Food Quality and Preference 76 (2019) 100–108

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Does environment matter? Assessments of wine in traditional booths compared to an immersive and actual wine bar

Mackenzie Hannum[a], Sheri Forzley[b], Richard Popper[b], Christopher T. Simons[a,*]

[a] Department of Food Science & Technology, The Ohio State University, 2015 Fyffe Rd., Columbus, OH 43210-1007, United States
[b] P&K Research, 6323 North Avondale Avenue, Chicago, IL 60631, United States

ARTICLE INFO

Keywords:
Context
Immersive technology
Consumer testing
Wine
Hedonics

ABSTRACT

Immersive environments may restore relevant context during consumer sensory testing and hence, could yield better discrimination and reliability in acceptance tests compared to traditional methods. To date, no study has compared hedonic data from these settings to those obtained in an actual consumption environment. Presently, sixty-two red-wine consumers evaluated the same 4 wines in 3 environments—a traditional sensory booth, an immersive wine bar, and an actual wine bar. For each wine, subjects evaluated overall liking, future consumption habits, and price estimation. Interestingly, wine liking did not differ across the three environments ($p = 0.076$) nor was there a significant wine by environment interaction ($p = 0.955$). However, at the individual level, wine liking was less stable. For each subject, the magnitude of difference in liking scores for each wine was calculated between two environments. On average, the greatest difference in liking scores occurred between the traditional booths and the actual wine bar (1.7 ± 0.1) and was significantly greater than the difference in liking scores between the booths and immersive wine bar (1.4 ± 0.1, $p = 0.008$); the liking difference between the immersive and actual wine bar (1.6 ± 0.1) was intermediate. Consumption behavior was also differentially impacted by environment. Subjects were more willing to order the wines at a wine bar when evaluating in the actual environment compared to the traditional booths ($p = 0.035$). However, environment did not influence the subject's willingness to purchase the wine to drink at home ($p = 0.064$). Generally, consumers were able to accurately differentiate price amongst the wines ($p < 0.001$) but estimated the wines at a higher price point in the actual environment compared to the traditional ($p = 0.049$). Overall, results suggest wine liking scores across environments were stable within the population, whereas greater variability was observed in individuals when comparing liking scores between traditional booths and the actual environment. Due to the marginal effect of the environment additional research is needed to evaluate the use of immersive technologies and better understand under which conditions context is important.

1. Introduction

As consumers seek the newest innovations and trends in an already saturated market, the risk for a failed product launch remains high (Costa & Jongen, 2006, 2010). Though multiple factors contribute to the increasing threat of product failure, poor predictive validity from a consumer sensory evaluation may also contribute. Intentionally designed, traditional sensory testing isolates the consumer's liking of the product by removing as many potential confounds as possible, including social interaction, light, color, sound, and packaging (Lawless & Heymann, 2010). The consumer's experience during a traditional sensory evaluation lacks resemblance to their typical experience (Stelick & Dando, 2018). The absence of relevant contextual information and ecological validity in these traditional testing environments might limit the predictive validity of the data collected during those evaluations, and in turn, could misinform the final decisions on product launch, ultimately resulting in product failure.

To overcome this lack of context in traditional sensory evaluations, companies organize home-use-tests to restore ecological validity. However, the introduction of context in home-use-tests coincides with a loss of experimental control and poses a logistical execution threat. The experimenter can no longer ensure a consistent testing environment and product evaluation for all participants, thus potentially decreasing the reliability of the data collected in a home-use-test. Therefore, to overcome the lack of context and maintain experimental control by limiting the number of confounding variables, we have proposed an intermediate form of context inclusion using immersive technologies (Bangcuyo et al., 2015; Hathaway & Simons, 2017). Prior studies have

* Corresponding author at: The Department of Food Science & Technology, The Ohio State University, 2015 Fyffe Rd., Columbus, OH 43210-1007, United States.
E-mail address: simons.103@osu.edu (C.T. Simons).

https://doi.org/10.1016/j.foodqual.2019.04.007
Received 10 February 2019; Received in revised form 15 April 2019; Accepted 15 April 2019
Available online 16 April 2019
0950-3293/

shown that using immersive technologies to restore relevant contextual information during product evaluations enhances subject engagement during testing and improves the discriminability and reliability of consumer acceptance data compared to data obtained in traditional environments (Bangcuyo et al., 2015; Hathaway & Simons, 2017; Holthuysen, Vrijhof, de Wijk, & Kremer, 2017; Kim, Lee, & Kim, 2016; Sester et al., 2012; Sinesio et al., 2018). A similar effect was shown for perception data using virtual reality goggles (Stelick, Penano, Riak, & Dando, 2018). There is increasing interest in the use of immersive and virtual technologies to restore context and ensure a more representative response during consumer product acceptability studies (Jaeger & Porcherot, 2017; Stelick & Dando, 2018).

Immersive and virtual technologies recreate an artificial environment and not the actual environment a consumer uses a product in. Results from these studies hinge on the premise that the immersive technologies restore some of the context of an actual eating or drinking environment, but it is not a comprehensive translation. To date, no study has compared hedonic or perceptual data from an immersive or virtual environment to those obtained in an actual consumption environment. One study, however, compared a traditional sensory lab environment, a physical re-created airplane environment deemed "semi-real life", and an actual in-flight meal (Holthuysen et al., 2017). In this particular study, due to logistical concerns, different consumers evaluated the products in the different environments, which might have caused different motivations surrounding their initial participation and potentially confounded the conclusions.

Presently, we designed an experiment to compare data collected from an immersive setting to a traditional testing environment, devoid of all context, and to an actual environment, replete with full context. Data obtained from an immersive environment may better reflect those obtained in a natural consumption environment and thus, may provide a more reliable indicator of product acceptability in real-world situations than a traditional testing environment. These comparisons are necessary to fully vet the use of immersive technologies in consumer acceptance testing and determine if results obtained in immersive environments reflect those obtained in actual settings. It is important to identify the potential limitations of immersive technologies since it is unfeasible for these environments to fully recreate a real-life consumption experience.

In this research, we investigated whether data obtained in an immersive wine bar recapitulated liking from the same consumers when tested in an actual wine bar better than when collected in traditional sensory booths. Wine was selected as the stimulus because it is known to be a highly involved product (Kim et al., 2016) and thus, perception of the wine is likely to be easily manipulated by context. We hypothesized the hedonic appreciation of red wines obtained from subjects in an immersive wine bar would better reflect liking of the same products evaluated in an actual wine bar. In addition, we expected wines in the immersive or actual environments to receive higher magnitude liking ratings compared to the traditional environment due to halo effects associated with the positive effect and greater enjoyment from these environments. We also included multiple parameters in the product evaluation to assess differences between the three environments: liking, likelihood to consume at home or in a restaurant, and price estimation.

