Sensory quality of wine : quality assessment by merging ranks of an expert-consumer panel

The Sensory Quality of Wine:

Quality Assessment by Merging Ranks of an Expert-Consumer panel

Guido De Mets[1], Peter Goos[1, 4, 5], Maarten Hertog[1], Christine Peeters[2], Jeroen Lammertyn[1], Bart M. Nicolaï[1,3]

[1]BIOSYST-MeBioS, University of Leuven, Leuven, Belgium

[2]Expertise Unit Teaching and Support, University of Leuven, Belgium

[3]Flanders Centre of Postharvest Technology, Leuven, Belgium

[4]Leuven Statistics Research Centre (LSTAT), University of Leuven

[5]Department of Engineering Management, University of Antwerp

Correspondence: Bart Nicolaï, BIOSYST-MeBioS, University of Leuven, Willem de Croylaan 42, Leuven B-3001 Belgium.

E-mail: bart.nicolai@biw.kuleuven.be

Version 12/03/2017

Abstract

Despite being a major focus in wine production, there are currently no standard procedures to measure the overall sensory quality of wine. While abundant, ratings from specialized guides and magazines lack scientific and statistical foundation and may confound preference with intrinsic quality. The presented method aims to bridge this gap by providing a 'quality assessment by merging ranks of an expert-consumer panel (QAMREC)' procedure, which ranks wines on a quantitative scale according to their sensory quality. While the methodology is essentially a preference testing method, by confining the sample space to wines with a similar origin and vinification and the recruitment of an expert panel, the effect of individual differences in preferences can be reduced and the resulting ranking is believed to provide a better representation of their intrinsic quality as valued by educated consumers who use an unconscious rationale for their judgement. It also takes into account human limitations and organizational constraints. Expert-consumers, consumers familiar with the rating of the sensory characteristics of wine, were selected as panellists with well-defined criteria including a high wine involvement profile. An optimized incomplete block design was deployed to guarantee a balanced tasting sequence. By applying the rank-order logit model, incomplete rankings obtained were converted into utility values for each wine compared to a reference wine. These utility values are an approximation of the intrinsic quality of the individual wines as judged by experts. The method was applied in ten tasting sessions each comprising 9 wines from one particular origin and 12 panellists. Most sessions, with the exception of the Pouilly Fumé and the Graves tasting, resulted in the identification of several wines with statistically different utilities. These findings introduce QAMREC as a valid approach for assessing the sensory quality of wines.

Keywords: wine, quality, preference, rank-order logit model, utility

1 Introduction

The quality of wine refers mainly to the sensory enjoyment it offers to consumers who are used to savour wine in a conscious way. During consumption, wine invokes various sensory sensations resulting in a temporary feeling of enjoyment. This delight is the result of a holistic perception of its overall sensory quality and combines visual, gustative, tactile and olfactory impressions (Shepperd, 2006; Guillaumie et al, 2013). Relevant sensory characteristics of wine are its colour hue and depth, its brown coloration state and clarity, its sweet, sour and bitter taste, oral tactile observations like astringency, viscosity, velvety, prickling and the feeling of heat caused by ethanol and olfactory characteristics including aromatic intensity, complexity, finesse and duration (Jackson, 2002).

The anticipated pleasure that wine might evoke during consumption is a major buying motivation and is, therefore, an important focus for both producer and consumer. Wine producers mainly count on their expertise, experience and intuition to judge the sensory quality of their wine. Consumers on the other hand are mainly driven by familiarity, taste, labels, price or availability and sometimes rely on wine scores and recommendations from quality guides, magazines and competitions. Examples of these are the universal 'Parker Guide' (Parker, 2015), the French 'Guide Hachette' (Hachette Pratique, 2014), the Italian 'Gambero Rosso' (Fabrizio et al., 2014), the Spanish 'Guía Peñin (Grupo Peñin, 2015), Decanter' (Decanter, 2015), 'La Revue du Vin de France' (La Revue du Vin de France, 2015) and the 'International Wine Challenge' (The International Wine Challenge, 2015). However, the methodology to obtain their scores is usually unclear, the sensory sessions that are used to assign the scores are rarely organized according to good experimental practices and they are often based on subjective assessments of too few experts. The nature of human processing of sensory information puts further constraints on the proper utilization of human experts as measuring devices for sensory attributes. Hominids assess the sensory aspect of their environment by relative comparison and their senses are ill suited to provide absolute scores (Barth et al, 2012). In addition, the absolute rating of sensory properties requires the provision of identical and optimal conditions taking also in account that humans are easily distracted (Niedenthal and Kitayama, 1994). As human senses are quickly fatigued, the number of samples that can be evaluated in a short time period is limited (Haarmann and Usher, 2001). Humans also cannot switch off individual senses which imply that human sensory evaluations are multisensory and holistic. Finally, persons not only differ in the anatomical and physiological details of their sensory apparatus, their judgment is also highly affected by personal memories and experiences (Hirsh et al, 2013). For all these reasons, results obtained by human experts

will always demonstrate variability. The capability to learn may improve observation skills and introduce standardization through appropriate training (Jackson, 2002; Lawless and Heymann, 2010).

A scientific evaluation approach would cover discrimination testing, the determination of attribute rating scales and consumer preference analysis. Several books (Meilgaard et al., 2006; Kemp et al., 2009; Lawless and Heymann, 2010; Stone et al., 2012;) and ISO standards (e.g. ISO 13299, 2003; ISO 6658, 2005; ISO 5492, 2008) were published to guide applicants in a standardised, scientific direction. In general, two types of testing can be distinguished. Objective testing or sensory profiling involves the assessment of sensory attributes of a product by a selected or trained panel. On the other hand, reactions of consumers are measured during subjective testing including consumer preference determination.

Sensory profiling is used extensively in the food industry during product development and quality inspection. Multiple applications involve the sensory profiling of wine. First, a set of relevant sensory attributes is determined. For wine, properties are typically defined for vision, taste, mouthfeel and aroma. Each parameter is further associated with a consistent vocabulary and an evaluation scale. The latter can be a numeric category or point-scale (Sáenz-Navajas et al., 2010; D'Alessandro and Pecotich, 2013; Cetó et al., 2015) or a line scale (Cliff et al., 2007; Ou et al., 2010; Harbertson et al., 2011; Kallithraka et al., 2011; Parker et al., 2012; McRae et al., 2013). Standard mixtures and wines are adopted for each attribute to train the panellists. The size of the panel is rather small and varies from seven (Callejon et al, 2009) to thirty (Sáenz-Navajas et al, 2010). The selected panellists often follow an appropriate training program during which they learn to employ the correct terminology and rating scale in order to optimize their performance. The presentation of the wine corresponds with an experimental design to guarantee a randomized and counterbalanced judgment (Cliff et al., 2007; Esti et al., 2010; Kallithraka et al., 2011). The results are typically processed with an ANOVA procedure and multivariate techniques like PCA and PLS. The statistical outcome is then integrated in a descriptive analysis of the product and presented by a radar plot, a PCA biplot or a PLS-DA correlation loading plot.

Consumer reaction testing involves the selection of a representative consumer panel. Multiple examples of its application concern the consumer's opinion of presented wines. In order to obtain sufficient statistical power, the panel consists often of more than fifty or even hundred members (Bindon et al., 2014). These panel members are untrained consumers and rather than appraising individual product properties, they are asked to make global judgements about the presented goods. This includes acceptance testing, discriminating product variants, providing hedonic ratings, indicating binary preference and preference ranking. According

to the dataset that is produced, the proper statistical procedure is applied to evaluate the targeted effect. They cover parametrized procedures, like discrimination testing, ANOVA and Chi squared based comparison, and non-parametrized techniques including the Mann-Withney U test, the Spearman's rank correlation coefficient and Friedmann two-way analysis of variance (Meilgaard et al, 2006; Kemp et al., 2009; Lawless and Heymann, 2010; Stone et al., 2012).

Although scientifically sound, both sensory profiling and consumer reaction testing focus not directly on the overall quality of wine and need therefore be adapted for that purpose. A sensory profiling based method, however, faces major technical and practical hurdles. The technical issues are highlighted by a wine quality rating method, developed by a team of twelve wine experts, presented by Etaio and his colleagues (Etaio et al., 2010). First, a consensus has to be attained to determine all relevant quality related sensory attributes. Second, the selected attributes should be clarified by a defining terminology, an unambiguous scoring system and one or more reference standards to illustrate the evaluation criteria and train the panel. The final problem to overcome is assigning the proper weights to each individual rating to calculate the overall sensory quality. The aforementioned Spanish method (Etaio et al., 2010) has selected the following quality defining parameters, which were rated on a 7-point scale, and relative weights that total to 100%: odour intensity 12%, odour complexity 18%, aroma or retronasal intensity 10%, aroma or retronasal complexity 15%, balance and body 25%, global aroma persistence 10%, colour hue 6%, colour intensity 4%. Both attributes and weights are defined by consensus of the panel rather than scientific experiments. Uniformity of the results is further obtained by averaging the individual scores and ignoring the values which differ more than 2 points from the mean as outliers. A criticism to this consolidating approach may well be that the individual variance is in itself an important factor in the evaluation of wine quality. The team used young red Rioja Alavesa wines to mature their method and suggested that the quality rating process might be specific for each wine type. The broad liking range of both wine experts and consumers during the appraisal of Californian Cabernet sauvignon wines supports this idea (Hopfer and Heymann, 2014). Although efforts were made to improve the defining capability of the proposed vocabulary, the correspondence between observation and score remains a sensitive area. In addition, the provision of appropriate standards for each property was incomplete in the Spanish experiment. Furthermore, in order to obtain statistically useful performance for judging only one parameter, selected panel members need to follow an intensive training program making this a time consuming and expensive operation. For wine quality, multiple parameters need to be judged which amplifies this already extensive effort. For the mentioned attributes body, balance and complexity, it might be very difficult, nearly impossible, to design an appropriate training program. Taking all these problems in account, more research is needed before parameter based

wine quality rating by a trained sensory panel can be even considered as the basis for a scientific assessment of the sensory quality of wine. In addition, such a method will dictate one particular view on sensory quality to everybody, and ignores the observable preference differences between individuals.

