
Denton Marks
Department of Economics, University of Wisconsin-Whitewater, USA
E: marksd@uww.edu

ABSTRACT

Purpose: Fine wines challenge the consumer to know them and form a meaningful willingness to pay (WTP) for them. Given the wine consumer’s problem of knowing what she is buying, Nelson’s (1970) idea of “experience goods” might apply—goods that must be “experienced” to be accurately evaluated. Lacking own knowledge, consumers might use expert opinion to guide their WTP. However, the evidence of correlation between expert opinion and WTP is uneven. We question whether fine wine is an experience good and whether it can be known vicariously through ratings. The idea that one can know a fine wine by buying one—like knowing refrigerators by buying one (Nelson’s example)—seems misplaced, and the experience of others, including the seller, is also of limited value.

Approach: The paper considers and critiques the expert information available to the consumer as well as other factors that complicate the consumer’s knowing what s/he will find in the bottle. It provides a detailed critique of the variety and structure of wine ratings.

Findings: The critique of wine ratings—each implicitly a weighted average score of often unidentified characteristics with unspecified weights—demonstrates the difficulty of comparing or even understanding them. That result and the philosophical problem of knowing wine raise doubt about treating them as knowledge.

Practical implications: Given the considerable expense often involved, consumers are eager to reduce their perceived risk by knowing what they are buying before they arrive at WTP for fine wine. Experts offer to fill that void, but consumers should be skeptical. The paper closes with a re-characterization of the consumer’s wine purchase decision.

Keywords: wine markets, consumer knowledge, fine wine, experience goods, expert ratings.

© Louis Denton Marks, Jr.
INTRODUCTION

A core idea in our understanding of the demand for wine is the consumer’s willingness to pay--but what determines one’s willingness to pay (WTP)? This question is distinct from what determines the price of a good because it focuses upon the demand side of a transaction, and price—what is offered in exchange for the good—reflects the influence of other influences such as supply and, in the case of wine, government regulation and taxation of alcoholic beverages. The economic theory of the consumer recognizes a number of influences upon WTP, and our focus is two of them: the consumer’s taste for the good and the prices of related goods. That is, the requirements for a rational determination of WTP would seem to involve knowledge of the product and knowledge of related products and the prices one must pay for them.

The analysis begins with a discussion of the consumer’s problem in knowing a wine as a product and of our attempts to capture consumer knowledge as a component of WTP. A primary focus of that discussion is the interpretation of expert opinions about wines which are often used as a representation of consumer knowledge in empirical studies of the determinants of wine prices, following the assumption that wine is an “experience good”. This leads to the larger question of the consumer’s ability to know a wine which is distinct from her access to information about it.

The question of consumer knowledge is intrinsically interesting but also bears upon a variety of other issues. One of the most important is the impact of consumers’ lack of knowledge upon both the size and composition of the demand for wines, especially fine wines. One expects that a variety of factors affect whether consumers have an interest in wine—location (e.g., country), level of education, economic status, age, gender—but, perhaps surprisingly, there is no comprehensive study of that subject. Moreover, many empirical studies of wine consumer behavior draw upon “convenience samples” which have little value in providing reliable, robust evidence of consumer buying patterns (Lockshin and Corsi 2012: p. 19).

The complexity of knowing what one is buying—one of the “perceived risks” of wine purchases (e.g., Lockshin and Hall 2003: pp. 8-9)—likely suppresses demand by risk-averse consumers for relatively expensive fine wine and deters many consumers from taking a serious interest in wine or consistently buying any but the most basic wines. Among the six U.S. market segments identified in Constellation Brands’ path-breaking Genome Project study of wine consumer behavior, the largest segment—about one quarter (23%)—were described as the “Overwhelmed” (Veseth 2008), and another third (34%) had limited knowledge and a limited interest in wine per se: the Image Seekers (20%) who see wine as a status symbol and the Satisfied Sippers (14%) who are brand loyal and see wine as little more than an everyday beverage. Those who likely have the most serious interest in wine and are perhaps willing to tackle the complexity are a small share of the market. For example, Gallup, Inc., has reported that about two-thirds of Americans acknowledged drinking alcohol in a 2012 poll, averaging about four drinks per week, but only one-third of those (so 22% of all consumers) chose wine as their preferred drink. Among those who drink wine, women are 2.5 times as likely (about 16% of all consumers) to drink wine relative to men (Saad 2012)—the remaining 6%. However, while we know little about gender differences in purchases of “luxury wines” (Lockshin and Corsi 2012: p. 18)—which would include fine wine—casual observation (e.g., attendance at live wine auctions) and some reporting (e.g., Hanson 2008) indicate that men are much more likely to be wine collectors: that does not mean that men are much more likely to be fine wine drinkers, but it suggests that this small share of the population is the one making the purchase decisions. It might be larger if wine were easier to know.
We cannot know that the fine wine market is smaller because of this component of perceived risk or, if it is, how much; but the resources spent on wine promotion and education and the prominent place of expert opinion in fine wine marketing and the effort to persuade buyers of pedigree, quality, and rarity suggests that sellers want buyers—and potential buyers—to know more or think they do. Of course, the risk may attract some buyers, but it is difficult to know the importance of this market segment—and the Genome Project did not find them.

