AMERICAN ASSOCIATION OF WINE ECONOMISTS

AAWE WORKING PAPER
No. 230
Economics

DOES BLIND TASTING WORK? INVESTIGATING THE IMPACT OF TRAINING ON BLIND TASTING ACCURACY AND WINE PREFERENCE

Qian Janice Wang and Domen Prešern

Jun 2018

www.wine-economics.org

AAWE Working Papers are circulated for discussion and comment purposes. They have not been subject to a peer review process. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the American Association of Wine Economists AAWE.

© 2018 by the author(s). All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Does Blind Tasting Work?
Investigating the Impact of Training on Blind Tasting Accuracy and Wine Preference

Qian Janice Wang[1] and Domen Prešern[2]

[1] Department of Experimental Psychology, University of Oxford, Oxford, UK. Ex-President (2015-2017), Oxford University Blind Tasting Society.

[2] Theoretical Chemistry Group, Department of Chemistry, University of Oxford, Oxford, UK. President, Oxford University Blind Tasting Society.

ABSTRACT

Blind wine tasting refers to the practice of tasting a wine without seeing the label, and deducing the grape variety, location of origin, and vintage of the wine based on what one perceives and infers from the glass. We had the opportunity to analyse data from Oxford University Blind Tasting Society's 2018 training season, where 15 participants attended 18 training sessions over 36 days. A total of 212 wines (104 white, 108 red) were tasted and a total of 2271 tasting notes were processed. The aim of this study was to assess whether blind tasting training can improve accuracy, both in terms of within-participant accuracy of guesses and wine structural elements, and in terms of group-wide accuracy (e.g. variance of guesses within a group). The results showed that over time, guesses for grape variety increased in terms of accuracy as well as within-group agreement. On the other hand, guesses for vintage decreased in terms of accuracy as well as within-group agreement. No change in accuracy was observed for place of origin (country and region) guesses. Moreover, it was demonstrated that for grape variety, country, region, and vintage, the chances of the most common within-group guess being correct was significantly higher than the underlying frequency distribution, e.g., if participants had just guessed the most frequently occurring grape/country/region/vintage. Structurally, participants' estimation of acidity level increased in accuracy over time, while the average error in acidity and alcohol estimations were not statistically different from zero. Finally, we assessed how wine preference is related to wine attributes as well as the taster's wine experience. It was demonstrated that, overall, wine preference was positive correlated with wine age, acidity, sweetness, and colour (red wine was preferred to white). Over time, we observed a shift in preference towards older wines, and a decrease in the importance of wine colour. Those with little initial blind tasting training also experienced a change in preference towards greater acidity and alcohol, and decreased their preference for oak. These observations have important implications for the acquisition of wine expertise, and for growing wine markets with an increasingly educated consumer population.

1. Introduction

There is a wide body of literature on differences between wine novices and experts when it comes to perceptual sensitivity, the use of descriptions, and wine preference (D'Alessandro & Pecotish, 2013; Gawel, 1997; Hughson & Boakes, 2014; Parr et al., 2002; Royet et al., 2013; Weil, 2007). What is relatively lacking, however, is an assessment of how the experience of wine tasting changes with the acquisition of expertise. As the Oxford University Blind Tasting Society prepares for its annual varsity blind tasting match against Cambridge, we have a unique opportunity to follow the progress of a group of relatively novice tasters as they go through an intensive five-week training program.

Is blind tasting just nonsense? Rosenheck (2017) performed an analysis of guesses made at the 2017 varsity blind tasting match and deduced that most wines were deduced at a rate better than chance. By collecting participants' tasting sheets and guesses over their training program, we aim to assess the impact of blind tasting training in terms of group accuracy (variability of guesses amongst the group as a whole) as well as individual accuracy (including accuracy in determine structural elements such as acid, alcohol, presence of oak, etc.). We will also use natural language processing to analyse changes in language usage in participants' tasting notes as they become more experienced in blind tasting. Finally, we will keep track of any changes in wine preference throughout the training session, to study how the acquisition of expertise may shape purchase intentions. While wine preference is not explicitly a part of the blind tasting match, we believe it is an important question to investigate. Given the rapid rise of knowledgeable wine tasters in developing markets (e.g. China, see Li & Bardaji, 2017), it is crucial to understand how increasing expertise can influence preference for specific wine styles (e.g., less fruit-driven, higher acid, more bottle maturation, etc.). Because the wines will be given blind, we will be able to track preference for only wine-intrinsic attributes without the participants being biased by the label, price, or origin information (Almenberg & Dreber, 2011; Cordeaus et al., 2013; D'Alessandro & Pecotish, 2013; Mueller & Szolnoki, 2010).

1.1 The Oxford-Cambridge varsity blind tasting match

The varsity match, now in its 65th year, is the oldest university based wine competition in the world (Segal et al., 2013). The competition involves two flights – one white, one red - of six

wines each. All wines are served at room temperature in ISO glasses from numbered clear glass bottles (the original bottles are hidden, rather than merely covered with a wine sock, as to not give away information based on their shape). Competitors are given 40 minutes per flight, and must guess the grape variety, country, region, sub-region, and vintage of each wine in addition to giving justifications for each guess in terms of a brief tasting note. The wines can originate from anywhere in the world, as long as they are accessible on the UK market and have a dominant variety of at least 70%. While sweet wines can be included in the match, rosé, sparkling, and fortified wines are not allowed in the match. For the most part, the wines come from major wine producing countries and are no more than twenty years old.

1.2 Influence of wine expertise and training on the wine tasting experience

Training improves people's ability to discriminate flavours when tasting wine (Owen & Machamer, 1979). However, that is possibly because trained panellists and experts can adapt an analytical strategy that helps them to distinguish different components of wine flavours, when compared to untrained panellists (Arvisenet et al., 2016). Furthermore, several neuroimaging studies involving wine have been conducted with the goal of pinpointing the influence of expertise on multisensory integration in wine evaluation. Sommeliers activate those brain regions that are involved in high-level cognitive processes such as working memories and behavioural strategies when they taste wine – unlike novices who activate the primary gustatory cortex and emotional processing areas more (Castriota-Scanderbeg et al., 2005). In a follow-up study focused on the effect of expertise during the different phases of tasting (i.e., during vs. after tasting), Pazart et al. (2014) observed that wine experts activated those brain regions responsible for sensory integration immediately during the wine tasting phase, whereas for control participants they were only activated during the after tasting phase. This result implies that experts are able to analyse the sensory properties of wine more efficiently than untrained participants.

In terms of odours, expertise has been shown to increase sensitivity and discrimination (see Royet et al., 2013, for a review), possibly giving rise to structural reorganisation in olfactory brain regions (Delon-Martin et al., 2013). However, there seems to be no evidence that wine experts experience increased sensitivity when it comes to wine tasting. In fact, there seems to be

no differences in sensitivity to odours in general – either those typically found in wine or otherwise (Brand & Brisson, 2012; Parr et al., 2002).

In terms of language usage, wine experts use a different vocabulary when describing wines, using analytical terms whereas non-experts use holistic terms (Challoet & Valentin, 2000; Weil, 2007). Moreover, there is evidence that experts are more consistent in terms of vocabulary usage due to an alignment of sensory concepts (Hughson & Boakes, 2002; Ishii & O'Mahony, 1991), not only in terms of concrete descriptors but also of complex and technical terms such as malolactic fermentation (Gawel, 1997).

Taken together, then, the evidence that has been published to date suggests that training in wine tasting results in a different way of thinking about and describing wine. Wine experts are trained to categorise and look for specific flavours or combination of flavours in a wine, and develop their own prototypes for wine representation that includes sensory as well as inferred (such as referring to the winemaking process) and hedonic information (Brochet & Dubourdieu, 2001).