2. Materials and methods

2.1. Subjects

Eligible subjects who passed the screener questionnaire and qualified for the study were contacted to participate in the study. Eligibility was determined by the following classifications: consumed red wine at least twice a week, consumed wine at home and/or in a restaurant within the past month, previously never visited the experimental testing location for the actual environment (Wine on High, Columbus, OH), and had not participated in a sensory evaluation in the previous month. Sixty-eight eligible subjects (21 male and 47 female) were recruited through The Ohio State University Consumer Sensory Testing Center's recruitment database and consented to participate in the experiments. Data from six subjects were removed due to failure to attend all three experimental sessions and/or incomplete data, resulting in a final sample of sixty-two subjects (18 male and 44 female) ranging in age from 23 to 59 years old. All protocols were approved by the OSU Institutional Review Board (Protocol #2017B0303). Subjects were informed that the complete experiment would take place over three weeks, with one testing session per week at approximately the same time and day each week. All subjects experienced the experimental conditions (sessions) in a randomized order (counterbalanced across all subjects). Two of the sessions, located in the traditional testing environment and the immersive testing environment, lasted approximately 30 min and subjects were compensated $10 at the conclusion of each respective testing session. The third testing session, located at Wine on High, lasted approximately 60 min to accommodate for added logistical constraints (parking, checking-in and breathalyzing multiple panelists); however, actual sample evaluation time was maintained. Subjects were compensated $20 at the conclusion of this session. Subjects were asked to refrain from wearing fragrant perfume or cologne, avoid smoking two hours prior and alcohol consumption six hours prior to each testing session. All responses were recorded on paper ballots. Subjects were instructed to turn their phones off for the duration of the study to ensure proper testing conditions and limit distractions.

2.2. Stimuli

Four commercial cabernet sauvignon wines were purchased: Black Box (Black Box brand, 2016 Chile), 1895 (Norton brand, 2016, Mendoza, Argentina), St. Michelle (Chateau St. Michelle brand, 2014 Columbia Valley), and Freakshow (Michael David Winery brand, 2014 Lodi, California). Wines were composed of cabernet sauvignon varietal grapes, grown from different geographical regions and selected to represent a range of prices and perceived differences. Selection of the appropriate wines was determined in part by intentionally selecting wines that spanned across the spectrum of retail price. St. Michelle and Freakshow were considered the higher-priced wines in the stimuli set, selling in local grocery stores around $15-$25 depending on the season, and 1895 and Black Box were selected to represent the lower-priced wines, both ranging from $5 to $12 per 750 mL bottle. Stimulus selection was based on a preliminary assessment to verify the wines were differentiable from one another and varied across the hedonic continuum. No apparent color differences were observed between the wines and therefore all testing occurred under normal light conditions. Wines were stored undisturbed in a temperature-controlled room (22 °C). Each day testing occurred, a bottle of each wine was opened approximately 60 min prior to the first session. Wine samples were allocated in 1 oz. pours into a clear wine glass (Riedel Glassware, Kufstein, Austria), and delivered at room temperature to the subject in sequential monadic presentation (Ishii, Stampanoni, & O'Mahony, 2008). Each sample received a 3-digit blinding code written with a gold marker (Wine Glass Writer, Fairfax, California) on the wine glass. The presentation order of the wine was randomized and counterbalanced across all subjects, however, the presentation order was maintained within each subject to minimize the influence of order effects on hedonic or other scores (Mead & Gay, 1995). Four and a half hours was the maximal time a wine bottle was open prior to serving. New bottles were opened to compensate for any perceptual sensory changes that might have occurred over the course of the day.

2.3. Procedure

At the beginning of all testing sessions, subjects were read a script

that contained general information about what to expect during the test. Subjects were given instructions to signal when they were finished with the ballot, so the investigator could begin preparing their next sample. To eliminate potential bias from attending numerous sessions and recall of their prior ratings, when subjects returned for their second and third session, they were informed that the wines could be different or the same from what they had tasted previously. There was an opportunity for subjects to ask any questions, and afterwards they confirmed their understanding of the testing session and were guided into the testing environment. Consent, demographics, and typical consumption information was obtained during their first session. At the beginning of all testing sessions, their blood alcohol content (BAC) level was assessed with a breathalyzer (BACtrack® Select S80, San Francisco, CA). All subjects were required to report a 0.00% BAC at the beginning of the session to participate. At the conclusion of every testing session subjects were required to report below a 0.06% BAC before leaving the facility. If they did not meet those parameters, they were instructed to sit in a waiting room for 15 min, drinking water, until their BAC was below 0.06%. For each wine, subjects rated their overall liking assessed on a 9-point hedonic scale (Peryam & Pilgrim, 1957), and level of agreement (5-point Likert scale) to the following two statements: "I would order this wine at a wine bar or restaurant," and "I would purchase this wine to drink at home." They also answered the question, "How would you expect this wine to be priced (one bottle, 750 mL)?" by selecting one of the following price ranges: a) below $5, b) $5–$10, c) $10–$20, d) $20–$50, and e) $50+. During all testing sessions, subjects were instructed to take their time during the evaluation, to ensure full flavor perception when swallowing (or spitting if so inclined), and to relax as to replicate a typical wine consumption experience. Testing always occurred between the hours of 2:00 pm and 6:00 pm. In between each wine assessment, subjects were instructed to rinse with water thoroughly for 1.5 min in order to mitigate carryover effects.

2.3.1. Testing environment – traditional

Testing occurred in a private sensory booth located at The Ohio State University (Fig. 1A). The sensory booth consisted of a 5′ × 7′ room with a counter top and spit sink, pass-through window and door. Subjects were seated at the counter and given instructions regarding what to expect for the session. On the counter was a styrofoam cup for water (Dart Container Corp., Mason, MI) and a paper napkin, both of which were different from the immersive and actual environment. After instructions were read and a breathalyzer reading was taken, the door to the booth was closed and samples and their respective paper ballots were delivered via the pass-through window. Once a subject was done with a sample, they verbally notified the investigator. There was only one subject in the traditional sensory booth during a session.