The distinction between information and knowledge is important in the discussion that follows. We can relate the two simply by thinking of knowledge as information that the individual has processed, synthesized, and retained as true. The focus of the discussion is the examination of wine as an experience good and the related question of capturing a measure of consumer product knowledge as a determinant of WTP. This is related to a philosophical question of the identification and measurement of taste, and a brief discussion of this appears later in the paper. The analysis closes with some suggestions for next steps.

The analysis necessarily draws upon the idea of expertise. Throughout the discussion, use of the term “expert” means nothing more than someone who seems widely known and influential among consumers according to published evidence of popularity (e.g., circulation numbers for a newsletter or magazine, regular appearance over time of citations in widespread advertising).

**WINE AND TYPES OF GOODS**

At the heart of any transaction is the good involved (perhaps both goods involved if we allow for barter). What is it that defines the nature of a good beyond the tautological definition of that which is transacted? Because it often helps us understand the qualitative differences among various types of transactions in the marketplace (e.g., role of advertising, nature of those transacting, source of supply), economists have identified different types of goods along a variety of dimensions—for example, their relationship to changes in market parameters (e.g., complements/substitutes, inferior/normal), to the nature of the parties transacting (final goods, intermediate goods), to the capacity to confine the benefits and costs of a transaction to a single buyer and seller (public/private), and to the transactors’ information about and knowledge of the good (experience and credence goods). Moreover, beyond all these categories, goods can fall into more than one—for example, an inferior good that is a substitute for some other good.

Most of the world’s wine production falls easily into existing categories because that seems to be the primary goal in its production. “Commodity wine” is designed to be easily understood and identifiable, much like most of the beverages with which it competes such as mass market beer or soda. While it might potentially reflect some of the variations in characteristics that can contribute to the complexity of wine (e.g., differences in blends, differences reflecting changes in vintage and source), the goal is often consistency and homogeneity so that consumers know what they are buying. Some have derisively referred to it as “cola wine” (e.g., Aylward 2008).

At the other extreme of the variety of wines available, a small proportion of global production challenges consumers to know what they are buying. As with commodity wines, we have no clear or accepted definition of these wines, but we shall call these “fine wines”, following terminology often used to advertise major wine auctions (e.g., Sotheby’s, Christie’s) and appearing in some of the academic literature (e.g., Smith 2007c: 48-51). The range of wines available in a large, wine-consuming economy fall all along the spectrum from “commodity” to “fine” so the relevance of the discussion to any particular wine will be a matter of degree.
To some extent, understanding fine wines involves issues of asymmetric information (AI): the producer knows considerably more about a wine’s origins, production, and final characteristics than the consumer; and the consumer hopes to learn that. Beyond that, however, some wines challenge the consumer’s ability to know them and therefore to form a WTP for them based upon personal knowledge. Of course, the role of such personal knowledge depends upon one’s motives for purchase. To the extent that WTP is a reflection of what one expects others to pay—that is, the return one expects from an investment rather than the utility one expects from own consumption—then external indicators that correlate highly with past value (e.g., producer, vintage, scarcity, ratings, returns on alternative investments) are perhaps most or all of what one needs to know since they represent most or all of what many others consider in making their purchases. Also, to the extent that one’s utility depends upon the opinions of others whose valuation is similarly determined by external indicators rather than direct consumption—that is, Veblen or positional goods—then they (the Image Seekers?) may also affect WTP significantly.

Perhaps most challenging, however, is arriving at WTP based simply upon own consumption, that is, the enjoyment one expects from drinking what is in the bottle, presumably divorced from wine’s alcoholic effect, though apparently that is not possible (Postman 2011).