1.3 Aims and contributions

The present study aims to assess, first and foremost, whether training can improve blind tasting accuracy, which entails amongst other things, accuracy in assessing structural elements in a wine. Notably, unlike previous studies on wine expertise, where there are already separate groups of novice and expert participants, the present study aims to get a longitudinal view of expertise acquisition over an intense period of blind tasting training where participants attend multiple tastings a week over a period of 36 days. This gives us the opportunity to study within participant development as well as group-wide accuracy as a whole (e.g. variance of guesses within a group). We present five hypotheses below, related to questions of whether and how training can improve blind tasting accuracy, and, from a commercial prospective, how wine preferences change with increased wine-tasting experience.

Hypothesis 1: Training will reduce within group variability in guesses (Ishii & O'Mahony, 1991).

Hypothesis 2: Training will make participants more accurate in their guesses (Hughson & Boakes, 2002; Owen & Machamer, 1979; Parr et al., 2002).

Hypothesis 3: Training will make participants more accurate in their structural assessment of wines (Arvisenet et al., 2016; Royet et al., 2013).

Hypothesis 4: Training will change the way participants write tasting notes, leading to the usage of more specific descriptors and a different body of vocabulary (Brochet & Debourdieu, 2001; Gawel, 1997; Langlois et al., 2011; Weil, 2007).

Hypothesis 5: Training will change the tasters' wine preference, leading them to prefer less fruit-forward, more expensive, and/or more structured (higher acid/tannin to alcohol ratio) wines (D'Alessandro & Pecotish, 2013; Goldstein et al., 2008; Weil 2001, 2005).

2. Methods and materials

2.1 Participants

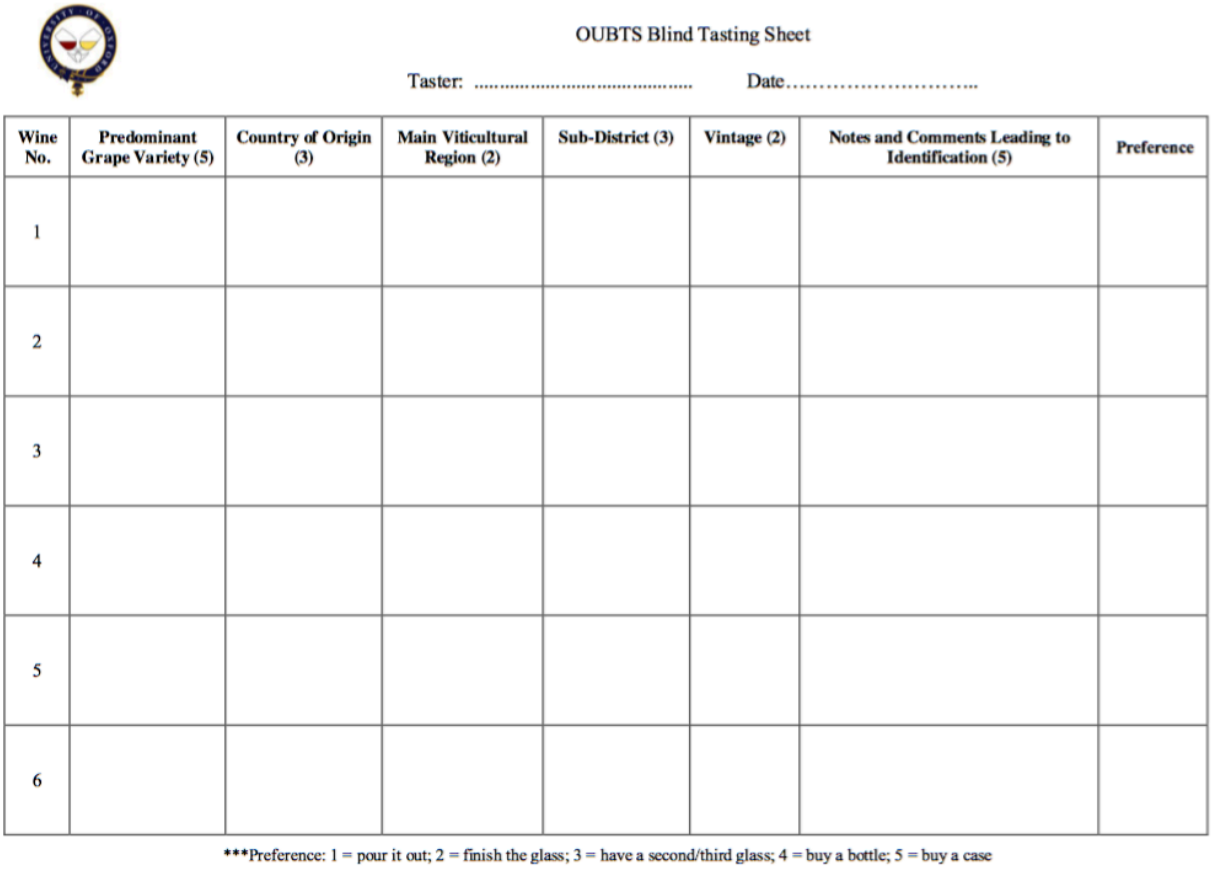
A total of 23 people participated in the blind tasting training scheme, recruited from the Oxford University Blind Tasting Society. After the 36-day training period, we chose to analyse the tasting sheets of 15 participants (3 women, 12 men), who participated in at least 9 out of the 18 training sessions held. Out of the 15 participants, 9 had very limited exposure to formal wine tasting (hereafter labelled as "novice"), consisting of attendance at several beginner tasting sessions held in Autumn 2017. The remaining 6 participants (hereafter labelled as "experienced") were either returning team members or had a significant history of involvement with wine societies (~10 years). The participants ranged from 19-32 years in age, with an average of 25.2 years (SD=3.97).

2.2 Wine samples

Still white and red wines from any wine-producing region in the world with a dominant (at least 70%) grape variety can be used for training, under the condition that the wine is available on the UK market. The wines are generally in the £15-£40 price range (see Appendix D) and are from within the last 13 years (see Appendix B). The wines were selected by the presenter at each tasting session and were unknown to the experimenters ahead of time. See Appendix A for frequency distributions of grape variety and place of origin of all wines used for training.

2.3 Experiment Design:

The overall training scheme consists of a longitudinal study over 5 weeks (January 13 – February 17), with a total of 18 training sessions. Tasting sheets from participants were collected and analysed from all sessions (see Figure 1 for tasting sheet format).



OUBTS Blind Tasting Sheet

Taster: ..... Date.....

Wine No.	Predominant Grape Variety (5)	Country of Origin (3)	Main Viticultural Region (2)	Sub-District (3)	Vintage (2)	Notes and Comments Leading to Identification (5)	Preference
1							
2							
3							
4							
5							
6							

***Preference: 1 = pour it out; 2 = finish the glass; 3 = have a second/third glass; 4 = buy a bottle; 5 = buy a case

Figure 1. Tasting sheet used by participants at each blind tasting session. Participants were asked to provide a tasting note as well as guess the grape variety, country, region, sub-region, and vintage. Furthermore, they were asked to give a preference score (1-5) for the wine.

2.4 Procedure

The training sessions were conducted at the University of Oxford in meeting rooms at various colleges (Magdalen, Merton, Exeter) and well as at the Department of Experimental Psychology. Each session lasted approximate 2-3 hours and will be held at common wine-consumption times

(either around noon or in at 6:30 PM). In each session, the coach of the blind tasting society or another external presenter presented 10-12 wines to be tasted at the session. Participants tasted the wines blind (i.e., without seeing the bottle), served in flights of 5-6 wines at a time. The wines were served at room temperature in 30 mL samples, in standard clear 215 mL ISO glasses. In each session, all participants were presented with the same wines in the same order, although they can choose the order of tasting within each flight. Participants were given 30 minutes per flight to taste the wines and fill out the tasting sheet. All participants tasted the wines in silence. For each wine, participants were instructed to give a guess for the dominant grape variety, place of origin (country, region, and sub-region) and vintage, as well providing a tasting note backing up their guess and rate their preference for the wine. Notably, preference was measured via a 5-point scale (1 = pour it out; 2 = finish the glass; 3 = have a second/third glass; 4 = buy a bottle; 5 = buy a case).

