Journal of Wine Economics, Volume 16, Number 3, 2021, Pages 321–342
doi:10.1017/jwe.2021.15

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

Wine Ratings: Seeking a Consensus among Tasters via Normalization, Approval, and Aggregation

Olivier Gergaud [a], Victor Ginsburgh [b] and Juan D. Moreno-Ternero [c]

Abstract

The modern era of wine journalism has provided abundant information about wines and widespread use of numerical rating systems. A tiny difference, especially at the top of the distribution of ratings, may have striking consequences on wine sales and investment returns. This article provides a general framework to obtain a consensus among tasters' opinions (reflected as numerical wine ratings) via three subsequent stages: *normalization*, *approval,* and *aggregation*. It is inspired by contributions in political science, social choice, game theory, and operations research. We apply it to the Judgment of Paris as well as to rank 2018 *en-primeur* Bordeaux wines, rated by five international experts. (JEL Classifications: C18, L15, L66)

Keywords: global wine score, Judgment of Paris, ratings, 2018 *en-primeur* Bordeaux wines, wines.

I. Introduction

Wine consumption has been part of civilization for more than 8,000 years. During most of that time, consumption was essentially local. In 1920, Europe (including Algeria, a French territory at the time) accounted for 95 percent of the world's wine production and exported only 5 percent to other countries (Anderson and Pinilla, 2018). This changed dramatically in recent years. In 2019, according to

We thank Karl Storchmann (Editor of this *Journal*) and an anonymous referee for helpful comments and suggestions. The third author acknowledges financial support from the Spanish Ministry of Economy and Competitiveness (ECO2017-83069-P).

[a] KEDGE Business School (Bordeaux campus), 680 Cours de la Libération, 33405 Talence Cedex, France; e-mail: olivier.gergaud@kedgebs.com (corresponding author).
[b] ECARES, Université Libre de Bruxelles and CORE, Université catholique de Louvain, 50 avenue F. Roosevelt 1050, Brussels, Belgium; e-mail: vginsbur@ulb.ac.be.
[c] Department of Economics, Universidad Pablo de Olavide, Carretera de Utrera, Km.1. 41013. Seville, Spain; e-mail: jdmoreno@upo.es.

© The Author(s), 2021. Published by Cambridge University Press on behalf of American Association of Wine Economists

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

Statista,[1] just the top three exporters (Italy, Spain, and France) shipped almost 20 percent of the world's wine production.

Similarly, although the literature on wines has existed throughout the history of wine, the widespread use of numerical rating systems is associated with the modern era of wine journalism. In 1978, Robert Parker launched the *Wine Advocate*. He popularized the [50 – 100]-point rating system, which is now widely used in the wine world.[2] As of today, it is difficult to find wine reviews without numerical ratings. This proliferation of ratings entails a new problem, as wine stores and consumers may face different assessments of the same wine. Amerine and Roessler (1983) were probably the first to raise the importance of seeking a consensus among tasters, exploring procedures to do so. We follow their steps here, aiming to provide a formal and comprehensive framework to obtain consensuses, building on contributions within political science, social choice, game theory, and operations research.

The choice of a specific procedure to reach a consensus might have important implications. An early instance in which this can be exemplified is the Judgment of Paris. This was a competition between American and French wines, organized in Paris in 1976 by British wine merchant Steven Spurrier. He invited 11 competent French wine connoisseurs who judged two flights of ten white wines and ten red 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 from 0 to 20. The 11 rates for each wine were added, resulting in a ranking of the ten wines. A Californian wine (Stag's Leap Wine Cellars) was at the top of the ranking. This happened at a time in which, according to Taber (2005, p. 17), "France ruled the world." The outcome challenged the French supremacy, and helped to introduce American wines to the worldwide market.