2.3.2. Testing environment – immersive

Testing occurred in the Immersive Technologies Laboratory at The Ohio State University (Fig. 1B). The testing space consists of two sections, an immersive section and an investigator section, separated by a door and a one-way mirror. Subjects were seated at a high-top table in the immersive section, directly facing the video wall composed of nine 48-inch high-definition LCD screens (model 47VS20-BAA, LG Electronics, Seoul, Korea). The 8′ × 10′ immersive testing space is equipped with surround sound tile speakers (Quam-Nichols, Chicago, IL) and adjustable lighting. The room was darkened except for the light emanating from the video footage. Video displayed on the video wall was recorded in high definition (1080 × 1920 resolution) using a Nikon D5100 video camera (Tokyo, Japan). Video footage depicting sights and sounds experienced regularly at Wine on High (Columbus, OH), a local wine bar and retail establishment that was used for the actual testing environment condition, was displayed via computer feed. The video footage was edited to a final length of 40 min to ensure no breakage in presentation while the subject was evaluating all of the wines. Subjects were given instructions as to what to expect for their tasting session while standing outside of the Immersive Technologies Laboratory. When they had no remaining questions, they were directed into the immersive section and instructed to take a seat. The video footage was already playing at the time of entry to ensure full immersion upon entrance. Subjects were led to a high-top table which held a metal spittoon, glass of water, and a black cloth napkin, all present to represent an immersive context similar to what was experienced in the actual environment setting. The investigator left the immersive section and shut the door, to allow the subject to acclimate to the environment and attend to all the contextual information for one minute. Upon the delivery of their first sample the subject would proceed to fill out the paper ballot. Once they had completed their assessment of that sample, they raised their hand. The investigator removed their wine glass and paper ballot for that sample and subsequently returned with their next sample after the 1.5-minute break was over. There was only one subject in the immersive setting during a session.

2.3.3. Testing environment – actual

Testing occurred onsite at Wine on High (Columbus, OH), a local wine bar and retail store (Fig. 1C). Wine on High was closed to the public throughout the duration of the experiment to ensure a controlled environment across all sessions. Subjects were greeted at the door and instructed to sit at their assigned spot at the bar. There was a maximum of 10 subjects present during a session. All subjects sat at the bar top, facing the bar area. At each of their seats there was a glass of water, metal spittoon, and black cloth napkin (all of which were also included in the immersive environment). There were two additional investigators present during the study, alongside the lead investigator, and the head sommelier and general manager of Wine on High. The general manager was responsible for pouring the 1-oz. samples. The investigators served as runners and delivered the correct samples to each subject. Once all subjects were present, the lead investigator stood behind the bar, facing the subjects and gave them instructions regarding what to expect during the testing session. The investigator encouraged subjects to talk and engage with one another to simulate a

Fig. 1. Subject perspective in all three environments: A) Traditional environment; B) Immersive environment equipped with audio and visual aspects typically associated with the actual environment location; C) Actual location at Wine on High (Columbus, OH).

typical experience at a wine bar, however, only during times when they were not answering the ballot and samples were not present in front of them. They were not allowed to discuss the wines, or anything associated with the testing. The investigators were present throughout the duration of the testing to ensure subjects complied with the instruction. Samples were prepared behind the bar in front of the subjects, resembling what they would experience if they were there ordering wine during normal business hours. However, the wine labels were hidden and blinded with 3-digit codes. Once they were finished with their sample, subjects were instructed to wait until everyone was done. To resemble an accurate actual environment, the investigators would rinse and wash the wine glasses intermittingly throughout the sensory evaluation behind the bar and engage with the subjects, acting similar to a bartender.

2.4. Data analysis

All statistical analyses were performed using SPSS 20 software (IBM, Armonk, NY). Hedonic and likelihood to consume data are reported as means ± standard error (SE). Price estimation data are reported as frequencies within each categorical answer. Figures displaying data were created using GraphPad Prism 5 software (GraphPad Software Inc., La Jolla, CA). A 3-way, repeated measures Analysis of Variance (ANOVA) was used to assess differences in overall liking, and occasion consumption scenarios; subject, testing environment, and wine were treated as main effects. When appropriate, post-hoc Tukey's test was conducted to resolve the direction of the difference. Price estimation data were analyzed using Friedman's test with Wilcoxon signed-rank test used post hoc to identify significant differences between price assessment in the environments and across the wines. An $\alpha \leq 0.05$ was taken as significant.

To assess the stability of subject's wine liking across the testing environments, the absolute difference in a subject's liking scores for a specific wine was calculated across two environments. Additionally, this analysis allows for the determination of whether one wine was more prone to environmental differences than others. This was repeated for all four wines. Levene's test of homogeneity was conducted to ensure the variance in liking scores was similar across the three environments. Such an analysis ensures any change in wine liking between two environments is not simply a reflection of heteroscedastic data. A 3-way ANOVA with Tukey's post hoc analysis was conducted to assess whether the change in liking scores for each wine differed between environments; subjects, wine and environments were main effects. An $\alpha \leq 0.05$ was taken as significant.

3. Results

3.1. Typical consumption environment

As part of their demographic questionnaire, at the conclusion of the first testing session, each subject provided information regarding their typical red wine consumption habits and level of knowledge of wine. Over 75% of subjects reported most often drinking red wine at home compared to a restaurant, bar, or other environment. When asked to rank their typical consumption environment in terms of frequency (Fig. 2), either by themselves, with friends, with food, or for a special occasion/even; over half of the subject's indicated they consume wine "most often with friends", followed by "with food", "special occasion", and "by themselves". Over 80% of subjects reported possessing having "some knowledge about wine" compared to only 7% knowing a lot of information and 10% not feeling knowledgeable about wine. These are important parameters to consider in the interpretation of the data and the effect of context.

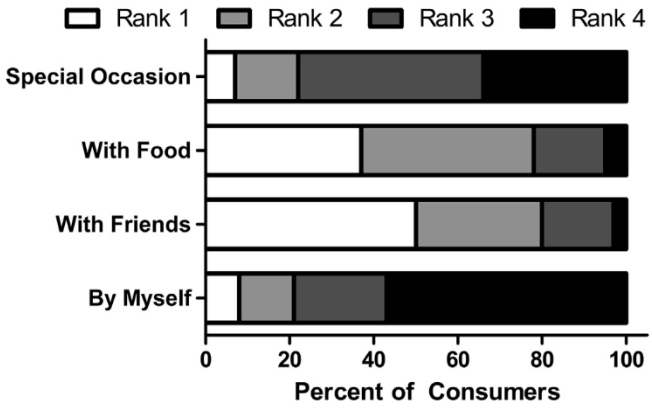

Fig. 2. Typical consumption scenario when asked to rank the above aspects from Rank 1-most often to Rank 4-least often.

