Journal of Wine Economics, Volume 12, Number 1, 2017, Pages 59–83
doi:10.1017/jwe.2016.27

Dimensions of Expertise in Wine Evaluation*

Robert H. Ashton [a]

Abstract

This article explores the question of what distinguishes novices from experts in wine evaluation. Is it experts' superior sensory abilities related to taste and smell, their superior cognitive abilities related to knowledge and memory, or a combination of both—and if a combination, which of the two dimensions of expertise, sensory or cognitive, seems to be more important? I address these issues by considering what has been learned in the past 30+ years from research concerning the sensory and cognitive dimensions of expertise in wine evaluation. The research examines expert/novice differences at both the chemical component level (detecting, discriminating among, and describing wine-relevant chemical components) and the holistic level (hedonic evaluation of wine as an integrated manifestation of its components). (JEL Classification: C93)

Keywords: cognitive experts, expertise, expert opinion, sensory experts, wine evaluation.

I. Introduction

Expertise in wine evaluation involves both sensory and cognitive dimensions, unlike expertise in many other domains (e.g., financial analysis, economic policy, and the natural sciences) that lack a sensory component. In these latter domains, the expert's task typically involves acquiring, weighting, and integrating verbal and/or numerical information to make a judgment, recommendation, or choice and thus involves only cognitive mechanisms. In wine evaluation, however, much of the information the expert uses is acquired by the senses, thus fundamentally changing the nature of the expert's task. Given this complexity, and recognizing that wine evaluation is also done by novices, a fundamental question arises: To what extent do acknowledged wine experts possess superior sensory abilities with respect to smell and taste, and/or superior cognitive abilities with respect to knowledge and

* I am indebted to two anonymous reviewers and to the editor, Karl Storchmann, for suggestions that have improved the development and exposition of this paper.
[a] Fuqua School of Business, Duke University, 100 Fuqua Drive, Box 90120, Durham, NC 27708; e-mail: robert.ashton@duke.edu.

memory, than those possessed by novices? I consider this question by reviewing and analyzing the rapidly growing body of research on wine evaluation.

This article is related to Storchmann's (2012) analysis of the wine economics literature. Storchmann deftly brings together three important research streams: (1) wine and finance, including the monetary returns to investing in wine relative to other asset classes; (2) wine and the weather, including the influence of climate change on wine quality and prices; and (3) wine and expert opinion, including the inconsistencies and biases in wine critics' ratings of wine quality. I focus on the third area, expert opinion, by exploring what distinguishes those who are considered experts in wine evaluation from those who are not.

Elucidating the nature of expertise in wine evaluation is challenging because wine evaluation involves judgments of sensory quality, which is subjective, situation-specific, and intertemporally dynamic (Lawless, 1995). Wine has no aroma, taste, or mouthfeel until it is sampled by a person, and people vary enormously in their physiological characteristics related to the olfactory, gustatory, and tactile senses. Moreover, what is considered a quality wine for a picnic would likely not be viewed as suitable for a gourmet meal with wine connoisseurs. Finally, with repeated exposure to different wines people often change their ideas about sensory quality.

Thus, one might argue that wine consumers, not wine experts, should be the arbiters of quality and that their idea of quality will be revealed by what they purchase. This argument ignores the fact that wine consumers have budget constraints that limit the range of wines they purchase and consume. Assuming that price and quality are positively related, few consumers will be able to sample the entire range of quality that is available. Moreover, it is unlikely that the mass-produced wines that sell best to the public will be judged as high quality by people who have experienced a greater range of wines. As Lawless (1995, p. 192) puts it, "We would not ask the public to score the quality of the performance of a figure skater or rate the participants in a gymnastics exhibition, so why would we have them be the arbiters of what a fine cheese or great wine should be?"

The importance of expert opinion is echoed in formal analyses by economic theorists who distinguish among search goods, experience goods, and credence goods (Darby and Karni, 1973; Nelson, 1970, 1974). Search goods are products or services whose quality can be evaluated prior to purchase (e.g., an article of clothing). In contrast, the quality of experience goods can be evaluated only after purchase and consumption (e.g., canned tuna). Credence goods are a special case of experience goods whose quality is difficult or impossible to evaluate immediately after purchase (e.g., automobile repairs) as time may be required before quality can be ascertained. Wine is a classic example of an experience good. It can also be considered a credence good because its quality when mature is difficult to assess upon early consumption (Ashton, 2014b). Consequently, economic theory holds that people often turn to expert opinion (along with reputation, advertising, warranties, and price) to judge the quality of experience goods such as wine.

https://doi.org/10.1017/jwe.2016.27 Published online by Cambridge University Press

The following section discusses the sensory and cognitive mechanisms involved in wine evaluation. The relevant research literature is then reviewed. This literature includes studies conducted at the chemical component level, which emphasizes expert/novice differences in detecting, discriminating among, and describing chemical components found in wine, and the holistic level, which emphasizes expert/novice differences in evaluating wine "as a whole," recognizing that wine is more than the sum of its chemical components. A brief discussion and conclusion follows.

II. The Nature of Wine Evaluation

Wine evaluation rests on both sensory and cognitive mechanisms (Morrot, 1999; Parr, 2002). Sensory mechanisms are physiologically based, pertaining not only to taste and smell but also to the visual and tactile senses. Cognitive mechanisms are experience based, including wine-related knowledge, memory, expectations, and the context in which evaluation occurs. Sensory mechanisms entail "bottom-up" processes, reflecting the influence of the sense organs. Cognitive mechanisms entail "top-down" processes, reflecting knowledge, expectations, beliefs, desires, and motives. Cognitive and sensory mechanisms are inextricably linked, as top-down processes often bias the interpretation of sensory information acquired by bottom-up processes.[1]

A. Sensory Mechanisms

Wine evaluation involves assessment of a wine's flavor, a combination of smell and taste. Although smell/olfaction and taste/gustation are distinct physiological systems (Firestein, 2001; Lindemann, 2001), their roles in flavor assessment cannot be disentangled. It is possible to smell a wine without tasting it, but it is not ordinarily possible to taste a wine without smelling it (Goode, 2014). Thus, in terms of physiology, the combination of smell and taste distinguishes one wine from another, with substantial practical implications: "winemakers know full well that small differences in aroma or taste can mean the difference between a premium, gold medal winner and an 'also-ran' table wine" (Swiegers, Chambers, and Pretorius, 2005, p. 109).

Research on the physiology of taste illustrates the difficulty of developing expertise in wine evaluation. For example, although the human tongue has on average a few hundred taste buds per square centimeter, each of which contains up to 100 taste cells

[1] My exploration of sensory and cognitive mechanisms abstracts from incentive-based mechanisms that can lead to intentional bias in wine evaluations. Incentives are, of course, relevant to a more complete understanding of wine evaluation, as documented by Reuter's (2009) finding that, controlling for quality, wineries that advertise in *Wine Spectator* receive higher ratings than do nonadvertisers, and by Goldstein's (2008) receipt of a *Wine Spectator* Award of Excellence for a fictitious restaurant (followed by the magazine's solicitation of a paid advertisement to appear along with its announcement of the award).

or receptors, there is substantial variability across individuals in both the number of taste buds and their sensitivity to different tastes (Bartoshuk, 1993; Lindemann, 2001; Zuniga et al., 1993). Moreover, it was previously believed there were only four basic tastes—sweet, sour, bitter, and salty—and that each could be detected only in a certain region of the tongue, resulting in the notion of a "tongue map" (which maintained that sweet is tasted on the tip, bitter on the back, etc.). It is now recognized that a fifth basic taste exists (umami)[2] and that all five can be detected in any region of the tongue where taste receptors are located (O'Mahony and Ishii, 1986).[3] The variability across individuals in taste sensitivity, combined with the fact that wine quality is considered to depend on a balance of at least three of the five basic tastes (sweet, from sugar; sour, from tartaric and malic acids; and bitter, from alcohol and polyphenols), underscores the difficulty of developing sensory expertise.

Research has examined the possibility that genetic variation in both the number of taste buds and their sensitivity to particular tastes is so great that some people may be "supertasters"—that is, they experience the sense of taste with far greater intensity than the average person (e.g., Bartoshuk, 1993; Bartoshuk et al., 1992, 1993; Miller and Reedy, 1990a, 1990b). In spite of such genetic variation, however, research has not shown that expert tasters are born instead of made. Bartoshuk, who coined the term "supertasters," observes that her own research has not always produced consistent results, and that the documented genetic variation in number of taste buds would lead one to expect greater differences in taste sensitivity across individuals than are actually found (Bartoshuk, 1993; also see Goode, 2008, 2014). Thus, genetic differences in taste sensitivity do not appear to distinguish experts from novices, suggesting that experience-based learning may be more important.