Adding rates across tasters is an obvious way to reach a consensus among them. But it is admittedly naive and flawed. For instance, each expert comes with her cultural and personal preferences, which might influence the outcome beyond desire. Ashenfelter and Quandt (1999, p. 170) suggested that "converting rates to ranks guarantees that each judge has the same influence on the outcome." This is essentially the so-called Borda's proposal in voting theory, which tallies the score of a candidate x by counting $n-1$ points for each voter for whom x was the best candidate, $n-2$ points for each voter who ranked x as second, and so on, until 0 points were given for each voter ranking x last.[3] In the case of the Judgment of Paris, this means awarding points from 0 to 10 to each of the wines, based on the rankings

[1] https://www.statista.com/statistics/240649/top-wine-exporting-countries-since-2007/

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

[3] Jean-Charles de Borda introduced his proposal almost two and a half centuries ago (Borda, 1784). Charles L. Dodgson, also known as Lewis Caroll, re-introduced it almost a century later (Dodgson, 1873). And it took another century to obtain its normative foundations in a formal analysis (Young, 1974).

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

provided by the judges. It turns out that Borda's proposal would have changed the *Judgment*'s conclusion, leading to a tie between American Stag's Leap Wine Cellars and French Château Montrose (see Table 1, two last rows).[4]

As of today, wine rankings are closely scrutinized by wine consumers and wine investors alike. The impact of expert reviews on the demands for wine and wine returns has been highlighted in many publications (Jaeger, 1981; Ali, Lecocq, and Visser, 2008; Friberg and Grönqvist, 2012; Fogarty and Sadler, 2014; Masset, Weisskopf, and Cossutta, 2015; Ashton 2016; Cyr, Kwong, and Sun, 2017). A tiny difference in terms of ranking at the top of the distribution may have striking consequences on wine sales and investment returns. The previous story on the Judgment of Paris, together with the fact that wine investors aim to diversify their portfolio, illustrates the importance of scrutinizing the alternative procedures to seek a consensus among tasters. We aim to do so in this article, providing a formal and comprehensive framework, which rests on previous contributions in political science, social choice, game theory, and operations research. Our starting point is a group of 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 matrix, the rows of which are experts ($i = 1, 2, \ldots, I$), and the columns are wines ($j = 1, 2, \ldots, J$). Element a_{ij} of this matrix represents the rating of wine j by expert i. The output will be a column that represents a consensual rating for each wine, which we suggest obtaining after three stages: *normalization*, *approval,* and *aggregation*.

The first stage (*normalization*) converts absolute ratings into relative (i.e., quantile-based) ratings. This is meant to have ranges of ratings equalized across experts and ratings themselves homogenized. In doing so, we address Ashenfelter and Quandt's argument stated previously. It is also what Cardebat and Paroissien (2015) suggested to standardize scores, which is also endorsed by *Global Wine Score*.[5]

The second stage (*approval*) determines a quantile that can be interpreted as a *threshold* that wines have to meet to be approved.[6] 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 wine, 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.[7] We suggest

[4] Hulkower (2009), who scrutinized the use of Borda's proposal for the Judgment of Paris, noted that Château Montrose would be the only winner, provided we dismiss the votes of the two non-official judges.
[5] See https://www.globalwinescore.com, last consulted March 2021.
[6] Obviously, if the threshold is the lowest possible one, all wines would be approved, and this stage would be irrelevant.
[7] 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, last consulted in March 2021. One might say that wines below 90 are simply ignored (i.e., not *approved*) by a large number of consumers.

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

Table 1
The Judgment of Paris: 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.

that only approved wines pass to the next stage with a positive weight. In other words, non-approved wines are all equally discriminated. As for approved wines, we allow for two options: to discriminate among them (using their normalized ratings as their weights in the next stage) or not (assigning the same positive weight to all approved wines). On the one hand, it is frequently argued that dichotomous decisions to signal whether a wine is approved or not are much easier to make and may be sufficient, while ratings are much more difficult. On the other hand, it might be considered unfair not to distinguish among (approved) wines with very different ratings once they are available. Thus, we shall consider both options (which will obviously yield different outcomes).

The third stage (*aggregation*) concludes by aggregating the information obtained from each expert after the previous stages. The first option is to simply add the ratings (obtained after the previous two stages). The second option is to introduce a second normalization before aggregating. More precisely, we divide each rating by the sum of the ratings each judge confers to all wines, and then we aggregate. As before, the choice between both options results from a value judgment: in this case, deciding whether the ratings given by an expert should depend on the number of wines that passed the approval stage or not. Thus, we shall consider both options too (which will obviously yield different outcomes as well).

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

Our three-stage procedure is inspired by contributions in other fields.