3.2. Overall wine liking and liking stability

Data revealed that environment had only a marginal impact on the hedonic ratings of the wines ($F_{2,366} = 2.592$, $p = 0.076$). Generally, hedonic ratings were slightly higher in the actual condition compared to the traditional condition ($p = 0.068$, Fig. 3A), indicating a minor halo effect experienced in the actual environment compared to the

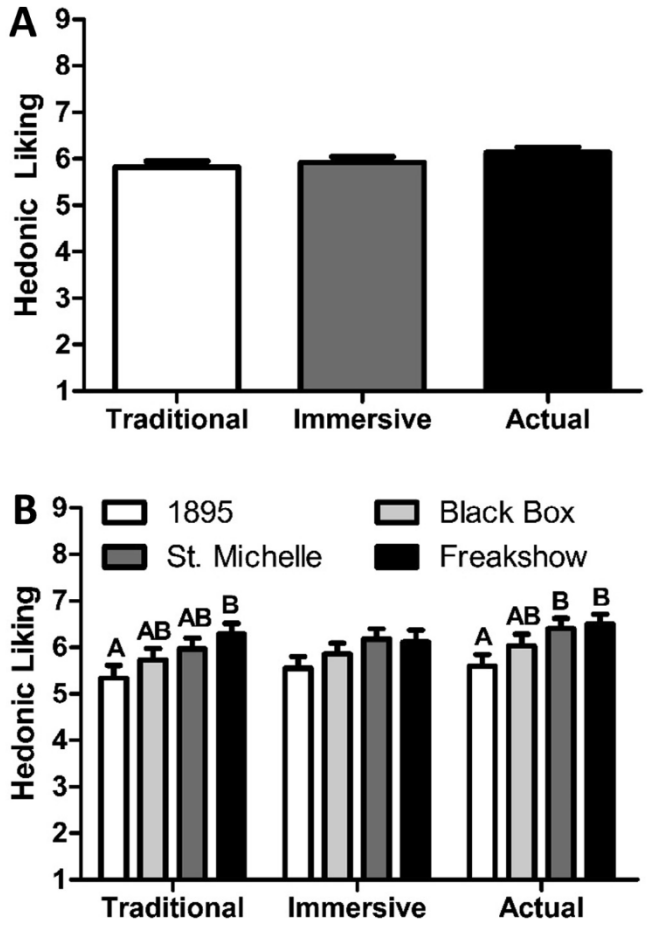

Fig. 3. Liking scores for wines across all three environments. A) Overall wine liking in each of the three testing environments. B) Liking for each wine in each environment. Note, the order of wine liking was consistent across environments although in the immersive environment no significant differences between wines were found. Letters above the bars indicate significant differences ($p < 0.05$) between each wine as determined by Tukey's post hoc analysis.

traditional, with the immersive environment as intermediate. Overall, there was a significant wine effect ($F_{3,366} = 10.213$, $p < 0.001$): St. Michelle and Freakshow received significantly higher overall liking scores than 1895 ($p < 0.001$, < 0.001, respectively). Subjects rated the wines similarly across all three environments as there was not a significant wine*environment interaction ($F_{6,366} = 0.260$, $p = 0.955$). Therefore, the average ratings for each wine were consistent across all three testing environments (Fig. 3B).

There was a significant wine effect ($F_{3,183} = 2.952$, $p = 0.034$) in the traditional environment that tracked similarly with the aggregated hedonic scores across all environments; Freakshow was liked significantly more than 1895 ($p = 0.023$). There were no significant product differences in the immersive environment in the present study ($F_{3,183} = 1.808$, $p = 0.147$). There was a slight improvement in resolution with a significant effect of wine in the actual environment ($F_{3,183} = 3.850$, $p = 0.011$); St. Michelle and Freakshow were liked significantly more than 1895 ($p = 0.035$, 0.014, respectively) with Black Box intermediate. Therefore, the addition of context in the actual environment improved hedonic discrimination of the wines and resolved two product differences (St. Michelle and Freakshow versus 1895) while the traditional only resolved one product difference (Freakshow versus 1895). However, the addition of context in the immersive environment did not resolve any product differences.

To assess the stability of subjects' liking scores across the environments, the absolute difference in liking scores for each wine was calculated and compared across all three environments. The variance in liking scores within each environment was similar ($F_{2,741} = 1.336$, $p = 0.264$) indicating that any potential differences in wine liking between two environments are not due to heteroscedastic data. There was a significant effect of environment ($F_{2,366} = 4.682$, $p = 0.010$, Fig. 4A), subject ($F_{61,366} = 3.922$, $p < 0.001$) and wine ($F_{3,366} = 3.129$, $p = 0.026$, Fig. 4B) on the absolute difference in liking scores, suggesting higher inconsistency among some environments, among some subjects, and among some wines. On average, the greatest change in wine liking occurred between the traditional booths and the actual wine bar (1.7 ± 0.1) and was significantly greater ($p = 0.008$) than the difference in liking scores obtained from the booths and immersive wine bar (1.4 ± 0.1). The liking difference between the immersive and actual wine bar (1.6 ± 0.1) were intermediate and not statistically different from the other comparisons. Hedonic ratings for St. Michelle were most stable whereas those for Black Box were most variable. Across all environmental comparisons, Black Box had a significantly greater difference in liking scores (1.7 ± 0.1) compared to St. Michelle (1.3 ± 0.1, $p = 0.024$); 1895 (1.5 ± 0.1) and Freakshow (1.6 ± 0.1) were intermediate and not statistically different from the other wines. The trend that St. Michelle was most stable and Black Box was least stable was consistent across all three environmental comparisons as the wine*environments effect was non-significant ($F_{6,366} = 1.388$, $p = 0.218$).

3.3. Consumption occasion appropriateness

When asked to rate their level of agreement on a 5-point scale, ranging from strongly disagree to strongly agree, to the statement, "I would order this wine at a wine bar or restaurant" environment had a significant effect (Fig. 5A, $F_{2,366} = 3.380$, $p = 0.035$). Unsurprisingly, subjects reported a significantly higher level of agreement to order the wines at a wine bar when they were in the actual environment compared to the traditional booth setting. Wine also had a significant effect ($F_{3,366} = 7.472$, $p < 0.001$); subjects were more willing to order St. Michelle and Freakshow at a wine bar than 1895 ($p = 0.016$, < 0.001, respectively), aligning with the overall liking findings. Overall, the wines selected for consumption in a wine bar or restaurant were consistent in all three testing environments as there was not a significant wine*environment interaction ($F_{6,366} = 1.111$, $p = 0.355$). However, upon closer examination, wines were not resolved in subject's