The sense of smell is even more complex than the sense of taste. Although there are only five basic tastes, people can distinguish hundreds of odors (Firestein, 2001; Jackson, 2008). Part of olfaction's complexity stems from the fact that odorous compounds are sensed both directly via the nostrils (orthonasal olfaction) and indirectly via the back of the throat (retronasal olfaction). The latter is especially important in the perception of flavor (i.e., the combined sensation derived from smell and taste).

Odor strongly influences the perception of flavor (e.g., Engen, 1982; Jackson, 2008). Some research even finds that simply imagining an odor has a similar effect (e.g., Djordjevic, Zatorre, and Jones-Gotman, 2004). Moreover, research suggests that the senses in general are not independent of each other but "work together" to influence our perception of objects in the environment (Gilbert, Martin, and Kemp, 1996; Marks, 1978). Although the extent to which the senses are unified is

[2] Lindemann (2001) observes that umami (derived from the Japanese *umai*, or "delicious") is the dominant taste of food containing the amino acid L-glutamate (e.g., chicken broth, meat extracts, and ageing cheese).
[3] Bartoshuk (1993) describes the idea of a tongue map as "an enduring scientific myth" that likely results from the simplicity of the idea and its repeated inclusion in elementary-school textbooks.

not well understood, the basic idea can be illustrated by the common description of a smell as sweet (e.g., "the sweet smell of a strawberry"). Sweet is not a smell (it is one of the five basic tastes), but linking smells with tastes is such an automatic process that it is natural to speak of a "sweet smell" (Goode, 2014).

A demonstration of the critical role of smell in flavor perception is provided by Mozell et al. (1969). Participants tasted 20 common flavors twice, once with their nasal passages blocked and once with them open, and tried to identify each flavor. The flavors included red wine, coffee, chocolate, garlic, dill pickle juice, and several fruit juices. When participants' nasal passages were blocked, 11 flavors were correctly identified by none of the participants, including coffee, chocolate, and cherry (all of which were correctly identified by at least 77% of participants when their nasal passages were open). Across all 20 flavors, there were 60% correct identifications when the nasal passages were open (range = 40–95). When the nasal passages were blocked, correct identifications fell to 10% (range = 0–25).

There is substantial variability across individuals in the sensitivity to odors, and, as with taste, individual differences appear to result from experience-based learning as research has not documented convincing evidence for what might be termed "supersmellers" (Engen, 1982), akin to the lack of convincing evidence for "supertasters." For example, Chrea et al. (2004) had American, French, and Vietnamese participants categorize fruit and flower odors on the basis of perceived similarity. They found that "French and American participants clearly separated fruit from flower odors whereas this separation was nonexistent for Vietnamese participants" (Chrea et al., 2004, p. 669) and explained this result in terms of differences in participants' earlier exposure to fruit and flower odors.

The importance of early exposure is demonstrated in a study of unintentional learning conducted in Germany (Haller et al., 1999). Some years ago, bottle milk that was fed to babies in Germany was flavored with vanilla. Haller et al., asked 133 people (ages 12–59) who were attending a Frankfurt street fair several questions concerning their food habits, among which was: "Were you breast fed or bottle fed as a newborn?" Participants then tasted two samples of ketchup, one flavored with a small amount of vanilla (0.5 g of vanilla per 1 kg of ketchup), and indicated which sample they preferred. Two-thirds of those who were bottle fed preferred the ketchup with added vanilla, whereas 71% of those who were breast fed preferred the normal ketchup, a highly significant difference. Thus, participants' very early olfactory and gustatory experiences were predictive of their preferences much later in life.[4]

[4] Research has established that both taste and smell diminish with age (Bartoshuk et al., 1986). Although there is substantial variability across people in the extent of diminution (Stevens, Cain, and Burke, 1988), smell generally fades to a much greater extent than taste (Stevens and Cain, 1993; Stevens, Bartoshuk, and Cain, 1984).

B. Cognitive Mechanisms

Sensory judgments are only one part of wine evaluation. The other part is cognitive: The signals resulting from contact of a wine with the taster's olfactory and gustatory receptors is transformed by cognitive processes that generate a representation of odor and taste that is unique to the taster. Shanteau's (1987, p. 99) distinction between sensory experts (whom he calls "perceptual" experts) and cognitive experts is relevant: Sensory experts rely on "highly developed sensory skills [to] perceive differences that are not apparent to others," whereas cognitive experts rely on superior knowledge and ability to "think through problems [to] discover relations not found by others." Wine experts must be both sensory and cognitive experts.

Experts and novices in wine evaluation are likely to differ greatly in knowledge and related memory structures—for example, knowledge of grape varieties, wine-producing regions, and characteristic styles of wines (Solomon, 1991). A potential benefit of experts' superior knowledge and memory is the ability to categorize wines and their components at a "deeper" level, such as grape variety, whereas novices may be attuned only to a wine's "surface" features, such as sweetness or fruitiness (Chi, Feltovich, and Glaser, 1981). Categorization-based experience may enable experts to identify certain features that are common to a particular category. For example, red wines made from the Pinot Noir grape may be seen to bear a "family resemblance" (Rosch and Mervis, 1975) to other members of the category such that an expert has in mind a prototype of a Pinot Noir whereas a novice does not. With greater experience, the expert may develop more differentiated prototypes for Pinot Noirs from France, New Zealand, Oregon, and so forth.

Of critical importance are the expectations of the taster. Expectations have been shown in many contexts to exert a strong influence on the interpretation of subjective experiences, including both perceptions of and memories about those experiences (e.g., Klaaren, Hodges, and Wilson, 1994). The source of such expectations could be the taster's own experiences, the reports of others, or the presence of certain cues in a particular setting. Moreover, cues that influence the taster's expectations could be intrinsic (e.g., sensory cues encountered during previous tasting experiences) or extrinsic (e.g., nonsensory cues such as the price and prestige of the wine).[5]

A complicating factor is that nonsensory cues can enhance or degrade the pretasting perception of a wine, which, in turn, can influence sensory perception during tasting (Cardello, 1994; Deliza and MacFie, 1996). Research shows that when tasters are told wine A is more expensive than wine B, they perceive wine A to taste significantly better, even though the two wines are actually identical

[5] A wine's prestige reflects attributes such as country of origin (e.g., France vs. Chile), region or subregion within country (e.g., California's Napa Valley vs. New Jersey), the chateau or producer (e.g., Chateaux Lafite Rothschild vs. Yellowtail), whether the wine is a "reserve" or "regular" bottling, and the wine's vintage or age. Prestige and price are related because high-prestige wines are generally higher priced, but price and prestige are distinct concepts (Ashton, 2015).

(Plassmann et al., 2008). Other research finds that tasters perceive a white wine to taste better when it is described as American instead of Chilean, and even better when it is described as French, although they are tasting the same wine (Veale and Quester, 2008). Still other research shows that when a French red wine is poured from a bottle bearing a prestigious ("*grand cru*") label, tasters describe it as full, complex, and balanced, but when the identical wine is poured from a bottle bearing an unprestigious ("table wine") label, the same tasters describe it as light, unbalanced, and faulty (Brochet, 2001). In contrast, when tasters are unaware of price and prestige cues (i.e., the wines are tasted blind), several studies show that price and prestige effects disappear (Ashton, 2015).

Finally, the context in which wine evaluation occurs is also important. Contextual factors include the purpose of the evaluation (e.g., to provide advice to consumers vs. producers vs. retailers or to judge wines in a head-to-head competition), whether the wine is to be consumed with food, and the social setting surrounding its consumption. The typical study is silent concerning context, and the possibility exists that different evaluators assume different contexts. The importance of context is demonstrated by Sauvageot (1999), who had enology students at the University of Dijon blind taste the same three Burgundy wines in a university classroom and a nearby winery (where the wines were poured by the winemaker himself). Half of the participants tasted first in the classroom and half in the winery, and participants were not told they would taste the same wines later. All three wines were perceived to taste significantly better in the winery. Sauvageot (1999, p. 69) invoked the notion of expectations to explain this result, suggesting that "in a winery, the students would think that the wines presented by a wine-grower are necessarily of good quality, even if they are part of a series of analyses."