The normalization stage mentioned previously was already proposed by Cardebat and Paroissien (2015) to standardize wine scores. It 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, Herrero, and Pinto, 2002; Herrero and Moreno-Ternero, 2009), or the evaluation of scientific performance (Albarrán et al., 2010, 2011).

As for the other two stages, political scientists suggested using *Cumulative Voting* more than half a century ago (Glasser, 1959; Sawyer and MacRae, 1962). This 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 (currently in practice in some U.S. local elections, as well as to elect officers in numerous professional organizations). It allows each voter to cast a vote for as many candidates she wishes; each positive vote is counted in favor of the candidate. The votes are then added candidate by candidate, and the winner is the one who gets the largest number of votes.[8] Under plausible assumptions, *Approval Voting* compares favorably with both the plurality rule or Borda's rule (Weber, 1995).

Another intuition is derived from the *Shapley value* (Shapley, 1953), 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. Ginsburgh and Zang (2003) used this concept to study the problem of sharing the total revenue collected from selling museum passes, which give access to several museums.[9] Later on, Ginsburgh and Zang (2012) applied the same theoretical model to aggregate ratings from the Judgment of Paris.[10] In their case, the protocol is different from ours because, instead of a threshold, they (exogenously) fixed 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 that 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,

[8] Both Approval Voting and Cumulative Voting can be seen as members of a family of voting procedures called *Size Approval Voting*, which is characterized by Alcalde-Unzu and Vorsatz (2009).
[9] See also Bergantiños and Moreno-Ternero (2015).
[10] Ginsburgh, Moreno-Ternero, and Weber (2017) and Alcalde-Unzu, Moreno-Ternero, and Weber (2021) also explored similar ideas to rank languages.

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

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 ratings until the budget is exhausted, then knapsack voting is a partially truthful mechanism (Goel et al., 2019). In the context of wine, Conrad, Gómez, and Lamadrid (2011) posed several knapsack problems to select the optimal subset of wines subject to budget and quantity constraints.

Our proposal is, therefore, a unifying framework to accommodate a variety of options and provide a consensus among wine tasters. It ranges from the basic approach followed by *Global Wine Score*, which aggregates scores after standardizing them as suggested by Cardebat and Paroissien (2015), to more sophisticated approaches arising after setting thresholds to approve wines and weighing them proportionally to the ratings of approved wines. It is preferable to propose a unique rule because, as is well known, no single consensus rule is flawless (see Arrow (1951), Gibbard (1973), and Satterthwaite (1975)).

The rest of the article is organized as follows. In Section II, we formalize the three stages discussed earlier. Section III illustrates these methods using the data from the Judgment of Paris. Section IV turns to apply them to a set of 114 en primeur (early) Bordeaux wines rated in 2018 by five international experts. Section V concludes.

II. The Model

We consider a set of judges or experts who rate wines. Formally, there exists a matrix A, the rows of which are experts ($i = 1, 2, \ldots, I$), and the columns are wines ($j = 1, 2, \ldots, J$). 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 column 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 A to C). Subsection D illustrates the computations using a very simple example.

A. Normalization Stage

As suggested by Cardebat and Paroissien (2015), 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

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

expert's ratings. This reduces 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 only 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 the original 95 and 90, respectively.

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

After this normalization, Cardebat and Paroissien (2015) simply suggest averaging the normalized ratings that each wine receives from all experts. That is, for each $j = 1, 2, \ldots, J$,

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

Instead of endorsing the above proposal, we suggest replacing averaging with two additional stages: the *approval stage* and the *aggregation stage*.

B. Approval Stage

This second stage sets a quantile π that can be interpreted as a *threshold* that wines have to meet in order to be *approved* or to avoid being *censored*.[11] This stage converts the matrix of (normalized) ratings A^n into a new matrix in which the rating of a *censored* wine (i.e., a rating below the quantile π) is replaced by zero. As for *approved* wines, we consider two options:

a) Either ratings are replaced by a constant (1 without loss of generality); this leads to a matrix $A^{1\pi}$, in which $a_{ij}^{1\pi} = 1$ if $a_{ij}^n \geq \pi$ and 0 otherwise,

b) or ratings are stored in a matrix A^π such that $a_{ij}^\pi = a_{ij}^n$ if $a_{ij}^n \geq \pi$ and 0 otherwise.