Consumer preference testing might be a better starting point. This comprehensive method with making relative comparisons is more appropriate for the sensory evaluation capabilities of humans. Furthermore, equating samples is less demanding in controlling test conditions and the unnecessity for training increases the applicability of this practice. However, three major issues need to be resolved. The first deals with the size of the panel. Normally, a large number of consumers, often more than 50, is required to obtain reliable data about the preference of a consumer population. This includes not only the recruitment of a considerable number of people, but might also require the organization of multiple tasting sessions. However, in order to be easily applicable, it is preferable to fit the evaluation in one event with a panel of limited extent. As the main goal of the QAMREC method is to assess sensory quality instead of estimating preference, this problem can be overcome using a much smaller expert consumer panel. Hereby, the assumption can be made that the consumer preference of a wine expert consumer panel, due to the experience of its members, approaches their assessment of the sensory quality of wine. The remaining influence of personal preference over sensory quality can be further reduced by presenting comparable wine samples which are similar in origin, style and major vinification aspects. Potential lacks of expertise in a wine group of one panellist is compensated by the diversity of wine experience of the other panel members. As only regularly served wine groups were presented, a possible lack of expertise of a panellist for a particular wine group was highly improbable. Temporary fatigue and adaptation of the human sensory ability and limited short-term memory performance is a second problem that requires attention. This can be handled by restricting the number of simultaneously presented samples for one panel member to four or less for mutual ranking. The overall rating of the panel results from a combination of the individual ranking results. The accuracy of the panel estimation and the total number of samples in the evaluation set can be increased by organizing multiple presentation rounds. The third obstacle concerns the scientific soundness of the method and requires the incorporation of statistical techniques based on the exploded logit model (Allison and Christakis, 1994; Johnson et al., 2008). This approach expresses the panel's preference for each wine as a utility value toward a reference wine, which could be used in all other sessions evaluating wines of the same wine group to obtain a completer image. This utility reflects the intrinsic quality of the wine assigned by the panel of which its members base their judgement on certain quality rules of which they are not consciously aware. It approximates the sensory quality of wine by a preference test in a population

represented by expert-consumers under the assumption that individual differences in preferences are removed as much as possible by restricting the sample space to wines of equal origin, style and vinification.

The objective of this work was to develop a framework for a practical, realistic and statistically sound method for determining the sensory quality of wine regarded as a matter of preference likelihood within a group of expert-consumers. The proposed method is entitled 'Quality Assessment by Merging Ranks of an Expert-Consumer Panel' and is further referred to with the abbreviation 'QAMREC'.

2 Materials and method

2.1 Wine samples

Ten QAMREC sessions, reflecting ten origins, were organized to determine the comparative sensory quality of nine wines per session. Specialized guides like "Le Guide Hachette", the "Gamberro Rosso" and the "Guía Peñin" or local wine contests including the 'Trophée des grands crus de Graves 2014' cannot afford to present bad wines and were consulted to ensure the selection of good wines. In future sessions, a wine of poor quality could be included as a negative control. The different wine groups were chosen to cover a wide range of wine types with regard to grape variety and vinification approach. With regard to European wines, this equivalence is bound up with its origin due to the corresponding appellation rules. However, using New World wines, where vinification style and origin is less tied, the principle of similarity can also be applied. The chosen wine groups comprised 5 white and 5 red dry wine types. The white wine sets included the AOP Pouilly Fumé 2012, the AOP Graves 2012, the DOC (G)'s Verdicchio di Castelli di Jesi and Verdicchio di Matelica, the AOP Pouilly Fuissé 2012 and dry Riesling wines from the Rheinpfalz. The red wine group consisted of the AOP Saint-Chinian, the AOP Moulis-en-Médoc 2010, the AOP Mercurey 2012, the Doc Rioja Reserva 2009 and the AOP Gigondas 2012. All wines were assigned a number with a randomizing procedure. S.1 lists the rated wines of all sessions.

2.2 The expert-consumer Panel

Members of the expert-consumer panel were recruited as to meet a least one of the following criteria: visiting a restaurant during which wine is consumed at least once a week, being an active member of one or more wine clubs and regularly participating in blind tastings, being a wine professional active in production, commerce, journalism, science, education or catering. To evaluate

159 the quality of the panel it was decided to determine the wine involvement profile (WIP) of our 20 panellists following the
160 methodology described by Bruwer et al (2014). This approach is essentially based on 13 scale items with which an individual can
161 agree or disagree on a 7 point scale. The WIP scale therefore theoretically ranges from 13-91, and the midscore of 52 is used to
162 segment the panellist into low- and high-involvement consumers.

163 The expert-consumer panel consisted of twelve members for each session. Although most panellists participated in multiple
164 sessions, the composition of the entire panels varied slightly across sessions. Twenty panellists, among them seven females,
165 participated in at least one session. The age groups were represented as follows: three between 30 and 40, two between 40 and
166 50, seven between 50 and 60, seven between 60 and 70 and one between 70 and 80. Six were frequent restaurant visitors, 13
167 were members of a wine club and eight were professionals. Each time, a random number was assigned to each panellist. The WIP
168 score of the panellist ranged between 56 and 86 (Supplementary table S.8), so that all of them could be considered as high-
169 involvement consumers.

170 ## 2.3 Tasting sessions

171 Because a human can only evaluate a limited number of samples in a short time, a variant of the randomized incomplete block
172 design described in ISO standard 29842 was applied to organize the wine tasting session (ISO 29842, 2011). Per session, exactly
173 nine different wines were tasted in three series of four wine samples per panellist. The design met the following criteria:

174 - In each session, each wine had to be presented at least once to every panel member;
175 - Each series of four wines included a replicate to monitor the repeatability of the experts' assessments;
176 - Each wine was used equally often in any given session;
177 - The number of times any pair of wines appears together in a series of four wines had to be as balanced as possible.

178 First, a set of suitable designs was created using a backtracking algorithm, which assigned wines to each expert and each series
179 such that the above criteria were met. Thereafter, the design that minimized the difference between the minimum and maximum
180 occurrence of all pairwise sample comparisons was selected in order to obtain a balanced and randomized presentation scheme.

181 In the resulting design (see S.4), during each session, 9 wines were judged by 12 panel members during 3 presentation rounds of 4
182 wines per presentation. All sessions were organized early during the evening at 7 pm and a diner was provided after the tasting

session. Further, each panel member evaluated each wine at least once and three of them twice, each wine was presented sixteen times and each replicate set four times, and all wines were compared with each other at least four and at most six times (see S.5).

The tasting sessions were carried out in tasting booths, preventing communication and visual contact between panel members. As the appreciation of the visual characteristics of the wine is part of the entire evaluation, the booths were illuminated with standard fluorescent white light. Before being served, the wines were stored in an incubator set to 8 °C (white wines) or 15 °C (red wines) to guarantee appropriate and equal consumption temperatures. The samples were prepared in a different room and were presented in ISO standardized wine glasses (ISO 3591, 1977). Water and bread was supplied to the panellists during the breaks between the presentation series to neutralize the palate and reduce carry-over effects.

Every panellist received all four wines of a series simultaneously. The different series were separated by a break of at least ten minutes. Before the start of each session, a brief presentation of about five minutes was given to the panel members informing about the generic geographic and vinification aspects of the wine group evaluated in the current session, the results of the previous session organized more than two months ago and the instructions for the tasting protocol. As this procedure was easy to perform, no training was required. Three cl of each sample, selected according to the calculated design, was presented to each panellist for evaluation and could be handled by the panellists in their favoured way. The panellists were only asked to rank the wine samples in each series according to their preference. The comparison operator '>' was used to indicate the preference relations between the wine samples. For example, '4>2' meant that the panellist preferred sample 4 above sample 2.'. Although there are models like the one of Breslow to handle tied rankings, they are mathematically very complex and mostly not handled by standard statistical software (Skrondal and Rabe-Hesketh, 2003). For this, the granting of equal ranks was not permitted. As each panellist evaluated three series of four wine samples, he contributed 3 ranks resulting into 36 incomplete rankings per session. The panellists were instructed not to make assumptions on the future quality of the wine and asked to base their rankings on the question 'which wine would I take home to consume this evening? To avoid fatigue, only one session was organised on a single day.

2.4 QAMREC converts incomplete rankings into wine utilities

The 36 incomplete rankings obtained by a QAMREC session were converted in a quality score for each of the nine wines. To this end, ordinal preference data needed to be transformed into a quantitative measure of wine quality. To do so in a statistically justified way, the rank-order logit model (also known as the exploded logit model), proposed by Punj and Staelin (Punj and Staelin,

1978), Beggs et al. (Beggs et al., 1981) and Chapman and Staelin (Chapman and Staelin, 1982) for analysing preference rankings in marketing and economics was used. The model was further developed by Hausman and Ruud (Hausman and Ruud, 1987). More recent applications of the model can be found in Allison and Christakis (Allison and Christakis, 1994), Kumar and Kant (Kumar and Kant, 2007) and Azucena and de-Magistris (Azucena and de-Magistris, 2016). The design of factorial experiments for estimating rank-order logit models was discussed by Vermeulen et al. (Vermeulen et al., 2011).

The rank-order logit model generalizes the conditional logit model and is quite similar to standard logistic regression. The model is based on the fact that a ranking of several items can be viewed as a series of choices, for each of which the standard conditional choice probability is derived. For example, suppose that a panellist evaluates four wine samples identified by the numbers 1, 2, 3 and 4, and that the panellist's preference is given by 3 > 4 > 1 > 2. Then this ranking can be viewed as a series of three choices:

1. Wine sample 3 is the preferred sample in the set {1, 2, 3, 4}.
2. Wine sample 4 is the preferred sample in the set of remaining wine samples, {1, 2, 4}.
3. Wine sample 1 is the preferred sample in the pair of remaining wine samples, {1, 2}.