III. Research in Wine Evaluation

There are two traditions of wine-evaluation research. One focuses on the chemical components found in wines, whereas the other focuses on the holistic evaluation of wine as a manifestation of its components (i.e., on the hedonic evaluation of a wine's "quality"). Wine contains hundreds of chemical components, some 20 of which appear in concentrations significant enough to influence its odor, taste, or mouthfeel (Goode, 2014; Thorngate, 1997). These components result from the grapes themselves, the metabolism of yeasts during fermentation, and the manner in which wines are aged (e.g., the geographic origin and species of oak used in barrel aging). Stated differently, wine's chemical components reflect a combination of grape variety, local climate conditions, soil composition, and vineyard location (e.g., altitude and sun exposure), as well as the wine producer's viticultural and enological practices.

Research in this tradition examines whether experts possess greater sensitivity than novices in (1) detecting chemical components found in wines, (2) discriminating among types or intensities of different components, and (3) describing the detection and discrimination of such components in meaningful ways (i.e., whether experts have developed a shared language for communicating wine-related sensory perceptions). To illustrate, Parr (2002) investigated whether wine experts are better than novices at recognizing and identifying odors such as anise, and if so, whether their superiority is due to greater physiological sensitivity to the odor of anise, better memory of the odor of anise, or better ability to name anise. Research of this type tends to be the province of chemists, biologists, and sensory scientists. In contrast, research in the holistic tradition elicits responses about a wine's perceived quality, or simply the extent to which the taster likes or enjoys it. Tasters must integrate the information provided by the interaction of their senses and the wine's chemical components with the knowledge, memories, expectations, and so forth that they bring to the task and provide an overall hedonic evaluation of the wine.

In both research traditions, studies that compare expert versus novice performance have used widely varying criteria for designating who is an expert. Indeed, the same authors sometimes define expertise differently in different studies (e.g., Hughson and Boakes, 2001, 2002). Typically, however, research participants are designated as experts if they satisfy any of the following criteria: (1) winemaker, wine writer, or wine retailer, distributor, or broker; (2) formal training in wine tasting or holder of a professional wine-related credential; (3) judge in wine competitions; (4) professor or graduate student of viticulture or enology; or (5) history of extensive wine "involvement" such as regular participation in formal tastings, owning an extensive wine cellar, or being active in organizations dedicated to wine appreciation (e.g., Hughson and Boakes, 2001; Lawless, 1984; Melcher and Schooler, 1996; Parr, White, and Heatherbell, 2004). People who meet none of these criteria but who consume wine regularly are designated as novices.

A. Evidence from Detection and Discrimination Tasks

Research on the sensitivity that people exhibit in detecting and discriminating among wine-relevant chemical components, which are often studied together, relies on the notion that experience separates experts from novices. The role of experience in perceptual learning was observed long ago by Gibson and Gibson (1955), who used wine evaluation as an example. In essence, those deemed experts are thought to have the experience necessary to make fine chemosensory discriminations, whereas those deemed novices are not.

Evidence on expert/novice differences in chemosensory discrimination is mixed. Berg et al. (1955) investigated 16 chemical components contained in wine (e.g., tartaric acid and potassium bitartrate), finding that wine experts do not have lower absolute thresholds than novices for detecting wine-related chemosensory stimuli. Bende and Nordin (1997) found the same result with respect to chemosensory stimuli in general (i.e., odors unrelated to wine). Bende and Nordin (1997) found, however, that wine experts are superior to novices in a binary discrimination task:

The experts could better discriminate between an odor of cloves and an odor mixture of cloves and lemon (the latter being a common odor in white wines). More generally, research has established that experts find it extremely difficult to distinguish odors in a mixture containing more than two constituent odors (Laing and Francis, 1989; Livermore and Laing, 1996), and that task-specific training does not help. By implication, it seems unlikely that wine experts would exceed this level of performance. In contrast, Parr, Heatherbell, and White (2002) and Parr, White, and Heatherbell (2004) found wine experts to outperform novices in tasks involving the recognition of some two dozen previously sampled wine-relevant odors. However, none of the studies described thus far involved actual wines—only chemical substances that are found in wines.

In studies that examine actual wines, Parr and colleagues investigated the ability of experts and novices to discriminate sensory attributes of wines produced from the Sauvignon Blanc grape grown in New Zealand and France (Parr et al., 2007, 2010; Urdapilleta et al., 2011). Among the results are the following: (1) When French wine experts evaluated French and New Zealand wines by orthonasal olfaction (smelling but not tasting), they successfully discriminated the French from the New Zealand versions. Moreover, they discriminated regional (within-country) differences in odor profiles for the French but not the New Zealand wines, demonstrating the importance of superior knowledge and memory concerning the French version. (2) When New Zealand wine experts performed a similar task for New Zealand wines (except that both smelling and tasting were allowed), they exhibited substantial agreement concerning aroma and taste characteristics that most clearly typify a New Zealand Sauvignon Blanc. (3) When experts and novices with respect to New Zealand wines sorted the sensory properties of New Zealand wines into a hierarchical tree structure, experts exhibited substantial agreement, but novices did not. The structures that emerged from experts did not depend on whether they tasted the wines before sorting (thus using bottom-up processes) or relied on their memories of the wines' sensory characteristics (thus using top-down processes).

Ballester and colleagues reported similar studies with respect to wines made from the Chardonnay grape (Ballester et al., 2005, 2008). In one study, experts and novices judged the extent to which 48 French whites were representative of Chardonnays. Twenty-nine of the wines actually were Chardonnays, whereas 19 were not. Novices could not reliably distinguish the Chardonnays from the non-Chardonnays. There was reasonably good agreement among experts that the actual Chardonnays were more representative of Chardonnays than were the non-Chardonnays, but sizable overlap in categories existed even for experts. Experts were better able to discriminate between Chardonnays and non-Chardonnays if they tasted the wines in addition to smelling them, but novices were not. A study that focused on just Chardonnay and Muscadet wines investigated expert/novice differences in sorting wines based on orthonasal olfaction. Experts could discriminate the two grape varieties, whereas novices could not. In sum, the studies by Ballester, Parr, and colleagues demonstrate the critical role played by cognitive processes related to knowledge and memory in wine evaluation.

The biasing effect of a wine's color on perceptions of its aroma and taste has also been studied. Research in many food and beverage contexts demonstrates the effects of color on the detection, discrimination, and description of odors. Effects include the perception that odors are present when they are not, and the effect of color on recognition of odors, perception of odor intensity, and liking of odors (e.g., Engen, 1982; Gilbert, Martin, and Kemp, 1996; Sakai et al., 2005; Spence et al., 2010; Zellner and Kautz, 1990; Zellner, Bartoli, and Eckard, 1991). Other research documents similar biasing effects of color on taste (e.g., Garber, Hyatt, and Starr, 2000; Hoegg and Alba, 2007; Maga, 1974).

In wine evaluation, Morrot, Brochet, and Dubourdieu (2001) had novices smell a white wine that had been colored red by an odorless dye and then choose from a list of aroma descriptors those that best described the wine. Participants overwhelmingly chose red-wine descriptors (e.g., raspberry, cherry, prune, and tobacco) over white-wine descriptors (e.g., lemon, grapefruit, pear, and mango). Parr, White, and Heatherbell (2003) had novices and experts perform a similar task; experts outperformed novices but nevertheless exhibited a color-induced olfactory bias. Experts performed better when the wine was presented in opaque glasses, as they could not see the wine before smelling it. Pangborn, Berg, and Hansen (1963) found experts to perform worse than novices in a tasting task. Pangborn et al., altered a single dry white wine to simulate Sauterne, sherry, rosé, Bordeaux, and Burgundy by coloring it yellow, brown, pink, red and purple, respectively. Participants indicated the degree of sweetness they perceived in each sample. Experts perceived, for example, that a white wine colored to resemble a rosé wine tasted sweeter than the same white wine colored to resemble a red wine. Clearly, these studies reflect tasters' expectations of aroma and taste that result from first seeing a wine's color, suggesting a strong prototype that tasters have developed with experience. Thus, the integration of visual information with that provided by smell and taste is a top-down, cognitive-based process.