In the first option, the ratings given by experts to approved wines are not used to reach a consensus, whereas, in the second, they are. 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 be sufficient, while ratings are much more difficult. Difficulties to exactly express quality by a score would support (a) although the method may end up with many ties in the aggregation stage that follows. On the other hand, it might seem unfair not to

[11] Note that experts are not involved in this approval stage, although they rated the wines.

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

distinguish among (approved) wines with very different ratings once they are available, which would support (b).

C. Aggregation Stage

The last stage of our procedure aggregates the information from matrices $A^{1\pi}$ or A^{π}, which both construct approval ratings. The first option is to simply compute averages across wines from one or both matrices, thus mimicking what we did for matrices 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 matrices $A^{1\pi}$ 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 judgment. It amounts to the critical issue of deciding whether the ratings given by an expert should depend on the number of wines that passed the approval stage or not.

D. An Example of the Computations

Consider the following example, in which five judges rate three wines, as described in the following matrix.

$$A = \begin{pmatrix} 100 & 95 & 95 \\ 95 & 100 & 100 \\ 96 & 97 & 98 \\ 96 & 99 & 99 \\ 99 & 100 & 99 \end{pmatrix}.$$

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

We assume that the judges rated many more wines and, thus, we can obtain the CDFs of their ratings. Suppose the resulting matrix from A, after normalizing ratings making use of their CDFs, is the following:

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

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

- Usual Protocol (1):

$$U(1) = \frac{1}{5}(100 + 95 + 96 + 96 + 99) = 97.2$$

$$U(2) = \frac{1}{5}(95 + 100 + 97 + 99 + 100) = 98.2$$

$$U(3) = \frac{1}{5}(95 + 100 + 98 + 99 + 99) = 98.2$$

- Normalized Protocol (2):

$$N(1) = \frac{1}{5}(96 + 90 + 96 + 96 + 94) = 94.4$$

$$N(2) = \frac{1}{5}(90 + 96 + 97 + 99 + 96) = 95.6$$

$$N(3) = \frac{1}{5}(90 + 96 + 98 + 99 + 94) = 95.4$$

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

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

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

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

- Approval Protocol (3):

$$A(1) = \frac{1}{5}(1 + 0 + 1 + 1 + 0) = 0.6$$

$$A(2) = \frac{1}{5}(0 + 1 + 1 + 1 + 1) = 0.8$$

$$A(3) = \frac{1}{5}(0 + 1 + 1 + 1 + 0) = 0.6$$

- Proportional Approval Protocol (4):

$$PA(1) = \frac{1}{5}(96 + 0 + 96 + 96 + 0) = 57.6$$

$$PA(2) = \frac{1}{5}(0 + 96 + 97 + 99 + 96) = 77.6$$

$$PA(3) = \frac{1}{5}(0 + 96 + 98 + 99 + 0) = 58.6$$

The last two protocols in which matrices $A^{1\pi}$ 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):

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

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

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

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

III. Application 1. The Judgment of Paris

The Judgment of Paris alluded to in the Introduction, invited 11 well-known French experts to evaluate 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 how 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.[12]

We now turn to the alternative protocols considered in this article. 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.[13]

Table 2 contains all the rankings described in this article. The rankings yield similar results. Indeed, wines A, B, and C appear among the first three in all cases, in some cases they are tied, in other cases they are not. Wine D is almost always number 4, except that it is tied once with wine F. Wines H, I, J are always the last ones. Differences mostly occur in the middle, where rankings switch among wines.

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 protocols) breaks. Wine C (French) actually falls to fourth place, whereas wine D (French) is third. As for wines A (American) 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), and H (French).

IV. Application 2. Bordeaux 2018 Future Wines

We now consider the more interesting and recent tasting of the 2018 Red Bordeaux called *Bordoverview*.[14] Bordoverview contains ratings for 114 (Bordeaux 2018 future)

[12] See Taber (2005) for details.
[13] 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.
[14] See https://www.bordoverview.com/?RP=98.1.