After each flight was tasted, participants compared tasting notes and offered their guesses. The discussion, led by the society coach or the external presenter, is much of where the pedagogical element came into play, as people debated why certain guesses were correct or incorrect (for instance, a low acid wine is unlikely to be Riesling).

2.5 Data Analysis

First, we assessed within-group variability in terms of people's guesses for grape, location, and vintage, over 5 weeks (Hypothesis 1). For each participant, we analysed the accuracy of the participants' guesses for grape, country, region, and vintage (Hypothesis 2). In addition, we analysed their accuracy in terms of assessed acid and alcohol levels (Hypothesis 3). Acidity was coded from 1-5 (1=low acid, 3=medium acid, 4=crisp acid, 5=high acid). Alcohol was coded from 1-3 (1=low alcohol, 2=medium alcohol, 3=high alcohol).

Natural language processing (Python version 2.7.10, nltk toolkit) was utilised to analyse participants' tasting notes, specifically paying attention to the length of tasting notes and the size of vocabulary pool used by participants on a session-by-session basis (Hypothesis 4).

Finally, we analysed patterns in wine preference over time (Hypothesis 5), by analysing session by session correlation values of wine preference with wine price, vintage, climate (new or old world), and structure (levels of acidity, alcohol, sweetness, and oak). Sweetness was coded from

0-2 (0=dry, or < 5 g/L residual sugar; 1=off-dry, or < 50 g/L residual sugar; 2=sweet, or 50 g/L residual sugar). Oak was coded from 0-2 (0=no new oak; 1=some new oak usage; 2=all new oak usage).

3. Results and Discussion

Overall, 196 tasting sheets (2271 tasting notes) were analysed from 15 participants who participated in 18 tasting sessions spanning 36 days. On average, 10.9 participants attended per tasting session (*SD* = 1.7). A total of 212 wines were tasted (104 whites, 109 reds) ranging in vintages from 2005-2017. In terms of country, France was by far the most represented with 82/213 wines (detailed information regarding frequency analysis of grape varieties, places of origin, and vintage can be found in Appendix A and B). In terms of grape variety, Chardonnay was the most common white grape (21/104 wines) and Syrah/Shiraz was the most common red grape (13/109 wines). On average, the wines cost £15.38 (SD = 5.08).

3.1 Within-group variability (Hypothesis 1)

We hypothesised that training should reduce within-group variability of guesses – in other words, we expected participants to agree more in terms of grape variety, place of origin (country, region), and vintage as time went on. First, we calculated the average number of people who agree on a guess of grape variety, country, region, and vintage (for each wine, the number of different guesses of grape variety/country/region/vintage divided by the number of people who tasted the wine). Pearson's correlation between time (in terms of day number, from 1-36) and average number of people per guess of grape variety, country, region, and vintage showed a positive correlation between time and the number of people per grape variety guess at the $p<.10$ level (Table 1).

Table 1. Pearson's correlation coefficients (N=212) between time (ranging from 1-36) and the average number of people who agree on a guess for grape variety, country, region, and vintage for each wine. * indicates significance at 0.1 level.

	Grape variety	**Country**	**Region**	**Vintage**
Time	0.13 *	0.05	0.10	0.05

To get a sense of specific differences in within-group agreement over time, a multivariate analysis of variance (MANOVA) was conducted with time (we divided the sessions into the first half – sessions 1 to 9 - and the second half – sessions 10 to 18) as independent factor and the average number of people who agree on a guess of grape variety, country, region, and vintage as dependent measures. Results revealed that time indeed had a significant overall effect ($F(4,207) = 2.91$, $p = .02$, *Wilk's Lamba* = .95), with further univariate ANOVAs showing the significant influence of time on the number of people who share the same guess for grape variety ($F(1, 210) = 6.13$, $p = .01$, $\eta^2 = 0.03$) and vintage ($F(1,210) = 4.57$, $p = .03$, $\eta^2 = 0.02$). Pairwise comparisons with Bonferroni corrections revealed that, as predicted, the number of people who agree on the same guess for grape variety significantly increased from the first half to the second half of the training period (Table 2). However, the number of people who shared the same vintage guess decreased rather than increased with time, which implies that participants diverged in their vintage guesses as opposed to coming together.

Table 2. Average number of people who agree on a guess for grape variety, country, region, and vintage for each wine, with time as independent factor. Standard error shown in parentheses. Different letters indicate significant differences according to a Bonferroni test (* indicates significance at level 0 < .10, ** indicates significance at level p < .05).

		Grape variety	Country	Region	Vintage
Time	**Sessions 1-9**	1.98 (0.11) a **	2.84 (0.15) a *	1.74 (0.06) a	3.27 (0.10) a **
	Sessions 10-18	2.35 (0.11) b **	3.20 (0.15) b *	1.87 (0.06) a	2.93 (0.11) b **

We were also interested in whether implicit group-think is accurate – in other words, how often is the most commonly guessed grape variety, country, region, and vintage in fact the correct one? We calculated the instances of the most common guess amongst participants for each wine being correct (see Table 3), and, by way of comparison, used the chances of getting the variety/country/region/vintage correct simply by using the statistically most frequently shown attributes (based on Appendix A and B). We chose not to compare participant accuracy to simply random chance since the wines shown were not uniformly distributed amongst grape varieties/country/region/vintage. Chi squared tests of independence showed that for all four

categories, the most common guess from participants performed significantly better than merely the most frequently occurring value.

Table 3. Instances where the most common guess amongst the participants for each wine, by categories of grape variety, country, region, and vintage was in fact the correct one. % of most common guess is compared against chances of being correct simply by guessing the most frequently occurring grape/country/region/vintage (for grape and region, the most frequently occurring red and white grape and regions are added). Chi squared test of independence was conducted to compare the accuracy of the most common guess against the baseline condition of occurrence frequency. * indicates significance at 0.05 level, and ** indicates significance at 0.01 level.

	Grape variety	**Country**	**Region**	**Vintage**
Most common guess being correct	94	105	71	78
% correct	44%	50%	33%	37%
% correct by simply guessing the most frequently occurring grape/country/region/vintage)	16%	39%	11%	27%
χ^2	40.3 **	5.06 *	31.5**	4.33*

3.2 Accuracy in guessing grape variety, place of origin, and vintage (Hypothesis 2)

Given that, as a group, the most commonly occurring guesses for grape variety, country, region, and vintage were significantly more accurate than chance distribution (Section 3.1), the next question is whether participants become more accurate with training.

Pearson's correlation coefficients were calculated between session day (ranging from 1-36) and the number of grape variety, country, region, and vintage that participants got correctly on average (see Table 4). While accuracy for grape variety did increase with time (see Figure 2), we did not observe any improvements in accuracy with regard to country and region guesses. Furthermore, the observation that grape variety accuracy is significantly correlated with place of origin accuracy (country and region) might be explained by the fact that grape varieties are not uniformly distributed geographically.

Surprisingly, vintage accuracy decreased with time. This might be due to changes in the participants' anecdotal strategy, whereby novices begun by guessing the same vintage for all the wines. Given that most wines were very recent and usually from the same two vintages (2015 or 2016 for white wine, 2014 or 2015 for red wine, see Appendix B), the strategy of guessing the same vintage for all wines was usually fairly successful. Over time, as participants became more confident, they begun varying vintage guesses which led to more error.

Table 4. Pearson's correlation coefficients (N=193) between time (ranging from 1-36) and the average number of accurate grape variety, country, region, and vintage guesses per session. * indicates significance at .05 level, and ** indicates significance at the .01 level.