Other sensory discrimination research has employed the triangle test (Amerine and Roessler, 1983), which presents tasters with three wines, two of which are similar (or identical) along some dimension and one of which is different. Tasters must identify which wine differs from the other two. Thus, by chance, one-third of the tasters should make correct identifications. Ashton (2014a) and Weil (2001, 2005) found that novices perform at or near chance levels in the triangle test. Ashton (2014a) had novices blind taste wines in two different triangles, each containing two red wines from California and one from New Jersey. In both triangles, tasters performed at exactly chance level (i.e., 33.3% correct identifications). Weil (2001) had novices blind taste wines from vintages rated as "excellent" versus "average to appalling," and Weil (2005) had novices blind taste "reserve" bottlings versus "regular" bottlings from the same vineyard or producer. Correct identifications were made 41% of the time in both Weil studies, significantly better than random guessing but rather modest performance in an absolute sense. However, none of these studies involved experts, only novices.

Solomon (1990) and Melcher and Schooler (1996) compared experts to novices in discrimination tasks. Solomon (1990), using the triangle test, presented tasters with three "cheap white Bordeaux, all very bland and virtually identical in visual appearance" to make the task "as difficult as possible in order to magnify potential group differences" (p. 509). Tasters performed the task eight times with different combinations of wines, so chance performance is 2.7 correct identifications. Novices performed at almost exactly this level (mean = 2.8), whereas experts performed significantly better (mean = 5.5). Melcher and Schooler (1996) had participants first taste a "target" red wine and, four minutes later after a brief distractor task, try to identify the target from a four-wine array. Again, experts significantly outperformed novices. Finally, Solomon (1990) had experts and novices rank five wines in order of the intensity of each of 12 attributes (acidity, tannin, fruit, balance, etc.). Experts (novices) agreed at better than a chance level for 3 (1) of the 12 attributes. Thus, experts outperformed novices, but the absolute performance level of both groups was poor.

B. Evidence from Description Tasks

Much research on describing the odor and taste of wines has been reported, as language is considered critical to the development of expertise. Knowledge of a "wine vocabulary" is seen as essential not only for communicating about wines, but also for a high level of odor and taste discrimination (e.g., Hughson and Boakes, 2001). The language tasters use to describe a wine may shape their perceptions of it (Rabin and Cain, 1984).

In 1975, the linguist Adrienne Lehrer published an article (and later two books: Lehrer, 1975, 1983, 2009) that investigated the structural analysis of the language people use to describe and communicate about wines. Lehrer reported several studies of words and phrases that novices use to describe wines. The key findings were that novices (1) exhibit little agreement on the words and phrases used and (2) cannot match wine descriptions to wines at better than a chance level, regardless of whether the descriptions were written by experts, other novices, or the participants themselves in an earlier research session.

Lehrer's research was followed by several attempts to determine if experts would perform better (e.g., Brochet and Dubourdieu, 2001; Gawel, 1997; Hughson and Boakes, 2002; Lawless, 1984; Melcher and Schooler, 1996; Morrot 1999; Sauvageot, Urdapilleta, and Peyron, 2006; Solomon, 1990, 1997; Weil, 2007). Later researchers were often skeptical about whether Lehrer's results would hold for experts because of the more "disciplined" nature of experts' tasting procedures (e.g., Lawless, 1985).

Lehrer's results for novices were replicated several times. Concerning expert/novice differences, subsequent research found the following: (1) Experts outperform novices in matching word descriptions to wines, but only when they have been trained in the task or are matching descriptions written by themselves or other experts, not by novices (Gawel, 1997; Lawless, 1984; Lehrer, 2009). Even then, matching performance is modest. Lawless (1984), for example, using a set of 6 white wines, found that experts (novices) correctly matched descriptions to only 2.6 (1.8) of the 6 wines. (2) Experts generate about twice as many descriptive terms as novices (Lawless, 1984; Solomon, 1990), although the absolute number tends to be small (about three to seven, on average). (3) Experts generate a greater proportion of "concrete" or more specific terms (Gawel, 1997; Lawless, 1984; Solomon, 1997). A concrete term is one with an external reference standard (such as oak, berry, or vanilla), and a more specific term might be citrus or peach as opposed to simply fruit. In contrast, novices generate more abstract or less specific terms—such as interesting or drinkable. (4) Experts recall more wine-related words in both incidental and intentional learning conditions, but in the latter condition only when the words are grouped in "meaningful" ways—for example, to form possible descriptions of actual wines (Hughson and Boakes, 2002). (5) The expert's advantage stems from a superior ability to compare a wine to a number of prototypes in memory, as experts have in memory a group of "reference wines" (Morrot, 1999). (6) Experts agree somewhat better in both the terms they use to describe wines and their ranking of wines on attributes such as sweetness or acidity (Melcher and Schooler, 1996; Solomon, 1990).

To summarize, experts perform better than novices in various tasks that rely on the development of a wine vocabulary. Much of the advantage seems to result from experts' development of prototypes (e.g., Gawel, 1997; Lawless, 1985; Solomon, 1997). Experts use knowledge about attributes that tend to cluster together in wine produced from different grape varieties, and this knowledge, which is not likely to be possessed by novices, affects their perception and description of a particular wine.

Experts' demonstrated superiority is relatively small, however, and there is substantial variability in performance among them. For example, in the Lawless (1984) study, where experts outperformed novices in matching descriptions to wines and in generating a larger number of descriptors, there was little agreement in the descriptors generated: experts "almost never agreed on the more subtle … characteristics of the wines [and] the vast majority of terms were only used by one individual for a given wine" (p. 108). Similar results were found by Brochet and Dubourdieu (2001). Unlike other researchers, they analyzed descriptors included in thousands of published tasting notes written by four expert wine critics. Only about 30% of the descriptors were common between pairs of critics. Unfortunately, it is not clear whether this level of disagreement should be ascribed mostly to the experts' sensory or cognitive characteristics or simply to the inherent difficulty of communicating sensory perceptions (Goode, 2014).

C. Evidence from Holistic Judgment Tasks

The judgments described thus far concern detecting, discriminating, and describing chemical components of wines. The second tradition of wine-evaluation research views wine as an integrated manifestation of its components and has much in common with judgment/decision-making research in other domains (see, e.g., Hogarth, 1987). In these domains, the focus is on cognitive processes, and experts' judgments are evaluated using performance-based criteria such as accuracy, validity, calibration, and coherence. In sensory domains, criteria such as accuracy (correspondence with an objectively measured external criterion that is independent of the expert) typically cannot be employed, and attention turns to surrogate criteria such as reliability and consensus. Reliability concerns the similarity of repeat evaluations of the same wine by an individual judge, whereas consensus concerns the similarity of the evaluations of a particular wine between/among two or more judges working independently. Reliability and consensus are considered necessary but not sufficient for establishing expertise.

Both reliability and consensus are considered extremely important by wine researchers. Concerning reliability, Hodgson (2009b, p. 241) states that reliability is the most important quality for professional wine judges "for if a judge cannot closely replicate a decision for an identical wine served under identical circumstances, of what value is his/her recommendation?" Concerning consensus, Hodgson (2008, p. 106) says simply "good judges agree with each other." Several reliability and consensus studies have been reported, in which experts in blind tastings independently evaluate the quality of several wines on numerical rating scales. As in other fields, correlational measures of reliability and consensus are employed in almost all of these studies.

(1) Reliability of Wine Evaluation

The results of 12 reliability studies, all involving experts, were analyzed by Ashton (2012). Mean reliability ranged from 0.31 to 0.74 across studies, with a grand mean of 0.50. Within studies, tremendous variability existed across judges, ranging from –0.49 to +1.00. To benchmark this level of performance, Ashton relied on an earlier analysis that reviewed 41 reliability studies across six fields—meteorology, medicine, clinical psychology, personnel management, business, and auditing (Ashton, 2000). Mean reliability ranged from 0.70 to 0.91 across fields, with a grand mean of 0.80. Thus, perhaps not surprisingly given the sensory nature of wine evaluation, the reliability of wine experts is substantially below that of professionals in other fields.

One reliability study reviewed by Ashton (2012) included novices, thus allowing a direct comparison with experts. Lawless, Liu, and Goldwyn (1997) had three expert groups and one novice group taste—and retaste less than an hour later—14 different wines. Each of the expert groups displayed substantially greater reliability (means across group members of 0.42, 0.53, and 0.61) than the novices (mean = 0.31).