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

Table 2
The Judgment of Paris: 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 Judgment of Paris: 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.

wines produced by five (international) experts: Jancis Robinson (JR), Tim Atkin (TA), *Revue du Vin de France* (RVF), *Decanter Magazine* (DM), and Parkers' *Wine Advocate* (WA).[15] 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.[16]

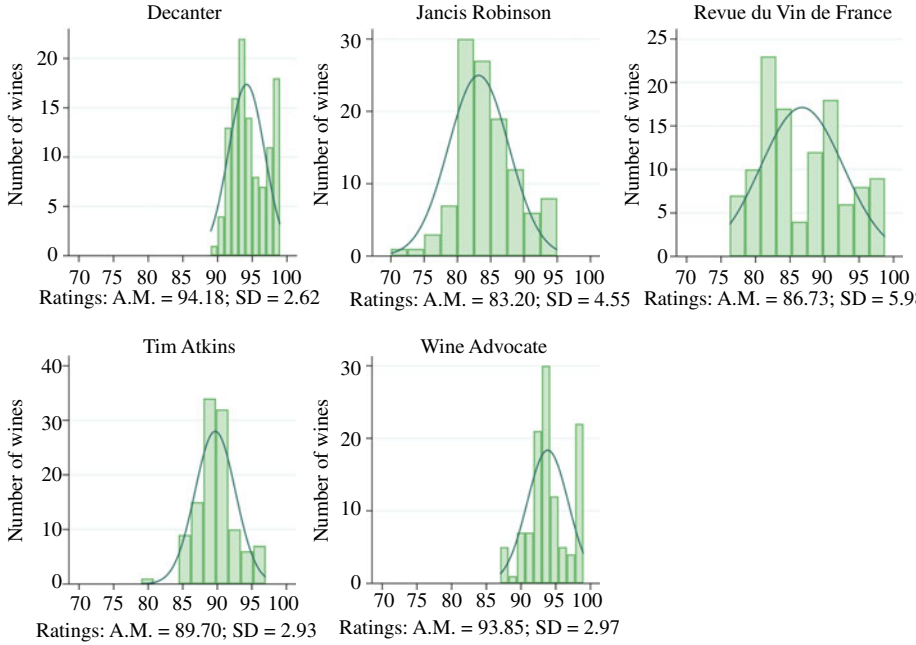
[15] See Appendix 1 for some details on the five experts.

[16] Some caveats are in order regarding the ratings associated with WA. Sometimes their ratings are not specific but rather intervals representing an estimated rating range (for instance, 90 – 93). In those cases, we considered the midpoint of the interval as the specific rating for that wine in our analysis. WA also includes sometimes a plus sign following a rating (for instance, 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 the wine an extra 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

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

Figure 1

Rating Distributions of Experts



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, from Figure 1, we can observe how different their rating distributions are.

Table 4 shows the ratings for the six protocols introduced earlier when the (normalized) threshold is set at 90.[17] This threshold eliminates 76 of the 114 wines that are thus given a zero rating (with an exception for the first two protocols, which do not consider the approval stage), and 38 pass the test and are "approved." Table 4 also contains the associated rankings for each of the six protocols and leads us to some interesting conclusions.

First, the top of the rankings seems to be quite robust. Lafite-Rothschild comes first in all cases (although tied with nine other wines in the third and fifth protocols, which, as mentioned earlier, may generate many ties). Léoville-Las-Cases seems to be a solid second follower, except for the first protocol, where it appears as number three (and the

in this task, although it has been argued that his ratings are not the ones showing the highest correlation with those of Parker (Cyr, Kwong, and Sun, 2019).

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

Table 4
Ratings and Rankings According to Each Protocol

	Protocol (1)		*Protocol (2)*		*Protocol (3)*		*Protocol (4)*		*Protocol (5)*		*Protocol (6)*	
Wines	*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

Continued

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

Table 4
Continued

	Protocol (1)		*Protocol (2)*		*Protocol (3)*		*Protocol (4)*		*Protocol (5)*		*Protocol (6)*	
Wines	*Rates*	*Ranks*	*Rates*	*Ranks*	*Rates*	*Ranks*	*Rates*	*Ranks*	*Rates*	*Ranks*	*Rates*	*Ranks*
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

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

caveat for the ties in the third and fifth protocols). The first protocol actually awards the second position to Palmer, although it goes to eighth place after normalization. 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 to the fifth position for protocols after normalization.[18] Other specific and somewhat striking differences between the second, fourth, and sixth protocols can be observed; for instance, 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 previously, the third and fifth protocols introduce many ties, as the rates are dichotomous (1 for approved wines, 0 for all others). 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 fewer 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, except for the first (non-normalized) protocol, in which Palmer comes second and Léoville-Las-Cases third.