Recognizing the challenge to the wine consumer’s knowledge of what she is buying, some of the most careful research on determinants of wine prices has turned to Nelson’s (1970) concept of “experience goods”—goods whose characteristics and quality are sufficiently costly to discover or to know through search that, in effect, “experiencing” them is the only reasonable way to evaluate them. Purchasing and using them is the cost-effective way to know them: “The consumer has a simple alternative to search: he can use experience, that is, he can determine the quality of brands by purchasing brands and then using them” (p. 327). Nelson uses examples like canned tuna fish and home appliances: “To evaluate brands of tuna fish…the consumer would almost certainly purchase cans of tuna fish for consumption…[and] determine from several purchases which brand he preferred” (p. 312). He also allows for the importance of accessing expert opinion such as Consumer Reports.

Treating wine as an experience good seems well established in the literature. In a recent survey, one researcher states flatly: “…wine is an experience good” (Storchmann 2012: 22). However, for those who have tried a wide variety of fine wines, it seems self-evident that, as a problem in determining WTP, fine wines are considerably more challenging to know than canned tuna fish and refrigerators.

How do we represent that component of WTP that reflects the consumer’s innate enjoyment of the wine? A popular answer to that question in the literature is to represent wine quality—or what we assume consumers take to represent wine quality—by one or more expert opinions of the wine.

Much of the research on wine price determinants predicated upon the idea of experience goods looks at the relationship between price and some external measure of quality such as chateau reputation or expert opinion (e.g., Thornton 2013: 250-263; Hadj Ali and Nauges 2007; Hadj Ali, Lecocq, and Visser 2008): unable to know the wine before buying it, consumers use this external information to guide their WTP with better ratings correlating positively with price. However, while this information may influence WTP, it does little to improve the consumer’s knowledge significantly. This may help explain why the evidence of the impact of expert opinion upon wine prices is mixed and positive correlations tend to be weak (Storchmann 2012: 22).
24): this information may contribute little to the consumer’s knowledge of what she is buying so its influence upon WTP is direct but small.

One might argue that wine consumers do not buy wines that they do not know, and their WTP reflects what they know: somehow they manage to taste a wine through friends or tastings, and those are the only wines they buy. If we limit our attention to fine wines, the proliferation of ratings, tasting notes, and “shelf talkers” and the dependence upon sales staff suggest that simply is not true. Data on the proportion of fine wine sales by consumers who have never tasted the wine are unavailable, but most—perhaps the large majority—of such transactions likely fall into that category (e.g., Thornton 2013: pp. 241-2). Given the virtually unlimited variety of fine wines and, among them, the further variability due to factors such as vintage and provenance, most purchases reflect use of something other than own knowledge to determine WTP.

With that said, what do consumers know? They may have degrees of experience with a particular producer and thus have a sense if its style, but for most transactions, the information available about what is in the bottle is often nothing more than one or more expert ratings.

Given this challenge to the consumer, perhaps instead of the fine wine transaction being perceived simply as an exchange of goods for money, it either is or should be treated as an example of service-dominant marketing (e.g., Vargo and Lusch 2004): an ongoing relationship between seller and buyer, “collaborating with and learning from consumers and being adaptive to their individual and dynamic needs (p. 6).” At least at the level of the retail transaction, for example, one can certainly see a role for this among novice wine buyers who can benefit from an ongoing relationship with a supplier who both educates the buyer and learns both her particular preferences and her willingness to experiment. Perhaps over time the knowledge exchange from the relationship reduces considerably the buyer’s perceived risk of the purchase. However, the degree of collaboration needed here could be considerable and would be subject to risks and constraints of its own: the departure of the seller from the market, continued compatibility of seller and buyer, frequent enough transactions that the knowledge exchange remains fresh, the seller’s own ignorance of what is in any given bottle either at the time of purchase or over time, the seller’s potential conflict of interest in selling her own product.

The service-dominant model also looks toward something closer to the ultimate goal of a purchase: for example, the consumer wants not a furnace but toasty comfort during cold weather, and sellers offer the ongoing provision of that environment (p. 13). In the case of fine wine, does one want to drink a glass of red Bordeaux, or does one want to be transported in one’s mind to the vineyards along the Gironde in early October? It is difficult to fit fine wine into that model where the seller targets the buyer’s near-ultimate goal. Also, if the difficulty were simply the buyer learning what the seller knows, the service-dominant model would fit better, but the issue of knowledge—or ignorance—of fine wine is often shared on both sides of the transaction, and neither knows more than what the experts say.

WINE RATINGS AS KNOWLEDGE

Are such ratings a good proxy for own knowledge and, in effect, vicarious experience that drives the consumer’s WTP? That seems unlikely for a number of reasons.

First, all prominent experts who produce ratings insist that one must consider the non-quantifiable tasting notes (TNs) to begin to understand the expert opinion. For example, Parker writes: “Scores, however, do not reveal the important facts about a wine...The written commentary that accompanies the ratings is a better source of information regarding the wine's style and personality, its relative quality vis-à-vis its peers, and its value and aging potential than...
any score could ever indicate.” (Parker 2013). The lesson is: The TN increases one’s knowledge more than the rating. But quantifying the relationship between TN and WTP is unlikely.

