ECARES

Wine Ratings

Olivier Gergaud
KEDGE Business School

Victor Ginsburgh
ECARES, Université libre de Bruxelles

Juan D. Moreno-Ternero
Universidad Pablo de Olavide

September 2020

ECARES working paper 2020-38

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Wine Ratings*

Olivier Gergaud[†]

Victor Ginsburgh[‡]

Juan D. Moreno-Ternero[§]

September 7, 2020

Abstract

Wine ratings are extremely important for the wine industry but, nowadays, they have flourished to the extent that they complicate wine merchants', stores' and consumers' decisions. There is, thus, an increasing need to compromise among them. This paper explores alternative ways to do so, inspired by contributions in political science, social choice, game theory and operations research. We apply our methods to rank 2018 *en-primeur* Bordeaux wines, rated by five international experts.

Keywords: Wines; Ratings; Global Wine Score; 2018 *en-primeur* Bordeaux wines; Judgement of Paris;

JEL codes: D63; L66;

*The third author acknowledges financial support from the Spanish Ministry of Economy and Competitiveness (ECO2017-83069-P).

[†]KEDGE Business School, Bordeaux, France.

[‡]ECARES, Université Libre de Bruxelles, Belgium and CORE, Université catholique de Louvain, Belgium.

[§]Department of Economics, Universidad Pablo de Olavide, Seville, Spain

1 Introduction

Wine consumption has been part of civilization for over 8,000 years, but, during most of that time, each regional wine was essentially consumed locally. It is only much later that things started to change, but imports and exports were still small compared to what they are now. In 1920, Europe (including Algeria that at the time was a French territory) accounted for 95 percent of world's wine production, but only five percent were exported to other countries (Anderson and Pinilla, 2018). This has again dramatically changed nowadays, and the boom in the wine retailing industry is partly attributed to the popularity of wine ratings, which have become essential, especially in the United States where consumers expect more guidance.

The famous *Judgment of Paris*, a competition between American and French wines, organized in Paris in 1976 by British wine merchant Steven Spurrier, invited eleven competent French wine connoisseurs, who judged two flights of ten white wines and ten red wines each. In this paper, we shall focus on red wines only, but the same story could be told for white wines. Among the ten red wines, four came from France and six from the United States. The judges had to taste and rate each of them on a scale between 0 and 20. The rates were simply added, and a ranking, based on the ratings of the eleven experts was computed. A Californian wine (Stag' s Leap Wine Cellars), was ranked first at a time in which, according to Taber (2005, p. 17) "France ruled the world." This was obviously a blow against the French supremacy, and helped introducing American wines to the worldwide market.

Two years later, Robert Parker launched the *Wine Advocate*. He popularized the $[50-100]$-point rating system, which is now widely used in the wine world.[1] As of today, one can hardly find wine reviews without numerical ratings. This proliferation of ratings entails a new problem, as wine stores and consumers are now faced with (many) different assessments of wines. We try to address this problem here.

Our starting point is a group of (possibly, international) experts who assign (numerical) ratings to the same set of wines. The objective is to produce a *consensual* rating (and ranking). The inputs consist of a table the rows of which are experts ($i = 1, 2, ..., I$) and the columns are wines ($j = 1, 2, ..., J$). Element a_{ij} of this table represents the rating of wine j by expert i. The output will be a column that represents a consensual rating for each wine, using various

[1] *The Wine Spectator* and *The Wine Enthusiast* followed. British wine expert, Jancis Robinson still rates on a $[0-20]$-scale.

ways to compromise among individual ratings. This seemingly obvious problem encompasses, nevertheless, a variety of difficult issues that need to be addressed.

One issue is that each expert comes with her cultural and personal tastes. French and Japanese experts do not necessarily have the same tastes and rating habits, though they may have to rate the same wines. Some judges may be strict, do not give high rates, and use a smaller interval between the best and the worst wine, while others are more generous. Some would start with high ratings, others with low ones. They may also choose more or less distance between their rates. In the *Judgment of Paris*, Clos du Val Winery was rated 16.5/20 by Christian Vanneque, 5 by Aubert de Villaine and 2 by Odette Kahn. Pierre Bréjoux and Claude Dubois-Millo thought that the Château Mouton-Rothschild offered in the flight was worth 16, Pierre Tari rated it 11 only (see Table 1).

Some twenty-five years later, Ashenfelter and Quandt (1999, p. 170) suggested that "converting rates to ranks guarantees that each judge has the same influence on the outcome," as otherwise very generous or very stingy judges exercise more influence than others.[2] This simple remark would have changed the *Judgment*'s story. Ashenfelter and Quandt show that adding rankings, based on the ratings of each expert, produces another outcome which would have led to a tie between American Stag's Leap Wine Cellars and French Château Montrose (See Table 1, two last rows). This may have prevented the outcome that "changed the world" of wines. This prompted us to construct alternative methods that may lead to a larger consensus. All of them are subject to a common prior stage, which addresses some of the above issues raised by Ashenfelter and Quandt's argument: Instead of using absolute ratings, we suggest comparing relative (that is, quantile-based) ratings.

The paper is organized as follows. In Section 2, we formalize the three stages implied by our procedure. Section 3 illustrates these methods using the data from the *Judgement of Paris*. Section 4 turns to applying them to a set of 114 en primeur (early) Bordeaux wines rated in 2018, by five international experts. Section 5 concludes.

[2]Note that ranking makes the distance between any two wines equal to one, which reduces the expert's freedom to leave more 'space' between them.

2 The model

We consider a set of judges or experts who rate wines.[3] Formally, there exists a table A, the rows of which are experts ($i = 1, 2, ..., I$) and the columns are wines ($j = 1, 2, ..., J$). Element a_{ij} thus represents the rating of wine j by expert i. Our objective is to summarize the information from matrix A into a unique row containing the consensual rating for each wine. An obvious one, which we call the *usual rating consensus*, would simply take the average rating of all experts. That is, for each $j = 1, 2, \ldots, J$,

$$U(j) = \frac{1}{I} \sum_{i=1}^{I} a_{ij}. \tag{1}$$

We try to improve this (somewhat naive, but widely used) method in several ways. To do so, we introduce a procedure that involves three stages (subsections 2.1 to 2.3). Subsection 2.4 illustrates the computations using a very simple example. Section 2.5 provides some sources that helped us to derive our method.

2.1 Normalization stage

Following *Global Wine Score*,[4] original ratings are first normalized so that the ranges of ratings are equalized across experts, and ratings themselves are homogenized. To do so, we consider the Cumulative Distribution Function (CDF) of each expert's ratings. This boils down to computing a certain number of quantiles, that is, proportions of wines rated above a given level, by each expert. If, say, expert 1 rated a wine 95, whereas expert 2 rated it 90, but both considered that 20% of the wines they rated themselves separately were better than this one (and 80% were worse), then the normalized rating for this wine will be 80 (for both experts). If, instead, expert 1 considered that 5% of the wines he rated were better than this one (and 95% were worse), whereas expert 2 considered 10% and 90%, respectively, then the normalized ratings for this wine will be 95, and 90 (as originally).

This first stage thus converts the original table A into the associated *normalized ratings table*, A^n, where each entry a^n_{ij} reflects the quantile associated to a_{ij} in the CDF of expert i.

After this normalization, *Global Wine Score* simply suggests averaging the normalized rat-

[3] We thus assume that each expert in the sample rates all the wines considered.

[4] See https://www.globalwinescore.com, last consulted July 2020.

ings that each wine receives from all experts. That is, for each $j = 1, 2, \ldots J$,

$$N(j) = \frac{1}{I} \sum_{i=1}^{I} a_{ij}^n. \tag{2}$$

We take normalization on board, but replace the *Global Wine Score* averaging by two additional stages, described next: the *approval stage* and the *aggregation stage*.