	Time	**Grape variety**	**Country**	**Region**	**Vintage**
Time	1.0	0.21 **	-0.01	-0.02	-0.16 *
Grape variety	-	1.0	0.54 **	0.74 **	-0.06
Country	-	-	1.0	0.80 **	-0.08
Region	-	-	-	1.0	-0.02
Vintage	-	-	-	-	1.0

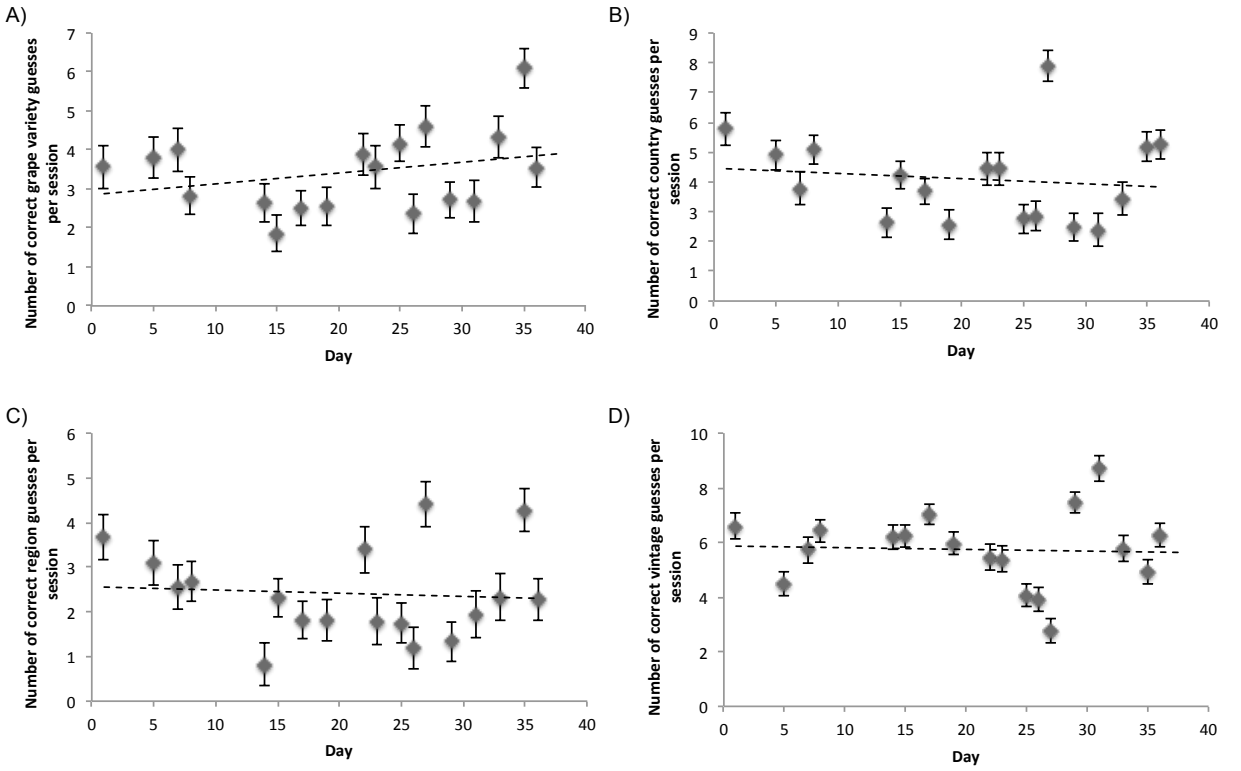


Figure 2. Plots showing accuracy of grape variety (A), country (B), region (C), and vintage (D) guesses over time. Error bars indicate the standard error of the means.

We were especially interested in the progress of novice tasters (those who began the study with no prior blind tasting training experience). Therefore, a MANOVA was conducted with prior experience level (novice or experienced) and time (sessions 1-9 or sessions 10-18) as independent factors, and the average number of correct grape variety, country, region, and vintage guesses per session as dependent measures. Results showed that prior experience level ($F(4,186) = 19.07$, $p < .0005$, *Wilk's Lamba* = .71) and time ($F(4,186) = 8.14$, $p < .0005$, *Wilk's Lamba* = .85) both had significant main effects, whereas we did not observe a significant interaction effect between prior experience level and time ($F(4,186) = 2.34$, $p = .06$). Further univariate ANOVAs revealed a significant effect of prior experience on guess accuracy of grape variety ($F(1,189) = 64.25$, $p < .0005$, $\eta^2 = 0.25$), country ($F(1,189) = 28.65$, $p < .0005$, $\eta^2 = 0.13$), and region ($F(1,189) = 55.56$, $p < .0005$, $\eta^2 = 0.23$); in addition, there was a significant effect of time on guess accuracy of grape ($F(1,189) = 28.76$, $p = .001$, $\eta^2 = 0.06$) and vintage ($F(1,189) = 20.19$, $p = .02$, $\eta^2 = 0.03$). Pairwise comparisons with Bonferroni corrections revealed that experts were significantly more accurate than novices when it comes to guessing

grape varieties and place of origin (see Table 5). Furthermore, the ANOVA confirmed what had been observed in the correlation analysis, that grape variety accuracy increased significantly, but vintage accuracy decreased significantly, in the second half of the training period (sessions 10-18) compared to the first half of the training period (sessions 1-9).

Table 5. Average number of correct guesses for grape variety, country, region, and vintage per session (12 wines per session), with prior experience and time as independent factors. Standard error shown in parentheses. Different letters indicate significant differences according to a Bonferroni test (at level $p < .05$).

		Grape variety	Country	Region	Vintage
Prior experience	**Novice**	2.56 (0.15) a	3.40 (0.19) a	1.62 (0.15) a	5.57 (0.18) a
	Experienced	4.41 (0.17) b	4.92 (0.21) b	3.33 (0.17) b	6.02 (0.21) b
Time	**Sessions 1-9**	3.10 (0.16) a	4.24 (0.20) a	2.55 (0.16) a	6.12 (0.19) a
	Sessions 10-18	3.88 (0.16) b	4.08 (0.20) a	2.40 (0.16) a	5.47 (0.20) b

Of course, it should be kept in mind that the choice of wine shown at each session was not designed by the experimenters. Therefore, the present study is closer to a field observation than a strictly controlled experiment per-se, since the difficulty of each tasting was not controlled for.

3.3 Accuracy in structural assessment of wines (Hypothesis 3)

With regard to structural elements in the wine, we decided to assess participants' accuracy in evaluating the acid and alcohol levels (see Appendix C for frequency distribution of acidity and alcohol levels for all wines used in the training sessions). For each wine, we first extracted the participants' assessment of acid and alcohol level from their tasting note. Next, we calculated the average assessment of acid and alcohol level per wine, and compared it to the actual acidity and alcohol level of the wine (given by the ABV % and total acidity information, when possible).

Figure 2 shows the distribution of estimation error for wines of different acidity (1=low, 3=medium, 4= crisp, 5=high) and alcohol levels (1=less than 12% ABV, 2 = 12%-13.5% ABV, 3 = more than 13.5% ABV). An ANOVA analysis showed a significant difference between participants' estimation of acidity level and actual acidity level ($F(2,209) = 222.03$, $p < .0005$, η^2

= 0.68). When the acidity level is low, participants tend to drastically overestimate the acidity (M_{low} = 43% error, SE = 2.7), and when the acidity level is high, participants tend to under-estimate the acidity level (M_{high} = -12% error, SE = 0.7), although to a lesser degree than the overestimation ($p < .0005$). Interestingly, participants tend to overestimate the acidity when acidity level is medium (M_{medium} = 11% error, SE = 1.3), but are accurate when acidity level is crisp (M_{crisp} = 1% error, SE = 0.6, error not significantly different from 0%, $t(115) = 1.76$, p =.08). Therefore, participants' estimation accuracy reflects the frequency of wines being shown (see Appendix C.1), with the most frequently appearing acidity level (crisp) being the most accurately assessed.