A reliability study that employed noncorrelational measures was reported by Hodgson (2008), who analyzed data from the California State Fair Wine Competitions of 2005–2008. Four triplicate samples of wines were judged by 16 panels of four judges each. Both repeat samples were tasted in the same tasting flight and were poured from the same bottle as the original sample. Judges assigned numerical scores and awarded medals to each wine (gold, silver, bronze, or no award), and the main results concern their ability to replicate their own awards. Hodgson (2008) says, "The overriding principle was to design the experiment to maximize the probability in favor of the judges' ability to replicate their scores" (p. 106). Judges awarded the same medal only 18% of the time—usually for wines that received no award. In many instances, judges awarded gold to one of the triplicate samples and either bronze or no award to another.

(2) Consensus of Wine Evaluation

Ashton (2012) also analyzed the results of 9 studies, all involving experts, in which pairwise measures of consensus were reported, comparing the results to those from 46 studies in the same six fields mentioned earlier. Mean consensus ranged from 0.11 to 0.58 across studies, with a grand mean of 0.34. Within studies, there was again tremendous variability, with pairwise correlations ranging from −0.62 to +0.99. Across the other six fields, mean consensus ranged from 0.37 to 0.75, with a grand mean of 0.57.[6]

Hodgson (2009a) reported consensus results from a study that employed noncorrelational measures. Hodgson analyzed data concerning 4,167 wines entered in 13 major U.S. wine competitions in 2003. One hundred six of the 375 wines entered in five competitions received gold medals in one competition, but only 20 of these 106 received a second gold medal, and only 6 of these 20 received a third. None of these 375 wines received gold medals in more than three competitions. In addition, only 132 of the 3,347 wines entered in two or more competitions received the same medal in all competitions entered (almost always in just two competitions). Finally, of the 2,440 wines entered in more than three competitions, 1,142 wines received at least one gold; however, 957 of these 1,142 wines failed to receive any medal in at least one competition. Hodgson (2009a, p. 5) concluded that "wine judges concur in what they do not like but are uncertain about what they do." An earlier study by Cliff and King (1997) had reached a similar conclusion.

A drawback of using consensus as a quality measure was vividly demonstrated in Ashton's (2014a) study of California and New Jersey wines: All of the experts could agree yet all be wrong. In addition to the triangle tests with novices described earlier, Ashton had experts with up to 25 years of experience in the wine business blind taste four California and two New Jersey red wines. In addition to providing enjoyment ratings, the experts indicated what they believed to be each wine's most likely place of origin. There was complete disagreement for one of the New Jersey wines —it was believed to be from Argentina, France, Italy, and Virginia. For the other

[6] It is interesting to note that mean consensus in the wine studies (0.34) is only slightly below that in clinical psychology (0.37).

New Jersey wine, there was complete agreement—all of the experts believed it was from California. Which of the two wines generated poorer performance is an interesting question.

The consensus studies described previously involved experts as that term has typically been defined—that is, experienced professionals who work in the wine industry or have significant training and practice in wine evaluation. There are, of course, degrees of expertise in any field. In wine evaluation, the top of the expertise distribution is believed to be occupied by a few world-renowned wine critics who evaluate thousands of wines each year and disseminate both numerical ratings and tasting notes. Robert Parker is often viewed as the very top of this distribution with Jancis Robinson a close second, though certainly not everyone would agree with this ranking (e.g., Nossiter, 2009). Such critics are extremely influential with wine consumers. Critics' ratings can be controversial (and confusing) when they disagree in their evaluations of the same wine. Some observers assert that critics' ratings lack any value whatsoever (e.g., Quandt, 2007). From a practical standpoint, however, to the extent that critics agree in their evaluations, it matters little which critic consumers choose to follow. When critics systematically disagree, consumers can identify a critic whose "taste profile" is similar to theirs and follow that critic's recommendations (Taber, 2011).

Research has analyzed the consensus among the ratings of several prominent wine critics. Ashton (2013) studied the ratings given to red Bordeaux wines from the 2004 to 2010 vintages by six critics. The wines had been rated while still in barrel during the spring following the fall harvest, 18–24 months before they were bottled and before wine producers had established prices in the Bordeaux futures market. The main analysis concerned the 98 wines rated by all six critics in all seven years, thus eliminating any variability resulting from different critics evaluating different wines.

All 15 possible pairwise correlations were calculated for each of the seven years, and all were positive and significant at the 0.01 level. The mean pairwise correlation was 0.60, which is considerably greater than the overall mean of 0.34 found in Ashton's (2012) analysis of consensus among experienced wine professionals. Averaged across the seven years, the 15 correlations ranged from 0.45 to 0.69. The lowest correlation was between the ratings of Robert Parker and Jancis Robinson, who are widely known to have divergent preferences with respect to wine "styles" (e.g., Burnham and Skilleas, 2008; McCoy, 2005; Taber, 2011; Voss, 2004).

Masset, Weisskopf, and Cossutta (2015) analyzed the ratings of 12 critics for 122 red Bordeaux wines over the 2003–2012 period. All 66 possible pairwise correlations were positive and significant at the 0.01 level. The mean pairwise correlation was 0.60, exactly as found by Ashton (2013), and the range was somewhat wider (0.38 to 0.77). The correlation between the ratings of Robert Parker and Jancis Robinson (0.44) was the third lowest of the 66 correlations.

Stuen, Miller, and Stone (2015) analyzed ratings published in four U.S. wine magazines (*Wine Advocate, Wine Enthusiast, Wine Spectator*, and *International Wine*

Cellar) for a sample of California and Washington State wines from the 2005 vintage. Multiple critics had provided ratings for each magazine. Ratings provided for both red and white wines were included in a single analysis. For the 45 wines rated by all four magazines, mean pairwise consensus was 0.47 with a range of 0.28 to 0.61.

A drawback of using consensus as a quality measure was noted earlier—all of the experts could agree yet all be wrong. In this regard, it is worth noting that Robert Parker's fame as an expert wine critic skyrocketed in 1983 when, as a relative newcomer to wine evaluation, his opinion of the 1982 Bordeaux vintage strongly disagreed with those of many established critics. As Ashton (2013, p. 233) puts it: "Whatever the field, one is unlikely to become known as an expert among experts by agreeing with everyone else." Parker described 1982 as one of the greatest vintages of the century and argued that the wines would age well and be extremely long lived; he predicted steep price increases over time and urged his readers to buy all they could afford in the futures market. Many other critics offered the opposite advice. For example, the well-established and well-regarded Robert Finigan gave 1982 negative reviews and urged his readers to buy the 1980 and 1981 vintages instead (McCoy, 2005).

Independent tastings of the 1982s over the past 30+ years are seen as validating Parker's view of the vintage, including a 2012 tasting that concluded the vintage was "historic" (Asimov, 2012). In 2006, 50 cases of one of the 1982s, Chateau Mouton Rothschild, sold at auction for US$1.05 million, or US$1,750 per bottle, about two-and-a-half times the amount the seller had paid nine years earlier (Taub, 2006). Ironically, although it was the lack of consensus that propelled Parker's status as an expert, the subsequent validation of his opinion as "correct" is based squarely on the consensus of experienced wine tasters and auction market participants.

In summary, the consensus demonstrated by world-renowned wine critics is greater than that demonstrated by wine experts of less renown, especially in the Bordeaux futures setting. It should be noted, however, that in this setting tastings are generally not blind, in contrast to the tastings done by experienced wine professionals in studies described previously. This difference surely favors the critics, but the extent of the advantage is unknown.[7]

[7] An important body of work, not reviewed here, examines statistical features of experts' ratings that underlie the level of consensus among them. Earlier studies in this stream examine the merits of using parametric versus nonparametric methods to identify statistically reliable differences across both raters and wines (Cicchetti, 2004b; Quandt, 2006). Later studies focus on statistical components of experts' ratings that may help explain the lack of consensus. For example, Bodington (2015) posits that the ratings of different experts reflect a mixture of random, common, and idiosyncratic components, and Cao (2014) and Cao and Stokes (2010) focus on randomness as well as systematic bias and discriminatory ability as components of judge performance.