2.2 Approval stage

This second stage, determines a quantile that can be interpreted as a *threshold* that wines have to meet in order to be approved.[5] The use of thresholds is widespread in many instances of real life. It is, for example, a common practice in education, where students pass a test or a course only if they reach a certain threshold. In the case of wines, the term *Parker effect* was coined to claim that a rating of 90 points or less in Parker's rating system can cause a tipping point for buyers.[6]

This stage converts the table of (normalized) ratings A^n into a new table of approved ratings in which the rating of a wine below the threshold (quantile π) is replaced by zero. For wines passing the threshold, we consider two options:

(a) Either approved ratings are replaced by a constant (1 without loss of generality); this leads to a table $A^{1\pi}$, in which $a_{ij}^{1\pi} = 1$ if $a_{ij}^n \geq \pi$ and 0 otherwise,
(b) or ratings of approved wines are stored in a table A^π such that $a_{ij}^\pi = a_{ij}^n$ if $a_{ij}^n \geq \pi$ and 0 otherwise.

The difference is that in the first option, one does not use the ratings given by experts, since there are only ones and zeroes, while in the second case, exact ratings contribute also to the consensus.

Note that experts are not involved in this approval stage, though they rated the wines.

Which option should one choose? On the one hand, it is frequently argued that dichotomous decisions (1 or 0) to signal whether a wine is approved or not are much easier to make and may

[5] Obviously, if the threshold is the lowest possible one, all wines would be approved, and this stage would be irrelevant.

[6] This effect is also called the '89-point curse', which means that a rating below 90 causes sluggish sales. See https://www.toptal.com/finance/market-sizing/wine-industry, consulted last in July 2020.

be sufficient, while ratings are much more difficult. What is really the subtle difference between 19/20, 18.5/20 and 18/20? This would then support (a) though the method may end up with many ties in the aggregation stage that follows. On the other hand, it might seem unfair not to distinguish among (approved) wines with very different ratings, once they are available, which endorses (b).

2.3 Aggregation stage

The last stage of our procedure aggregates the information from tables $A^{1\pi}$ or A^{π}, which both construct approval ratings. A first option is to simply compute averages across wines from one or both tables, thus mimicking what we did for tables A and A^n. We shall refer to them as the *Approval* and *Proportional Approval* consensus, respectively, which implies that for each $j = 1, 2, \ldots J$, we compute:

$$A(j) = \frac{1}{I} \sum_{i=1}^{I} a_{ij}^{1\pi}, \text{ and} \tag{3}$$

$$PA(j) = \frac{1}{I} \sum_{i=1}^{I} a_{ij}^{\pi}. \tag{4}$$

The second option is to introduce a new normalization in which, before aggregating, ratings are considered *relative*, rather than absolute: This means that in tables $A^{1\pi}$ and A^{π}, we divide each entry by the aggregate amount in the corresponding row (which represents the sum of the ratings each judge confers to all wines), and then we aggregate. We shall refer to them as the *Relative Approval* and *Relative Proportional Approval* consensus, respectively. Formally, for each wine $j = 1, 2, \ldots J$,

$$RA(j) = \sum_{i=1}^{I} \frac{a_{ij}^{1\pi}}{\sum_{k=1}^{J} a_{ik}^{1\pi}}, \text{ and} \tag{5}$$

$$RPA(j) = \sum_{i=1}^{I} \frac{a_{ij}^{\pi}}{\sum_{k=1}^{J} a_{ik}^{\pi}}. \tag{6}$$

Again, here the choice between both methods results from a value judgement. It boils down to the critical issue of deciding whether the ratings given by an expert should be restricted to the wines that passed the approval stage or not.

2.4 An example of the computations

Consider the following example, in which five judges rate three wines. Assume that the resulting tables are:

$$A = \begin{pmatrix} 100 & 95 & 95 \\ 95 & 100 & 100 \\ 96 & 97 & 98 \\ 96 & 99 & 99 \\ 99 & 100 & 99 \end{pmatrix} \qquad A^n = \begin{pmatrix} 96 & 90 & 90 \\ 90 & 96 & 96 \\ 96 & 97 & 98 \\ 96 & 99 & 99 \\ 94 & 96 & 94 \end{pmatrix}$$

Protocols (1) and (2) discussed above yield the following ratings from which one can also compute rankings:

- Usual protocol (1):

  $U(1) = 1/5(100 + 95 + 96 + 96 + 99) = 97.2$

  $U(2) = 1/5(95 + 100 + 97 + 99 + 100) = 98.2$

  $U(3) = 1/5(95 + 100 + 98 + 99 + 99) = 98.2$

- Normalized protocol (2):

  $N(1) = 1/5(96 + 90 + 96 + 96 + 94) = 94.4$

  $N(2) = 1/5(90 + 96 + 97 + 99 + 96) = 95.6$

  $N(3) = 1/5(90 + 96 + 98 + 99 + 94) = 95.4$

Suppose now that the threshold is set at $\pi = 95$. Tables $A^{1\pi}$ and A^{π} easily follow from table A^n:

$$A^{1\pi} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \qquad A^{\pi} = \begin{pmatrix} 96 & 0 & 0 \\ 0 & 96 & 96 \\ 96 & 97 & 98 \\ 96 & 99 & 99 \\ 0 & 96 & 0 \end{pmatrix}.$$

The analogue to the previous two protocols, but using tables $A^{1\pi}$ and A^{π} instead of tables A and A^n are:

- Approval protocol (3):

  $A(1) = 1/5(1 + 0 + 1 + 1 + 0) = 0.6$

  $A(2) = 1/5(0 + 1 + 1 + 1 + 1) = 0.8$

  $A(3) = 1/5(0 + 1 + 1 + 1 + 0) = 0.6$

- Proportional Approval protocol (4):

  $PA(1) = 1/5(96 + 0 + 96 + 96 + 0) = 57.6$

  $PA(2) = 1/5(0 + 96 + 97 + 99 + 96) = 77.6$

  $PA(3) = 1/5(0 + 96 + 98 + 99 + 0) = 58.6$

The the last two protocols in which tables $A^{1\pi}$ and A^{π} are normalized further, so that each entry is divided by the overall entry in the corresponding row are:

- Relative Approval protocol (5):

  $RA(1) = 1 + 0 + 1/3 + 1/3 + 0 = 10/6$

  $RA(2) = 0 + 1/2 + 1/3 + 1/3 + 1 = 13/6$

  $RA(3) = 0 + 1/2 + 1/3 + 1/3 + 0 = 7/6$

- Relative Proportional Approval protocol (6):

  $RPA(1) = 1 + 0 + (96/291) + (96/294) + 0 = 1.656$

  $RPA(2) = 0 + 1/2 + (97/291) + (99/294) + 1 = 2.318$

  $RPA(3) = 0 + 1/2 + (98/291) + (99/294) + 0 = 1.174$

2.5 The inspiration from other fields

The normalization procedure discussed above had already been used in diverse areas such as the design of equal-opportunity policies (Roemer, 1998; Moreno-Ternero, 2007), the economic evaluation of health care programs (Bleichrodt et al., 2002; Herrero and Moreno-Ternero, 2009) or the evaluation of scientific performance (Albarrán et al., 2010, 2011).

The other stages are also inspired by classical contributions in different fields, such as political science, game theory, social choice, and operations research. For instance, more than half a century ago, political scientists suggested using *Cumulative Voting* (Glasser, 1959; Sawyer and MacRae, 1962), which allows voters to distribute points among candidates in any arbitrary

way. An interesting case is the one in which every agent is endowed with a fixed number of votes that are evenly divided among all candidates for whom she votes. *Approval Voting* (Brams and Fishburn, 1978) is another voting method in modern social choice theory. It allows each voter to cast her vote for as many candidates she wishes; each positive vote is counted in favour of the candidate. The votes are then added candidate by candidate, and the winner is the one who gets the largest number of votes.[7]

The second intuition is derived from the game-theoretical concept, called the *Shapley value* (Shapley, 1953). The Shapley value yields a natural way to allocate the total surplus generated by the coalition of all players involved in a joint venture, or a cooperative game, based on the marginal contributions players produce. This led Ginsburgh and Zang (2003) to study the problem of sharing the total revenue collected from selling museum passes, which give access to several museums.[8] Later on, Ginsburgh and Zang (2012) precisely applied the same theoretical model to aggregate ratings from the *Judgment of Paris*.[9] In their case, the protocol is different from ours because, instead of a threshold, they fixed (exogenously) a number of wines for each judge.