The three successive choices can now be modelled using the standard conditional logit model. In that model, based on the 'Utility theory' (Aleskerov and Montjardin, 2002), every wine sample *i* evaluated is assumed to have a utility U_{ij}

$$U_{ij} = \mu_i + \varepsilon_{ij} \qquad \text{(Eq. 1)}$$

where μ_i represents the unknown actual utility or intrinsic quality of wine *i* and ε_{ij} is the random error made by panellist *j* when judging the intrinsic quality of wine *i*. The random errors are assumed to be Gumbel distributed (Ben-Akiva and Lerman, 1985). Applying the standard conditional logit model to the above example, stipulates that the probability that wine 3 is the most preferred one from the set {1, 2, 3, 4} for respondent *j* equals

$$P\left(U_{3j} > \max\left(U_{1j}, U_{2j}, U_{4j}\right)\right) = \frac{\exp(\mu_3)}{\exp(\mu_1) + \exp(\mu_2) + \exp(\mu_3) + \exp(\mu_4)} \qquad \text{(Eq. 2)}$$

while the probability that wine 4 is the most preferred one from the set {1,2,4} is

$$P\left(U_{4j} > \max\left(U_{1j}, U_{2j}\right)\right) = \frac{\exp(\mu_4)}{\exp(\mu_1) + \exp(\mu_2) + \exp(\mu_4)} \quad \text{(Eq. 3)}$$

229

230 and the probability that wine 1 is the most preferred one from the pair {1,2} is

$$P\left(U_{1j} > U_{2j}\right) = \frac{\exp(\mu_1)}{\exp(\mu_1) + \exp(\mu_2)} \quad \text{(Eq. 4)}$$

231

232 The rank-ordered logit model then assumes that the probability that respondent *j* ends up with the ranking 3 > 4 > 1 > 2 is equal to

$$L_{3412} = P\left(U_{3j} > \max\left(U_{1j}, U_{2j}, U_{4j}\right)\right) \times P\left(U_{4j} > \max\left(U_{1j}, U_{2j}\right)\right) \times P\left(U_{1j} > U_{2j}\right) \quad \text{(Eq. 5)}$$

$$L_{3412} = \frac{\exp\left(\mu_3\right)}{\exp(\mu_1) + \exp(\mu_2) + \exp(\mu_3) + \exp(\mu_4)} \times \frac{\exp\left(\mu_4\right)}{\exp(\mu_1) + \exp(\mu_2) + \exp(\mu_4)} \times \frac{\exp\left(\mu_1\right)}{\exp(\mu_1) + \exp(\mu_2)} \quad \text{(Eq. 6)}$$

233

234 The value L_{3412} is the likelihood of the ranking 3 > 4 > 1 > 2. It is a function of the intrinsic wine qualities μ_1, μ_2, μ_3 and μ_4. In a similar
235 fashion, the likelihood of all rankings of all panellists can be derived.

236 The rank-order logit model can be estimated using the maximum likelihood estimation technique (Aldrich, 1997; Mung, 2003). This
237 requires maximizing the total likelihood function with respect to μ_1, μ_2... μ_9. The values of μ_1, μ_2... μ_9 that maximize the total
238 likelihood are the maximum likelihood estimates. The maximum likelihood estimates are approximately normally distributed
239 (Allison and Christakis, 1994). The variances and covariances of the maximum likelihood estimates of the utilities can be obtained
240 from the asymptotic variance-covariance matrix, which is the inverse of the Fisher information matrix *F* (Ly et al., 2014). The
241 diagonal elements of the asymptotic variance-covariance matrix are the estimates' variances, while the off-diagonal elements are
242 the covariances between pairs of estimates. Using these variances and covariances, pairwise comparison between the wines can
243 be performed.

244 Since the type of wine studied is a 9-level categorical variable, most software packages treat one of the wine types as a reference
245 category. Which wine is treated as the reference is arbitrary (i.e., the model fit and the p-values of the statistical tests will not be

influenced by the choice), but it does affect the way in which all pairwise comparisons should be performed. Suppose, for example, that the ninth wine is the reference wine. This implies that the intrinsic quality of the ninth wine, μ_9, is set to zero when estimating the model. If the maximum likelihood estimates of μ_i and μ_j by $\hat{\mu}_i$ and $\hat{\mu}_j$ are denoted, respectively, then comparing any pair of the first eight wines requires using the test statistic

$$\frac{\hat{\mu}_i - \hat{\mu}_j}{\hat{\sigma}_{\cdot\,\hat{\mu}_i - \hat{\mu}_j}} = \frac{\hat{\mu}_i - \hat{\mu}_j}{\sqrt{\hat{\sigma}^2_{\hat{\mu}_i} + \hat{\sigma}^2_{\hat{\mu}_j} - 2\hat{\sigma}_{\cdot\,\hat{\mu}_j,\hat{\mu}_j}}} \qquad \text{(Eq. 7)}$$

where $\hat{\sigma}_{\hat{\mu}_i - \hat{\mu}_j}$, $\hat{\sigma}_{\hat{\mu}_i}$, $\hat{\sigma}_{\hat{\mu}_j}$ and $\hat{\sigma}_{\cdot\,\hat{\mu}_1,\hat{\mu}_2}$ are the standard error of the contrast $\hat{\mu}_i$ - $\hat{\mu}_j$, the standard error of $\hat{\mu}_i$, the standard error of $\hat{\mu}_j$ and the covariance between $\hat{\mu}_i$ and $\hat{\mu}_j$. Comparing the ninth wine with any one of the other wines, say the *i*th, is done using the test statistic

$$\hat{\mu}_i \Big/ \hat{\sigma}_{\hat{\mu}_i} \qquad \text{(Eq. 8)}$$

because μ_9 is set to zero. Each of these test statistics is approximately standard normally distributed, so that the resulting significance tests are *z*-tests. The maximum likelihood estimates μ_1, μ_2, ..., μ_9 allow the nine wines to be ranked from good to bad. The PHREG procedure (SAS, 2013), a statistical program in SAS Studio 3.2 designed to perform regression analysis of survival data (see S.2 and S.3), was used to estimate the intrinsic qualities of the wines studied and their standard errors based on the 36 incomplete rankings. These utilities represent the logarithm of the odds that the corresponding wine will be preferred over the lowest scoring wine in their group. For simplifying interpretation, the smallest utility was set to zero and the corresponding wine was considered as the reference wine; the other utilities were then expressed relative to this reference utility. A significantly higher value of the utility of a wine compared to that of the reference wine means that the panel considered this wine as better. The results of the pairwise comparisons using Wald tests and a significance level of 5% were summarized in a connecting letters report. Wines that do not share a letter in such a report have a significantly different intrinsic quality.

2.5 Panel member deviation and performance

To evaluate the ranking performance of each panellist, two different measures were used. The first measure quantified to what extent the ranking of an individual panellist influenced the ultimate ranking of the wines. For this measure, a Spearman rank correlation test compared the ranking results of the entire panel with those of the panel without the involved panellist and a coefficient threshold of 0.9 was applied to determine deviating panellists. The second measures quantified to what extent an individual panellist ranked the replicated wines within each series of four closer or further from each other. The second measure thus dealt with the internal consistency of the panellists.

For a certain respondent *i*, the first measure was computed by comparing the overall wine ranking obtained from all panellists to the ranking obtained after dropping the responses from respondent *i*. This comparison was made by calculating the Spearman's rank correlation coefficient between the ranking obtained from the entire panel and the ranking obtained after dropping panellist *i*'s responses. The lower this correlation coefficient, the more influential respondent *i* was and the more different his preferences were from the rest of the panellists. Ideally, all rank correlation coefficients should have been close to 1.

The second performance measure made use of the fact that this experimental design involved two identical wines in each series of four. Any good panellist should have either ranked the two wines as first and second, or as second and third, or as third and fourth. Consequently, the difference in rank between the two replicated wines should not have exceeded 1 in any of the series evaluated. To investigate which panellist was unable to achieve this ideal performance, the average difference in rank for the two identical wines over all series evaluated by any given respondent was calculated. Since the worst possible performance of a panellist was to assign rank 1 to one of the identical wine samples and rank 4 to the other, the maximum difference was 3.

3 Results

3.1 Quality Scores

Table 1 contains the utilities of the white wine sessions. The utilities varied from 0.00 to 0.74 and one letter group ($\alpha = 0.05$) in the Pouilly Fumé session (session 1); from 0.00 to 0.80 and one letter group in the Graves session (session 4); from 0.00 to 1.53 and two letter groups in the Verdicchio session (session 5), from 0.00 to 2.10 and four letter groups in the Pouilly Fuissé session (session 7) and from 0.00 to 2.36 and three letter groups in the Riesling session (session 9).

The results of the red wine sessions are summarized in Table 2. Their utilities varied from 0.00 to 3.42 and four letter groups in the Saint-Chinian session (session 2); from 0.00 to 3.96 and four letter groups in the Moulis-en-Médoc session (session 3); from 0.00 to 1.98 and three letter groups in the Mercurey session (session 6); from 0.00 to 1.53 and two letter groups in the Rioja session (session 8) and from 0.00 to 1.66 and two letter groups in the Gigondas session (session 10).

3.2 Panel member deviation and performance

Each panellist was identified by the abbreviation pm*n*, with *n* referring to the same panellist. The Spearman rank coefficient, obtained by comparing the complete wine ranking in a session of the entire panel with the panel where the panellist pmn was left out, is a measure for the deviance of that panel member in the envisaged session. These coefficients are outlined in Table 3. They ranged from 0.55 (panellist 9 in session 1) to 1. The general panellist deviance was obtained by averaging his coefficients over all the sessions. They varied from 0.78 to 0.96. The mean of the coefficients of each panel member in one session is a measure for the overall panellist deviance in that session. They ranged from 0.86 to 0.97.

The performance scores of each panellist are summed up in Table 4. They varied from 1 to 2.33. The mean of the performance values of a panel member over all sessions is a measure for the general error rate of each panellist. They ranged from 1.00 to 1.76. Averaging the performance values of all panellists in one session indicates the overall fault rate in that session, which varied from 1.28 to 1.56.

4 Discussion

4.1 The merits of QAMREC

QAMREC's discrimination resolution depends on the size of the wine set, the quality differences in the presented wine set, the size of the panel, the number of samples in each ranking and the number of rankings per panellist . As the judged wine collection was fixed, the discrimination power of a QAMREC test can only be improved by increasing the panel size, the number of samples per ranking and the number of rankings per panel member. These quantity settings are a trade-off between statistical optimization, practical considerations and human limitations. Using expert-panels with more than 20 panellists implies considerable logistic and organisational difficulties. Furthermore, human limitations restrict the number of samples per ranking and the number of rankings

per panellist. The applied configuration, in which 9 wines were judged by 12 panel members during 3 presentation rounds of 4 wines per presentation, can be considered as a balanced compromise.

A QAMREC session provided comparative quality scores (the utilities) of wines in a statistically sound way. Even if only good wines were evaluated, most sessions (sessions 2, 3, 5, 6, 7, 8, 9, 10) contain wines with significant quality differences. Only session 1 and 4 had only one letter group. Probably, the wines in these sessions were too similar to obtain significant differences applying this procedure. This problem could be solved by increasing the size of the panel. In any case, it can be expected that larger differences will be obtained if the whole quality range is included in a QAMREC session. Nevertheless, the lack of significant differences means that all wines are equally judged by the panel which is also information. In addition, it is easy and inexpensive to organize and feasible human beings. Processing of the results is based on standard statistical software accessible by everyone. Its scores may be used to evaluate the impact of agricultural and vinification interventions on the overall sensory quality of wine.

