	*VG*	*VG*	*G*	*VG*
	Wine *B:*									
Majority-grade	*Exc*	*Exc*	*VG*	*VG*	*G*	*VG*	*VG*	*VG*	*G*	*G*
	Wine *C:*									
Majority-grade	*Exc*	*Exc*	*VG*	*VG*	*VG*	*VG*	*VG*	*VG*	*G*	*VG*

Table 9. A jury's grades for three white wines, les Citadelles du Vin, 2006.

The *multi-criteria majority judgment procedure* replicates the grades of each criterion according to its weight and applies majority judgment to this extended set of grades. The data is exactly the same but in table 9 the word grades are inserted rather than their corresponding points.

The procedure is explained for wine *A*. Every grade assigned to the three wines is either *Excellent*, *Very Good*, *Good* or *Fair*. Letting the 4-tuples below be the numbers of *Excellent*, *Very Good*, *Good* and *Fair* attached to each attribute, the assignment of weighted grades is

VL (75,0,0,0)	VA (90,60,0,0)	NG (54,54,27,0)	NI (20,80,0,0)	NQ (20,180,60,0)
TG (27,54,54,0)	TI (20,60,20,0)	TP (60,60,30,0)	TQ (80,80,240,0)	HO (45,90,90,0)

This data allows the majority-grades of each attribute of the wines to be calculated as well, which is useful information (they are given in table 9). Note that the weighted majority-grade of a wine's attribute is necessarily the same as the majority-grade (without replication) of that wine's attribute.

The total count of grades for wine *A* is (531,778,521,0) or 531 *Excellent*, 778 *Very Good*, 521 *Good* and 0 *Fair*. *B*'s total count is (195,714,736,125) and *C*'s (165,980,625,0).

It would be laborious, to say the least, to write down in order 1,830 grades. There is a simpler procedure. An absolute majority is 916 or more. Therefore, *A*'s *majority-grade* is *Very Good* since there is less than a majority for *Excellent* and an absolute majority for at least *Very Good*.

A wine's *majority-gauge* is a triplet (*p*, majority-grade , *q*), where *p* is the number of grades above the majority-grade and *q* is the number of grades below the majority-grade. A " " is adjoined if there are more grades above the majority-grade than there are grades below it (if), and a "–" otherwise. Thus the three wines' majority-gauges are

A : (531, *Very Good* 521) *B* : (195, *Very Good* – , 861) *C* : (165, *Very Good* – 625)

Each of the wines is *Very Good*. The following rule determines the *majority-ranking*:

- When wine *X*'s majority-grade is above *Y*'s, ;
- When they have the same majority-grade, *X*'s with a and *Y*'s with a – , ;
- When they have the same majority-grade and both are or both are – , the biggest of the four associated numbers of grades decide: if it is one of *X*'s and it is the number of grades above the majority-grade, , whereas if it is one of *X*'s and it is the number of grades below the majority-grade, .

This rule gives exactly the majority-ranking obtained by the laborious procedure unless it produces a tie in the ranking (unlikely when there are many grades). If this occurs and there is a need to resolve the tie the laborious procedure may be used (which guarantees there can be no tie unless both wines have exactly the same numbers of each grade).

In conclusion

The foremost reasons why majority judgment should be used to amalgamate the opinions of judges to make jury decisions – to reach a jury consensus on the value of each wine and on how they should be ranked – have been presented. A much more complete and detailed theoretical argument together with proofs is presented in the book [1], where experimental evidence in diverse uses including the award of prizes and elections is described as well.

References

[1] M. Balinski and R. Laraki, *Majority Judgment: Measuring, Ranking and Electing*. Cambridge, MA: MIT Press, 2010.

[2] D. Black, *The Theory of Committees and Elections*. Cambridge, UK: Cambridge University Press, 1958.

[3] J.-C. le Chevalier de Borda, "Mémoire sur les elections au scrutin." *Histoire de l'Académie royale des sciences : Année* 1781, 657-665.

[4] Le Marquis de Condorcet, *Essai sur l'application de l'analyse à la probabilité des decisions rendues à la pluralité des voix*. Paris: l'Imprimerie royale, 1785.

[5] A. H. Copeland, "A 'reasonable' social welfare function." Seminar on Applications of Mathematics to the Social Sciences, University of Michigan, Ann Arbor, 1951.

[6] P. Dasgupta and E. Maskin, "The fairest vote of all." *Scientific American* 290, March 2004, 92-97.

[7] Sir Francis Galton, "One vote, one value." *Nature* 75 (February 28, 1907) 414.

[8] G. Hägele and F. Pukelsheim, "Llull's writings on electoral systems." *Studia Lulliana* 41 (2001) 3-38.

[9] G. Hägele and F. Pukelsheim, "The electoral systems of Nicolas of Cusa in the Catholic Concordance and beyond." In *The Church, the Councils and Reform: Lessons from the Fifteenth Century*, ed. By G. Christianson et al, pp. 229-249. Washington, DC: Catholic University of America Press, 2008.

[10] D.H. Krantz, R.D. Luce, P. Suppes, and A. Tversky, *Foundations of Measurement*. Vol. 1. New York: Academic Press, 1971.

[11] Émile Peynaud and Jacques Blouin, *Le gout du vin : le grand livre de la degustation*. Paris: Dunod, 2006.

[12] Émile Peynaud and Jacques Blouin, *Découvrir le gout du vin*. Paris: Dunod, 1999.

[13] Pliny the Elder (Gaius Plinius Secundus), *The Natural History*, Book 14, Chapter 8. http://www.perseus.tufts.edu/hopper/text?doc=Plin.+Nat.+toc&redirect=true

[14] R.E. Quandt, "Measuring and inference in wine tasting." *Journal of Wine Economics* 1 (2006) 7-30.

[15] S.S. Stevens, "On the theory of scales of measurement." *Science* 103 (1946) 677-680.

[16] Wilderness Emergency Medical Services Institute. "Mankoski pain scale." http://wemsi.org/painscale.html