The way to capture the knowledge content of a TN to identify a measurable effect upon wine knowledge is not obvious, but a reasonable approach would seem to be testing whether subjects can match wines with TNs: does this TN let one know enough about a wine that s/he knows it when s/he tastes it? The research design for such testing is challenging—for example, choosing wines that are different enough that matching a wine with a TN requires only an average level of sensory discrimination but similar enough that the matches are not obvious (e.g., not simply matching the red wine with the red TN and white wine with the white TN).

Storchmann (2012) reviews some of this research. As noted, the research design challenge is significant here, but the upshot from published research is that, among non-experts, the ability to identify wines after reading TNs is random (p. 25): the evidence is that they add nothing to our ability to know a wine. He cites the limited evidence for experts using TNs, and their success rate is only slightly better (Lawless 1984). It is important that the research on this subject is not extensive—it is not well established that TNs do not aid identification, perhaps because careful testing is likely complex and expensive—but the existing evidence is not encouraging. Also related to this is (a) the evidence of the notorious difficulty of identifying wines in blind tastings (e.g., Smith 2007c: 69-70; Robinson 2013)—experts are rarely able to identify correctly wines they have tasted before—and (b) Hodgson’s extensive research on the widespread inconsistency of judges in wine competitions (e.g., Hodgson 2009), leading to headlines such as “Wine Tasting: It’s Junk Science” (Derbyshire 2013).

One of the most vivid illustrations of the problem comes from George Taber’s reporting on the celebrated “judgment of Paris” in 1976 (Taber 2005) when the French judges famously got it wrong. In recounting Taber’s reporting, New York Times wine writer Eric Asimov captured the idea well when he wrote (2005): “‘Ah, back to France,’ one judge famously sighed after tasting a Napa Valley chardonnay, while another, sniffing a Batard-Montrachet, declared: ‘That is definitely California. It has no nose.’”

Not surprisingly, skepticism about the reliability of ratings and related TNs does not sit well with the experts. One of the central themes of Ashenfelter’s useful wine economics newsletter Liquid Assets was skepticism about expert ratings of wines and vintages (especially when new) and the implied existence of a “magic tongue” to which skepticism Robert Parker replied “That’s bullxxxx!” (Ashenfelter 1993 (December: cover page)). Ironically, a prominent econometrician used that same term to describe the content of TNs (Quandt 2007).

THE VARIETY OF RATINGS

Second, the variety of available wine ratings begs the question of their capacity to capture experience. Unlike the measurement of temperature or a characteristic like bottle format (regular, magnum, etc.), the measurement of a wine’s quality is not well defined. Consider Table 1 which provides some basic information about some of the most prominent sources of wine evaluation. For example, included are popular sources of ratings of wines from the global market for many prominent US “flash sale” retail wine websites (e.g., Wines Til Sold Out (WTSO), Cinderella Wine) as well as two well-known country-oriented (France, Italy) printed wine buying guides. Table 1 includes the three most widely circulated publications cited by Storchmann (2012: 22)—the others he cites have much smaller circulation (<25,000) and, for
two of the three, are California-centric by comparison—as well as some of the most popular sources that depend much more upon online access.

They have in common that they provide a combination of numeric ratings, which allow various logical analysis (e.g., ranking, quantitative quality comparison), and assorted TNs.

Beyond that, however, they vary widely and thereby raise many issues:

- They use different levels of discrimination, ranging from a minimum of 3 quality levels to a “maximum” of 51 (plus, for all, a default category of wines that were either missing or too poor to evaluate further—a blending of reasons that is unfortunate in itself), leading one to wonder whether this is an accurate reflection of differences in ability to judge wine quality;
- The maximum number of quality levels is actually greater than indicated through the use of ranges (e.g., 89-91), plusses and minuses, fractions, and so forth.
- The scores are also subject to revision of any size in either direction (within the limits of the rating system), a weakness emphasized by Ashenfelter in Liquid Assets, (e.g., September 1990).
- The source of the evaluation varies widely: some depend upon an expert individual, some depend upon a nominally stable “committee” of named experts, some depend upon an nominally stable anonymous committee, some depend upon an anonymous committee that may vary from location to location and vintage to vintage (perhaps even day to day—details are not available). Unless committees are either stable or large enough to represent the population reliably (they are not), then consistency in any committee-based evaluation is unlikely.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Sources of Wine Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratings Source</td>
<td>Access</td>
</tr>
<tr>
<td>Cellar Tracker</td>
<td>O</td>
</tr>
<tr>
<td>Decanter</td>
<td>B</td>
</tr>
<tr>
<td>Gambero Rosso</td>
<td>B?</td>
</tr>
<tr>
<td>Guide Hachette</td>
<td>P</td>
</tr>
<tr>
<td>International Wine Cellar</td>
<td>O</td>
</tr>
<tr>
<td>Jancis Robinson</td>
<td>O</td>
</tr>
<tr>
<td>Wine Advocate (R Parker)</td>
<td>B</td>
</tr>
<tr>
<td>Wine Spectator</td>
<td>B</td>
</tr>
</tbody>
</table>