(3) Composite Judgments in Wine Evaluation

Research has also demonstrated that composites, often simple averages of the ratings of two or more wine experts, outperform the experts themselves. More generally, in settings where "correct answers" exist, and thus accuracy can be evaluated, research in many fields has established that composite judgments are almost always more accurate than those of the individuals (including experts) on whom they are based. In settings where correct answers do not exist, rendering the evaluation of accuracy impossible, research shows that the reliability and consensus of composites are superior to that of individuals. It has long been known that in repetitive tasks the accuracy of a composite must be equal to or greater than the mean accuracy of the individuals on whom it is based (Dawes, 1970; Hogarth, 1978; Larrick and Soll, 2006). The advantages of composites have been demonstrated empirically in business, economics, and other settings as varied as population size, football game outcomes, and livestock prices (see reviews by Armstrong [2001] and Clemen [1989]). It would be surprising if results in the wine-evaluation setting were different, and indeed they are not.

Three studies have demonstrated the superiority of wine-related composite judgments. Lawless et al. (1997) investigated reliability and consensus with three groups of experts and one group of novices. In addition to the usual (individual) notion of intrajudge reliability, Lawless et al., calculated each group's mean rating of each wine, on both the initial and repeat tastings, and determined the correlation between these mean ratings. The resulting correlations were much higher than the mean of the group members' individual reliability values. For the novices, the correlation between the mean ratings on the initial and repeat tastings was 0.52, whereas the mean of the individual reliability values was 0.31. For the three expert groups, the respective values were 0.90 versus 0.53, 0.89 versus 0.61, and 0.82 versus 0.42. A similar analysis of mean ratings found the three expert groups to agree much more with each other than with the novices.

Composite judgments were also the focus of Ashton's (2011) analysis of experts' ratings from the "Judgment of Paris," the famous 1976 tasting of French and California wines. Nine renowned experts, all members of the French wine establishment who were largely unfamiliar with California wines, blind tasted six California and four French reds (and six California and four French whites), rating each on a 20-point scale. To enable "accuracy" to be evaluated, Ashton designated the ratings of one expert as "correct," based on certain statistical features of his ratings, and calculated the correlation between these "correct" ratings and those of each of the other experts (as well as equally weighted composites of the other experts). This approach is consistent with the view of some wine researchers that it is useful to designate one expert in a group as the standard for the purpose of training others to mimic him or her (e.g., Cicchetti, 2004a).

For the reds, forming composites increased accuracy from 0.33 (the mean accuracy of the individual experts) to 0.76 (when all experts were included in the composite). For the whites, the comparable accuracy levels were 0.43 and 0.59. The larger increase in accuracy for the reds reflects the fact that the average interjudge consensus for reds was substantially below that for whites (0.16 vs. 0.44). Thus, the inclusion of more experts in the composite reduced average interjudge variability to a greater extent for reds. Consistent with analyses of composites in other settings, composites of two to four experts were quite effective for improving accuracy.

Finally, Ashton's (2016) analysis of the impact of wine critics' ratings on Bordeaux *en primeur* prices is relevant to the power of composites. Ratings for more than 1,700 wines provided by Robert Parker and Jancis Robinson were analyzed. Ashton first added Parker's and Robinson's rating separately to a regression model containing control variables known to influence *en primeur* prices, finding that adding Parker's ratings increased the explanatory power of the model significantly more than adding Robinson's ratings. However, when both critics' ratings were added, the explanatory power of the model was significantly greater than when only Parker's ratings were included. Thus, as found in many other contexts, the impact of even the best expert in a group can be improved by simply averaging his or her judgments with those of another expert, especially when the two experts disagree.[8]

IV. Discussion and Conclusion

To summarize, research documents that experts generally outperform novices in tasks involving detection, discrimination, and description of wine-related chemical components, although the demonstrated magnitude of experts' superiority is relatively small—perhaps because of the inherent difficulties of researching phenomena with a significant sensory component. Research also shows that prominent wine critics agree much better in holistic wine-evaluation tasks than experts of less renown. The consensus exhibited by both groups is modest, however, again perhaps reflecting the subjective nature of sensory evaluations. Finally, in wine evaluation as in other domains, combining the holistic judgments of two or more experts results in improved judgments.

Considering the sensory and cognitive dimensions of wine-evaluation expertise has many implications for research and practice. Because of space limitations, I will briefly mention only two ideas. A fundamental research issue is how the sensory and cognitive dimensions combine (or should combine) to best characterize expertise in wine evaluation. An additive combination of sensory and cognitive abilities, reflecting a weighted average of the two in which strengths in one dimension

[8] Other research, not reviewed here, has begun to apply more theoretically sophisticated methods than simple averaging to the aggregation of wine experts' opinions. An important subset of this research focuses on the aggregation of ranks and includes the application of Borda counts (Hulkower, 2009), Condorcet methods (Borges et al., 2012), Shapley rankings (Ginsburgh and Zang, 2012), and rank-sum methods (Quandt, 2006).

compensate for weaknesses in the other, is certainly a possibility. Perhaps a better possibility is a multiplicative model, as in the general conceptualization of performance as a multiplicative function of ability and motivation that is one of the oldest tenets in psychology (e.g., Maier, 1955). A multiplicative model easily captures the notions that both sensory and cognitive abilities are necessary for expertise, that neither alone is adequate, and that a sufficiently low level of one will strongly mitigate against expertise regardless of the level of the other. An even better possibility might be a noncompensatory model (e.g., Einhorn, 1970), specifically a conjunctive model in which some minimum level of both sensory and cognitive abilities is required for expertise.

On the practical side, I mentioned earlier the possibility of identifying an expert and training others to mimic his or her approach to wine evaluation. A related possibility involves establishing a "standard" approach to wine evaluation and training novices in that approach. Efforts related to standardization have been underway for years in the form of the "wine aroma wheel" (Noble et al., 1984, 1987), a decision aid that groups wine aromas in a three-tier structure. The inner circle of the wheel (tier 1) contains 12 high-level categories of wine aromas (e.g., fruity, woody, earthy, and spicy). These first-tier categories are divided into 29 second-tier categories, which are further divided into 94 third-tier categories. According to Jackson (2008), only rarely are the dozens of terms included in the wine aroma wheel accurate representations of wine aromas, but the wheel may nevertheless serve the practical purpose of prodding a taster's memory for wine aroma descriptors.

Expertise in wine evaluation is a major research area within the new field of wine economics (Storchmann, 2012). Although such research began decades ago, our understanding of wine-evaluation expertise is now accelerating, thanks in no small part to the 2006 launch of the *Journal of Wine Economics*. Future research should continue to focus on the two dimensions of wine-evaluation expertise, sensory and cognitive, and especially on the intricate relationship between them.

References

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

Armstrong, J.S. (2001). Combining forecasts. In Armstrong, J.S. (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, MA: Kluwer Academic, 417–439.

Ashton, R.H. (2000). A review and analysis of research on the test-retest reliability of professional judgment. *Journal of Behavioral Decision Making*, 13, 277–294.

Ashton, R.H. (2011). Improving experts' wine quality judgments: Two heads are better than one. *Journal of Wine Economics*, 6(2), 160–178.