Three types of QAMREC outcomes can be used for that purpose. The first two are respectively the ordinal utilities and the complete ranks. A third result interpretation can be derived from the significantly different wines summarized in the connecting letters reports. For QAMREC sessions with three or more letter groups (sessions 2, 3, 6, 7 and 9), the wines belonging only to the first or last group without overlaps in other groups contain wines for which there is a big consensus among the panellists concerning their sensory quality. The non-overlapping wines in the first group could be considered as the best wines of that session, while the non-overlapping wines in the last group can be regarded as of clearly inferior quality than the others. In contrast, the quality evaluation of the middle group comprising the remaining overlapping wines depend more on the composition of the panel. The interpretation of the two letter group sessions (sessions 5, 8 and 10) is less straightforward and probably depends on the size of the non-overlapping letter groups.

The choice of the panellist is decisive for the targeted population which is in this experiment the educated consumer, both professional and not-professional. As a panel represents a certain population, their judgements can differ. We did not compare panels but its highly probably that the results of panels representing different populations, for example experts and non-experts, might diverge.

4.2 Assessing the panel of QAMREC sessions

QAMREC allowed assessing both the panellist deviance and performance. The deviance of a panel member from the entire panel in a particular tasting session is represented by a Spearman rank correlation coefficient expressing the similarity between the complete ranks of the entire panel and the complete ranks of the panel without the envisaged panellist. A Spearman rank correlation coefficient ≥ 0.9 indicates a very high correlation and a low rating divergence of the panellist towards the full panel. The closer its value approaches zero, the smaller the correlation and the larger the deviance. When only panellists were considered which participated in at least three sessions (pm1 to pm13 + pm17), only pm9 (0.89) had a mean coefficient lower than 0.9. This was, however, mainly due to his extremely low coefficient (0.55) in session 1 which could be considered as an outlier, especially as there are no significantly different groups in that session. In general, no systematic divergent rating behaviour was noted for a particular panellist over all sessions. This is also illustrated by the frequency distribution of all coefficients illustrated in Fig.1. In addition, the mean coefficient per session can be considered as an indication for the consensus level of the panel in the session. The agreement of the panel was highest in the Moulis-en-Médoc (0.97), the Pouilly Fuissé session (0.97) and the Mercurey session (0.96). The lower mean coefficients for the Rioja session (0.86), the Pouilly Fumé session (0.87) and the Verdicchio session (0.87) suggest less accordance among the panellist in these sessions.

The performance of a panellist is described by a performance value between 1 and 3. The higher this value, the more errors are made. A value of 1.67, allowing two small or one large misclassification per panellist in one session, could function as a threshold for identifying bad performing panellists and compromised sessions. Using this criterion and considering only panellists who participated at least three times, only the performance of pm8 (1.76) is questionable. No sessions had a mean performance that surpassed the threshold value of 1.67. The error rate was lowest in the Verdicchio session (1.28) and highest for the Graves and Gigondas sessions (1.56). This is illustrated by the histograms in Fig.2 and Fig.3. A higher session error score might also suggest that the wines in the corresponding sessions were more similar.

The ‘leave one out’ technique discloses how much a panellist deviates from the global panel. In addition, the consequent application of replicates offers the possibility to measure the performance of each individual panel member. However, it is not advisable to communicate this information to the panel as it might increase the risk that members will alter their rating behaviour in an attempt to avoid being a bad performer or outlier. It is preferable to convince panellists that the method is robust against

individual errors and that QAMREC scores are the result of the panel as a group. Only when a panel member regularly demonstrates poor performance and deflection over multiple sessions, it might be desirable to intervene in the panel's composition.

4.3 QAMREC scores and guide ratings

Literature suggests two perspectives on quality. In the first view, the quality is considered good if the product's properties comply with predefined conditions (Crosby, 1979). Such a binary approach distinguishes the good wines from the ones that do not meet the required criteria, but provides no information about the wine's excellence. Considering quality as 'the degree fitness for purpose', where fitness is defined by the consumer, is probably better suited for QAMREC sessions (Juran and De Feo, 2010). However, as every panellist has his personal sensory preferences, the resulting utility of a wine cannot be used to predict the evaluation behaviour of an individual. Instead, QAMREC results only reflect the dominant trend of the population represented by the panel. The choice of the panellist is decisive for the targeted population which is in this experiment the educated consumer, both professional and not-professional. This favourizes a group-based quality classification approach above representing the sensory quality by one number. With this in mind, a guide using four quality categories (a citation, one star, two stars, three stars) like 'The Guide Hachette" is probably more appropriate than the ones using values to indicate the wine's sensory quality like "The Guía Peñin", which scores up to 100 %.

From a review of the six sessions (1, 2, 3, 7, 8 and 10) that allow for a full comparison (see S.6 and S.7), it is obvious that QAMREC scores differed from the associated guide evaluations. Although some of the guide ratings were preserved in QAMREC (equal), there were also notable differences (under- or overestimated by the guide compared with the QAMREC results). These differences could be partially explained by the parameter based system that is mostly applied to obtain guide ratings and might overvalue or depreciate certain sensory characteristics. Moreover, such a parameter based system suggests that every panellist uses the same criteria to evaluate the sensory quality of wine, which might not be the case. Furthermore, each guide panellist needs to evaluate often more than 15 wines in a short time which may exceed human capabilities and possibly results in a significant influence of the presentation order. By selecting only wine professionals, the perceptions of the educated customer are excluded from the guide ratings, which might also produce different results. In addition, local jury members are probably more tolerant for borderline sensory properties than a more generic audience. In this context, higher levels of minerality in Pouilly Fumé wines, some reductive aromas in Saint-Chinian and Mercurey wines and a more profound tannin structure in Moulis-en-Médoc wines might be considered

as qualities by locals but are less appreciated by the average taster. Finally, the judges might also integrate assumed future quality projections in their ratings, which is explicitly not the case in QAMREC sessions. In any case, those differences show that the results partly depend on the applied method.

Wine contests and guides typically deal with a large number of wines which exceeds the number of wine samples evaluated in the various sessions of this experiment. Increasing the numbers of wines and the size of the panel might seem the proper approach but will comprise major practical challenges and probably surpass human tasting capabilities. First, a checklist could be used to select the good wines. Secondly, an appropriate design, using the same reference wines over multiple sessions, should be developed to bundle a series of QAMREC sessions into an overall quality rating when dealing with larger quantities of good wines.

5 Conclusion

What is considered as the sensory quality of wine? Most wine rating systems see quality as the sum of weighted ratings of a set of predefined properties. QAMREC on the other hand considers the sensory quality of wine as a matter of preference and can be considered as a promising approach to objectively assess this likelihood view of a wine-educated population on sensory quality of wines with sufficient resolution. However, as preference and sensory quality are intrinsically different concepts, QAMREC, which is essentially a preference measurement technique, can only narrow the gap between preference and sensory quality if the appraised samples are comparable from a sensory point of view and the large enough rating panel comprises sufficient expertise. QAMREC meets the required statistical significance for being employed in scientific experiments. Clearly the homogeneity of the panel will determine the variability of the utility or intrinsic quality scores and, therefore, their usefulness – a heterogeneous panel will likely not be able to produce a meaningful ranking of the wines. However, as the method is generic and has a sound statistical basis, this would naturally be revealed as a lack of significant differences between the wines. Future trials may assess the effect of agricultural, production and storage operations and conditions on the overall sensory quality of wine with QAMREC. A series of QAMREC sessions, presenting the same reference wines, can also be clustered to compute the sensory quality of wine in wine contests and guides.

It should be clear that, although QAMREC is developed to tackle the problem of scientific wine rating, the method is not restricted to wine only. It may be applied for estimating the sensory quality of any complex product of which the different quality parameters and their weights are unknown.

6 List of tables

7 List of figures

A. Tables

Table 1.Results of the white wine sessions. Utilities, standard errors (SE) and odds of each wine compared to a reference wine for all white wine sessions. The wines are presented in descending utility order. The reference wine of each session is the wine with utility equal to 0. Wines belonging to the same column (labelled 'Gx') are not significantly different ($\alpha = 0.05$).

Session 1. Pouilly Fumé 2012								
ID	**Wine**	**Utility**	**Odds**	**SE**	**GA**	**GB**	**GC**	**GD**
2	Tabordet 2012	0.74	2.09	0.50	A			
4	Bailly 2012	0.67	1.95	0.51	A			
6	Bardin 2012	0.64	1.90	0.50	A			
3	Eclat 2012	0.58	1.78	0.52	A			
5	Tracy principale 2012	0.56	1.76	0.51	A			
8	Séguin 2012	0.48	1.62	0.53	A			
1	Rabichattes 2012	0.20	1.22	0.51	A			
7	Champeau principale 2012	0.10	1.10	0.53	A			
9	Séguin Prestige 2012	0.00	1.00		A			
Session 4. Graves 2012								
ID	**Wine**	**Utility**	**Odds**	**SE**	**GA**	**GB**	**GC**	**GD**
6	Pont de Brion 2012	0.80	2.22	0.55	A			
4	Haut Selve 2012	0.78	2.17	0.56	A			
7	Gaubert 2012	0.62	1.86	0.54	A			
5	Bourgelat cuvée Caprice 2012	0.50	1.65	0.53	A			
8	Villa Bel Air 2012	0.33	1.39	0.51	A			
2	La Rose Sarron 2012	0.23	1.26	0.55	A			
3	Chantegrive 2012	0.23	1.26	0.56	A			
9	Floridène 2012	0.11	1.12	0.54	A			
1	Des Places 2012	0.00	1.00		A			
Session 5. Verdicchio								
ID	**Wine**	**Utility**	**Odds**	**SE**	**GA**	**GB**	**GC**	**GD**
9	Balciana 2011	1.53	4.61	0.58	A			
5	Terravignata 2012	0.84	2.32	0.53	A	B		
3	Cambrugiano 2011	0.80	2.23	0.57	A	B		
8	Villa Bucci riserva 2010	0.78	2.19	0.55	A	B		
1	Pallio di San Floriano 2013	0.78	2.17	0.56	A	B		
4	Mirum 2012	0.69	1.99	0.56	A	B		
2	Ylice 2012	0.38	1.46	0.57	A	B		
7	Alarico 2013	0.37	1.45	0.58		B		
6	Colle Stefano 2013	0.00	1.00			B		
Session 7. Pouilly Fuissé 2012								
ID	**Wine**	**Utility**	**Odds**	**SE**	**GA**	**GB**	**GC**	**GD**
8	Vessigaud Veilles Vignes 2012	2.10	8.17	0.64	A			
9	Corsin 2012	1.66	5.26	0.63	A	B		
3	Soufrandise 2012	1.44	4.22	0.65	A	B	C	
4	Sève 2012	1.39	4.03	0.61	A	B	C	