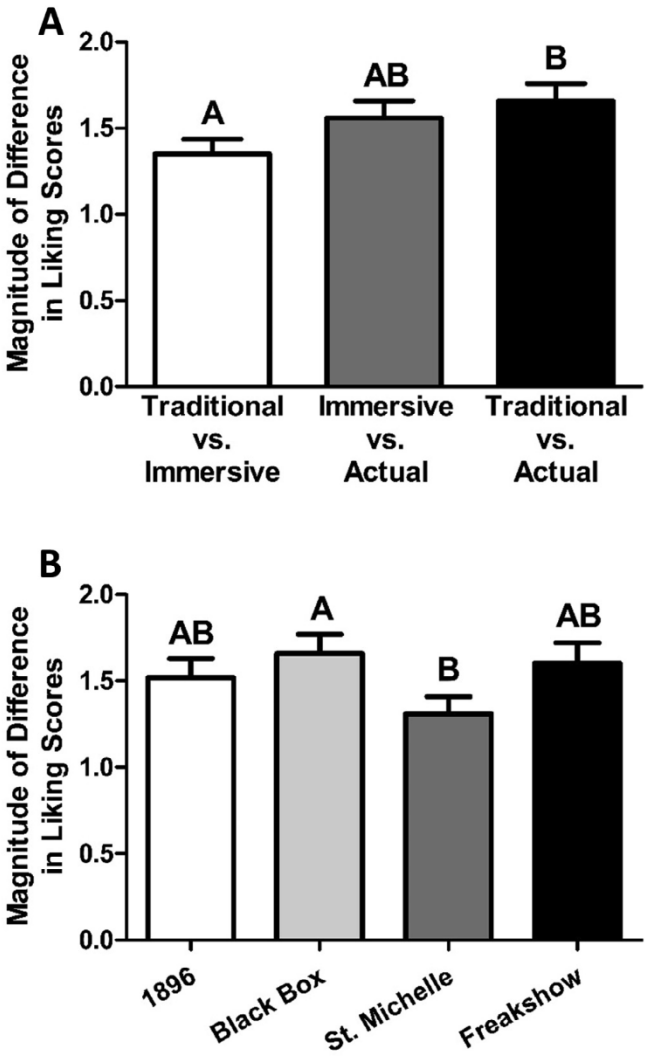

Fig. 4. Absolute magnitude of difference in liking scores between two respective environments. A) Magnitude of difference in liking scores across respective environments averaged over all wines. B) Magnitude of difference in liking scores of each wine averaged across the environments. Letters above the bars indicate significant differences ($p < 0.05$) between each environment and wine as determined by Tukey's post hoc analysis.

willingness to order them in the traditional and actual environments (Fig. 5B, $F_{3,183} = 2.218$, $p = 0.088$ and $F_{3,183} = 2.206$, $p = 0.089$, respectively). Unlike in the hedonic results, for purchase willingness there was improved resolution in the immersive environment compared to the other environments, with a significant wine effect that tracked with overall liking (Fig. 5B, $F_{3,183} = 3.176$, $p = 0.025$); subjects were more willing to purchase St. Michelle compared to Black Box ($p = 0.038$).

When asked if they would purchase the same wine to drink at home, environment was the only factor that did not significantly impact the average agreement ratings (Fig. 6A, $F_{2,366} = 2.775$, $p = 0.064$). Panelist ($F_{61,366} = 4.674$, $p < 0.001$) and wine ($F_{3,366} = 7.970$, $p < 0.001$) both significantly impacted agreement ratings. In trend with the other context statement, there was not a significant wine*environment interaction ($F_{6,366} = 0.340$, $p = 0.916$), indicating wines selected to purchase for home use were consistent in all three testing environments. Overall, subjects were more willing to purchase Freakshow and St. Michelle to drink at home, assigning them significantly higher agreement ratings than 1895 ($p < 0.001$, 0.006, respectively) and aligning with the overall hedonic ratings. Upon closer examination, there was a non-significant effect of wine within each of the traditional

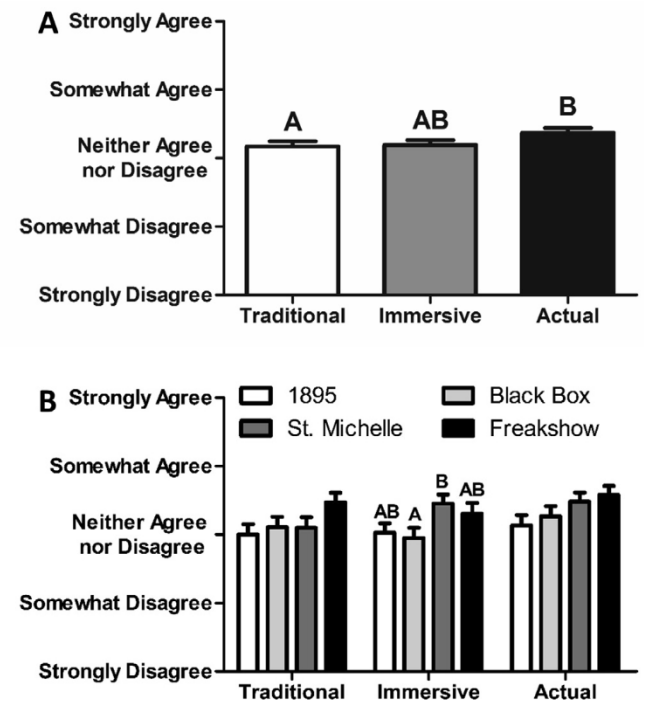

Fig. 5. Subject's average level of agreement to the question "I would order this wine at a wine bar or restaurant." A) Level of agreement in each environment averaged over all wines. B) Level of agreement in each environment for each wine individually. Letters above the bars indicate significant differences ($p < 0.05$) between each wine as determined by Tukey's post hoc analysis.

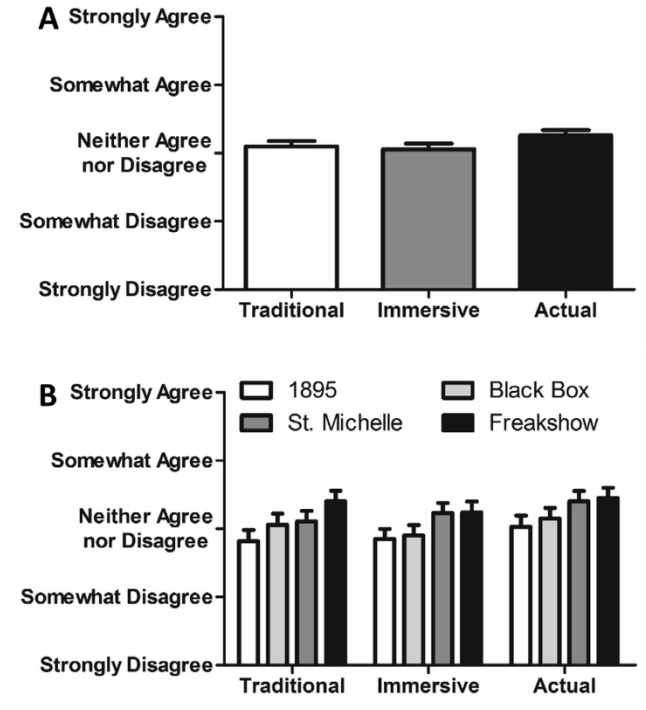

Fig. 6. Subject's average level of agreement to the question "I would purchase this wine to drink at home." A) Level of agreement in each environment averaged over all wines. B) Level of agreement in each environment for each wine individually. No significant differences were noted between environments.