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

Ashton, R.H. (2013). Is there consensus among wine quality ratings of prominent critics? An empirical analysis of red Bordeaux, 2004–2010. *Journal of Wine Economics*, 8(2), 225–234.
Ashton, R.H. (2014a). "Nothing good ever came from New Jersey": Expectations and the sensory perception of wines. *Journal of Wine Economics*, 9(3), 304–319.
Ashton, R.H. (2014b). Wine as an experience good: Price versus enjoyment in blind tastings of expensive and inexpensive wines. *Journal of Wine Economics*, 9(2), 171–182.
Ashton, R.H. (2015). Price, prestige, and the sensory perception of underdog wines. *Wine & Viticulture Journal*, 30, 63–66.
Ashton, R.H. (2016). The value of expert opinion in the pricing of Bordeaux wine futures. *Journal of Wine Economics*, 11(2), 261–288.
Asimov, E. (2012). From 1982, glasses more than full. *New York Times*, March 21, D1.
Ballester, J., Dacremont, C., Le Fur, Y., and Etiévant, P. (2005). The role of olfaction in the elaboration and use of the Chardonnay wine concept. *Food Quality and Preference*, 16, 351–359.
Ballester, J., Patris, B., Symoneaux, R., and Valentin, D. (2008). Conceptual vs. *perceptual wine spaces: Does expertise matter?* Food Quality and Preference, 19, 267–276.
Bartoshuk, L.M. (1993). The biological basis of food perception and acceptance. *Food Quality and Preference*, 4, 21–32.
Bartoshuk, L.M., Conner, E., Grubin, D., Karrer, T., Kochenbach, K., Palcso, M., Snow, D., Pelchat, M., and Danowski, S. (1993). PROP supertasters and the perception of ethyl alcohol. *Chemical Senses*, 18, 526–527.
Bartoshuk, L.M., Fast, K., Karrer, T.A., Marino, S., Price, R.A., and Reed, D.A. (1992). PROP supertasters and the perception of sweetness and bitterness. *Chemical Senses*, 17, 594.
Bartoshuk, L.M., Rifkin, B., Marks, L.E., and Bars, P. (1986). Taste and aging. *Journal of Gerontology*, 41, 51–57.
Bende, M., and Nordin, S. (1997). Perceptual learning in olfaction: Professional wine tasters versus controls. *Physiology & Behavior*, 62, 1065–1070.
Berg, H.W., Filipello, F., Hinreiner, E., and Webb, A.D. (1955). Evaluation of thresholds and minimum difference concentrations for various constituents of wines. I. Water solutions of pure substances. *Food Technology*, 9, 23–26.
Bodington, J.C. (2015). Evaluating wine-tasting results and randomness with a mixture of rank preference models. *Journal of Wine Economics*, 10(1), 31–46.
Borges, J., Real, A.C., Sarsfield Cabral, J., and Jones., G.V. (2012). A new method to obtain a consensus ranking of a region's vintages' quality. *Journal of Wine Economics*, 7, 88–107.
Brochet, F. (2001). Chemical object representation in the field of consciousness. Application presented for the Grand Prix of the Académie Amorim following work carried out towards a doctorate from the Faculty of Oenology, University of Bordeaux, Bordeaux, France.
Brochet, F., and Dubourdieu, D. (2001). Wine descriptive language supports cognitive specificity of chemical senses. *Brain and Language*, 77, 187–196.
Burnham, D., and Skilleas, O.M. (2008). You'll never drink alone: Wine tasting and aesthetic practice. In Allhoff, F. (Ed.), *Wine & Philosophy: A Symposium on Thinking and Drinking*. Malden, MA: Blackwell, 157–171.
Cao, J. (2014). Quantifying randomness versus consensus in wine quality ratings. *Journal of Wine Economics*, 9(2), 202–213.
Cao, J., and Stokes, L. (2010). Evaluation of wine judge performance through three characteristics: Bias, discrimination, and variation. *Journal of Wine Economics*, 5(2), 132–142.
Cardello, A.V. (1994). Consumer expectations and their role in food acceptance. In MacFie, H.J.H., and Thomson, D.M.H. (Eds.), *Measurement of Food Preferences*. London: Blackie Academic and Professional, 253–297.

Chi, M.T.H., Feltovich, P.J., and Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121–152.

Chrea, C., Valentin, D., Sulmont-Rossé, C., Mai, H.L., Nguyen, D.H., and Abdi, H. (2004). Culture and odor categorization: Agreement between cultures depends upon the odors. *Food Quality and Preference*, 15, 669–679.

Cicchetti, D.V. (2004a). On designing experiments and analysing data to assess the reliability and accuracy of blind wine tastings. *Journal of Wine Research*, 15, 221–226.

Cicchetti, D.V. (2004b). Who won the 1976 blind tasting of French Bordeaux and U.S. *Cabernets? Parametrics to the rescue.* Journal of Wine Research, 15, 211–220.

Clemen, R.T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5, 559–583.

Cliff, M.A., and King, M.C. (1997). The evaluation of judges at wine competitions: The application of eggshell plots. *Journal of Wine Research*, 8, 75–80.

Darby, M.R., and Karni, E. (1973). Free competition and the optimal amount of fraud. *Journal of Law and Economics*, 16, 67–88.

Dawes, R.M. (1970). *An Inequality Concerning Correlation of Composites vs. Composites of Correlations.* Methodological Note, vol. 1, no. 1. Eugene, OR: Oregon Research Institute.

Deliza, R., and 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, 103–128.

Djordjevic, J., Zatorre, R.J., and Jones-Gotman, M. (2004). Effects of perceived and imagined odors on taste detection. *Chemical Senses*, 29, 199–208.

Einhorn, H.J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, 73, 221–230.

Engen, T. (1982). *The Perception of Odors.* New York: Academic Press.

Firestein, S. (2001). How the olfactory system makes sense of scents. *Nature*, 413, 211–218.

Garber, L.L., Hyatt, E.M., and Starr, R.G. (2000). The effects of food color on perceived flavor. *Journal of Marketing Theory and Practice*, 17, 59–71.

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

Gibson, J.J., and Gibson, E.J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, 62, 32–41.

Gilbert, A.N., Martin, R., and Kemp, S.E. (1996). Cross-modal correspondence between vision and olfaction: The color of smells. *American Journal of Psychology*, 109, 335–351.

Ginsburgh, V., and Zang, I. (2012). Shapley ranking of wines. *Journal of Wine Economics*, 7 (2), 169–180.

Goldstein, R. (2008). What does it take to get a Wine Spectator Award of Excellence? *Blindtaste*, August 15, http://blindtaste.com/2008/08/15/what-does-it-take-to-get-a-wine-spectator-award-of-excellence/.

Goode, J. (2008). Experiencing wine: Why critics mess up (some of the time). In Allhoff, F. (Ed.), *Wine & Philosophy: A Symposium on Thinking and Drinking*. Malden, MA: Blackwell, 137–154.

Goode, J. (2014). *The Science of Wine: From Vine to Glass.* Berkeley: University of California Press.

Haller, R., Rummel, C., Henneberg, S., Pollmer, U., and Köster, E.P. (1999). The influence of early experience with vanillin on food preference later in life. *Chemical Senses*, 24, 465–467.

Hodgson, R.T. (2008). An examination of judge reliability at a major U.S. wine competition. *Journal of Wine Economics*, 3, 105–113.

Hodgson, R.T. (2009a). An analysis of the concordance among 13 U.S. wine competitions. *Journal of Wine Economics*, 4, 1–9.

Hodgson, R.T. (2009b). How expert are “expert” wine judges? *Journal of Wine Economics*, 4, 233–241.

Hoegg, J., and Alba, J.W. (2007). Taste perception: More than meets the tongue. *Journal of Consumer Research*, 33, 490–498.

Hogarth, R.M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance*, 21, 40–46.

Hogarth, R.M. (1987). *Judgement and Choice: The Psychology of Decision*. New York: Wiley.

Hughson, A.L., and Boakes, R.A. (2001). Perceptual and cognitive aspects of wine expertise. *Australian Journal of Psychology*, 53, 103–108.

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

Hulkower, N.D. (2009). The Judgment of Paris according to Borda. *Journal of Wine Research*, 20, 171–182.

Jackson, R.S. (2008). *Wine Science: Principles and Applications*, 2nd ed. New York: Elsevier.

Klaaren, K.J., Hodges, S.D., and Wilson, T.D. (1994). The role of affective expectations in subjective experience and decision-making. *Social Cognition*, 12, 77–101.

Laing, D.G., and Francis, G.W. (1989). The capacity of humans to identify odors in mixtures. *Physiology & Behavior*, 46, 809–814.

Larrick, R.P., and Soll, J.B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, 52, 111–127.

Lawless, H.T. (1984). Flavor description of white wine by “expert” and nonexpert wine consumers. *Journal of Food Science*, 49, 120–123.

Lawless, H.T. (1985). Psychological perspectives on winetasting and recognition of volatile flavours. In Birch, G., and Lindley, M. (Eds.), *Alcoholic Beverages*. London: Elsevier Applied Science.

Lawless, H.T. (1995). Dimensions of sensory quality: A critique. *Food Quality and Preference*, 6, 191–199.

Lawless, H., Liu, Y., and Goldwyn, C. (1997). Evaluation of wine quality using a small-panel hedonic scaling method. *Journal of Sensory Studies*, 12, 317–332.

Lehrer, A. (1975). Talking about wine. *Language*, 51, 901–923.

Lehrer, A. (1983). *Wine and Conversation*. Bloomington: Indiana University Press.

Lehrer, A. (2009). *Wine & Conversation*, 2nd ed. New York: Oxford University Press.

Lindemann, B. (2001). Receptors and transduction in taste. *Nature*, 413, 219–225.

Livermore, A., and Laing, D.G. (1996). Influence of training and experience on the perception of multicomponent odor mixtures. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 267–277.