To have a more general view on the differences in the protocols (except for the third and fifth, which, as we already said, have too many ties), we calculated Pearson correlations, as well as Spearman rank correlation coefficients on rankings (see Table 5).

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 observe that Protocol (2) (normalization) is not very different from Protocol (1) (between 0.93 and 0.97) which implies that normalizing does not change much. In addition, Protocols (1) and (6) (the most sophisticated ones) 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 eighth in the usual protocol, may strongly benefit, getting from rank eighth to rank fourth, in all other protocols. It also makes a very large difference being among the top 10 and the top 20. 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 complex relationship

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

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

Table 5
Correlations Between Protocols

Pearson Correlation Coefficients on Ratings				
	Protocol (1)	*Protocol (2)*	*Protocol (4)*	*Protocol (6)*
Protocol (1)	1			
Protocol (2)	0.94	1		
Protocol (4)	0.91	0.81	1	
Protocol (6)	0.92	0.82	0.99	1
Spearman Correlation Coefficients on Ranks				
	Protocol (1)	*Protocol (2)*	*Protocol (4)*	*Protocol (6)*
Protocol (1)	1			
Protocol (2)	0.97	1		
Protocol (4)	0.89	0.86	1	
Protocol (6)	0.94	0.91	0.95	1

between talent, or quality, essentially driven by experts, and prices are illustrated by Ali, Lecocq, and Visser (2008).

In short, changes of ranks (even from one to two, or five to six) are quite 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 prestige because one of them, Stag's Leap Wine Cellars, overshadowed three prestigious 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 always come first and second. Margaux is third in all cases except when the threshold climbs to 95, where it falls to 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.

V. Discussion

Numerical wine ratings are very popular. The scientific literature has paid attention to aspects such as (1) inconsistency of ratings in blind tastings (Lindley, 2006; Hodgson, 2008; Bodington, 2017, 2020), (2) consensus among experienced wine experts (Ashton, 2012, 2013; Cao, 2014; Luxen, 2018), (3) variations in the severity of experts (Masset, Weisskopf, and Cossutta, 2015; Stuen, Miller, and Stone, 2015), and (4) consumers' demand for wine ratings (Ashenfelter and Jones, 2013; Marks, 2015, 2020).

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

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 (with the exception of Column 1) there are ties among wines (in italics).

In this article, we aimed at providing a comprehensive framework to reach a consensus among tasters' opinions (expressed via wine ratings). We consider three (consecutive) stages: *normalization*, *approval,* and *aggregation*. The first stage converts absolute ratings into relative ratings, as suggested by Cardebat and Paroissien (2015). The second one puts thresholds on normalized ratings, below which wines get no credit. The third one shares the credit from each expert among approved wines, equally or proportionally. The last two stages

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

are inspired by classical contributions in political science, social choice, game theory, and operations research.

Our framework can accommodate six different protocols, depending on the stance one takes in each of the three stages. We illustrate our protocols using the Judgment of Paris and 2018 *en-primeur* Bordeaux wines, rated by five international experts.[19] Our analysis concludes that how 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 article. We argued that normalizing ratings are 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 the choices of experts much easier.

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 considered: 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 (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 protocols, the opinions of experts were equally weighted. But it may be the case that certain experts are more influential than others. The protocols could easily be extended to account for unequal weights of experts.

References

Albarrán, P., Crespo, J., Ortuno-Ortin, I., and 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(1), 329–344.

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

[19] Both instances have received attention in the literature (Cicchetti, 2006; Hulkower, 2009).

Alcalde-Unzu, J., Moreno-Ternero, J., and Weber, S. (2021). The measurement of the value of a language. Centre for Economic Policy Research, Discussion Paper No. 15770. Available from https://repec.cepr.org/repec/cpr/ceprdp/DP15770.pdf.