($F_{3,183} = 2.601$, $p = 0.054$), immersive ($F_{3,183} = 2.114$, $p = 0.100$), and actual ($F_{3,183} = 1.944$, $p = 0.124$, Fig. 6B) environments. Thus, providing contextual elements of a wine bar during the evaluation had

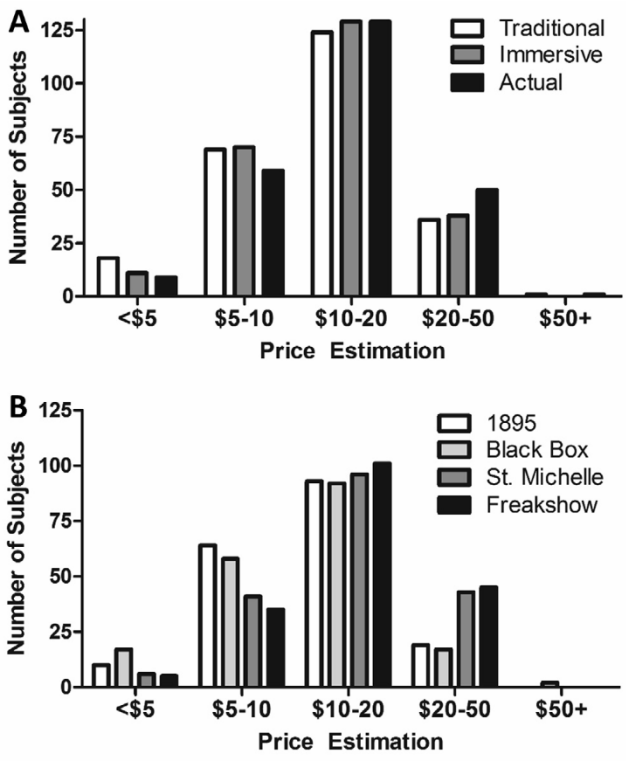

Fig. 7. Price estimation of wines if offered as a standard bottle of wine (750 mL). A) Price estimation summed over wines in each testing environment. Note, the halo effect experienced in the actual environment where a significant increase in price ratings is observed compared to the traditional environment ($p = 0.049$). B) Price estimation for each wine summed across testing environments. Overall subjects accurately discriminated the price of the higher-retail wines compared to the lower-retailed wines ($p < 0.001$). Significant differences ($p < 0.05$) between the environment and each wine were determined by Wilcoxon signed-rank test.

a specific effect on intended purchasing behavior, but was only relevant for the wine bar consumption question and only when asked in the immersive environment.

3.4. Price estimation

Subjects were asked to estimate the cost of the wine if it was being offered in a 750 mL bottle by selecting one of the following price ranges: a) below $5, b) $5–$10, c) $10–$20, d) $20–$50, and e) $50+. As shown in Fig. 7A, there was a significant effect of environment ($\chi^2 = 10.641$, $p = 0.005$) as subjects tended to estimate wine prices higher in the actual environment compared to the traditional environment ($p = 0.049$); the immersive environment was intermediate. Overall, there was a significant wine effect ($\chi^2 = 54.175$, $p < 0.001$), and subjects were able to discriminate between the higher priced wines on the market and the lower priced wines (Fig. 7B). Subjects significantly rated Freakshow and St. Michelle at higher price points than they did for 1895 and Black Box.

4. Discussion

In general, the effect of context on wine liking was minimal with similar findings observed across all three testing environments. However, when subjects were asked about ordering the wine at a wine bar and to estimate the price, increasing context resulted in a higher willingness to order the wine and at a higher estimated price. This latter finding is not surprising as others have noted the addition of context tends to increase positive disposition to the products (Hersleth, Mevik,

Næs, & Guinard, 2003; King, Weber, Meiselman, & Lv, 2004; Meiselman, Johnson, Reeve, & Crouch, 2000). The addition of external cues in both the immersive and actual environment could have raised the subjects' expectations prior to evaluating the wines, which in turn positively influenced their general perceptions of the wine, resulting in the demonstrated halo effect (Deliza & Macfie, 1996). Additionally, the higher incentive paid for attending the actual wine bar may also have contributed to this result. Interestingly, this same halo effect was not observed when subjects were asked if they would purchase the wine for home consumption, suggesting the level of congruency between the testing environment and questions asked may be an important factor to consider.

Overall, results from the consumption occasion questions tracked with overall liking, as subjects were more willing to purchase the wines that received a higher overall liking score. Similarly, subjects perceived St. Michelle and Freakshow, the two most liked wines, to cost more than 1895 and Black Box. While the current study was blinded, this confirms recent research that suggests actual price correlates positively with experienced pleasantness, both in subjective ratings and fMRI scans (Plassmann, O'Doherty, Shiv, & Rangel, 2008). However, other research suggests the correlation between price and overall rating is minor when modeled over 6000 blind wine tastings, and the average consumer prefers a less expensive wine over a pricier option compared to experts (Goldstein et al., 2008). For this study, wines were intentionally selected to span a range of retail prices. Subjects were able to accurately discriminate the price of wines in all three environments, due possibly to their familiarity with the product category. This discrimination might not have been observed with a respondent sample consisting of less frequent red wine users.

4.1. Effect of context

The effect of context on wine liking was marginal overall, aligning with recent research which found no effect of the environment on the flavor perception, liking, or emotional responses on spiked cabernet wines (Jiang, Niimi, Ristic, & Bastian, 2017). However, data assessed on an aggregated level relies on the assumption that context exerts a consistent effect on all products in a set. More research is needed to better understand what occurs on an individual level (Yackinous et al., 1999). In our experiment we noticed consumers changed their liking ratings of the wines between environments even though wines were presented in the exact same order during each testing session which controlled for order effects. This is consistent with previous findings that demonstrate individuals display inconsistency and variability in their preferences across repeated assessments (Chapman & Lawless, 2005; Koster, Couronne, Leon, Levy, & Marcelino, 2003). However, our data begs the question of how much variability in a subject's response is due to natural individual inconsistency versus the varying levels of context influencing their hedonic appreciation? Previous studies using immersive technologies have demonstrated the effect of context actually stabilizing individual ratings across replicates (Bangcuyo et al., 2015; Hathaway & Simons, 2017). The present results extend these findings by indicating that providing context consistent with an actual consumption scenario elicits liking scores that are more consistent with those obtained in actual environments. Further research is needed to understand the concept of individual variability and how different individuals might be influenced more or less by the addition of context.