The widely varying quality scales demonstrate the level of our ignorance. It seems likely that, given 100 wines of some general type (e.g., red, white, sparkling), most consumers would be able to divide them into 4-5 categories: certainly 2-3—like or dislike or indifferent—and then perhaps 1 more between indifference and each of the endpoints. If we turn to the literature on
consumers’ ability to discriminate, then the widespread use of the Likert scale suggests 5 or 7 categories and pain research suggests 11 (e.g., Farrar et al. 2001). Almost half of the sources in Table 1 limit themselves to 11 or fewer categories.

Adopting a scale of 20 to 100 or more degrees of discrimination may be good for marketing—it suggests a remarkable capacity to discriminate—but there is no evidence validating it, and a lack of replication and validation is a structural weakness for all of these sources. No one from any of these sources has ever been able to replicate the original scores assigned by tasting blind the same wines again, assigning a rating, and checking the consistency of the latter score with the former. Imagine how persuasive such replication and validation would be; its absence from the marketplace says something about the likelihood of its ever occurring. The notes from some sources such as Wine Spectator that a wine was tasted one or more additional times “with consistent notes” represents an implied acknowledgement that such replication and validation would be compelling—but they do not represent replication and validation.

The impression from this is not one of precision and consistency. It is, instead, that the vacuum—the absence of knowledge—attracts attempts both to influence consumers and supply the knowledge that consumers would value and perhaps be willing to purchase. Consumers are certainly reluctant to accept the difficulty of knowing what is in the bottle: they want TNs and ratings to be important, especially for more expensive wines, because that is all they have. As with the knowledge itself, we cannot identify some shortfall from a maximum attainable that entrepreneurs will attempt to fill, but the existence of these resources also does not indicate that anyone knows much more with them than without them.

A concern hidden in this information is exactly what lies behind the construction of the ratings produced. What is the model that experts follow, or act as if they follow, in producing a wine rating?

WINE RATINGS AS WEIGHTED AVERAGE SCORES (WAS)

All well-established wine ratings are single numbers. Considering the content of the notes that often accompany them, it would seem that a rating is actually a weighted average of the scores of a wine along a variety of dimensions. It could be used as an index number—“a statistic which assigns a single number to several individual statistics in order to quantify trends” (http://mathworld.wolfram.com/IndexNumber.html) —but that use of ratings has not emerged. To the extent that wine experts explain their methods, very few indicate explicitly that they are providing a weighted average score (WAS), but a model of wine rating would suggest that they almost necessarily are. That an expert is acting as if s/he is evaluating the significant dimensions of a wine—for example, appearance, aroma, flavor, finish, overall impression—seems almost self-evident, even though we typically see a single numeric score and we learn nothing of either the dimensions measured, the scores assigned to each, or the weights assumed.

The wide array of evaluation forms used to score wines at wine tastings provides evidence that this is at least the implied model (http://www.google.com/search?q=wine+scoring+sheet&tbm=isch&tbo=u&source=univ&sa=X&ei=RymxUYPFOeTh4AOytIHwBw&sqi=2&ved=0CCoQsAQ&biw=1173&bih=606). One of the simplest examples of this approach is the rating (“tasting”) sheet used by the Oxford University Wine Society (“Bacchus”) (http://users.ox.ac.uk/~bacchus/docs/tasting_sheet.pdf) which provides space for a wine description (facts about the wine); notes on appearance, nose, and palate; and “conclusions”. It discriminates among three most general sensory characteristics
of the wine itself and says nothing about scoring. On the other hand, the UC-Davis Wine Evaluation Chart, also used by the American Wine Society (AWS), (http://www.americanwinesociety.org/associations/10474/files/Wine%20Evaluation%20chart%202010.pdf) includes a detailed “Aroma Wheel” with 12 categories of aroma, most with sub-categories and specific aromas with each sub-category; and 5 rating attributes, each with its own allowable points with detailed description of each point value (Appearance (4), Aroma and Bouquet (7), Taste and Texture (7), Aftersave (4), and Overall Impression (3)). Since each attribute can be scored 0, the maximum possible score is 20.