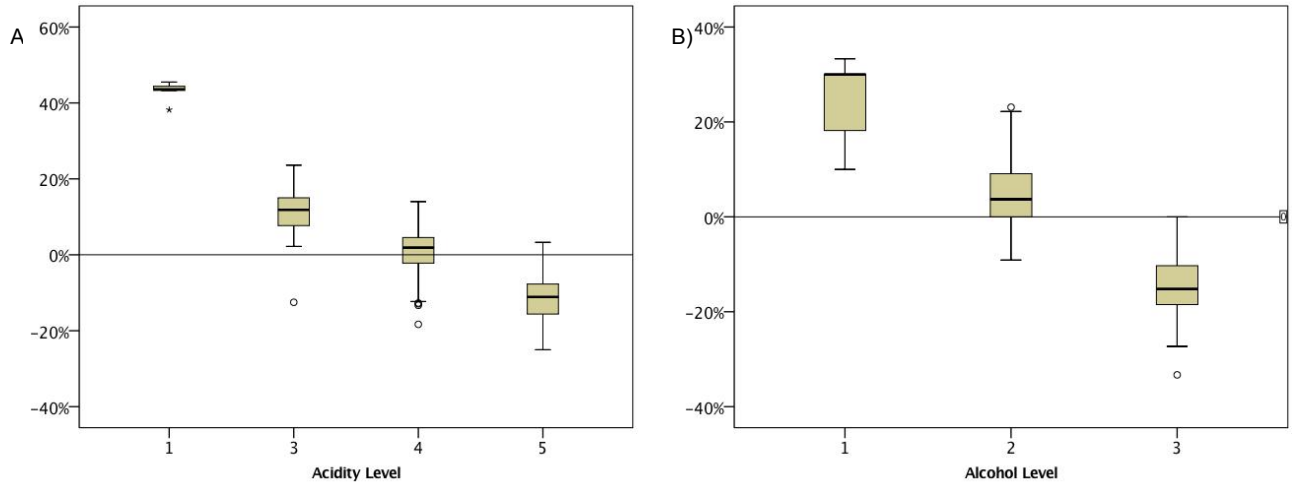


Figure 2. Plots showing accuracy of acidity (A) and alcohol (B) estimations for wines of different acidity (1=low, 3=medium, 4= crisp, 5=high) and alcohol (1=less than 12% ABV, 2 = 12%-13.5% ABV, 3 = more than 13.5% ABV) levels.

Similarly for alcohol estimation, an ANOVA analysis showed a significant difference between participants' estimation of alcohol level and actual alcohol level ($F(2,209) = 222.03$, $p < .0005$, $\eta^2 = 0.68$). When the alcohol level is low, participants tend to overestimate the alcohol (M_{low} = 24% error, SE = 3.0), and when the alcohol level is high, participants tend to under-estimate the alcohol level (although to a lesser extent than for the low alcohol case, M_{high} = -15% error, SE = 0.9, p = .006). Once again, participants' estimation accuracy reflects the frequency of wines being shown (see Appendix C.2), with the most frequently appearing alcohol level (medium) being the most accurately assessed.

Since we were interested in whether acidity and alcohol estimations became more accurate over time, Pearson's correlation coefficients were calculated between estimation error percentages (both signed and unsigned, to take into account the magnitude of error) and session day number (ranging from 1-36). In terms of acidity, we observed a significant decrease in error magnitude over time (r_{18} = -.53, p = .02), lead by an increase in estimation error over time (r_{18} = .77, p < .0005). In other words, participants gradually went from underestimating acidity to overestimating acidity, but with the net effect of increasing accuracy over time (see Figure 3). In terms of alcohol level assessment, we did not observe any significant changes in error magnitude (r_{18} = -.10, p = .68) or estimation error (r_{18} = .03, p =.92). This is possibly because the estimation categories (<12%, 12-13.5%, >13.5%) are fairly broad and easy to learn, which means the estimation error is already fairly small for alcohol. In fact, one-sample t-tests showed that average acidity and alcohol errors are not significantly different from 0 (for acidity: $t(17)$ = -1.45, p = .17; for alcohol: $t(17)$ = 0.94, p = .36).

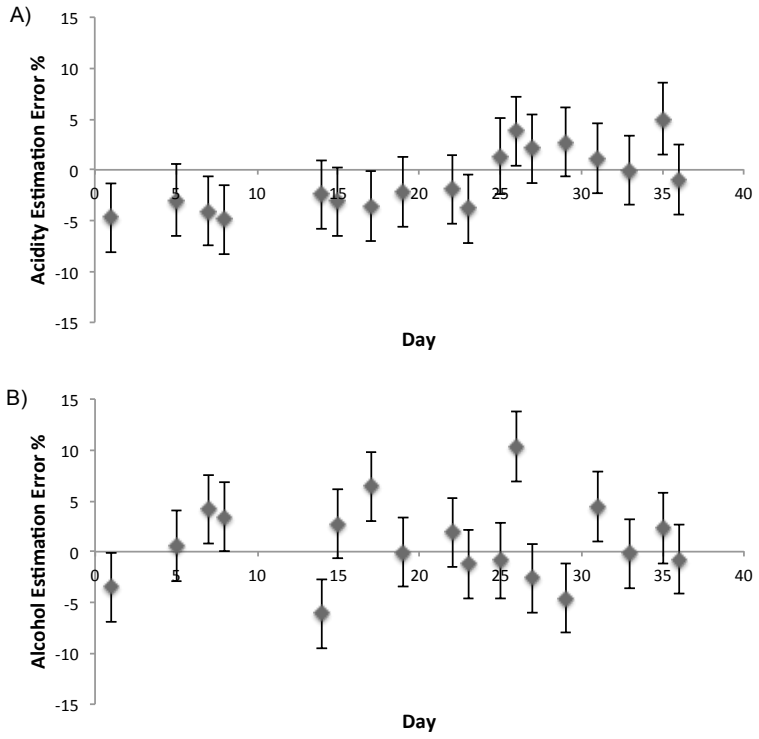


Figure 3. Plots showing accuracy of acidity (A) and alcohol (B) estimations over time. Error bars indicate the standard error of the mean.

3.4 Tasting note (Hypothesis 4)

For each participant, we calculated the number of words per tasting note. An ANOVA with prior experience (novices or experienced tasters) and wine preference (1-5) as independent factors and number of words as the dependent measure showed significant effects of prior experience ($F(1,2109) = 284.14$, $p < .0005$, $\eta^2 = 0.12$) and wine preference ($F(4,2109) = 15.15$, $p < .0005$, $\eta^2 = 0.03$) on the number of words. In terms of prior experience, novices write shorter tasting notes compared to those of more experienced participants ($M_{novice} = 22.37$, $SE = 0.23$, $M_{experienced} = 27.59$, $SE = 0.21$, $p < .0005$). In terms of wine preference, tasting notes for more preferred (those at levels 4 and 5, corresponding to 'willing to buy a bottle' and 'willing to buy a case') wines are significant longer than tasting notes for less preferred wines (those at levels 1-3, corresponding to 'pour it out', 'finish the glass', and 'have another glass') wines ($p < .005$ for all comparisons, see Figure 4).

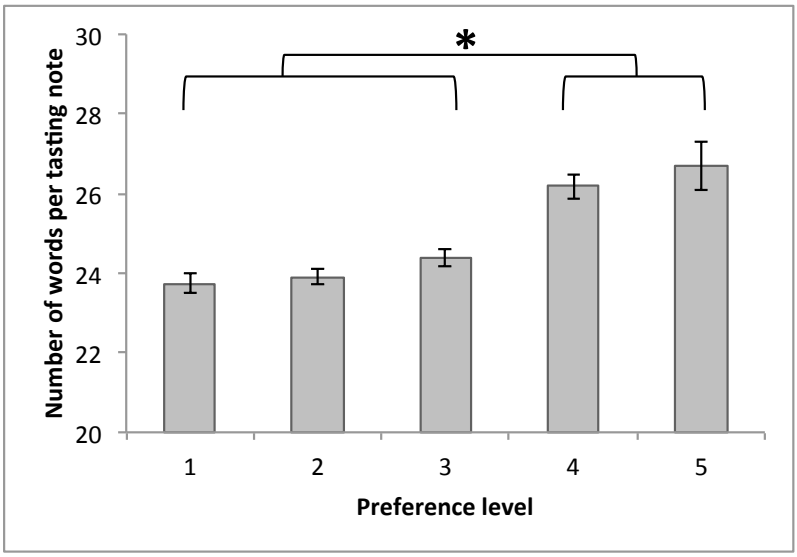


Figure 4. Mean values of number of words per tasting note for different wine preference levels (1=toss it out, 2=finish the glass, 3=have another glass, 4=buy a bottle, 5=buy a case). Overall, tasting notes for wines at preference levels 4 and 5 are significantly longer than tasting notes at preference levels 1-3. Asterisk '*' indicates statistical significant at $p < .005$.