6	Le Manoir du Capucin Aux Morlays 2012	0.98	2.67	0.61	A	B	C	D
5	Chateau de Vergisson 2012	0.90	2.45	0.59		B	C	D
7	Feuillarde Veilles Vignes 2012	0.42	1.52	0.61			C	D
1	Chateau de Chaintré 2012	0.21	1.24	0.62				D
2	Chateau de Lavernette 2012	0.00	1.00					D

Session 9. Riesling Pfalz 2013

ID	Wine	Utility	Odds	SE	GA	GB	GC
1	Forster Elster 2013	2.36	10.54	0.71	A		
3	Müller-Catoir Haardt 2013	1.13	3.08	0.57	A	B	
6	Von Winning Grainhübel 2013	0.85	2.35	0.55		B	C
7	Koehler-Ruprecht Saumagen 2013	0.71	2.04	0.58		B	C
4	Bürklin-Wolf Wachenheimer 2013	0.43	1.54	0.58		B	C
2	Grosser Durst 2013	0.26	1.29	0.60		B	C
9	Wehrheim Kastanienbusch 2013	0.07	1.07	0.56		B	C
5	Odinstal 350 NN 2013	0.05	1.05	0.59		B	C
8	Knipser Steinbückel 2013	0.00	1.00				C

423

Table 2.Results of the red wine sessions. Utilities, standard errors (SE) and odds of each wine compared to a reference wine for all red wine sessions. The wines are presented in descending utility order. The reference wine of each session is the wine with utility equal to 0. Wines belonging to the same column (labelled 'Gx') are not significantly different ($\alpha = 0.05$).

Session 2. Saint-Chinian								
ID	**Wine**	**Utility**	**Odds**	**SE**	**GA**	**GB**	**GC**	**GD**
8	Best of Belot 2011	3.42	30.46	0.76	A			
1	Cuvée de Penelle 2011	2.58	13.22	0.72	A	B		
7	Karrimour 2011	2.36	10.61	0.70	A	B	C	
6	Les Schistes 2011	2.29	9.91	0.72	A	B	C	
9	La Sentenelle 310 2011	2.17	8.80	0.72	A	B	C	
2	Maurerie Veilles Vignes 2011	2.02	7.57	0.75		B	C	
3	Prieuré des Mourges Tradition 2009	1.31	3.70	0.69			C	D
5	Haut Coup De Foudres 2010	0.47	1.61	0.69				D
4	Servelière Tradition 2011	0.00	1.00					D
Session 3. Moulis-en-Médoc 2010								
ID	**Wine**	**Utility**	**Odds**	**SE**	**GA**	**GB**	**GC**	**GD**
8	Branas Grand Poujeaux 2010	3.96	52.54	0.86	A			
5	Poujeaux 2010	2.77	16.04	0.71	A	B		
7	Chemin Royal 2010	1.93	6.90	0.68		B	C	
4	Bouqueyran 2010	1.72	5.59	0.67		B	C	
1	Lestage Darquier 2010	1.60	4.94	0.67			C	
3	Granins Grand Poujeaux 2010	1.44	4.21	0.65			C	
9	Myon de L Enclos 2010	1.27	3.55	0.65			C	
6	La Mouline 2010	1.04	2.83	0.67			C	D
2	Pomeys 2010	0.00	1.00					D
Session 6. Mercurey 2012								
ID	**Wine**	**Utility**	**Odds**	**SE**	**GA**	**GB**	**GC**	
8	Michel Juillot 1e cru Clos des Barraults 2012	1.98	7.25	0.66	A			
6	Berthoux Les Chavances 2012	1.75	5.76	0.64	A			
9	Guillot 1e cru Les Veley 2012	1.67	5.29	0.63	A			
7	De la Monette 2012	1.31	3.72	0.65	A	B		
2	Milan 1e Cru Les Crets 2012	1.27	3.57	0.64	A	B		
4	G. Clos de la Charmée 2012	1.21	3.36	0.62	A	B	C	
1	G. et J. Meunier 1e cru 2012	1.02	2.77	0.62	A	B	C	
3	Theulet Juillot 1e cru Les Combins 2012	0.39	1.48	0.64		B	C	
5	Vincent Meunier 1e cru C. d. F. 2012	0.00	1.00				C	
Session 8. Rioja Reserva 2009								
ID	**Wine**	**Utility**	**Odds**	**SE**	**GA**	**GB**		
9	Remelluri Reserva 2009	1.53	4.60	0.58	A			
3	Caecus tinto reserva 2009	1.24	3.47	0.59	A			
6	Viña Pomal cent. reserva 2009	0.85	2.35	0.57	A	B		
8	Gaudium Gran Vino reserva 2009	0.72	2.05	0.56	A	B		
1	Imperial tinto reserva 2009	0.64	1.89	0.56	A	B		
2	Murrieta reserva 2009	0.57	1.76	0.55	A	B		

7	La Vicalanda Reserva 2009	0.56	1.74	0.54	A	B
5	Gonzalo De Berceo Reserva 2009	0.55	1.73	0.55	A	B
4	Ijalba Reserva 2009	0.00	1.00			B
Session 10. Gigondas 2012						
ID	**Wine**	**Utility**	**Odds**	**SE**	**GA**	**GB**
4	Cuvée Costeveille 2012	1.66	5.27	0.59	A	
3	Tourbillon 2012	1.53	4.64	0.60	A	
7	Cuvée de Beauchamps 2012	1.44	4.21	0.59	A	
8	Gour de Chaulé 2012	1.36	3.88	0.59	A	
6	Bouïsssière 2012	1.34	3.82	0.60	A	
5	Cuvée Cécile 2012	0.97	2.64	0.59	A	B
2	Combe Sauvage 2012	0.85	2.33	0.57	A	B
9	Terrasses de Montmirail 2012	0.72	2.06	0.61	A	B
1	Coteau de mon rève 2012	0.00	1.00			B

427

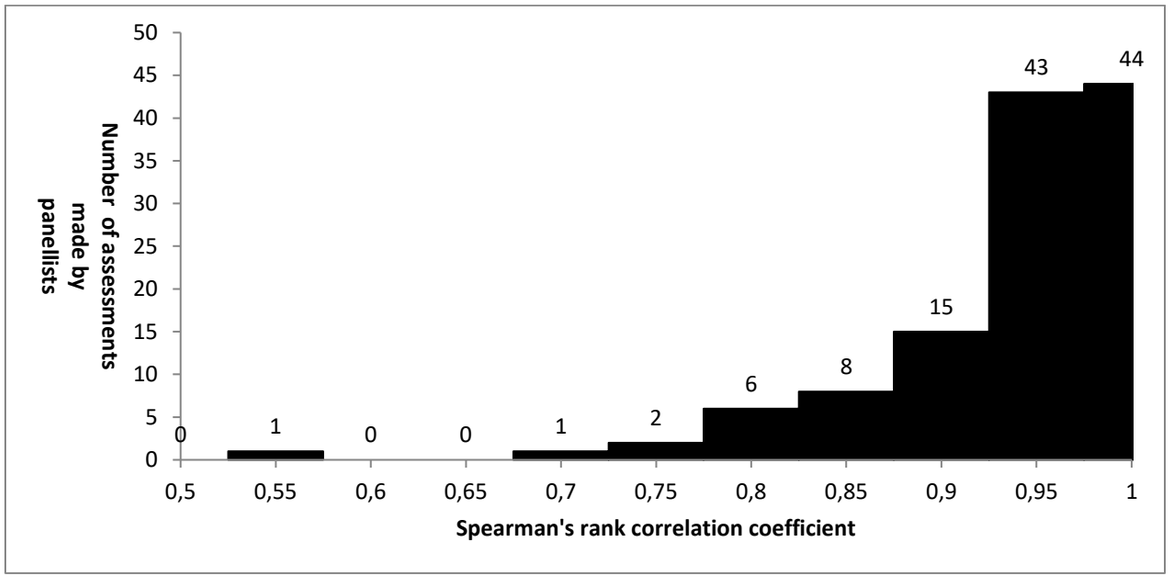
Table 3.Panel member deviance expressed by Spearman rank coefficients. For each panellist and session, the Spearman rank correlation coefficients, obtained by comparing the complete ranks of the entire panel with those of the panel where the mentioned panel member was left out, is shown. The left out panellist is identified by the abbreviation pm followed by a number and represents the same person. In addition, the mean coefficients per panellist and session are also displayed.

	Pouilly Fumé	**Saint-Chinian**	**Moulis-en-Médoc**	**Graves**	**Verdicchio**	**Mercurey**	**Pouilly Fuissé**	**Rioja**	**Riesling**	**Gigondas**	**Mean**
pm1	0.95	0.88	0.95	0.82	0.93		0.98		0.95	0.9	**0.92**
pm2	0.93	0.92	0.95	0.82			0.98	0.87	0.98		**0.92**
pm3	0.8	0.98	0.98	0.95	0.77	1	0.97	0.8	0.95	0.93	**0.91**
pm4	0.83	0.88		0.95	0.83	0.98	0.97	0.93	0.93	0.97	**0.92**
pm5	0.9	0.9	0.98	0.95		0.95	0.98	0.87	0.97	0.98	**0.94**
pm6	0.93	1	0.98			0.98		0.75		0.92	**0.93**
pm7	0.87	0.98	1	0.93	0.8	0.95	0.97	0.93	0.9	0.95	**0.93**
pm8	0.97	0.93	1	0.97		0.97		0.92	0.92		**0.95**
pm9	0.55	0.97	0.95	0.92	0.85	0.93	0.97	0.82	0.97	0.98	**0.89**
pm10	0.8	0.88	0.97	0.93	0.87	0.95	0.98	0.93	0.93	0.72	**0.90**
pm11	0.97		0.97	0.97		0.85	0.95	0.7	0.97	0.98	**0.92**
pm12	0.98	1	0.97	0.88	0.98	1	0.97	0.93	0.97	0.92	**0.96**
pm13		1		0.92	0.93				0.93	0.93	**0.94**
pm14					0.78						**0.78**
pm15					0.9		0.93				**0.92**
pm16					0.92						**0.92**
pm17					0.88	0.95	0.98			0.87	**0.92**
pm18						0.95					**0.95**
pm19			0.92								**0.92**
pm20								0.82			**0.82**
Mean	**0.87**	**0.94**	**0.97**	**0.92**	**0.87**	**0.96**	**0.97**	**0.86**	**0.95**	**0.92**	

Table 4.Panel performance per session. For each panellist and session, the performance scores are shown. These values vary from 1.00 (no errors) to 3.00 (maximum errors). The involved panellist is identified by the abbreviation pm followed by a number and represents the same person. In addition, the mean performance scores per panellist and session are also shown.