Of course, such a score represents a form of WAS with implied weights and different degrees of differentiation of the attributes it recognizes: think of the possible points as weights and the impression from that attribute as falling somewhere between 0 and 1—from worst to best. According to the model of wine quality implied by the Davis/AWS chart, appearance and aftertaste are equally important; and bouquet and taste are equally important and 75 percent more important than appearance and aftertaste. One’s overall impression can add only as much as 15 percent more to one’s rating.

The relevance of a WAS model is highlighted by its use without acknowledgement by one of the foremost wine critics, Robert Parker. His model is illuminating. In his presentation of the Wine Advocate Rating System (2013), he provides the weights for his rating index: “…my scoring system gives every wine a base of 50 points. The wine's general color and appearance merit up to 5 points…The aroma and bouquet merit up to 15 points…The flavor and finish merit up to 20 points…Finally, the overall quality level or potential for further evolution and improvement—aging—merits up to 10 points.” That is, bouquet is three times as important as appearance and three-fourths of the importance of flavor and finish. Parker allows 40 percent of his points to go to everything after the bouquet and 20 percent to go to overall impression; while AWS allows 45 percent for everything after bouquet but only 10 percent to overall impression. Table 2 below summarizes the different WAS models from these two examples:

### Table 2

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Parker (% of total)</th>
<th>UC-Davis/AWS (% of total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appearance</td>
<td>5 (10%)</td>
<td>3 (15%)</td>
</tr>
<tr>
<td>Aroma and Bouquet</td>
<td>15 (30%)</td>
<td>6 (30%)</td>
</tr>
<tr>
<td>Taste and Texture</td>
<td>20 (40%) with finish</td>
<td>6 (30%)</td>
</tr>
<tr>
<td>Aftersave</td>
<td></td>
<td>3 (15%)</td>
</tr>
<tr>
<td>Overall Impression</td>
<td>10 (20%)</td>
<td>2 (10%)</td>
</tr>
<tr>
<td>Total</td>
<td>50 (100%)</td>
<td>20 (100%)</td>
</tr>
</tbody>
</table>

A strikingly different rating system is Chebnikowski’s “Winespider” (CW) which identifies 16 attributes, each of which can earn a score from 0 to 10 (Cicchetti and Cicchetti 2009). The “spider” is the visual presentation of a wine’s ratings as a 16-spoked wheel (so 15 equal sides at its perimeter) with the score of each dimension indicated by a point and the points connected by lines, giving the impression of a spider web (all 0s appears as a point at the center of the wheel, all 10s appears as a pentadecagon, and all ratings in between appear as a “thread” weaving its
way around the 16 dimensions). The visual presentation facilitates comparisons among wines since one can see the relative strengths and weaknesses of the wines by seeing where each is inside or outside another along any given dimension.

Comparing a CW rating to either a Parker or a Davis/AWS rating highlights the differences among different ratings. CW uses considerably more disaggregated dimensions, but each has equal weight, and it is not apparent how one would aggregate the 16 CW attributes into the 4 or 5 from Parker and Davis/AWS. Cicchetti and Cicchetti explain how one can collapse a CW score into a 100-point scale (p. 76), but the comparability between a score of 86 between, say, Parker and CW is unknown: they could mean very different evaluations of the wine.

If a consumer does not agree with his priorities, then Parker’s ratings will be misleading. Considering only the differences between the AWS weights and the Parker weights given in the table above, one can see how the same wine may yield different scores if for no other reason than the different weights. To take an extreme example, if a wine is all color and nose (1,1,0,0,0), Parker gives 40 percent of a perfect score. AWS gives 45 percent. If it is all nose and finish (0,1,0,0,1), Parker awards 50 percent of perfect while AWS awards only 40 percent. These differences in scores are enough to give an impression of a risky wine if one looks for “consistent” ratings before settling upon a purchase.

Thinking of ratings as a WAS, it is perhaps more unsettling that wines with the same scores may mean very different impressions to the different experts. A score of 75 percent of possible points could mean excellent color and oral experience to one and great aroma, color, and potential to the other. On the assumption that experts rate wines independently of each other—perhaps a strong assumption since it is difficult to validate—consumers who are skeptical of ratings but are unfamiliar with a wine may favor wines that have two or more favorable ratings from prominent experts, thinking that they know “enough” about a wine that receives similar ratings from multiple experts (where a greater number of similar ratings is taken to mean more reliable knowledge). Wine advertising regularly responds to this instinct by noting that a wine is a “double 90-point” wine or a “triple-90s” wine. In the same vein, auction catalogues regularly provide a rating and tasting note if it is complimentary and from a prominent expert; subject to space limitations, they add additional ratings as long as they reinforce the positive image.