Furthermore, there is a need to assess under what conditions and for what product-environment combinations context is a key factor to consider. Subjects reported home to be the most prominent location where they consume red wine, which potentially dampened the effect of the wine bar context on their evaluations. We only required that consumers ordered wine at a wine bar or restaurant at least one time a month, therefore, consumption in a wine bar is not necessarily representative of their typical experience. Even more so, subjects might have only consumed wine at a restaurant in the past month as opposed to having visited a wine bar. The generality of the recruit with regard to away-from-home wine consumption may have resulted in a respondent sample for whom the wine bar environment was less familiar and therefore, a less impactful context. If the immersive and actual context in this study had been more similar to their home environment, we might have seen a more pronounced effect of context. Since a product's actual consumption context is not always logistically practical for inclusion in immersive technologies (Hein, Hamid, Jaeger, & Delahunty, 2012), further work needs to investigate when it is both practical and meaningful to provide context through immersive testing.

Similarly, there is a need to better understand for which questions context is an important consideration. The addition of context may not be relevant for all types of assessments and/or questionnaires exemplified in the consumption occasion questions. When considering the same wine, subjects answered the consumption occasion questions differently. Responses to the wine bar consumption occasion question were dependent on the environment, whereas there was no effect of the environment when subjects were asked if they would drink that wine at home.

4.2. Power

When comparing liking of the same products, previous studies in our immersive laboratory indicated a consistent increase in the power of the immersive environment to resolve differences in product liking when compared to the traditional environment (Bangcuyo et al., 2015; Hathaway & Simons, 2017). However, in the present study, there was a decrease in resolution power in the immersive environment compared to both the traditional and actual environment. From a consumer goods industry perspective, presently, the same business conclusion would be drawn from the immersive environment as from the actual and traditional environment (St. Michelle and Freakshow were most liked), though without significant statistical evidence to support that conclusion. In fact, the traditional environment was a sufficient proxy for the actual environment and yielded the same conclusion with similar power as the actual environment, suggesting that context (whether immersive or actual) provided no additional benefit in assessments of the acceptability of the wines in the present study.

Since we did not observe improved resolution for liking in the immersive environment, we expected to see a similar trend in the other parameters. However, when subjects were asked if they would purchase the wine to drink in a wine bar, more significant differences between the products were observed in the immersive environment compared to either the traditional or actual environment. In the immersive environment, the discrimination pattern of the wines followed the overall trend of liking, suggesting the immersive environment was still capable of resolving differences amongst the wines for the consumption question in a comparable manner to the hedonic data in the other two environments. The reason why the liking and context questions yielded differential levels of sensitivity depending on environment is not clear.

4.3. Potential limitations

The study was designed to provide a validation metric for the use of immersive technologies by comparing data to that also collected in an actual environment. There remains an extremely limited amount of research exploring immersive technologies, and even less comparing these immersive environments to a truly realistic consumption environment. This is likely due to the considerable logistical challenges of executing a controlled study in an actual consumption environment. The actual location used in this study provided experimental control and ensured each consumer was subjected to a similar experience when evaluating the wines. As a consequence, while in the actual environment, consumers were still interacting with the product in a manner that was not fully representative of their normal consumption by limiting distractions through regimenting the amount consumed,

instructing when they could taste the wine, and requiring their responses to various questions on a paper ballot. An alternative approach could be to benchmark immersive environments to an actual environment in which these controls are not exercised, allowing free response as much as possible, with minimal experimental influence.

The element of social context and interaction is a major difference between all three environments. In the traditional environment, subjects were completely isolated. In the immersive environment they were in the presence of other wine drinkers, however, this remained a static social interaction as the wine drinkers were only present in the video recording. In the actual location they were surrounded by other sensory subjects and encouraged to engage with one another to improve ecological validity. The addition of social context in the actual wine bar might be one reason for the slight increase in liking in the actual environment versus the traditional environment. Other studies have found that serving food and a social atmosphere had a positive effect on liking for the wine (Hersleth et al., 2003). The social environment and desire for social affiliation (see Cruwys, Bevelander, & Hermans, 2015) may also have changed perceptions in other ways, contributing to the greater differences in wine liking observed in the actual and immersive environments compared to the traditional environment. The modulation of social interaction is an aspect of context that merits further exploration, specifically in immersive technologies.

Another potential limitation of the study was the general level of excitement that consumers had towards the product category. This might have overcompensated for the effect caused by the addition of context in the immersive and actual environment, however, their excitement could have varied across the different treatments. Subjects anecdotally reported feeling happy and excited to participate in the test and evaluate wines. Consequently, they may have been more engaged, across all environments, than a typical subgroup of consumers in a sensory test. The high interest associated with wine tasting and the elevated levels of subject engagement might have mitigated the context effect seen in previous immersive studies (Bangcuyo et al., 2015; Hathaway & Simons, 2017).

4.4. Future of immersive testing

Immersive technologies are a rising trend in sensory testing, and it is important to fully vet the usage and application in food and drink evaluations. However, so much is still unknown regarding the true mechanism by which context exerts its effect, meriting further examination of the technology. The objective of the present study was to further probe the questions: Under what product-scenario combinations is context an important variable to consider and how do results obtained in an immersive environment compare to those from an actual location? We wanted to better understand the limitations of immersive technology and where it lies on the continuum between traditional and actual environments for sensory testing. Specifically, we wanted to assess the potential limitations of immersive technologies and the boundaries of applying this technology in consumer acceptability studies by comparing data collected from an immersive setting to that collected in a traditional testing environment and an actual environment. While the effect of context was marginal in this specific study, there were still differences between the three environments. Halo effects were discovered in the actual environment compared to the traditional, similar to previous studies (Kim et al., 2016; Sinesio et al., 2018). On an overall aggregated level, the variables measured were relatively stable across all three environments, however, upon further examination this stability started to break down as we looked closer at our subjects' individual responses. Future studies are needed to understand the effect of context at the individual respondent level and to investigate whether (and which) contexts (immersive, actual) can act to stabilize subjects' responses. A replicate trial was not included in this original study; however, it is advised to use one in the future to better assess subject stability within each environment. Consistent with our previous findings, we hypothesize that consumers would provide a more stable metric when in a contextually relevant environment across the two replicates compared to evaluating in a traditional booth. In summary, a marginal effect of the environment was observed evaluating wine liking; however, more research is needed to evaluate the use and application of immersive technologies and to better understand under which conditions context is important.

Acknowledgements

Special thanks to Meghan Hardgrove at Wine on High (Columbus, OH) for her willingness to collaborate on the study and usage of the space for sensory evaluations. The authors thank Alex Pierce-Feldmeyer and Brittany Miles for technical execution assistance, as well as the rest of the Food Perception and Liking Lab at the Ohio State University. Research support provided by P&K. Authors RP and SF were paid employees of P&K Research and contributed to the design of the study and interpretation of results.