Finally, there is a connection with the classical *knapsack problem* in operations research (Martello and Toth, 1990). It refers to a camper who would like to carry objects which have different utilities and different weights. His optimal choice results from maximizing his total utility under the constraint that he can carry only a given total weight. This integer problem has many applications. In participatory budgeting, for example, it can be formulated as follows. A (local) government asks residents to vote on proposals for how a certain fraction of their total budget should be spent (Cabannes, 2004). Each voter can specify a subset of these projects, such that the total cost in the subset is bounded by the total budget (regardless of how many projects are in the subset). Thus, each voter has to solve an individual knapsack problem. An advantage of *knapsack voting* is that, if the algorithm rates each project by the number of votes it receives, and chooses projects greedily in descending order of rating until the budget is exhausted, then knapsack voting is a partially truthful mechanism (e.g., Goel et al., 2019).

[7]Both Approval Voting and Cumulative Voting can be seen as members of a family of voting procedures called *Size Approval Voting*, which are characterized by Alcalde-Unzu and Vorsatz (2009).

[8]See also Bergantiños and Moreno-Ternero (2015).

[9]Similar ideas are explored too by Ginsburgh et al. (2017) to rank languages.

3 The Judgement of Paris as illustration

The *Judgement of Paris* alluded to in the introduction, invited 11 well-known French experts to test ten red wines. The detailed results of the contest are shown in Table 1. Each wine appears in a column and each expert in a row. The flight consists of four French wines (denoted by F), and six American wines (US). The results of the tasting are reported in the last rows 'average rating' of each wine, and 'final ranks' based on the average ratings. This is the way in which wine tastings usually end the performance. We added one row (the very last one, in bold characters, which is the ranking that one obtains by transforming the ratings of each expert into ranks and compute average ranks, as suggested by Ashenfelter and Quandt (1999). The result is quite different: (a) American wine A ties with French wine C; (b) the ranks of wines F and G change and (c) so do wines H, K and J.[10]

Insert Table 1 about here

We now turn to the alternative procedures considered in this paper. The number of wines tasted in this context is too small to normalize ratings. We thus go immediately to stage 2, and (arbitrarily) set the threshold at 10.[11]

Table 2 contains all the rankings described in the paper. The rankings are pretty close. Indeed, wines A, B, C appear among the first three in all cases, sometimes they are tied, other times they are not. Wine D is almost always number 4, except that it is tied once with wine F. Wines H, I, J appear always as being the last ones. The differences mostly occur in the middle where the rankings switch between wines.

Insert Table 2 about here

To conclude, changing the threshold in the previous analysis leads to some minimal changes (Table 3). For instance, if the threshold moves up to 15, the tie among wines A, B, and C (which occurred before for the third and fifth protocol) breaks. Wine C (French) actually goes down to fourth place, whereas wine D (French) comes third. As for wines A (American)

[10] See Taber (2005) for details.

[11] Ginsburgh and Zang (2012) considered related alternatives for the *Judgment of Paris*. In the first, they ran three simulations assuming that each judge would have chosen a unique wine, or two wines, or three wines. Next, they concentrated on the number of wines chosen by each judge at random. Finally, they selected, for each judge, the wines that were rated before a gap of two points occurred in his ratings.

and B (French), each one comes first with one of the two protocols. As for protocol (6), it would also have Wines F (French), I (American) and J (American) tied at the bottom (with no votes), whereas the rest of the wines would be ranked (with no ties) as follows: B (French), A (American), D (French), C (French), E (American), G (American), H (French).

Insert Table 3 about here

4 An Analysis of ratings for Bordeaux 2018 future wines

We now consider the more interesting and recent tasting of 2018 Red Bordeaux called *Bordoverview*.[12] Bordoverview contains ratings for 114 (Bordeaux 2018 future) wines produced by five (international) experts: Jancis Robinson, Tim Atkin, *Revue du Vin de France* (RVF), *Decanter Magazine*, and Parkers' *Wine Advocate* (WA).[13] Appendix 2 provides the list of 114 wines by alphabetic order (first column). The next five columns contain the ratings given by each expert, while the five last columns contain the normalized rates (See Section 2.1).[14]

A first aspect to notice is the striking differences that exist among some of these experts' ratings, which indicates how important the normalization stage is. More precisely, we can observe from Figure 1 how different their rating distributions are.

Insert Figure 1 about here

Table 4 shows the ratings for the six protocols introduced above, when the (normalized) threshold is set at 90.[15] This threshold eliminates 76 of the 114 wines, and only 38 pass the

[12] See https://www.bordoverview.com/?RP=98.1

[13] See Appendix 1 for some details on the five experts.

[14] Some caveats are in order regarding the ratings associated to the WA. Sometimes their ratings are not specific but rather intervals representing an estimated rating range (e.g., $(90-93)$). In those cases, we considered the midpoint of the interval as the specific rating for that wine in our analysis. The WA also includes sometimes a plus sign following a rating (e.g., 95+), indicating "a wine that the reviewer believes has the potential to improve over a period of time in bottle and may warrant a higher score in a subsequent/future tasting." In those cases, we gave to the wine an extra of 0.5 points. Additional relevant information regarding WA is that Robert Parker announced in 2015 that he would no longer rate en primeur wines, which created some uncertainty as some had purposely designed their production with his palate and preferences in mind. Neal Martin was named by Parker to be his successor in this task, although it has been argued that his ratings are not the ones showing the highest correlation with those of Parker (Cyr et al., 2019).

[15] To ease comparisons, we multiply by 100 the ratings from the last four protocols.

test and are approved, while the remaining 76 wines are given a zero rating (with the exception for the first two protocols, which do not consider the approval stage). Table 4 also contains the associated rankings for each of the six protocols and leads us to some interesting conclusions.

Insert Table 4 about here

First, the top of the rankings seems to be quite robust. Lafite-Rothschild comes first in all cases (although tied with eight other wines in the third and fifth protocol, which, as mentioned above, may generate many ties). Léoville-Las-Cases seems to be a solid second follower, with the exception of the first protocol, where it appears as number three (and the caveat for the ties in the third and fifth protocol). The first protocol actually awards the second position to Palmer, which comes down to the eighth place after normalizing. This is probably the first interesting difference among protocols. Another interesting case, in the opposite direction, is Ausone, which is ranked eighth in the first protocol, but goes up to the fifth position for protocols after normalization.[16] Other specific and somewhat striking differences between the second, fourth and sixth protocols occur, for instance, for le Pin, ranked number 24 in the second protocol, 11 in the fourth and 13 in the third. And, according to the fourth and sixth protocols, Margaux and Vieux Château Certan switch positions.

As mentioned above, the third and fifth protocols introduce many ties, since the rates are dichotomous (1 for approved wines, 0 for all other). One may argue that rating wines is a difficult business and that decisions should be simplified and simply be dichotomous: one likes a wine or not. If one accepts this position, only the third and fifth protocols would be valid. Breaking the ties requires distinguishing among approved wines, as the remaining protocols do.

Note however that if one sets a higher threshold for letting wines pass, then less wines will be approved, and some ties vanish. To illustrate this aspect, if the threshold is set at 95 (instead of 90), then only two wines (instead of eight) tie for the first place: Lafite-Rothschild and Léoville-Las-Cases. They also happen to be the top two wines in other protocols, with the exception of the first (not-normalized) protocol, in which Palmer comes second and Léoville-Las-Cases third.

To have a more general view on the differences in the protocols (with the exception of the third and fifth, which, as we already said, have too many ties) we calculated Pearson correlations

[16]Both Ausone and Palmer achieve the first position, tied with other eight wines, for the third and fifth protocols.

on ratings as well as Spearman rank correlation coefficients on rankings.

Insert Table 5 about here

It is remarkable to see that the correlations between Protocol 1 (the usual average ratings of all five experts) and Protocols 2, 4 and 6 are all quite large (between 0.89 and 0.97). It is also noteworthy to see that Protocol 2 (normalization) is not very different from Protocol 1 (0.93 and 0.97) which implies that normalizing does not change much. In addition, Protocols 1 and 6 (the most sophisticated protocol) are very close to each other. Still, these observations do not mean that these small changes are innocuous. There is indeed a large difference in the perception of which wine is first -note that, against all odds, Lafite Rothschild is always number 1 or 2-, but Château Ausone, ranked 8 in the usual protocol, may strongly benefit, getting from rank 8 to rank 4, in all other protocols. It also makes a very large difference being among the top ten and the top twenty. This is also so for prices: a small difference in ranking may have large effects on the decisions made by many buyers, and thus on prices. This convex relationship between talent, or quality, essentially driven by experts, and prices is illustrated by Ali et al. (2008).