Rarely does a vendor attempt to provide a comprehensive list of ratings, usually opting for the most favorable ones available from popular experts. One vendor that seems to provide relatively full disclosure is the flash sale site “Cinderella Wine” (http://cinderellawine.com/), sponsored by the Wine Library (Springfield NJ USA). It is unusually thorough in providing a link to consumer comments and ratings on “Cellar Tracker” (http://www.cellartracker.com/default.asp) as well as numerous ratings of wines offered, although lower scores, if any, appear at the end of the advertisement and without the accompanying TNs. It also provides a complete list of past offers (beginning October 2009). Among major online retailers, this would seem to be one whose marketing is relatively service dominant.

The fundamental difficulty, however, is that comparing ratings among experts who use different scales is not straightforward. How to compare a two-star rating (out of three possible) from Guide Hachette with a Parker score of 88 points is not clear. More subtle is the likely inconsistency of comparing a Parker 88 with a Stephen Tanzer 88 when, relative to Parker, we know very little about Tanzer’s underlying model of rating—or, as noted above, a CW 88 where the rating system, while stated explicitly, is so different.
We must bear in mind that most experts do not disclose such details—either the attributes in their weighting scheme, the weights assigned to each, or the quality level assigned to each attribute to arrive at the single rating. Perhaps one could extract an implied model by analysis of an expert’s ratings; but that would require identifying the attributes considered significant by the expert (perhaps by an analysis of accompanying TNs if available) as well as the relative importance of each (implied weights) and the scores assigned to each attribute. This would likely require considerable modeling.

Thinking of a wine rating as a WAS suggests a number of concerns. We shall note at the outset two assumptions. First, one’s willingness to assign a numeric rating to a wine suggests that one would also approve of assigning a numeric rating to any sensory component of a wine if one could agree that a given dimension is legitimate (e.g., bouquet). Second, a wine rating is intended as a sensory evaluation of the wine. Its other characteristics such as grape content, alcohol level, age, producer, and packaging format are objectively observable, at least potentially.

If a rating is a WAS across a number of attributes, how do we know the sensory attributes included in the rating and, more fundamentally, the extent of disaggregation of sensory attributes? Do we all know what bouquet means? Does bouquet have an initial impression? A length? A finish (like taste)? Can one distinguish as many degrees of finish as one can distinguish flavor? Is one a 3-point scale and the other a 10-point scale?

Among those who reveal something of their model of tasting, some provide more details than others. Parker identifies the attributes he judges and the points allocated to each. AWS indicates not only the attributes but also the number of points allocated to each and the meaning of each point.

Do the weights and the attributes differ by wine or “peer group” since some are known to be more aromatic than others or have better “prospects” (more ageworthy) than others? Does a Beaujolais Nouveau deserve a less-than-perfect score from Parker on “potential” because it is unlikely to improve? What is the “perfect” color of a red Burgundy? If one observes the varieties of behavior among wine consumers, it seems clear that we care about different dimensions—some look for a long time, some smell for a long time, some savor for a long time, and so forth. How does a rating tell a consumer what s/he wants to know?

**IS “VALUE” AN ATTRIBUTE?**

A particular concern is the relevance of price to ratings. If one is “grading” wines, then it seems reasonable to include “value” as one of its attributes: give a higher rating to wines that are better values in the eyes of the expert. In the context of experience goods, it might seem sensible to allow for the idea that the consumer also knows that a wine is a good value.

None of the experts listed in Table 1 state explicitly that they consider price in determining ratings, and some claim that they do not (e.g., Parker, Wine Spectator). It seems likely, however, that some do. For example, aside from the default category of wines not listed, the two largely country-specific buying guides Guide Hachette (GH) for France and Gambero Rosso (GR) for Italy have essentially four or three categories of quality respectively, and they assign each wine to one of a few price categories (e.g., in its 2014 edition, GH has, in euros, <5, 5-8, 8-11, 11-15, 15-20, 20-30, 30-50, 50-75, 75-100, + de 100 (Rosa 2013: 8); GR uses similar categories). It seems unlikely that, among red Bordeaux with a GH review, three-star wines are interchangeable whether they are in the “<5 euro” category or the “+ de 100 euro” category.
However, GR clouds the issue by highlighting in red the price category of a wine that is a particularly good value.