Maga, J.A. (1974). Influence of color on taste thresholds. *Chemical Senses and Flavor*, 1, 115–119.

Maier, N.R.F. (1955). *Psychology in Industry*. Boston, MA: Houghton-Mifflin.

Marks, L.E. (1978). *The Unity of the Senses: Interrelations among the Modalities*. New York: Academic Press.

Masset, P., Weisskopf, J.-P., and Cossutta, M. (2015). Wine tasters, ratings, and *en primeur* prices. *Journal of Wine Economics*, 10(1), 75–107.

McCoy, E. (2005). *The Emperor of Wine: The Rise of Robert M. Parker, Jr. and the Reign of American Taste*. New York: Ecco.

Melcher, J., and Schooler, J. (1996). The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language*, 35, 231–245.

Miller, I.J., and Reedy, F.E. (1990a). Quantification of fungiform papillae and taste pores in living human subjects. *Chemical Senses*, 15, 281–294.

Miller, I.J., and Reedy, F.E. (1990b). Variations in human taste bud density and taste intensity perception. *Physiology & Behavior*, 47, 1213–1219.

Morrot, G. (1999). Can we improve taster performance? In Wine tasting, special issue, *Journal International des Sciences de la Vigne et du Vin*, 29–35.

Morrot, G., Brochet, F., and Dubourdieu, D. (2001). The color of odors. *Brain and Language*, 79, 309–320.

Mozell, M.M., Smith, B.P., Smith, P.E., Sullivan, R.L., and Swender, P. (1969). Nasal chemoreception in flavor identification. *Archives of Otolaryngology*, 90, 367–373.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78, 311–329.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy*, 81, 729–754.

Noble, A.C., Arnold, R.A., Buechsenstein, J., Leach, E.J., Schmidt, J.O., and Stern, P.M. (1987). Modification of a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture*, 38, 143–146.

Noble, A.C., Arnold, R.A., Masuda, B.M., Pecore, S.D., Schmidt, J.O., and Stern, P.M. (1984). Progress towards a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture*, 35, 107–109.

Nossiter, J. (2009). *Liquid Memory: Why Wine Matters.* New York: Farrah, Straus, and Giroux.

O'Mahony, M., and Ishii, R. (1986). The umami taste concept: Implications for the dogma of four basic tastes. In Kawamura, Y., and Kare, M.R. (Eds.), *Umami: A Basic Taste.* New York: Marcel Dekker, 75–93.

Pangborn, R.M., Berg, H.W., and Hansen, B. (1963). The influence of color on discrimination of sweetness in dry table-wine. *American Journal of Psychology*, 76, 492–495.

Parr, W.V. (2002). Exploring the nature of wine expertise. *Australian and New Zealand Wine Industry Journal*, 17, 32–36.

Parr, W.V., Green, J.A., White, K.G., and Sherlock, R.R. (2007). The distinctive flavor of New Zealand Sauvignon blanc: Sensory characterisation by wine professionals. *Food Quality and Preference*, 18, 849–861.

Parr, W.V., Heatherbell, D., and 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.

Parr, W.V., Valentin, D., Green, J.A., and Dacremont, C. (2010). Evaluation of French and New Zealand sauvignon wines by experienced French wine assessors. *Food Quality and Preference*, 21, 56–64.

Parr, W.V., White, K.G., and Heatherbell, D.A. (2003). The nose knows: Influence of colour on perception of wine aroma. *Journal of Wine Research*, 14, 79–101.

Parr, W.V., White, K.G., and Heatherbell, D.A. (2004). Exploring the nature of wine expertise: What underlies wine experts' olfactory recognition memory advantage? *Food Quality and Preference*, 15, 411–420.

Plassmann, H., O'Doherty, J., Shiv, B., and Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 1050–1054.

Quandt, R.E. (2006). Measurement and inference in wine tasting. *Journal of Wine Economics*, 1(1), 7–30.

Quandt, R.E. (2007). On wine bullshit: Some new software? *Journal of Wine Economics*, 2(2), 129–135.

Rabin, M.D., and Cain, W.S. (1984). Odor recognition: Familiarity, identifiability, and encoding consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 316–325.

Reuter, J. (2009). Does advertising bias product reviews? An analysis of wine ratings. *Journal of Wine Economics*, 4, 125–151.

Rosch, E., and Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.

Sakai, N., Imada, S., Saito, S., Kobayakawa, T., and Deguchi, Y. (2005). The effect of visual images on perception of odors. *Chemical Senses*, 30, 244–245.

Sauvageot, F. (1999). Sensory evaluation. In Wine tasting, special issue, *Journal International des Sciences de la Vigne et du Vin*, 61–69.

Sauvageot, F., Urdapilleta, I., and Peyron, D. (2006). Within and between variations of texts elicited from nine wine experts. *Food Quality and Preference*, 17, 429–444.

Shanteau, J. (1987). Psychological characteristics of expert decision makers. In Mumpower, J. L., Phillips, L.D., Renn, O., and Uppuluri, V.R.R. (Eds.), *Expert Judgment and Expert Systems.* Berlin: Springer-Verlag, 289–304.

Solomon, G.E.A. (1990). The psychology of novice and expert wine talk. *American Journal of Psychology*, 103, 495–517.

Solomon, G.E.A. (1991). Language and categorization in wine expertise. In Lawless, H., and Klein, B. (Eds.), *Sensory Science Theory and Applications in Foods.* New York: Marcel Dekker, 269–294.

Solomon, G.E.A. (1997). Conceptual change and wine expertise. *Journal of the Learning Sciences*, 6, 41–60.

Spence, C., Levitan, C.A., Shankar, M.U., and Zampini, M. (2010). Does food color influence taste and flavor perception in humans? *Chemosensory Perception*, 3, 68–84.

Stevens, J.C., Bartoshuk, L.M., and Cain, W.S. (1984). Chemical senses and aging: Taste versus smell. *Chemical Senses*, 9, 167–179.

Stevens, J.C., and Cain, W.S. (1993). Changes in taste and flavor in aging. *Critical Reviews in Food Science and Nutrition*, 33, 27–37.

Stevens, J.C., Cain, W.S., and Burke, R.J. (1988). Variability of olfactory thresholds. *Chemical Senses*, 13, 643–653.

Storchmann, K. (2012). Wine economics. *Journal of Wine Economics*, 7(1), 1–33.

Stuen, E.T., Miller, J.R., and Stone, R.W. (2015). An analysis of wine critic consensus: A study of Washington and California wines. *Journal of Wine Economics*, 10(1), 47–61.

Swiegers, J.H., Chambers, P.J., and Pretorius, I.S. (2005). Olfaction and taste: Human perception, physiology and genetics. *Australian Journal of Grape and Wine Research*, 11, 109–113.

Taber, G.M. (2011). *A Toast to Bargain Wines: How Innovators, Iconoclasts, and Winemaking Revolutionaries Are Changing the Way the World Drinks.* New York: Scribner.

Taub, D. (2006). Fifty cases of "perfect" 1982 Bordeaux sell for $1.05 million. *Bloomberg*, November 20. http://www.bloomberg.com

Thorngate, J.H. (1997). The physiology of human sensory response to wine: A review. *American Journal of Enology and Viticulture*, 48, 271–279.

Urdapilleta, I., Parr, W., Dacremont, C., and Green, J. (2011). Semantic and perceptive organisation of Sauvignon blanc wine characteristics: Influence of expertise. *Food Quality and Preference*, 22, 119–128.

Veale, R., and Quester, P. (2008). Consumer sensory evaluations of wine quality: The respective influence of price and country of origin. *Journal of Wine Economics*, 3(1), 10–29.

Voss, R. (2004). Robinson, Parker have a row over Bordeaux. *San Francisco Chronicle*, May 24, F2.

Weil, R.L. (2001). Parker v. 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.

Zellner, D.A., Bartoli, A.M., and Eckard, R. (1991). Influence of color on odor identification and liking ratings. *American Journal of Psychology*, 104, 547–561.

Zellner, D.A., and Kautz, M.A. (1990). Color affects perceived odor intensity. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 391–397.

Zuniga, J.R., Davis, S.H., Englehardt, R.A., Miller, J.I., Schiffman, S.S., and Phillips, C. (1993). Taste performance on the anterior human tongue varies with fungiform taste bud density. *Chemical Senses*, 18, 449–460.