In short, changes of ranks (even from 1 to 2, or 5 to 6) are extremely significant, in this profession as well as among wine experts. The Judgment of Paris that we discussed earlier is a good example of this very unusual behavior. American wines gained a lot of prestige due to the fact that one of their wines, Stag's Leap Wine Cellars, overshadowed three very famous French wines: Mouton-Rothschild, Montrose and Haut Brion.

To conclude, Table 6 illustrates the results of the ratings for the sixth protocol (which we find the most interesting) with thresholds ranging from 90 to 95. They show that there is consensus at the top of the ranking. Lafite-Rothschild and Léoville-Las-Cases come always first and second. Margaux is third in all cases except when the threshold climbs to 95, where it goes down to the sixth place. Vieux Château Certan goes in the other direction, climbing to the third position at the 95 threshold, whereas it ranks between fifth and tenth in the other protocols.

Insert Table 6 about here

5 Discussion

Numerical wine ratings are extremely popular nowadays. The scientific literature has paid attention to aspects such as (i) inconsistency of ratings in blind tastings (Hodgson, 2008, Bodington, 2017), (ii) consensus among experienced wine experts (Ashton, 2012, 2013), (iii) variations in the severity of experts (Masset at al., 2015, Stuen et al., 2015) and (iv) consumers' demand for wine ratings (e.g., Ashenfelter and Jones, 2013, Marks, 2015).

As experts sometimes disagree very strongly, we tried to find a compromise between ratings. We take on board the well-known normalization procedure, which *Global Wine Score* was the first to use in this setting. The second stage (approval) refers to putting thresholds on normalized ratings below which wines get no credit. The third one (aggregation) shares the credit from each expert among approved wines, equally or proportionally. The last two stages are inspired from classical contributions in political science, social choice, game theory and operations research.

We illustrated our protocols using the famous *Judgement of Paris* and 2018 *en-primeur* Bordeaux wines, rated by five international experts. Our analysis concludes that the way in which we build the consensuses is quite different from the usual simple aggregations of ratings or rankings. We nevertheless believe that it is important to notice that different (plausible) decisions to build a consensual ranking generate different outcomes.

All our protocols have pros and cons, some of which are described in our paper. We argued that normalizing ratings is important, which makes the first (standard) protocol unreliable. We also believe that a simple average of (normalized) ratings is unsatisfactory. The remaining four protocols, obtained by implementing both the approval and aggregation stages, teach us something. If one is ready to simplify ratings, converting them to dichotomous choices, then the third and fifth protocols are preferable, because they make experts' choices much easier: I like this wine or I do not not, instead of haggling with myself between 14, 14.5-, 14.5, 14.5+ and 15, as suggested by Ginsburgh and Zang (2012).

But as more precise numerical ratings are available (0 to 20, or 50 to 100) , we cannot ignore them just distinguishing approved and non-approved wines, and directly go for the fourth or the sixth protocol. We nevertheless believe that the sixth protocol is more satisfactory as it also involves a second normalization in which the ratings of each expert are taken into account: When an expert approves a set of wines, her credit is split proportionally to her ratings, among

those wines. This is in line with many other problems in real life where the proceeds from a joint venture are only allocated among those contributing to the venture (e.g., Bergantiños and Moreno-Ternero, 2015, 2020). In summary, we endorse protocol 6 as the most appropriate one.

A final comment is in order. In all the variants of the procedure we suggested, experts' opinions were equally weighted. But it happens that certain experts are considered to be more influential than others. Our procedures could easily be extended to account for unequal weights of experts.

References

[1] Albarrán, P., Crespo, J., Ortuno-Ortin, I., Ruiz-Castillo, J., (2010). A comparison of the scientific performance of the US and the European Union at the turn of the 21st century. Scientometrics 85, 329-344.

[2] Albarrán, P., Crespo, J., Ortuno-Ortin, I., Ruiz-Castillo, J., (2011). The skewness of science in 219 sub-fields and a number of aggregates. Scientometrics 88, 385-397.

[3] Alcalde-Unzu, J., Vorsatz, M., (2009). Size approval voting. Journal of Economic Theory 144, 1187-1210.

[4] Ali, H., Lecocq, S., Visser, M., (2008), The impact of gurus: Parker grades and en primeur wine prices. The Economic Journal 118, F158-F173. (Reprinted in Journal of Wine Economics (2010), 5, 22-39.)

[5] Anderson, K., Pinilla V., (eds), Wine Globalization: A New Comparative History. New York: Cambridge University Press.

[6] Ashenfelter, O., Jones G., (2013). The demand for expert opinion: Bordeaux wine. Journal of Wine Economics 8, 285-293.

[7] Ashenfelter, O., Quandt, R., (1999). Analyzing a wine tasting statistically. Chance 12, 16-20.

[8] Ashton, R.H., (2012). Reliability and consensus of experienced wine judges: Expertise, within, and between? Journal of Wine Economics 7, 70?87

[9] Ashton R.H., (2013). Is there consensus among wine quality ratings of prominent critics? An empirical analysis of red Bordeaux, 2004-2010. Journal of Wine Economics 8, 225-234.

[10] Bergantiños, G., Moreno-Ternero, J., (2015). The axiomatic approach to the problem of sharing the revenue from museum passes. Games and Economic Behavior 89, 78-92.

[11] Bergantiños, G., Moreno-Ternero, J., (2020). Sharing the revenues from broadcasting sport events. Management Science 66, 2417-2431.

[12] Bleichrodt, B., Herrero, C., Pinto, J. (2002). A proposal to solve the comparability problem in cost-utility analysis Journal of Health Economics 21, 397-403.

[13] Bodington, J., (2017). The Distribution of Ratings Assigned to Blind Replicates. Journal of Wine Economics 12, 363-369.

[14] Brams, S., Fishburn, P., (1978). Approval voting. American Political Science Review 72, 831-847.

[15] Cabannes, Y., (2004). Participatory budgeting: a significant contribution to participatory democracy. Environment and Urbanization 16, 27-46.

[16] Cyr, D., Kwong, L., Sun, L., (2019). Who Will Replace Parker? A Copula Function Analysis of Bordeaux En Primeur Wine Raters. Journal of Wine Economics 14, 133-144.

[17] Ginsburgh, V., Moreno-Ternero J., Weber, S., (2017). Ranking languages in the European Union: Before and after Brexit. European Economic Review 93, 139-151.

[18] Ginsburgh, V., Zang, I., (2003). The museum pass game and its value. Games and Economic Behavior 43, 322-325.

[19] Ginsburgh, V., Zang, I., (2012). Shapley ranking of wines. Journal of Wine Economics 7, 169-180.

[20] Glasser, G. (1959). Game theory and cumulative voting for corporate directors. Management Science 5, 151-156.

[21] Goel, A., Krishnaswamy, S., Sakshuwong, S., Aitamurto T., (2019). Knapsack voting for participatory budgeting. ACM Transactions on Economics and Computation 7, 1-27.

[22] Herrero, C., Moreno-Ternero, J., (2009). Estimating production costs in the economic evaluation of health care programs. Health Economics 18, 21-35.

[23] Hodgson, R., (2008). An Examination of Judge Reliability at a major U.S. Wine Competition. Journal of Wine Economics 3, 105-113.

[24] Martello, S., Toth, P., (1990). Knapsack problems. Wiley, NY.

[25] Marks, D., (2015), Seeking the veritas about the vino: Fine wine ratings as wine knowledge. Journal of Wine Research, 26, 319-335.

[26] Masset, P., Weisskopf, J., Cossutta, M., (2015). Wine Tasters, Ratings, and En Primeur Prices. Journal of Wine Economics 10, 75-107.

[27] Moreno-Ternero, J., (2007). On the design of equal-opportunity policies. Investigaciones Económicas 31, 351-374.