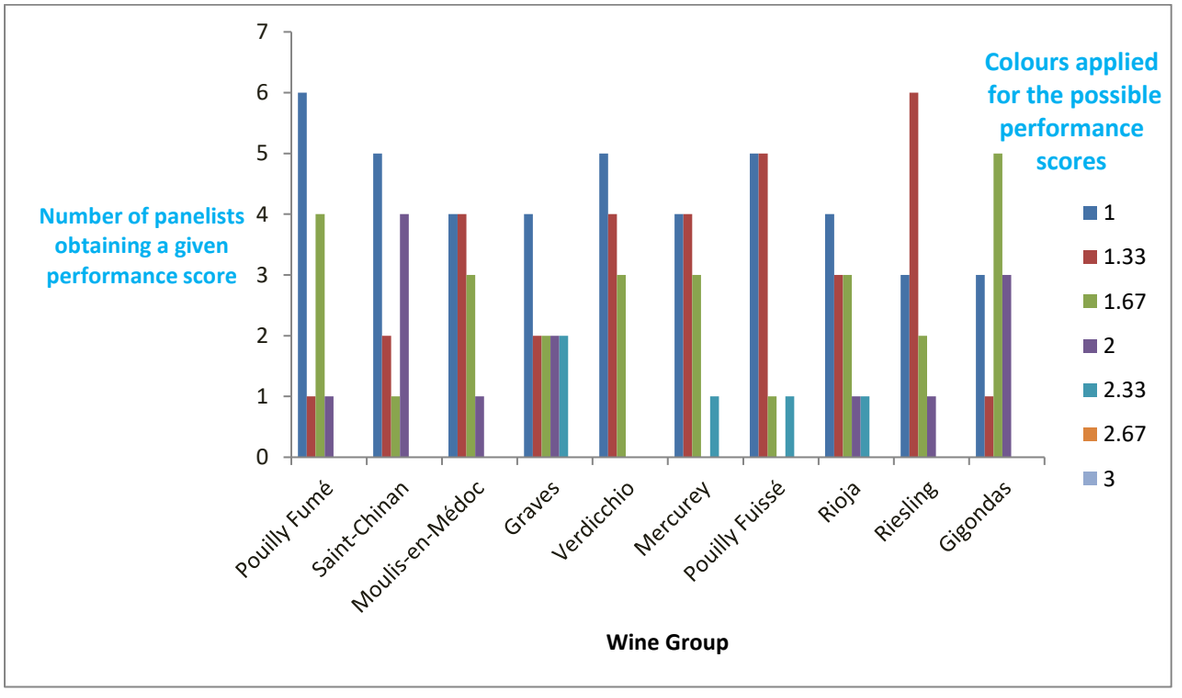
	Pouilly Fumé	Saint-Chinian	Moulis-en-Médoc	Graves	Verdicchio	Mercurey	Pouilly Fuissé	Rioja	Riesling	Gigondas	Mean
pm1	2.00	1.00	1.00	1.00	1.00		1.00		1.33	2.00	***1.29***
pm2	1.00	1.67	1.00	1.67			1.00	1.67	1.33		***1.33***
pm3	1.67	2.00	1.33	1.33	1.33	1.00	1.00	1.33	1.00	1.00	***1.30***
pm4	1.67	1.00		2.00	1.00	1.33	1.33	1.33	2.00	2.00	***1.52***
pm5	1.00	2.00	1.33	2.33		1.00	1.33	1.67	1.33	1.00	***1.44***
pm6	1.67	1.33	2.00			1.67		1.00		1.67	***1.53***
pm7	1.00	2.00	1.00	1.33	1.00	1.67	1.33	2.33	1.33	1.67	***1.47***
pm8	1.00	2.00	1.67	2.33		2.33		1.67	1.33		***1.76***
pm9	1.00	1.00	1.67	1.67	1.67	1.33	1.00	1.00	1.00	1.67	***1.30***
pm10	1.00	1.33	1.00	2.00	1.33	1.00	1.33	1.33	1.33	1.00	***1.27***
pm11	1.67		1.67	1.00		1.00	1.00	2.00	1.67	2.00	***1.50***
pm12	1.33	1.00	1.33	1.00	1.67	1.67	1.67	1.00	1.00	1.67	***1.33***
pm13		1.00		1.00	1.00				1.67	1.33	***1.17***
pm14					1.33						***1.33***
pm15					1.00		2.33			1.67	***1.67***
pm16					1.33						***1.33***
pm17					1.67	1.33	1.33				***1.44***
pm18						1.33					***1.33***
pm19			1.33								***1.33***
pm20								1.00			***1.00***
Mean	**1.33**	**1.44**	**1.36**	**1.56**	**1.28**	**1.39**	**1.30**	**1.44**	**1.36**	**1.56**	

438 B. Figures



439

440 Fig.1. *Histogram of all Spearman`s rank correlation coefficients obtained by all panellists in all sessions.*



441

442 Fig.2. *Histogram displaying the number of panellists obtaining a given performance in each session. Possible values were 1 (no*
443 *errors), 1.33, 1.67, 2 2.33, 2.67 and 3 (maximum errors).*

444

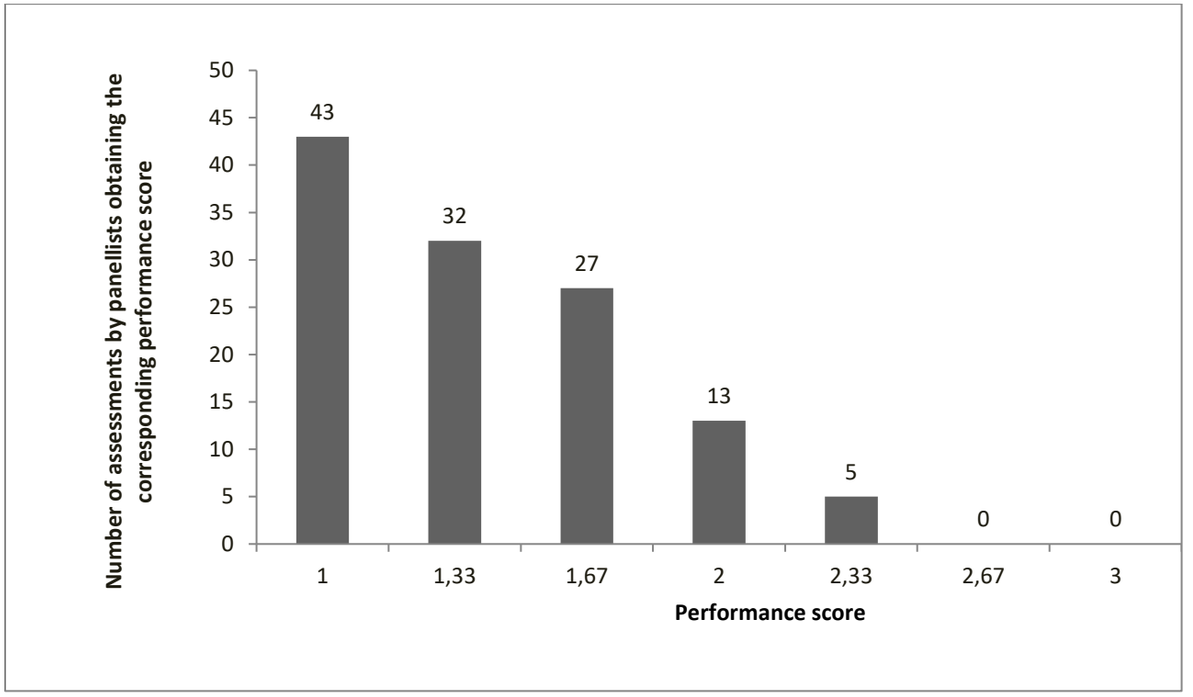


Fig.3. *Histogram of the number of assessments by panellists obtaining a given performance score over all sessions. Possible performance values were 1 (no errors), 1.33, 1.67, 2 2.33, 2.67 and 3 (maximum errors).*

C. References

Aldrich, J. (1997). R.A. Fisher and the Making of Maximum Likelihood 1912 – 1922. Statistical Science, 3, 162-176.

Aleskerow, F., Monjardet ,B. (2002). Utility Maximization, Choice and Preference. Heidelberg: Springer Verlag.

Allison, P.D., Christiakis, N.A., (1994). Logit Models for Sets of Ranked Items. Sociological Methodology, 24, 199-228.

Barth F.G., Humphrey J.A.C., Srinivasan M.V. (2012). Frontiers in Sensing, From Biology to Engineering, Springer-Verlag/Wien.

Beggs, S.,Cardell, S., Hausman J. (1981). Assessing the potential demand for electric cars. Journal of Econometrics, 17, 1-19.

Ben-Akiva, M., Lerman, S.R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand (Transportation Studies), The MIT press.

Bindon, K., Holt, H., Williamson, P.O., Varela, C., Herderich, M., Francis, I.L. (2014). Relationships between harvest time and wine composition in Vitis vinifera L. cv. Cabernet Sauvignon 2. Wine sensory properties and consumer preference. Food Chemistry, 154, 90-101.

Bruwer J., Burrows N., Chaumont S., Li E., Saliba A. (2014). Consumer involvement and associated behaviour in the UK high-end retail off-trade wine market. The International Review of Retail, Distribution and Consumer Research, 24-2.

Callejon, R.M., Clavijo, A., Ortigueira, P., Troncoso, A.M., Paneque, P., Morales, M.L. (2010). Volatile and sensory profile of organic red wines produced by different selected autochthonous and commercial Saccharomyces cerevisiae strains. Analytica Chimica Acta, 660, 68-75.

Cetó, X., González-Calabuig, A., Capdevilla, J., Puig-Pujol, A., Del Valle, M. (2015). Instrumental measurement of wine sensory descriptors using a voltammetric electronic tongue. Sensors and Actuators B: Chemical, 207, 1053-1059.

Chapman, R.G., Staelin, R. (1982). Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model. Journal of Marketing Research, 19, 288-301.

Cliff, M.A., King, M.C., Schlosser J. (2007). Anthocyanin, phenolic composition, colour measurement and sensory analysis of BC commercial red wines. Food Research International, 40, 92-100.