To understand which words in the tasting notes are more associated with greater wine preference, we conducted linguistic analysis using the Python Natural Language Toolkit (nltk, Bird et al., 2014). A Naïve Bayes classifier was written to categorise more from less preferred

wines based on the tasting note provided (more preferred = those with a preference level of 4 or higher). The classifier calculates the prior probability of each word occurring in a tasting note categorised in either the low or high wine preference group (Bird et al., 2014, Chapter 6). The prior probability distribution reveals which descriptors might be most informative (i.e. used more for tasting notes belonging to the high preference group rather than the low preference group, or vice versa). Table 6 below shows the top 15 most informative words for highly preferred wines for novices and experienced participants. Note that while novices preferred sweet wines ("botrytis, baking [spice], sweetness, toffee"), experienced participants preferred aged wines ("raisins, age, brick, honeyed, mushrooms") and wines with high acidity ("crunchy, austere"). See Section 3.5 for further analysis on wine preference.

Table 6. Left: Top 15 most informative words descriptors indicating whether a wine is more preferred or less preferred by novices, with its related likelihood ratios. Right: Top 15 most informative words descriptors indicating whether a wine is more preferred or less preferred by experienced tasters (with previous blind tasting training), with its related likelihood ratios.

Novice tasters			Experienced tasters		
Word	**Likelihood**	**Ratio**	**Word**	**Likelihood**	**Ratio**
Botrytis	High : low pref	12.9 : 1	Raisins	High : low pref	14.3 : 1
Baking	High : low pref	10.0 : 1	Lovely	High : low pref	14.3 : 1
Sweetness	High : low pref	7.7 : 1	Age	High : low pref	11.1 : 1
Stem	High : low pref	7.2 : 1	Hints	High : low pref	11.1 : 1
Toffee	High : low pref	7.2 : 1	Brick	High : low pref	11.1 : 1
Cider	High : low pref	7.2 : 1	Tar	High : low pref	10.6 : 1
Leaf	High : low pref	7.2 : 1	Honeyed	High : low pref	10.5 : 1
Honeysuckle	High : low pref	7.2 : 1	Mushrooms	High : low pref	10.5 : 1
Gentle	High : low pref	7.2 : 1	Crunchy	High : low pref	8.6 : 1
Opulent	High : low pref	7.2 : 1	Tending	High : low pref	8.6 : 1
Elegant	High : low pref	7.2 : 1	Farmyard	High : low pref	7.9 : 1
Perfume	High : low pref	6.0 : 1	Austere	High : low pref	7.9 : 1
Must	High : low pref	6.0 : 1	Precise	High : low pref	7.9 : 1
Cocoa	High : low pref	6.0 : 1	Botrytis	High : low pref	7.9 : 1

We then calculated Pearson's correlation coefficients between the number of words per tasting note and time (coded in terms of day number, from 1-36) to assess whether the number of words

used changes over time), for both novices and those with prior experience (Table 7). For the more experienced participants, there was a small negative correlation between the number of words and time, which suggests that participants might have started shortening their tasting notes slightly over time to be more efficient.

Table 7. Pearson's correlation coefficients between time (coded as day number, ranging from 1-36) and number of words & preference for all participants, those who started training as novices, and those who started training with significant prior blind tasting experience. * indicates significance at .05 level, and ** indicates significance at the .01 level.

	Overall (N=15)	**Novices (N=9)**	**Experienced (N=6)**
Number of words vs. time	-.10	.05	-.08 *
Preference vs. time	.06 **	.06 *	.05
N	2269	1277	992

Furthermore, we predicted that the vocabulary pool, in other words, the number of unique words utilised in the tasting notes, would increase per participant with training. To assess this hypothesis, we calculated the number of unique words utilised per session (see Table 8). An ANOVA analysis was conducted with timing condition (over the first half of the sessions, over the second half of the sessions) and initial expertise (those with no initial experience, and those with some initial experience) as independent measures and the number of unique words as dependent measure. No effect of timing condition was observed ($F(1,191) = 0.09$, $p = .76$). We did observe a significant effect of initial expertise ($F(1,191) = 20.05$, $p < .0005$, $\eta^2 = 0.10$), where more experienced tasters utilised more unique words than beginners ($M_{novice} = 98$, $SE = 2.2$, $M_{experienced} = 112$, $SE = 2.5$, $p < .0005$). This implies that wine experience does increase vocabulary size of tasting notes, but that the effect might take longer than an order of several weeks.

Table 8. The average number of unique words utilised per session, averaged over all participants, for novices only, and for experienced participants only. To understand the influence of training, we also present the number of unique words per session, calculated for the first half of the training sessions (1-9) and the second half of the sessions (10-18).

	Overall	Sessions 1-9	Sessions 10-18
All participants	104 words	104 words	104 words
Novice	98 words	97 words	99 words
Experienced	112 words	114 words	110 words

Qualitatively, we did find some evidence tasting notes became more detailed and descriptive over time for some of the 9 novice tasters. Here is an example of one participant's (ID 18) tasting note for two white Sancerre guesses, during session 2 (day 5), and session 18 (day 36).

Session 2*: Medium lemon in colour. Green apple, grass, without new oak. Crisp acid, medium alcohol, dry with a medium finish.*

Session 18*: Medium lemon in colour with green tinge. Nose of grass, bacon fat, gooseberry, smoke. No oak. On palate dry, 13% alcohol, bitter with crisp acid and long herb finish.*

Here is another example, of Participant 2's tasting notes for two Aconcagua (Chile) Cabernet Sauvignon guesses, during session 1 (day 1) and session 18 (day 36).

Session 1: *Dark purple. High acidity, high alcohol. Spicy aroma + taste. Full body. Dry.*

Session 18: *Deep purple. Cooked bell pepper, vanilla, and red currant. High acid, high alcohol, full body. High coarse tannins. Blackcurrant and earth on palate. Some new French oak.*

As we can see, tasting notes became more specific with respect to aromatic descriptors (from "spicy" aroma to the more descriptive "cooked bell pepper, vanilla, and redcurrant"). Furthermore, participants began paying attention to textural elements of the wine ("high coarse tannins") as well as temporal elements ("long herb finish").

3.5 Preference (Hypothesis 5)

In order to understand which factors drive participants' preference for wines, Pearson's correlation coefficients were calculated between preference and wine factors (Table 9). Since we

were specifically interested on the influence of training on novices, we repeated the preference analysis on the group of 9 novices.

Table 9. Pearson's correlation coefficients between wine preference (for overall group and for novices only) and vintage, acidity, alcohol, sweetness, oak, tannin, wine colour (red=1, white=0), region (new world = 1, old world = 0), and price. N = 212 for all pairs of variables except for tannin (N=108, red wines only). * indicates significance at 0.05 level, and ** indicates significance at 0.01 level.

	Vintage	**Acidity**	**Alcohol**	**Sweetness**	**Oak**	**Tannin**	**Red**	**New World**	**Price**
Preference	-.36 **	.26 **	-.02	.25 **	.09	.02	.19 **	-.01	.36 **
Preference (Novices only)	-.28 **	.06	.11	.31 **	.15 *	.06	.21 **	.16 *	.23 **

Interestingly, the price of the wines had a significant positive correlation with preference, even though the wines were tasted blind. This was true even for the group of novices, which implies that even novices can some how pick up qualities in the wine that is both preferred and more expensive. Furthermore, novices in general preferred wines with oak and from the new world, whereas the group overall had a preference for acidity (driven by the more expert participants).