[28] Roemer, J., (1998). Equality of Opportunity. Harvard University Press. Cambridge, MA.

[29] Sawyer, J., MacRae, D., (1962). Game theory and cumulative voting in Illinois: 1902-1954. American Political Science Review 56, 936-946.

[30] Shapley, L., (1953). A value for n-person games. In Kuhn, A. W., and Tucker, A. W. (eds.), Contribution to the Theory of Games. Vol. II. Princeton, NJ: Princeton University Press. 307-317.

[31] Stuen, E., Miller, J., Stone, R., (2015). An Analysis of Wine Critic Consensus: A Study of Washington and California Wines. Journal of Wine Economics 10, 47-61.

[32] Taber, G., (2005). Judgment of Paris: California vs. France and the Historic 1976 Paris Tasting That Revolutionized Wine. New York: Scribner.

Table 1. The Paris Judgment: Original Ratings

Wines *Origin*	A (US)	B (F)	C (F)	D (F)	E (US)	F (F)	G (US)	H (US)	I (US)	J (US)
Judges										
Pierre Brejoux	14	16	12	17	13	10	12	14	5	7
Aubert de Villaine	15	14	16	15	9	10	7	5	12	7
Michel Dovaz	10	15	11	12	12	10	11	11	8	14
Patricia Gallagher	14	15	14	12	16	14	17	13	9	14
Odette Kahn	15	12	12	12	7	12	2	2	13	5
Claude Dubois-Millot	16	16	17	13.5	7	11	8	9	9.5	9
Raymond Olivier	14	12	14	10	12	12	10	10	14	8
Steven Spurrier	14	14	14	8	14	12	13	11	9	13
Pierre Tari	13	11	14	14	17	12	15	13	12	14
Christian Vanneque	16.5	16	11	17	15.5	8	10	16.5	3	6
Jean-Claude Vrinat	14	14	15	15	11	12	9	7	13	7
Average ratings	14.14	14.09	13.64	13.23	12.14	11.18	10.36	10.14	9.77	9.45
Final ranks	1	2	3	4	5	6	7	8	9	10
Average rankings	**1.5**	**3**	**1.5**	**4**	**5**	**7**	**6**	**10**	**8**	**9**

Wines: A: Stag's Leap Wine Cellars, 1973; B: Château Mouton-Rothschild, 1970; C: Château Montrose, 1970; D: Château Haut Brion, 1970; E: Ridge Vineyards Monte Bello, 1971; F: Château Léoville Las Cases, 1971; G:Heitz Wine Cellars 1970; H: Clos du Val Winery, 1972; I: Mayacamas Vineyards, 1971; J: Freemark Abbey Winery, 1969.

Table 2. The Paris Judgment: Comparative Rankings

Wines *Origin*	A (US)	B (F)	C (F)	D (F)	E (US)	F (F)	G (US)	H (US)	I (US)	J (US)
(1) Average ratings	1	2	3	4	5	6	7	8	9	10
(2) Average rankings	1.5	3	1.5	4	5	7	6	10	8	9
(3) Approval ratings	2	2	2	4.5	6	4.5	7.5	7.5	9	10
(4) Proportional approval ratings	1	2	3	4	6	5	8	7	9	10
(5) Relative approval ratings	2	2	2	4	6	5	7.5	7.5	9	10
(6) Relative proportional approval ratings	1	2	3	4	6	5	8	7	9	10

Wines: A: Stag's Leap Wine Cellars, 1973; B: Château Mouton-Rothschild, 1970; C: Château Montrose, 1970; D: Château Haut Brion, 1970; E: Ridge Vineyards Monte Bello, 1971; F: Château Léoville Las Cases, 1971; G:Heitz Wine Cellars 1970; H: Clos du Val Winery, 1972; I: Mayacamas Vineyards, 1971; J: Freemark Abbey Winery, 1969.

Table 3. The Paris Judgment: Comparative Rankings (threshold 15)

Wines *Origin*	A (US)	B (F)	C (F)	D (F)	E (US)	F (F)	G (US)	H (US)	I (US)	J (US)
(1) Average ratings	1	2	3	4	5	6	7	8	9	10
(2) Average rankings	1.5	3	1.5	4	5	7	6	10	8	9
(3) Approval ratings	2.5	1	4.5	2.5	4.5	9	6	7	9	9
(4) Proportional approval ratings	3	1	5	2	4	9	6	7	9	9
(5) Relative approval ratings	1	3	4	2	5	9	6	7	9	9
(6) Relative proportional approval ratings	2	1	4	3	5	9	6	7	9	9

Wines: A: Stag's Leap Wine Cellars, 1973; B: Château Mouton-Rothschild, 1970; C: Château Montrose, 1970; D: Château Haut Brion, 1970; E: Ridge Vineyards Monte Bello, 1971; F: Château Léoville Las Cases, 1971; G:Heitz Wine Cellars 1970; H: Clos du Val Winery, 1972; I: Mayacamas Vineyards, 1971; J: Freemark Abbey Winery, 1969.