471 Crosby P. (1979). Quality is free, the art of making quality certain. New York: McGraw-Hill.

472 D'Allessandro, S., Pecotich, A. (2013). Evaluation of wine by expert and novice consumers in the presence of variations in quality,
473 brand and country of origin cues. Food Quality and Preference, 28, 287-303.

474 Decanter. (2015). http://www.decanter.com/ Accessed 20/2/2015

475 Esti, M., Airola, R.L.G., Monetta, E., Paperaio, M., Sinesio, F. (2010). Qualitative data analysis for an exploratory sensory study of
476 grechetto wine. Analytica Chimica Acta, 660, 63-67.

477 Etaio, I., Albisu, M., Ojeca, M., Gil, P.F., Salmerón, J., Elortondo, F.J.P. (2010). Sensory quality control for food certification: A case
478 study on wine. Method development. Food Control, 21, 533-541.

479 Fabrizio, G., Guerini , E., Sabellico, M. (2015). Gamberro Rosso, Italian Wines. Easthampton: Gamberro USA.

480 Gracia A., de-Magistris, T. (2016). Consumer preferences for food labeling: What ranks first?. Food Control, 61, 39-46.

481 Grupo Peñin. (2015). Guía Peñin de los vinos de España. Madrid:Pi & Erre comunicacion S.A.

482 Guillaumie S., Ilg A., Réty S., Brette M., Trossat-Magnin C., Decroocq S., Léon C., Keime C., Ye T., Baltenweck-Guyot R., Claudel P.,
483 Bordenave L., Vanbrabant S., Duchêne E., Delrot S., Darriet P., Hugueney P., Gomès E. (2013). Genetic Analysis of the Biosynthesis
484 of 2-Methoxy-3-Isobutylpyrazine, a Major Grape-Derived Aroma Compound Impacting Wine Quality. American Society of Plant
485 Biologists, 162, 604–615.

486 Haarmann, H., Usher, M. (2001). Maintenance of semantic information in capacity-limited item short-term memory. Psychonomic
487 Bulletin & Review, 8(3), 568-578.

488 Hachette Pratique. (2014). Le Guide Hachette des Vin. Paris:Hachette Livre.

489 Harbertson, J.F., Parpinello, G.P., Heyman, H., Downey, M.O. (2012). Impact of exogenous tannin additions on wine chemistry and
490 wine sensory character. Food Chemistry, 131, 999-1008.

491 Hausman, J. A., Ruud, P. A. (1987). Specifying and testing econometric models for rank-ordered data. Journal of Econometrics,
492 Elsevier, 34(1-2), 83-104.

493 Hirsh J., Mar R., Peterson J. (2013). Personal narratives as the highest level of cognitive integration. Behavioral and Brain Sciences,
494 Jun 2013, 36(3), 216-223.

495 Hopfer, H., Heymann, H. (2014). Judging wine quality: Do we need experts, consumers or trained s?. Food Quality and Preference,
496 32, 221-233.

497 ISO 3591. (1977a). Sensory analysis — Apparatus — Wine-tasting glass.

498 ISO 13299. (2003b). Sensory analysis, Methodology, General guidance for establishing a sensory profile.

499 ISO 6658. (2005c). Sensory analysis, Methodology, General guidance.

500 ISO 5492. (2008d). Sensory analysis, Vocabulary.

501 ISO 29842. (2011e). Sensory analysis, Methodology, Balanced incomplete block designs.

502 Jackson, R.S (2002). Wine Tasting, a Professional Handbook. London:Elsevier Ltd.

503 Johnson, J., Bruse, A., Jiejun, Y. (2008). The ordinal efficiency of betting markets: an exploded logit approach. Applied Economics,
504 42:29, 3703-3709.

505 Juran, J.M., De Feo J.A. (2010). Juran's Quality Handbook: The Complete Guide to Performance Excellence (edition 6). New York:
506 McGraw-Hill.

507 Kallithraka, S., Kim, D., Tsakiris, A., Paraskevopoulos, I., Soleas, G. (2011). Sensory assessment and chemical measurement of
508 astringency of Greek wines: Correlations with analytical polyphenolic composition. Food Chemistry, 126, 1953-1958.

509 Kemp, S.E., Hollowood, T., Hort, J. (2009). Sensory Evaluation, a Practical Handbook. Oxford: Wiley-Blackwell.

510 Kumar, S., Kant, S. (2007). Exploded logit modeling of stakeholders' preferences for multiple forest values. Forest Policy and
511 Economics, 9, 516-526.

512 La revue du Vin de France. (2015). http://www.larvf.com/ Accessed 20/2/2015

513 Lawless, H.T., Heymann, H. (2010). Sensory evaluation of food, Principles and Practices, second Edition. New York: Springer New
514 York.

515 Ly, A., Verhagen, J., Grasman, R, Wagenmakers, E-J. (2014). A Tutorial on Fisher Information. University of Amsterdam, 55 p.

516 Meilgaard, M.C., Carr, B.T., Civille, G.V. (2006). Sensory Evaluation Techniques (Fourth Edition). Boca Raton: CRC Press.

517 Myung, I.J. (2003). Tutorial on maximum likelihood estimation. Journal of Mathematical Psychology, 47, 90-100.

518 Niedenthal P., Kitayama, S. (1994). The heart's eye: Emotional influences in perception and attention. San Diego (Calif.): Academic
519 press.

520 Parker, R. (2015). The Wine Advocate Rating System. https://www.erobertparker.com/info/legend.asp Accessed 10/2/2015.

521 Parker, M., Osidacz, P., Baldock, G.A., Hayasaka, Y., Black, C.A., Pardon, K.H., Jeffery, D.W., Geue, J., Herderich, J., Francis, I.L. (2012).
522 Contribution of Several Volatile Phenols and Their Glycoconjugates to Smoke-Related Sensory Properties of Red Wine. Journal of
523 Agricultural and Food Chemistry, 60, 2629-2637.

524 Punj, G.N., Staelin, R. (1978), The Choice Process for Graduate Business Schools. American Marketing Association, 15:4, 588-598.

525 Sáenz-Navajas, M-P., Campo, E., Fernández-Zurbano, P., Valentin, D., Ferreira, V. (2010). An assessment of the effects of wine
526 volatiles on the perception of taste and astringency in wine. Food Chemistry, 121, 1339-1149.

527 SAS Institute Inc. (2014). SAS/STAT® 13.2 User's Guide. Cary, NC: SAS Institute Inc.

528 Shepperd M.G. (2006). Smell images and the flavour system in the human brain. Nature, 316-321

529 Skrondal A., Rabe-Hesketh S. (2003). Multilevel Logistic Regression For Polytomous Data And Rankings. Psychometrika, 68(2), 267-
530 287.

531 Stone, H., Bleibaum, R., Thomas, H.A. (2012). Sensory Evaluation Practices (Fourth Edition). Waltham: Academic Press.

532 The International Wine challenge. (2015). http://www.internationalwinechallenge.com/ Accessed 10/2/2015.

533 Vermeulen B., Goos P., Vandebroek M. (2011). Rank-order choice-based conjoint experiments: efficiency and design. Journal of
534 Statistical Planning and Inference, 141:8, 2519 - 2531.

D. Supplementary materials

S.1. List of evaluated wines

Dry white wines					
ID	**Pouilly Fumé 2012** Sauvignon blanc No barrel aging	**Graves 2012** Sauvignon blanc, Sémillon and optionally Muscadelle and/or Sauvignon gris Optional barrel aging	**Verdicchio** Verdiccchio optional barrel aging	**Pouilly Fuissé 2012** Chardonnay optional barrel aging	**Riesling Pfalz 2013** Riesling no barrel aging
1	Rabichattes 2012	Des Places 2012	Pallio di San Floriano 2013	Chateau de Chaintré 2012	Forster Elster 2013
2	Tabordet 2012	La Rose Sarron 2012	Ylice 2012	Chateau de Lavernette 2012	Grosser Durst 2013
3	Eclat 2012	Chantegrive 2012	Cambrugiano 2011	Soufrandise 2012	Müller-Catoir Haardt 2013
4	Bailly 2012	Haut Selve 2012	Mirum 2012	Sève 2012	Bürklin-Wolf Wachenheimer 2013
5	Tracy principale 2012	Bourgelat cuvée Caprice 2012	Terravignata 2012	Chateau de Vergisson 2012	Odinstal 350 NN 2013
6	Bardin 2012	Pont de Brion 2012	Colle Stefano 2013	Le Manoir du Capucin Aux Morlays 2012	Von Winning Grainhübel 2013
7	Champeau principale 2012	Gaubert 2012	Alarico 2013	Feuillarde Veilles Vignes 2012	Koehler-Ruprecht Saumagen 2013
8	Séguin 2012	Villa Bel Air 2012	Villa Bucci riserva 2010	Vessigaud Veilles Vignes 2012	Knipser Steinbückel 2013
9	Séguin Prestige 2012	Floridène 2012	Balciana 2011	Corsin 2012	Wehrheim Kastanienbusch 2013
Dry red wines					
ID	**Saint-Chinian** Syrah, Grenache and optionally Mourvèdre and Carignan optional barrel aging	**Moulis-en-Médoc 2010** Cabernet Sauvignon, Merlot and optionally Cabernet franc and Petit Verdot barrel aging	**Mercurey 2012** Pinot noir barrel aging	**Rioja Reserva 2009** Tempranillo and optionally Garnacha, Graciano and Mazuela barrel aging	**Gigondas 2012** Grenache, Syrah and optionally Mourvèdre and Cinsault optional barrel aging
1	Cuvée de Penelle 2011	Lestage Darquier 2010	G. et J. Meunier 1e cru 2012	Imperial tinto reserva 2009	Coteau de mon rève 2012
2	Maurerie Veilles Vignes 2011	Pomeys 2010	Milan 1e Cru Les Crets 2012	Murrieta reserva 2009	Combe Sauvage 2012
3	Prieuré des Mourges Tradition 2009	Granins Grand Poujeaux 2010	Theulet Juillot 1e cru Les Combins 2012	Caecus tinto reserva 2009	Tourbillon 2012
4	Servelière Tradition 2011	Bouqueyran 2010	G. Clos de la Charmée 2012	Ijalba Reserva 2009	Cuvée Costeveille 2012
5	Haut Coup De Foudres 2010	Poujeaux 2010	Vincent Meunier 1e cru C. d. F. 2012	Gonzalo De Berceo Reserva 2009	Cuvée Cécile 2012
6	Les Schistes 2011	La Mouline 2010	Berthoux Les Chavances 2012	Viña Pomal cent. reserva 2009	Bouïsssière 2012
7	Karrimour 2011	Chemin Royal 2010	De la Monette 2012	La Vicalanda Reserva 2009	Cuvée de Beauchamps 2012
8	Best of Belot 2011	Branas Grand Poujeaux 2010	Michel Juillot 1e cru Clos des Barraults 2012	Gaudium Gran Vino reserva 2009	Gour de Chaulé 2012
9	La Sentenelle 310 2011	Myon de L Enclos 2010	Guillot 1e cru Les Veley 2012	Remelluri Reserva 2009	Terrasses de Montmirail 2012