In addition, we performed linear regressions to understand how preference can be predicted based on wine factors that can be perceived during tasting. In other words, we excluded variables such as price and geographical region (new world vs. old world), since these cannot be perceived sensorially from the wine. Furthermore, to assess the effect of training, we conducted regressions separately on the first half of the sessions and the second half of the sessions. Presumably, differences between the two halves of the training period could be attributed to the influence of training (see Table 10). To ensure that our assessment wouldn't be biased by differences in wines between the two periods, we conducted a multivariate analysis of variance (MANOVA) with time period (first half or second half) as independent factor, and vintage, acidity, alcohol, colour (white or red), sweetness, and oak as measures. Results showed that there were no significant effect of time period on the measures ($F(6,205) = 1.93$, $p = .08$, *Wilks' Lambda* = .95).

Table 10. Fixed effects regressions of wine factors on preference for all 15 participants. * indicates significance at 0.05 level, and ** indicates significance at 0.01 level. Variables with no value are excluded from the model.

	Overall (Sessions 1-18)		First half (Sessions 1-9)		Second half (Sessions 10-18)	
	β	t-Stat.	β	t-Stat.	β	t-Stat.
Vintage	-0.28	-4.53 **	-	-	-0.39	-4.65 **
Acidity	0.30	5.13 **	0.30	3.58 **	0.35	3.99 **
Sweetness	0.29	4.63 **	0.36	4.13 **	0.25	2.82 **
Red (Red = 1, White = 0)	0.17	2.71 **	0.29	3.25 **	-	-
Adjusted R^2	0.27		0.26		0.28	
Num. Obs.	207		103		101	

Notably, from the first half of sessions (1-9) to the second half (10-18), we noticed an overall shift in the importance of vintage (with older wines being more preferred), and a decrease in red wine preference.

For the group of novices, it is especially interesting to observe that the preference for red wines and for oaked wines give way to preference for older wines, more acidic wines, and higher alcohol wines (Table 11).

Table 11. Fixed effects regressions of wine factors on preference for the group of 9 participants who began as wine novices. * indicates significance at 0.05 level, and ** indicates significance at 0.01 level. Variables with no value are excluded from the model.

	Overall (Sessions 1-18)		First half (Sessions 1-9)		Second half (Sessions 10-18)	
	β	t-Stat.	β	t-Stat.	β	t-Stat.
Sweetness	0.35	5.29 **	0.44	4.89 **	0.45	4.40 **

Red (Red = 1, White = 0)	0.25	3.69 **	0.26	2.75 **	0.15	1.47
Vintage	-0.16	-2.39 *	-	-	-0.22	-2.36 *
Oak	-	-	0.18	1.99 *	-	-
Acidity	-	-	-	-	0.30	3.09 **
Alcohol	-	-	-	-	0.22	2.14 *
Adjusted R^2	0.19		0.22		0.25	
Num. Obs.	208		103		99	

For both novices and the overall group, it is important to note the preference for sweetness. Of course, this could be explained by the relatively small number of high quality sweet wines shown in the tastings. Moreover, for both novices and the overall group, we observed over time a shift in preference towards older wines, and a decrease in the importance of wine colour. This has important implications for growing wine markets, where people may begin by enjoying young, oaky, red wines. But as they become more experienced wine consumers, they may start to prefer older wines of both colours, and begin to prefer structured wines with higher acidity and alcohol levels. Finally, Table 6 from Section 3.4 showed that, for those participants who started as wine novices, there was actually a significant positive correlation between the passage of time and overall wine preference. This is in line with the familiarity principle (Miller, 1976; Zajonc, 1968), which states that people's preference for a novel item increases with repeated exposure.

4. Conclusions

The results of the present study demonstrate that blind tasting is not a hoax, that participants can perform better than chance (even following the frequency distribution of grape/location/vintage), and that training can improve accuracy in terms of guesses as well as structural elements. In the span of 18 sessions (36 days), we observed an increase in accuracy and within-group agreement when it comes to grape variety, as well as an improvement in acidity estimation. However, we did not observe any improvements in accuracy or within-group agreement when it came to

guesses of country or region. That said, it should be noted that the acquisition of blind tasting accuracy is probably a long-term procedure that takes well over 4 weeks. This is demonstrated by the fact that those with previous blind-tasting training experience were significantly more accurate in terms of guessing grape variety, country, and region when compared to those with no previous training experience. Therefore, it is possible that while grape variety can be quickly discriminated (at least, to some extent), learning to correctly deduce country and region is a more nuanced task that takes greater training and experience.

Linguistic analysis of tasting notes revealed a similar pattern in terms of vocabulary size (represented by the number of unique words utilised per tasting session), where novices had a smaller vocabulary size, and wrote shorter tasting notes, when compared to experienced tasters. Like country and region guess accuracy, vocabulary size and tasting note length did not change over the course of the training session, implying longer-term learning may be needed. In addition, an interesting relationship emerged between length of note and wine preference, with more preferred wines having longer tasting notes. This may help to shed some light on the mental process behind writing tasting notes, where more preferred wines may enhance processing fluency (or that more easily described wines are in turn more preferred).

Analysis of wine preference demonstrated that, overall, preference was positive correlated with wine age, acidity, sweetness, and colour (red wine was preferred to white). More remarkably, when taking wine price into account, we demonstrated a significant positive correlation between wine price and preference, even when the wines were tasted blind. This gives limited evidence that more expensive[1] wines, broadly speaking, have sensorial properties which make them appealing to tasters. Over time, we observed a shift in preference towards older wines, and a decrease in the importance of wine colour. Those with little initial blind tasting training also experienced a change in preference towards greater acidity and alcohol, and decreased their preference for oak. These observations have important implications for growing wine markets with an increasingly educated consumer population, where the initial preference for heavily oaked, young red wines may shift in time towards wines with more maturity and structure.

[1] It should be kept in mind that the 212 wines used in the training sessions ranged from £6.95 to £35, and therefore is not representative of a wide spectrum of wine prices

ACKNOWLEDGEMENTS

The authors would like to thank the American Association of Wine Economists (AAWE) for presenting us with the research scholarship, and the members of the Oxford University Blind Tasting Society for their warm support and cooperation.

REFERENCES

Almenberg, J., & Dreber, A. (2011). When does the price affect the taste? Results from a wine experiment. *Journal of Wine Economics*, **6**, 111-121.

Arvisenet, G., Guichard, E., Ballester, J. (2016). Taste-aroma interaction in model wines: Effect of training and expertise. *Food Quality and Preference*, **52**, 211-221.

Banks, S. J., Sreenivasan, K. R., Weintraub, D. M., Baldock, D., Noback, M., Pierce, M. E., …, Leger, G. C. (2016). Structural and functional MRI differences in master sommeliers: A pilot study on expertise in the brain. *Frontiers in Human Neuroscience***, 10**:414, 1-12.

Bird, S., Klein, E., & Loper, E. (2014). *Natural language processing with Python*. Sebastopol, CA: O'Reilly Media.

Brand, G., & Brisson, R. (2012). Lateralisation in wine olfactory threshold detection: Comparison between experts and novices. *Laterality*, **17**, 583-596.

Brochet, F., & Dubourdieu, D. (2001). Wine descriptive language supports cognitive specificity of chemical senses. *Brain and Language***, 77**, 187-196.

Castriota-Scanderbeg, A., Hagberg, G. E., Cerasa, A., Committeri, G., Galati, G., Patria, F., et al. (2005). The appreciation of wine by sommeliers: A functional magnetic resonance study of sensory integration. *NeuroImage*, **25**, 570-578.

Chollet, S., & Valentin, D. (2000). Le degré d'expertise a-t-il une influence sur la perception olfactive? Quelques éléments de réponse dans le domaine du vin [Does the level of expertise influence olfactory perception? Some answers from the domain of wine]. *Annals of Psychology*, **100**, 11-36.

Corduas, M., Cinquanta, L., & Ievoli, C. (2013). The importance of wine attributes for purchase decisions: A study of Italian consumers' perception. *Food Quality and Preference*, **28**, 407-418.

Delon-Martin, C., Plailly, J., Fonlupt, P., Veyrac, A., & Royet, J.-P. (2013). Perfumers' expertise induces structural reorganization in olfactory brain regions. *NeuroImage*, **68**, 55-62.