Figure 1. Rating Distributions of Experts

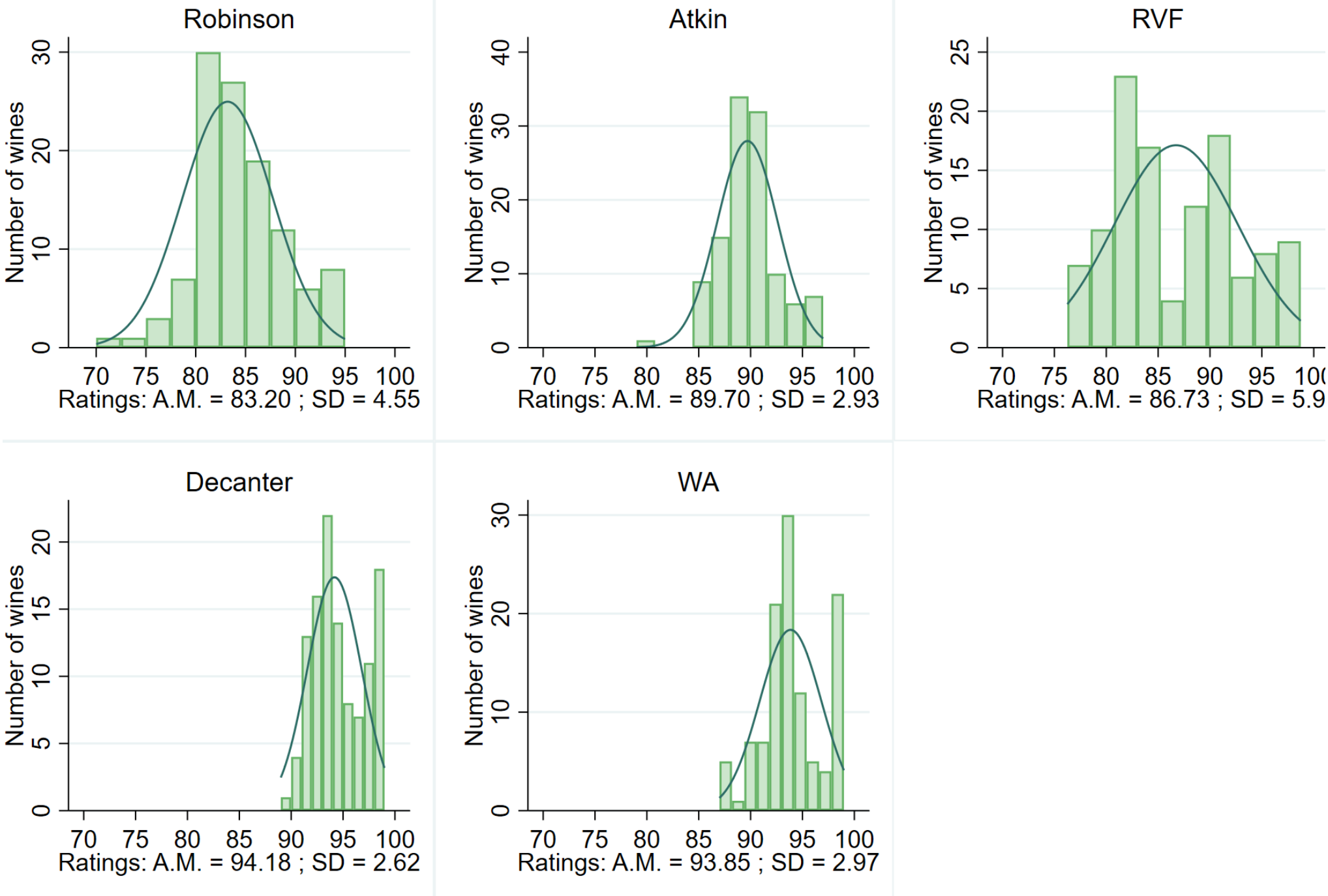

Table 4 Ratings and Rankings According to each Protocol

Wines	Protocol 1		Protocol 2		Protocol 3		Protocol 4		Protocol 5		Protocol 6	
	Rates	Ranks	Rates	Ranks	Rates	Ranks	Rates	Ranks	Rates	Ranks	Rates	Ranks
Lafite-Rothschild	97.50	1	99.11	1	100.00	1	99.11	1	23.92	1	24.83	1
Léoville-Las-Cases	96.40	3	97.98	2	100.00	1	97.98	2	23.92	1	24.55	2
Margaux	96.35	4	96.81	5	100.00	1	96.81	5	23.92	1	24.28	3
Ausone	95.55	8	96.83	4	100.00	1	96.83	4	23.92	1	24.26	4
Vieux Château Certan	96.20	5	96.92	3	100.00	1	96.92	3	23.92	1	24.25	5
Mouton-Rothschild	96.00	7	96.72	6	100.00	1	96.72	6	23.92	1	24.24	6
Lafleur	96.10	6	96.57	7	100.00	1	96.57	7	23.92	1	24.21	7
Palmer	96.45	2	96.21	8	100.00	1	96.21	8	23.92	1	24.13	8
Haut-Brion	95.15	10	94.46	10	100.00	1	94.46	9	23.92	1	23.68	9
Pichon-Longueville Baron	95.10	11	93.68	12	80.00	10	75.71	13	19.57	10	19.43	10
Cos d'Estournel	95.05	12	94.00	11	80.00	10	76.10	12	19.16	11	19.12	11
Cheval Blanc	95.35	9	94.74	9	80.00	10	77.61	10	18.66	12	18.94	12
le Pin	93.40	19	84.40	24	80.00	10	77.18	11	18.66	12	18.85	13
Figeac	93.10	21	88.96	19	80.00	10	75.59	14	18.66	12	18.43	14
Pontet-Canet	94.00	15	85.05	23	60.00	15	57.44	16	14.81	16	14.89	15
Montrose	94.40	14	91.70	15	60.00	15	56.36	19	15.03	15	14.79	16
la Mission Haut-Brion	93.90	16	91.44	16	60.00	15	56.95	18	14.57	17	14.48	17
Pichon-Longueville Comtesse	94.55	13	92.61	13	60.00	15	57.33	17	14.16	18	14.17	18
Angélus	93.65	18	92.43	14	60.00	15	57.52	15	13.89	19	13.94	19
Canon	91.70	27	81.93	28	60.00	15	54.50	20	13.66	20	12.97	20
Calon-Ségur	93.85	17	90.02	17	40.00	21	38.40	22	9.76	24	9.82	21
Rauzan-Ségla	92.95	22	86.86	22	40.00	21	37.43	23	9.81	22	9.62	22
Ducru-Beaucaillou	93.20	20	89.35	18	40.00	21	37.13	24	9.81	22	9.55	23
Pensées de Lafleur	91.25	31	71.20	34	40.00	21	36.06	25	10.03	21	9.48	24
Smith Haut Lafitte	92.95	22	88.56	20	40.00	21	38.62	21	8.89	25	8.98	25
Léoville-Barton	92.15	26	84.23	25	20.00	26	18.12	37	5.26	26	5.01	26
Lynch Bages	92.90	24	88.34	21	20.00	26	18.47	33	5.00	27	4.84	27
d'Issan	91.60	28	82.69	27	20.00	26	18.33	35	5.00	27	4.81	28
Domaine de Chevalier	91.55	29	79.93	29	20.00	26	18.70	29	4.76	29	4.65	29
Duhart-Milon	89.90	37	67.87	36	20.00	26	18.55	31	4.76	29	4.62	30
Grand-Puy-Lacoste	90.85	33	78.34	30	20.00	26	18.33	34	4.76	29	4.56	31
Beauséjour Duffau-Lagarrosse	90.85	33	69.59	35	20.00	26	20.00	26	4.35	32	4.54	32
Pavie-Macquin	91.00	32	75.98	33	20.00	26	18.89	27	4.35	32	4.29	33
les Carmes Haut-Brion	91.35	30	77.29	31	20.00	26	18.76	28	4.35	32	4.26	34
Troplong-Mondot	90.65	35	76.72	32	20.00	26	18.56	30	4.35	32	4.21	35
Clos Fourtet	89.95	36	63.21	37	20.00	26	18.50	32	4.35	32	4.20	36
Léoville-Poyferré	92.20	25	82.82	26	20.00	26	18.24	36	4.35	32	4.14	37

Table 5. Correlations between Protocols

Pearson Correlation Coefficients on Ratings

	Protocol 1	Protocol 2	Protocol 4	Protocol 6
Protocol 1	1	0,93	0,91	0,92
Protocol 2		1	0,81	0,82
Protocol 4			1	0,99
Protocol 6				1

Spearman Rank Correlation Coefficients on Rankings

	Protocol 1	Protocol 2	Protocol 4	Protocol 6
Protocol 1	1	0,97	0,89	0,94
Protocol 2		1	0,86	0,91
Protocol 4			1	0,95
Protocol 6				1

Table 6 The effects of Changing Approval thresholds on Consensus Rankings

Wines	Thresh. 90	Thresh. 91	Thresh. 92	Thresh. 93	Thresh. 94	Thresh. 95
Angélus	19	18	18	14	18	17
Ausone	4	5	5	4	4	7
Beauséjour Duffau-Lagarrosse	32	29	26	22	24	23
Calon-Ségur	21	20	20	19	17	15
Canon	20	34	*32*	*28*	*26*	*24*
les Carmes Haut-Brion	34	31	29	27	*26*	*24*
Cheval Blanc	12	11	11	10	11	10
Clos Fourtet	36	33	31	*28*	*26*	*24*
Cos d'Estournel	11	14	14	12	9	13
Domaine de Chevalier	29	24	24	23	*26*	*24*
Ducru-Beaucaillou	23	22	27	25	23	*24*
Duhart-Milon	30	26	25	*28*	*26*	*24*
Figeac	14	19	19	15	13	19
Grand-Puy-Lacoste	31	28	*32*	*28*	*26*	*24*
Haut-Brion	9	8	9	9	10	8
d'Issan	28	27	*32*	*28*	*26*	*24*
Lafite-Rothschild	1	1	1	1	1	1
Lafleur	7	7	6	6	6	4
Léoville-Barton	26	*36*	*32*	*28*	*26*	*24*
Léoville-Las-Cases	2	2	2	2	2	2
Léoville-Poyferré	37	35	*32*	*28*	*26*	*24*
Lynch Bages	27	25	23	*28*	*26*	*24*
Margaux	3	3	3	3	3	6
la Mission Haut-Brion	17	16	16	20	20	20
Montrose	16	13	13	17	21	22
Mouton-Rothschild	6	4	4	5	5	5
Palmer	8	9	7	7	7	9
Pavie-Macquin	33	30	28	26	25	*24*
Pensées de Lafleur	24	*36*	*32*	*28*	*26*	*24*
Pichon-Longueville Baron	10	10	8	16	15	14
Pichon-Longueville Comtesse	18	17	17	18	16	16
le Pin	13	12	12	11	12	11
Pontet-Canet	15	15	15	13	14	12
Rauzan-Ségla	22	21	21	24	22	21
Smith Haut Lafitte	25	23	22	21	19	18
Troplong-Mondot	35	32	30	*28*	*26*	*24*
Vieux Château Certan	5	6	10	8	8	3

Note. In most columns (wih the exception of column 1) there are ties among wines (in italics).