References

Bangcuyo, R. G., Smith, K. J., Zumach, J. L., Pierce, A. M., Guttman, G. A., & Simons, C. T. (2015). The use of immersive technologies to improve consumer testing: The role of ecological validity, context and engagement in evaluating coffee. *Food Quality and Preference, 41*, 84–95. https://doi.org/10.1016/j.foodqual.2014.11.017.

Chapman, K. W., & Lawless, H. T. (2005). Sources of error and the no-preference option in dairy product testing. *Journal of Sensory Studies, 20*(5), 454–468. https://doi.org/10.1111/j.1745-459X.2005.00039.x.

Costa, A., & Jongen, W. (2006). New insights into consumer-led food product development. *Trends in Food Science & Technology, 17*(8), 457–465. https://doi.org/10.1016/j.tifs.2006.02.003.

Costa, A., & Jongen, W. (2010). Designing new meals for an ageing population. *Critical Reviews in Food Science and Nutrition, 50*(6), 489–502. https://doi.org/10.1080/10408390802544553.

Cruwys, T., Bevelander, K. E., & Hermans, R. C. J. (2015). Social modeling of eating: A review of when and why social influence affects food intake and choice. *Appetite, 86*, 3–18. https://doi.org/10.1016/j.appet.2014.08.035.

Deliza, R., & Macfie, H. J. H. (1996). The generation of sensory expectation by external cues and its effect on sensory perception and hedonic ratings: A review. *Journal of Sensory Studies, 11*(2), 103–128. https://doi.org/10.1111/j.1745-459X.1996.tb00036.x.

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*(01), 1–9. https://doi.org/10.1017/S1931436100000523.

Hathaway, D., & Simons, C. T. (2017). The impact of multiple immersion levels on data quality and panelist engagement for the evaluation of cookies under a preparation-based scenario. *Food Quality and Preference, 57*, 114–125. https://doi.org/https://doi.org/10.1016/j.foodqual.2016.12.009.

Hein, K. A., Hamid, N., Jaeger, S. R., & Delahunty, C. M. (2012). Effects of evoked consumption contexts on hedonic ratings: A case study with two fruit beverages. *Food Quality and Preference, 26*(1), 35–44. https://doi.org/10.1016/j.foodqual.2012.02.014.

Hersleth, M., Mevik, B. H., Næs, T., & Guinard, J. X. (2003). Effect of contextual factors on liking for wine – Use of robust design methodology. *Food Quality and Preference, 14*(7), 615–622. https://doi.org/10.1016/S0950-3293(02)00190-8.

Holthuysen, N. T. E., Vrijhof, M. N., de Wijk, R. A., & Kremer, S. (2017). "Welcome on board": Overall liking and just-about-right ratings of airplane meals in three different consumption contexts—laboratory, re-created airplane, and actual airplane. *Journal of Sensory Studies, 32*(2), 1–9. https://doi.org/10.1111/joss.12254.

Ishii, R., Stampanoni, C., & O'Mahony, M. (2008). A comparison of serial monadic and attribute-by-attribute descriptive analysis protocols for trained judges. *Food Quality and Preference, 19*(3), 277–285. https://doi.org/10.1016/j.foodqual.2007.08.002.

Jaeger, S. R., & Porcherot, C. (2017). Consumption context in consumer research: Methodological perspectives. *Current Opinion in Food Science, 15*, 30–37. https://doi.org/10.1016/j.cofs.2017.05.001.

Jiang, W., Niimi, J., Ristic, R., & Bastian, S. E. P. (2017). Effects of immersive context and wine flavor on consumer wine flavor perception and elicited emotions. *American Journal of Enology and Viticulture, 68*(1) 1 LP-10. Retrieved from http://www.ajevonline.org/content/68/1/1.abstract.

Kim, S. E., Lee, S. M., & Kim, K. O. (2016). Consumer acceptability of coffee as affected by situational conditions and involvement. *Food Quality and Preference, 52*, 124–132. https://doi.org/10.1016/j.foodqual.2016.04.008.

King, S. C., Weber, A. J., Meiselman, H. L., & Lv, N. (2004). The effect of meal situation, social interaction, physical environment and choice on food acceptability. *Food Quality and Preference, 15*(7–8 SPEC.ISS.), 645–653. https://doi.org/10.1016/j.foodqual.2004.04.010.

Koster, E. P., Couronne, T., Leon, F., Levy, C., & Marcelino, A. S. (2003). Repeatability in hedonic sensory measurement: A conceptual exploration. *Food Quality and Preference,*

14(2), 165–176. https://doi.org/10.1016/S0950-3293(02)00075-7.
Lawless, H. T., & Heymann, H. (2010). *Sensory Evaluation of Food.* New York, NY: Springer New York doi: 10.1007/978-1-4419-6488-5.
Mead, R., & Gay, C. (1995). Sequential design of sensory trials. *Food Quality and Preference, 6*(4), 271–280. https://doi.org/10.1016/0950-3293(95)00029-1.
Meiselman, H. L., Johnson, J. L., Reeve, W., & Crouch, J. E. (2000). Demonstrations of the influence of the eating environment on food acceptance. *Appetite, 35*(3), 231–237. https://doi.org/10.1006/appe.2000.0360.
Peryam, D. R., & Pilgrim, F. J. (1957). Hedonic scale method of measuring food preferences. *Food Technology, 11*(Suppl.), 9–14.
Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences, 105*(3), 1050–1054. https://doi.org/10.1073/PNAS.0706929105.
Sester, C., Deroy, O., Sutan, A., Galia, F., Valentin, D., & Dacremont, C. (2012). Having a drink in a bar: an immersive approach to explore the effects of context on food choice. *Food Quality and Preference, 28*, 23–31. https://doi.org/10.1016/j.foodqual.2012.07.006.
Sinesio, F., Saba, A., Peparaio, M., Saggia Civitelli, E., Paoletti, F., & Moneta, E. (2018). Capturing consumer perception of vegetable freshness in a simulated real-life taste situation. *Food Research International, 105*, 764–771. https://doi.org/10.1016/j.foodres.2017.11.073.
Stelick, A., & Dando, R. (2018). Thinking outside the booth — the eating environment, context and ecological validity in sensory and consumer research. *Current Opinion in Food Science, 21*, 26–31. https://doi.org/10.1016/j.cofs.2018.05.005.
Stelick, A., Penano, A. G., Riak, A. C., & Dando, R. (2018). Dynamic context sensory testing–A proof of concept study bringing virtual reality to the sensory booth. *Journal of Food Science, 83*(8), 2047–2051. https://doi.org/10.1111/1750-3841.14275.
Yackinous, C., Wee, C., & Guinard, J.-X. (1999). Internal preference mapping of hedonic ratings for Ranch salad dressings varying in fat and garlic flavor. *Food Quality and Preference, 10*(4–5), 401–409. https://doi.org/10.1016/S0950-3293(98)00054-8.