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

Ali, H., Lecocq, S., and Visser, M. (2008). The impact of gurus: Parker grades and en primeur wine prices. *The Economic Journal*, 118(529), F158–F173. (Reprinted in 2010 in the *Journal of Wine Economics*, 5(1), 22–39.)

Amerine, M., and Roessler, E. (1983). *Wines: Their Sensory Evaluation*. New York: W. H. Freeman.

Anderson, K., and Pinilla V., eds. (2018). *Wine Globalization: A New Comparative History*. New York: Cambridge University Press.

Arrow, K. (1951). *Social Choice and Individual Values*. New York: John Wiley and Sons.

Ashenfelter, O., and Jones, G. (2013). The demand for expert opinion: Bordeaux wine. *Journal of Wine Economics*, 8(3), 285–293.

Ashenfelter, O., and Quandt, R. (1999). Analyzing a wine tasting statistically. *Chance*, 12(3), 16–20.

Ashton, R. H. (2012). Reliability and consensus of experienced wine judges: Expertise, within, and between? *Journal of Wine Economics*, 7(1), 70–87.

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(2), 225–234.

Ashton, R. H. (2016). The value of expert opinion in the pricing of Bordeaux wine futures. *Journal of Wine Economics*, 11(2), 261–288.

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

Bergantiños, G., and Moreno-Ternero, J. (2020). Sharing the revenues from broadcasting sport events. *Management Science*, 66(6), 2417–2431.

Bleichrodt, B., Herrero, C., and Pinto, J. (2002). A proposal to solve the comparability problem in cost-utility analysis. *Journal of Health Economics*, 21(3), 397–403.

Bodington, J. (2017). The distribution of ratings assigned to blind replicates. *Journal of Wine Economics*, 12(4), 363–369.

Bodington, J. (2020). Rate the raters: A note on wine judge consistency. *Journal of Wine Economics*, 15(4), 363–369.

Borda, J.-C. (1784). *Memoire sur les Elections au Scrutin*. Paris: Histoire de l'Academie Royale des Sciences.

Brams, S., and Fishburn, P. (1978). Approval voting. *American Political Science Review*, 72(3), 831–847.

Cabannes, Y. (2004). Participatory budgeting: A significant contribution to participatory democracy. *Environment and Urbanization*, 16(1), 27–46.

Cao, J. (2014). Quantifying randomness versus consensus in wine quality ratings. *Journal of Wine Economics*, 9(2), 202–213.

Cardebat, J., and Paroissien, E. (2015). Standardizing expert wine scores: An application for Bordeaux en primeur. *Journal of Wine Economics*, 10(3), 329–348.

Cicchetti, D. V. (2006). The Paris 1976 wine tasting revisited once more: Comparing ratings of consistent and inconsistent tasters. *Journal of Wine Economics*, 1(2), 125–140.

Conrad, J., Gómez, M., and Lamadrid, A. (2011). Wine in your knapsack? *Journal of Wine Economics*, 6(1), 83–110.