Appendix 1. Experts selected from Bordoverview

- Jancis Robinson is a British wine writer and critic who rose to fame in the mid-1980s after becoming the first MW (Master of Wine) outside the wine trade. She studied mathematics and philosophy at University of Oxford. She writes a weekly column for the *Financial Times*. See https://www.wine-searcher.com/critics-1-jancis+robinson or https://www.jancisrobinson.com/.

- Tim Atkin is a UK-based MW and wine journalist with an international following. After training in modern languages at the University of Durham, Atkin soon moved into a career of wine writing. See https://www.wine-searcher.com/critics-34-tim+atkin.

- The *Revue du Vin de France* is a monthly French wine publication which started in 1927. It specializes in French wines, and is highly regarded by the nation's wine industry.

- *Decanter* was established in 2004 by English wine critic Steven Spurrier (who was at the origin of the *Judgment of Paris*, discussed above) and awards trophies and medals, as well as wine ratings. See https://www.wine-searcher.com/critics-44-decanter+world+wine+awards.

- The *Wine Advocate* was created in 1978 by celebrated expert Robert Parker.

  See https://www.robertparker.com.

Appendix 2. Original and Normalized Rates (1)

Wine	Original rates					Normalized rates				
	Robinson	Atkin	RVF	Decanter	WA	Robinson	Atkin	RVF	Decanter	WA
Angélus	87.50	93.00	91.25	98.00	98.50	88.41	93.82	86.14	96.08	97.69
d'Armailhac	82.50	89.00	80.00	94.00	92.00	68.12	52.73	15.84	72.88	39.60
Ausone	87.50	97.00	96.25	98.00	99.00	94.20	99.64	94.55	96.41	99.34
Balestard La Tonnelle	80.00	86.00	82.50	91.00	92.00	19.20	17.09	38.12	17.65	49.17
Batailley	80.00	89.00	91.25	94.00	93.00	37.68	61.09	83.17	73.53	57.76
Beauregard	70.00	86.00	81.25	92.00	94.00	0.36	18.55	24.75	46.08	72.61
Beauséjour Duffau-Lagarrosse	77.50	89.00	92.50	99.00	96.25	6.88	65.82	88.12	100.00	87.13
Belgrave	82.50	91.00	85.00	92.00	90.00	42.39	86.18	53.96	38.56	14.85
Bellefont-Belcier	80.00	87.00	80.00	91.00	92.00	30.43	35.64	18.81	27.45	37.29
Bellevue	82.50	88.00	82.50	93.00	93.25	63.04	47.64	42.08	52.94	65.68
Berliquet	80.00	88.00	90.00	93.00	94.25	29.71	38.18	80.69	60.46	78.22
Beychevelle	80.00	90.00	88.75	94.00	95.25	28.62	72.36	72.77	76.14	82.84
le Bon Pasteur	75.00	88.00	80.00	91.00	93.25	5.80	44.36	19.31	22.55	66.67
Bouscaut	80.00	88.00	81.25	93.00	90.00	36.23	46.55	27.23	61.44	21.45
Branaire (Ducru)	82.50	89.00	91.25	93.00	93.00	52.17	56.36	81.68	51.96	52.81
Brane-Cantenac	80.00	91.00	90.00	95.00	93.00	33.33	81.82	76.24	81.05	62.71
Calon-Ségur	90.00	95.00	91.25	96.00	97.00	95.29	96.73	84.65	84.97	88.45
Canon	87.50	91.00	85.00	97.00	98.00	90.58	82.18	54.95	90.52	91.42
Canon-La-Gaffelière	82.50	87.00	90.00	94.00	95.00	57.61	35.27	78.71	65.36	79.87
Cantemerle	77.50	88.00	85.00	92.00	87.00	14.49	43.27	54.46	35.95	2.64
Cantenac-Brown	80.00	90.00	90.00	94.00	92.00	22.10	75.27	75.74	74.18	36.30
Cap de Mourlin	77.50	87.00	80.00	89.00	93.00	12.68	28.36	17.33	4.90	56.11
les Carmes Haut-Brion	82.50	91.00	90.00	98.00	95.25	50.72	80.73	76.73	93.79	84.49
Cheval Blanc	92.50	96.00	91.25	99.00	98.00	97.83	97.82	85.64	99.35	93.07
Clerc Milon	85.00	90.00	85.00	96.00	94.00	78.62	78.18	52.48	84.31	71.29
Clos du Marquis	85.00	89.00	90.00	91.00	93.00	73.91	65.45	77.72	25.49	62.05
Clos Fourtet	80.00	88.00	88.75	97.00	96.00	16.67	46.18	74.26	92.48	86.47

Appendix 2. Original and Normalized Rates (2)

Wine	Original rates					Normalized rates				
	Robinson	Atkin	RVF	Decanter	WA	Robinson	Atkin	RVF	Decanter	WA
la Clotte	80.00	91.00	82.50	94.00	94.25	38.04	86.55	38.61	72.22	77.89
Cos d'Estournel	87.50	96.00	96.25	97.00	98.50	89.49	98.91	94.06	90.85	96.70
Croizet-Bages	80.00	87.00	81.25	91.00	90.00	36.59	33.45	32.18	16.67	16.83
Dauzac	82.50	88.00	87.50	93.00	88.00	63.77	36.36	62.38	54.58	6.60
Desmirail	77.50	87.00	81.25	91.00	88.00	6.52	24.00	29.70	21.90	8.91
Domaine de Chevalier	87.50	89.00	90.00	96.00	95.25	93.48	58.18	78.22	85.62	84.16
Ducru-Beaucaillou	85.00	91.00	95.00	97.00	98.00	86.59	87.27	91.58	87.25	94.06
Duhart-Milon	87.50	88.00	85.00	95.00	94.00	92.75	39.64	51.98	80.07	74.92
Faugères	80.00	86.00	78.75	92.00	92.00	27.90	17.82	15.35	39.87	47.19
de Fieuzal	80.00	88.00	87.50	94.00	92.00	28.99	40.73	65.35	77.78	48.84
Figeac	90.00	92.00	87.50	98.00	98.00	96.01	90.18	66.83	97.39	94.39
Fonroque	80.00	79.00	81.25	92.00	92.00	30.80	1.45	23.27	42.16	40.92
Fourcas-Hosten	80.00	88.00	83.75	92.00	87.00	35.51	37.45	45.54	35.29	3.30
Franc-Mayne	77.50	87.00	77.50	90.00	91.00	10.51	25.82	9.90	13.40	25.74
Giscours	85.00	91.00	86.25	95.00	91.25	77.17	85.45	61.88	78.43	32.34
Gloria	82.50	89.00	85.00	93.00	93.00	43.84	67.27	56.44	53.27	63.70
Grand Corbin-Despagne	82.50	86.00	77.50	93.00	94.00	62.32	14.91	5.94	57.19	68.32
Grand-Pontet	75.00	89.00	85.00	91.00	90.00	2.54	54.55	49.50	26.47	20.46
Grand-Puy-Ducasse	82.50	90.00	82.50	92.00	92.00	63.41	73.09	41.09	37.25	35.97
Grand-Puy-Lacoste	87.50	91.00	87.50	95.00	93.25	91.67	84.36	69.31	81.70	64.69
Gruaud-Larose	80.00	90.00	92.50	94.00	96.00	26.45	74.91	88.61	75.82	84.82
Haut-Batailley	85.00	89.00	83.75	93.00	93.00	80.07	52.36	48.02	57.84	64.03
Haut-Brion	90.00	92.00	97.50	98.00	98.25	95.65	91.27	96.53	93.46	95.38
Haut-Marbuzet	80.00	87.00	87.50	91.00	92.00	40.94	23.27	63.37	20.92	39.93
d'Issan	85.00	92.00	90.00	96.00	95.00	72.83	91.64	80.20	86.93	81.85
Kirwan	82.50	89.00	83.75	93.00	92.00	46.38	59.64	43.56	52.29	37.95

Appendix 2. Original and Normalized Rates (3)