Gawel, R. (1997). The use of language by trained and untrained experienced wine tasters. *Journal of Sensory Studies*, **12**, 267-284.

Goldstein, R., Almenberg, J., Dreber, A., Emerson, J. W., Herschkowitsch, A., & Katz, J. (2008). Do more expensive wines taste better? Evidence from a large sample of blind tastings. *Journal of Wine Economics*, **3**:1, 1-9.

Hughson, A. L., & Boakes, R. A. (2002). The knowing nose: the role of knowledge in wine expertise. *Food Quality and Preference*, **13**, 463-472.

Ishii, R., & O'Mahony, M. (1991). Use of multiple standards to define sensory characteristics for descriptive analysis: Aspects of concept formation. *Journal of Food Science*, **56**, 838-842.

Landon, S. and Smith, C.E. (1997). The use of quality and reputation indicators by consumers: the case of Bordeaux wine. *Journal of Consumer Policy*, **20**, 289-323.

Langlois, J., Dacremont, C., Peyron, D., Valentin, D., & Dubois, D. (2011). Lexicon and types of discourse in wine expertise: The case of *vin de garde*. *Food Quality and Preference*, **22**, 491-498.

Lecocq, S. and Visser, M. (2006). What determines wine prices: objective vs. sensory characteristics. *Journal of Wine Economics*, **1**, 42-56.

Li, Y., & Bardaji, I. (2016). A new wine superpower? An analysis of the Chinese wine industry. American Association of Wine Economists Working Paper, June 2016, **198**.

Miller, R. L. (1976). Mere exposure, psychological reactance and attitude change. *Journal of Abnormal and Social Psychology*, **59**, 1-9.

Mueller, S., & Szolnoki, G. (2010). The relative influence of packaging, labelling, branding and sensory attributes on liking and purchase intent: Consumers differ in their responsiveness. *Food Quality and Preference*, **21**, 774-783.

Owen, D. H., & Machamer, P. K. (1979). Bias-free improvement in wine discrimination. *Perception*, **8**, 199-209.

Parr, W. V., Heatherbell, D., & White, K. G. (2002). Demystifying wine expertise: Olfactory threshold, perceptual skill and semantic memory in expert and novice wine judges. *Chemical Senses*, **27**, 747-755.

Pazart, L., Comte, A., Magnin, E., Millot, J.-L., & Moulin, T. (2014). An fMRI study on the influence of sommeliers' expertise on the integration of flavor. *Frontiers in Behavioral Neuroscience*, **8**:358.

Rosenheck, D. (2017). Think wine connoisseurship is nonsense? Blind-tasting data suggest otherwise. *The Economist*, **August 17th**. Accessed from https://www.economist.com/blogs/graphicdetail/2017/05/daily-chart-11.

Royet, J. -P., Plailly, J., Saive, A. -L., Veyrac, A., & Delon-Martin, C. (2013). The impact of expertise in olfaction. *Frontiers in Psychology*, **4**:928.

Segal, J., Pol Roger UK Ltd, & Waugh, H. (2013). *Red, Whites & Varsity Blues: 60 Years of the Oxford & Cambridge Blind Wine-Tasting Competition*. London, UK: Pavilion Books.

Weil, R. L. (2001). Parker vs. Prial: The death of the vintage chart. *Chance*, **14**, 27-31.

Weil, R. L. (2005). Analysis of reserve and regular bottlings: why pay for a difference only the critics claim to notice? *Chance*, **18**, 9-15.

Weil, R. L. (2007). Debunking critics' wine words: can amateurs distinguish the smell of asphalt from the taste of cherries? *Journal of Wine Economics*, 2(2), 136-144.

Zajonc, R. B. (1968) Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, **9**, Monongraph supplement No. 2, Part 2.

Appendix A

Frequency of wines shown during the training sessions, by grape (A.1), country (A.2), and region (A.3)

Table A.1

White wines		**Red wines**	
Grape variety	**Frequency**	**Grape variety**	**Frequency**
Chardonnay	21	Syrah/Shiraz	13
Riesling	13	Cabernet Sauvignon	12
Chenin Blanc	9	Pinot Noir	11
Sauvignon Blanc	8	Merlot	10
Gruner Veltliner	8	Grenache/Garnacha	10
Albarino/Alvarinho	8	Sangiovese	7
Pinot Gris/Grigio	8	Tempranillo	6
Gewurztraminer	4	Gamay	6
Melon de Bourgogne	3	Corvina	5
Semillon	3	Zinfandel/Primitivo	5
Marsanne	2	Malbec	4
Gros Manseng	2	Cabernet Franc	3
Garganega	2	Nebbiolo	3
Viognier	2	Touriga Nacional	3
Torrontes	2	Tannat	2
Picpoul de Pinet	2	Mencia	2
Assyrtiko	2	Pinotage	2
Verdejo	1	Carmenere	2
Godello	1	Mouvedre/Monastrell	1
Viura	1	Blaufrankisch	1
Grillo	1	Barbera	1
Vermentino	1		
Savagnin	1		

Table A.2

Country	Frequency
France	82
Italy	22
Spain	20
Australia	16
South Africa	13
Austria	12
USA	11
New Zealand	10
Chile	9
Germany	5
Argentina	5
Portugal	4
Greece	2
Uruguay	1

Table A.3

White wines			Red wines		
Country	**Region**	**Count**	**Country**	**Region**	**Count**
France	Burgundy	12	France	Burgundy	11
France	Loire	10	France	Rhone	10
Austria	Lower Austria	9	France	Bordeaux	9
France	Alsace	9	Italy	Tuscany	7
South Africa	Western Cape	7	Chile	Central Valley	6
Spain	Galicia	7	South Africa	Western Cape	6
New Zealand	South Island	6	USA	California	6
Australia	South Australia	4	Australia	South Australia	5

France	Rhone	4	Italy	Veneto	5
Australia	New South Wales	3	Spain	Rioja	5
Italy	Veneto	3	Italy	Piedmont	4
Argentina	Mendoza	2	Argentina	Mendoza	3
Austria	Vienna	2	France	Languedoc	3
France	Bordeaux	2	France	Loire	3
France	Jura	2	New Zealand	North Island	3
France	Languedoc	2	Portugal	Douro	3
France	Southwest	2	Australia	Western Australia	2
Germany	Mosel	2	Chile	Aconcagua	2
Germany	Rhine	2	France	Southwest	2
Greece	Santorini	2	USA	Oregon	2
USA	California	2	Australia	Victoria	1
Australia	Tasmania	1	Austria	Burgenland	1
Chile	Central Valley	1	Germany	Baden	1
France	Corsica	1	Italy	Puglia	1
Italy	Alto Adige Trentino	1	New Zealand	South Island	1
Italy	Sicily	1	Spain	Castilla y Leon	1
Portugal	Vinho Verde	1	Spain	Galicia	1
Spain	Castillo y Leon	1	Spain	La Mancha	1
Spain	Rias Baixas	1	Spain	Murcia	1
Spain	Rioja	1	Spain	Ribera del Duero	1
USA	Oregon	1	Uruguay	Canelones	1

APPENDIX B

Table of vintage distribution of all wines (212) used during the training sessions

White wines		**Red wines**	
Vintage	**Count**	**Vintage**	**Count**
2005	1	2005	1
2006	1	2006	1
2007	1	2007	2
2008	1	2008	1
2009	0	2009	4
2010	2	2010	2
2011	0	2011	4
2012	5	2012	5
2013	4	2013	17
2014	10	2014	30
2015	30	2015	28
2016	42	2016	12
2017	7	2017	1

APPENDIX C

Frequency tables of 1) acidity and 2) alcohol levels of all wines (212) shown during the training sessions.

Table C.1)

Acid Level	**Frequency**	**%**
Low (1)	5	2
Medium (3)	22	10
Crisp (4)	116	55
High (5)	69	33

Table C.2)

Alcohol Level	**Frequency**	**%**
Low (1)	5	2
Medium (2)	153	72
High (3)	54	26

APPENDIX D

Histogram of price distribution of all wines (212) used during the training sessions