539 S.2. Converting a ranking to an input table for the PHREG procedure.

< expression	PHREG data table
8 > 1 > 1 > 4	Ranking | Rank | wine 1 | 2 | 1 1 | 4 | 4 1 | 3 | 1 1 | 1 | 8

540

541 S.3. The SAS code implementing the PHREG procedure

SAS code	Explanation
PROC IMPORT DATAFILE="<data table file path>" OUT=guido DBMS=xls REPLACE; RUN;	Read the data table

proc phreg data=guido;
class wine(ref='2');
strata ranking;
model rank = wine;
contrast '8 vs 5' wine 0 0 0 -1 0 0 1 0 / estimate ;
contrast '8 vs 7' wine 0 0 0 0 0 -1 1 0 / estimate;
contrast '8 vs 4' wine 0 0 -1 0 0 0 1 0 / estimate;
....
RUN;
	Estimate the utilities
	Reference wine is wine 2
	
	Determine contrasts using the Wald test

542

543

S.4. The Session Design

round	panellist	glass 1	glass 2	glass 3	glass 4
1	1	wine 1	wine 4	wine 1	wine 8
1	2	wine 3	wine 5	wine 4	wine 3
1	3	wine 5	wine 5	wine 1	wine 7
1	4	wine 8	wine 6	wine 2	wine 8
1	5	wine 2	wine 4	wine 4	wine 9
1	6	wine 9	wine 7	wine 9	wine 6
1	7	wine 1	wine 4	wine 4	wine 5
1	8	wine 6	wine 5	wine 5	wine 4
1	9	wine 1	wine 1	wine 3	wine 9
1	10	wine 6	wine 1	wine 9	wine 6
1	11	wine 2	wine 9	wine 2	wine 5
1	12	wine 8	wine 5	wine 6	wine 5
2	1	wine 9	wine 5	wine 6	wine 9
2	2	wine 2	wine 2	wine 1	wine 6
2	3	wine 6	wine 6	wine 8	wine 3
2	4	wine 5	wine 5	wine 1	wine 9
2	5	wine 7	wine 5	wine 7	wine 8
2	6	wine 1	wine 1	wine 4	wine 8
2	7	wine 9	wine 7	wine 8	wine 8
2	8	wine 9	wine 8	wine 3	wine 9
2	9	wine 7	wine 6	wine 4	wine 6
2	10	wine 4	wine 4	wine 8	wine 2
2	11	wine 8	wine 8	wine 1	wine 3
2	12	wine 1	wine 2	wine 1	wine 7
3	1	wine 7	wine 7	wine 3	wine 2
3	2	wine 8	wine 7	wine 9	wine 8
3	3	wine 9	wine 2	wine 4	wine 4
3	4	wine 3	wine 7	wine 4	wine 3
3	5	wine 3	wine 6	wine 6	wine 1
3	6	wine 5	wine 2	wine 2	wine 3
3	7	wine 3	wine 2	wine 3	wine 6
3	8	wine 1	wine 7	wine 7	wine 2
3	9	wine 2	wine 8	wine 5	wine 2
3	10	wine 3	wine 7	wine 3	wine 5
3	11	wine 6	wine 7	wine 4	wine 7
3	12	wine 9	wine 3	wine 9	wine 4

S.5. Pairwise comparison overview. This table indicates the number of times two wines were compared during each wine session. E.g. wine 1 was 4 times compared with itself, wine 3 and 9; 5 times with wine 2, 5, 6 and 7 and 6 times with wine 4 and wine 8.

	WINE 1	WINE 2	WINE 3	WINE 4	WINE 5	WINE 6	WINE 7	WINE 8	WINE 9
WINE 1	4	5	4	6	5	5	5	6	4
WINE 2	5	4	5	6	6	4	5	5	4
WINE 3	4	5	4	5	5	6	6	4	5
WINE 4	6	6	5	4	5	4	4	4	6
WINE 5	5	6	5	5	4	5	5	4	5
WINE 6	5	4	6	4	5	4	5	5	6
WINE 7	5	5	6	4	5	5	4	6	4
WINE 8	6	5	4	4	4	5	6	4	6
WINE 9	4	4	5	6	5	6	4	6	4

S.6. White wine sessions: comparison between the QAMREC results and the corresponding guide ratings. The obtained QAMREC utilities are compared with the associated guide ratings for two white wine sessions. The remark column indicates whether the guide has underestimated (under), overestimated (over) or equally estimated (equal) the wine with regard to QAMREC.

Session 1. Pouilly Fumé 2012					
ID	**Wine**	**Utility**	**Odds**	**Guide Hachette**	**remark**
2	Tabordet 2012	0.74	2.09	*	under
4	Bailly 2012	0.67	1.95	*	under
6	Bardin 2012	0.64	1.90	citation	under
3	Eclat 2012	0.58	1.78	*	equal
5	Tracy principale 2012	0.56	1.76	citation	under
8	Séguin 2012	0.48	1.62	**	over
1	Rabichattes 2012	0.20	1.22	**	over
7	Champeau principale 2012	0.10	1.10	citation	equal
9	Séguin Prestige 2012	0.00	1.00	**	over
Session 7. Pouilly Fuissé 2012					
ID	**Wine**	**Utility**	**Odds**	**Guide Hachette**	**remark**
8	Vessigaud Veilles Vignes 2012	2.10	8.17	**	equal
9	Corsin 2012	1.66	5.26	*	under
3	Soufrandise 2012	1.44	4.22	*	under
4	Sève 2012	1.39	4.03	citation	under
6	Le Manoir du Capucin Aux Morlays 2012	0.98	2.67	*	equal
5	Chateau de Vergisson 2012	0.90	2.45	citation	under
7	Feuillarde Veilles Vignes 2012	0.42	1.52	**	over
1	Chateau de Chaintré 2012	0.21	1.24	citation	equal
2	Chateau de Lavernette 2012	0.00	1.00	citation	equal

S.7. Red wine sessions: comparison between the QAMREC results and the corresponding guide ratings. The obtained QAMREC utilities are compared with the associated guide ratings for four red wine sessions. The remark column indicates whether the guide has underestimated (under), overestimated (over) or equally estimated (equal) the wine with regard to QAMREC.

Session 2. Saint-Chinian					
ID	**Wine**	**Utility**	**Odds**	**Guide Hachette**	**remark**
8	Best of Belot 2011	3.42	30.46	*	under
1	Cuvée de Penelle 2011	2.58	13.22	**	equal
7	Karrimour 2011	2.36	10.61	citation	under
6	Les Schistes 2011	2.29	9.91	*	equal
9	La Sentenelle 310 2011	2.17	8.80	**	over
2	Maurerie Veilles Vignes 2011	2.02	7.57	citation	under
3	Prieuré des Mourges Tradition 2009	1.31	3.70	citation	equal
5	Haut Coup De Foudres 2010	0.47	1.61	*	over
4	Servelière Tradition 2011	0.00	1.00	citation	equal
Session 3. Moulis-en-Médoc 2010					
ID	**Wine**	**Utility**	**Odds**	**Guide Hachette**	**remark**
8	Branas Grand Poujeaux 2010	3.96	52.54	**	equal
5	Poujeaux 2010	2.77	16.04	**	equal
7	Chemin Royal 2010	1.93	6.90	citation	under
4	Bouqueyran 2010	1.72	5.59	citation	under
1	Lestage Darquier 2010	1.60	4.94	citation	under
3	Granins Grand Poujeaux 2010	1.44	4.21	*	equal
9	Myon de L Enclos 2010	1.27	3.55	citation	equal
6	La Mouline 2010	1.04	2.83	*	over
2	Pomeys 2010	0.00	1.00	**	over
Session 8. Rioja Reserva 2009					
ID	**Wine**	**Utility**	**Odds**	**Guía Peñin**	**remark**
9	Remelluri Reserva 2009	1.53	4.60	94	equal
3	Caecus tinto reserva 2009	1.24	3.47	88	under
6	Viña Pomal cent. reserva 2009	0.85	2.35	90	under
8	Gaudium Gran Vino reserva 2009	0.72	2.05	95	over
1	Imperial tinto reserva 2009	0.64	1.89	92	equal
2	Murrieta reserva 2009	0.57	1.76	93	over
7	La Vicalanda Reserva 2009	0.56	1.74	91	over
5	Gonzalo De Berceo Reserva 2009	0.55	1.73	89	equal
4	Ijalba Reserva 2009	0.00	1.00	87	equal
Session 10. Gigondas 2012					
ID	**Wine**	**Utility**	**Odds**	**Guide Hachette**	**remark**
4	Cuvée Costeveille 2012	1.66	5.27	*	under
3	Tourbillon 2012	1.53	4.64	citation	under
7	Cuvée de Beauchamps 2012	1.44	4.21	**	equal
8	Gour de Chaulé 2012	1.36	3.88	*	equal
6	Bouïsssière 2012	1.34	3.82	citation	under
5	Cuvée Cécile 2012	0.97	2.64	*	equal
2	Combe Sauvage 2012	0.85	2.33	citation	equal
9	Terrasses de Montmirail 2012	0.72	2.06	*	over
1	Coteau de mon rève 2012	0.00	1.00	citation	equal

559

S.8. Wine involvement profile information of all panellists and sessions.

Panelist	WIP		Session	mean WIP
pm1	74		Pouilly Fumé	74.58 ± 4.96
pm2	72		Saint-Chinian	75.00 ± 4.91
pm3	74		Moulis-en-Médoc	73.75 ± 4.43
pm4	86		Graves	76.17 ± 3.50
pm5	77		Verdicchio	72.75 ± 4.87
pm6	57		Mercurey	75.08 ± 4.95
pm7	76		Pouilly Fuissé	74.75 ± 3.04
pm8	86		Rioja	74.67 ± 4.96
pm9	80		Riesling	76.17 ± 3.50
pm10	74		Gigondas	74.00 ± 4.42
pm11	71			
pm12	68			
pm13	76			
pm14	64			
pm15	70			
pm16	56			
pm17	75			
pm18	77			
pm19	76			
pm20	75			
mean	73.2 ± 3.60			
maximum	86			
minimum	56			