Wine	Original rates					Normalized rates				
	Robinson	Atkin	RVF	Decanter	WA	Robinson	Atkin	RVF	Decanter	WA
Labégorce	80.00	87.00	82.50	93.00	92.25	39.13	34.18	34.65	63.40	50.83
Lafite-Rothschild	95.00	97.00	97.50	99.00	99.00	100.00	100.00	97.52	98.37	99.67
Lafleur	92.50	94.00	97.50	98.00	98.50	97.46	94.91	98.02	95.42	97.03
Lalande-Borie	82.50	87.00	80.00	90.00	89.00	51.45	32.00	20.30	14.05	10.56
Langoa-Barton	82.50	90.00	86.25	93.00	93.00	60.14	70.55	60.40	54.90	53.14
Larcis-Ducasse	82.50	89.00	82.50	95.00	97.00	47.10	56.00	36.63	82.68	87.79
Larmande	80.00	85.00	77.50	92.00	93.00	35.87	8.36	7.43	41.18	63.37
Larrivet-Haut-Brion	80.00	85.00	81.25	93.00	92.00	21.38	9.09	29.21	62.75	41.91
Lascombes	77.50	91.00	86.25	92.00	93.00	8.33	83.64	58.42	34.31	57.43
Latour-Martillac	85.00	87.00	85.00	94.00	94.00	80.43	27.27	51.49	73.86	75.91
Léoville-Barton	85.00	91.00	93.75	96.00	95.00	83.70	84.00	90.59	83.99	78.88
Léoville-Las-Cases	92.50	95.00	97.50	98.00	99.00	99.28	96.36	98.51	96.73	99.01
Léoville-Poyferré	87.50	90.00	91.25	97.00	95.25	88.77	68.00	82.67	91.18	83.50
Lilian Ladouys	80.00	87.00	83.75	92.00	91.00	31.16	27.64	47.52	31.05	28.05
la Louvière	75.00	86.00	76.25	90.00	93.00	3.62	16.73	2.97	13.73	59.74
Lynch Bages	85.00	93.00	92.50	97.00	97.00	81.88	92.36	89.11	89.22	89.11
Lynch-Moussas	80.00	91.00	81.25	91.00	92.25	27.17	88.00	26.24	25.82	50.17
Malartic-Lagravière	80.00	89.00	86.25	95.00	93.00	38.41	65.09	59.90	81.37	54.79
Malescot-Saint-Exupéry	80.00	90.00	91.25	93.00	95.25	29.35	73.82	83.66	60.13	83.17
Margaux	92.50	94.00	98.75	98.00	98.50	97.10	94.18	100.00	94.77	98.02
Marquis de Terme	82.50	90.00	87.50	91.00	92.00	57.25	78.55	65.84	23.20	45.21
la Marzelle	72.50	89.00	81.25	92.00	93.00	1.09	59.27	27.72	40.85	56.44
Meyney	80.00	91.00	87.50	94.00	90.25	23.91	84.73	69.80	66.99	25.08
la Mission Haut-Brion	87.50	91.00	95.00	97.00	99.00	93.84	82.91	92.57	89.54	98.35
Montrose	90.00	93.00	95.00	97.00	97.00	96.74	92.00	93.07	88.56	88.12
Mouton-Rothschild	90.00	94.00	98.75	99.00	98.25	94.93	95.27	99.01	99.67	94.72

Appendix 2. Original and Normalized Rates (4)

Wine	Original rates					Normalized rates				
	Robinson	Atkin	RVF	Decanter	WA	Robinson	Atkin	RVF	Decanter	WA
Nénin	85.00	91.00	80.00	92.00	94.00	81.52	83.27	19.80	47.06	70.63
Olivier	82.50	86.00	83.75	94.00	93.00	67.75	21.45	46.53	70.26	58.75
les Ormes de Pez	82.50	88.00	80.00	93.00	91.00	52.90	45.09	18.32	50.00	30.36
Palmer	92.50	94.00	98.75	99.00	98.00	98.19	94.55	99.50	98.04	90.76
Pavie-Macquin	82.50	91.00	87.50	98.00	96.00	48.55	86.91	63.86	94.44	86.14
Pédesclaux	82.50	90.00	90.00	93.00	93.00	53.62	70.91	79.70	61.11	55.12
Pensées de Lafleur	87.50	90.00	93.75	93.00	92.00	90.22	69.09	90.10	65.03	41.58
Petit Village	85.00	91.00	82.50	93.00	91.25	72.46	87.64	37.62	56.21	34.65
de Pez	85.00	87.00	80.00	92.00	92.00	77.90	30.18	16.83	49.35	46.20
Phélan-Ségur	85.00	88.00	87.50	94.00	94.00	71.38	48.73	70.79	68.30	73.27
Pibran	82.50	89.00	80.00	90.00	90.00	50.00	66.55	17.82	15.03	20.13
Pichon-Longueville Baron	90.00	93.00	97.50	97.00	98.00	96.38	92.73	97.03	89.87	92.41
Pichon-Longueville Comtesse	87.50	92.00	96.25	99.00	98.00	87.68	88.73	95.54	99.02	92.08
le Pin	92.50	96.00	82.50	98.00	98.00	98.55	98.18	36.14	95.75	93.40
Pontet-Canet	82.50	96.00	97.50	96.00	98.00	54.71	97.45	96.04	83.33	93.73
Potensac	85.00	90.00	81.25	91.00	91.00	84.42	78.91	22.77	27.78	31.35
Poujeaux	77.50	88.00	81.25	93.00	87.00	9.06	45.82	28.22	63.07	2.97
Prieuré-Lichine	85.00	89.00	83.75	94.00	92.00	82.61	61.82	48.51	66.01	43.56
Rauzan-Gassies	80.00	87.00	82.50	92.00	94.00	26.09	28.73	35.15	45.42	75.58
Rauzan-Ségla	82.50	92.00	95.00	97.00	98.25	69.20	89.09	92.08	88.89	95.05
Rouget	82.50	85.00	76.25	93.00	92.00	41.67	8.00	3.96	59.15	44.22
Saint-Pierre	80.00	89.00	88.75	95.00	95.00	32.97	62.18	75.25	79.41	82.51
Siran	80.00	89.00	83.75	93.00	92.00	33.70	54.91	44.06	58.82	36.96
Smith Haut Lafitte	85.00	91.00	92.50	98.00	98.25	71.74	88.36	89.60	97.06	96.04
Soutard	82.50	88.00	82.50	92.00	94.00	58.33	41.45	40.10	47.39	74.59
Talbot	82.50	89.00	91.25	94.00	92.00	69.57	61.45	85.15	69.93	48.18

Appendix 2. Original and Normalized Rates (5)

Wine	Original rates					Normalized rates				
	Robinson	Atkin	RVF	Decanter	WA	Robinson	Atkin	RVF	Decanter	WA
du Tertre	85.00	87.00	85.00	92.00	90.00	79.35	25.45	55.94	49.02	21.78
la Tour Figeac	82.50	88.00	81.25	91.00	93.00	42.03	49.09	26.73	17.32	55.78
Tourelles de Longueville	85.00	90.00	83.75	91.00	91.00	76.45	71.27	44.55	28.43	25.41
Tronquoy-Lalande	85.00	88.00	77.50	93.00	94.00	73.55	40.00	9.41	62.42	73.93
Troplong-Mondot	87.50	92.00	82.50	97.00	94.25	87.32	89.45	37.13	92.81	76.90
Trottevieille	80.00	90.00	90.00	95.00	92.00	23.19	73.45	77.23	79.08	37.62
Valandraud	82.50	90.00	81.25	96.00	96.25	68.48	77.09	33.17	83.66	87.46
Vieux Château Certan	92.50	96.00	95.00	99.00	98.50	98.91	98.55	91.09	98.69	97.36
Villemaurine	80.00	87.00	77.50	93.00	94.00	35.14	26.91	10.89	55.56	71.95

Note: If WA ratings were originally given as intervals representing an estimated score range (e.g., [90- 93]), we considered the midpoint of the interval (e.g., 91.5) as the specific rating for that wine in our tables. WA also includes sometimes a plus sign following a rating (e.g., 95+), indicating ``a wine that the reviewer believes has the potential to improve over a period of time in bottle and may warrant a higher score in a subsequent/future tasting''. In those cases, we associated to the wine an extra 0.5 in the specific rating for that wine in our tables (e.g., 95.5).