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

Cyr, D., Kwong, L., and Sun, L. (2017). An examination of tail dependence in Bordeaux futures prices and Parker ratings. *Journal of Wine Economics*, 12(3), 252–266.
Cyr, D., Kwong, L., and Sun, L. (2019). Who will replace Parker? A copula function analysis of Bordeaux en primeur wine raters. *Journal of Wine Economics*, 14(2), 133–144.
Dodgson, C. L. (1873). *A Discussion of the Various Methods of Procedure in Conducting Elections*. Princeton: Princeton University Library.
Fogarty, J., and Sadler, R. (2014). To save or savor: A review of approaches for measuring wine as an investment. *Journal of Wine Economics*, 9(3), 225–248.
Friberg, R., and Grönqvist, E. (2012). Do expert reviews affect the demand for wine? *American Economic Journal: Applied Economics*, 4(1), 193–211.
Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica*, 41(4), 587–601.
Ginsburgh, V., Moreno-Ternero J., and Weber, S. (2017). Ranking languages in the European Union: Before and after Brexit. *European Economic Review*, 93(C), 139–151.
Ginsburgh, V., and Zang, I. (2003). The museum pass game and its value. *Games and Economic Behavior*, 43, 322–325.
Ginsburgh, V., and Zang, I. (2012). Shapley ranking of wines. *Journal of Wine Economics*, 7(2), 169–180.
Glasser, G. (1959). Game theory and cumulative voting for corporate directors. *Management Science*, 5(2), 151–156.
Goel, A., Krishnaswamy, S., Sakshuwong, S., and Aitamurto T. (2019). Knapsack voting for participatory budgeting. *ACM Transactions on Economics and Computation*, 7(2), 1–27.
Herrero, C., and Moreno-Ternero, J. (2009). Estimating production costs in the economic evaluation of health care programs. *Health Economics*, 18(1), 21–35.
Hodgson, R. (2008). An examination of judge reliability at a major U.S. wine competition. *Journal of Wine Economics*, 3(2), 105–113.
Hulkower, N. (2009). The judgment of Paris according to Borda. *Journal of Wine Research*, 20(3), 171–182.
Jaeger, E. (1981). To save or savor: The rate of return to storing wine. *Journal of Political Economy*, 89(3), 584–592.
Lindley, D. V. (2006). Analysis of a wine tasting. *Journal of Wine Economics*, 1(1), 33–41.
Luxen, M. (2018). Consensus between ratings of red Bordeaux wines by prominent critics and correlations with prices 2004–2010 and 2011–2016: Ashton revisited and expanded. *Journal of Wine Economics*, 13(1), 83–91.
Marks, D. (2015). Seeking the veritas about the vino: Fine wine ratings as wine knowledge. *Journal of Wine Research*, 26(4), 319–335.
Marks, D. (2020). Erring experts? A critique of wine ratings as hedonic scaling. *Journal of Wine Economics*, 15(4), 386–393.
Martello, S., and Toth, P. (1990). *Knapsack Problems*. New York: Wiley.
Masset, P., Weisskopf, J., and Cossutta, M. (2015). Wine tasters, ratings, and en primeur prices. *Journal of Wine Economics*, 10(1), 75–107.
Moreno-Ternero, J. (2007). On the design of equal-opportunity policies. *Investigaciones Económicas*, 31(3), 351–374.
Roemer, J. (1998). *Equality of Opportunity*. Cambridge, MA: Harvard University Press.
Satterthwaite, M. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2), 187–217.
Sawyer, J., and MacRae, D. (1962). Game theory and cumulative voting in Illinois: 1902–1954. *American Political Science Review*, 56(4), 936–946.

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press

Shapley, L. (1953). A value for n-person games. In H. W. Kuhn and A. W. Tucker (eds.), *Contribution to the Theory of Games*, vol. II, 307–317. Princeton, NJ: Princeton University Press.
Stuen, E., Miller, J., and Stone, R. (2015). An analysis of wine critic consensus: A study of Washington and California wines. *Journal of Wine Economics*, 10(1), 47–61.
Taber, G. (2005). *Judgment of Paris: California vs. France and the Historic 1976 Paris Tasting That Revolutionized Wine*. New York: Scribner.
Weber, R. (1995). Approval voting. *Journal of Economic Perspectives*, 9(1), 39–49.
Young, H. P. (1974). An axiomatization of Borda's rule. *Journal of Economic Theory*, 9(1), 43–52.

Appendix 1. Experts Selected from Bordoverview

- Jancis Robinson is a British wine writer and critic who became the first Master of Wine (MW) outside the wine trade. She studied mathematics and philosophy at the 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. After training in modern languages at the University of Durham, Atkin soon moved into a career in wine writing. See https://www.wine-searcher.com/critics-34-tim+atkin.
- The *Revue du Vin de France* is a monthly French wine publication that started in 1927. It specializes in French wines.
- The *Decanter* magazine was founded in London in 1975. It is the oldest consumer wine publication in the United Kingdom. The Decanter World Wine Awards, established in 2004 by English wine critic Steven Spurrier (who was at the origin of the Judgment of Paris, discussed earlier), awards trophies and medals, as well as wine ratings. See https://www.wine-searcher.com/critics-44-decanter+world+wine+awards.
- The *Wine Advocate* is a bimonthly wine publication based in the United States, created in 1978 by Robert Parker, regarded as one of the most widely known and influential wine critics in the world. See https://www.robertparker.com.

https://doi.org/10.1017/jwe.2021.15 Published online by Cambridge University Press