But why does attention to value raise a concern? Consistent with the economic concept of maximizing benefit and, for every transaction, maximizing benefit per dollar spent, the idea of ratings—notwithstanding the problems of interpretation discussed earlier—is that they should vary directly with benefit (at least the benefit to the rater) and, by extension, one therefore should maximize “points per dollar” with every wine purchase. Indeed, more than most, many wine consumers seem to have taken the benefit-per-dollar concept seriously by focusing upon the idea of a “quality-to-price ratio”, or QPR, which apparently is a widely followed measure of value followed by wine consumers relative to all others because of the widespread availability of wine ratings and prices (http://en.wikipedia.org/wiki/Glossary_of_wine_terms). The value of the wine is, in effect, points per dollar.

It is also relevant that wine ratings are most meaningful relative to other wines that consumers would consider close substitutes: they are shopping for a type of wine, and one potential benefit of ratings is to help the consumer know relative quality among the competing wines. For example, one is unlikely to treat Champagne and a fine Australian Shiraz interchangeably, and one would expect to find differences in the list of attributes judged for each (e.g., the quality of effervescence, depth of color). In this sense, if the consumer is considering ratings, then Champagne scores will be largely irrelevant to the consumer’s consideration of the Shiraz scores she finds: again, what matters are the scores of the consumer’s perceived close substitutes.

With this said, and assuming that ratings are true and comparable indicators of relative quality among close substitutes—a strong assumption, we know by now—the consumer’s strategy seems straightforward: among the close substitutes (call them “peers”), choose the wine with the highest QPR, or points per dollar.

But what if value is already considered as an attribute in the rating—if the rater has somehow allowed for something directly related to her judgment of QPR in assigning the rating? We cannot know how one might allow for price—those who seem to consider it (e.g., GH, GR) do not explain how—but a simple, and oversimplified, assumption illustrates the problem. If the score is an estimate of QPR—pure quality rating (q) divided by estimated price (p) (so q/p)—but that is not known by the consumer, then the consumer’s own naïve calculation of quality divided by price ((q/p)/p) yields, in effect, quality divided by price squared (q/p²), and the consumer has a deflated estimate of the wine’s true QPR. Moreover, the distortion grows exponentially with price. The distortion for a $10 wine is that the QPR is 0.1 of its true value (in our simple example), but it is 0.01 its true value for a $100 wine. Thus, the distortion is relatively greater for the more expensive wine, and the less expensive wine appears to be a relative bargain.

Many who have bought relatively inexpensive wines that seemed comparably rated to more expensive wines have been disappointed (e.g., the example described above of the GH 3-star rating in two extreme price categories), probably because they do not know whether and how price has been considered in the rating and they do not know the peer group.

Consider the following example. Without knowing the meaning of peer group and assuming that price is not one of its determinants, one finds from Daniel Bolomey’s useful website “BordOverview” that Parker rated both the ’08 Ch. Duhart-Milon-Rothschild and the ’08 Ch. Margaux at 94 points—two wines that one could infer come from the peer group “08 red Bordeaux”. The futures price of the Duhart was 31 euros while the price for the Margaux was 175 euros (http://www.bordoverview.com/?q=Robert-Parker). Why would anyone pay almost
six times as much for a wine that s/he can “know” is the same quality? Yet many choose Margaux over Duhart. Are they crazy?

A first alternative is that they are or, more likely, that they feel that the ratings miss a considerable amount of the quality differential between the two wines. Two additional alternative explanations are, first, that price has been considered in the rating (Parker’s stated policy notwithstanding) so that—again, oversimplified—the two wines have perhaps comparable QPRs (strictly speaking, the Margaux is almost six times as good as the Duhart but it is also almost six times as expensive), and the consumer should not make some further adjustment for price. A second alternative is that we have misstated the peer group, and, in particular, peer group definition and average price may not be independent.

Many sources of ratings say something about the role of peer groups:

- Parker: “Scores are important for the reader to gauge a professional critic’s overall qualitative placement of a wine vis-à-vis its peer group” (https://www.erobertparker.com/info/legend.asp).
- Tanzer (IWC): “Wines are scored relative to their peer group based on their expected quality during their period of peak drinkability” (http://www.wineaccess.com/expert/tanzer/ratingscale.html).
- Wine Spectator: “Our tasting coordinators organize the wines into flights by varietal, appellation or region….The tasters are told only the general type of wine (varietal and/or region) and the vintage. No information about the winery or the price of the wine is available to the tasters while they are tasting.” (http://www.winespectator.com/display/show/id/tasting-format)

Of these three rating sources, only Tanzer omits commenting upon the role of price in setting his ratings so it is likely that experts understand the sensitivity of confounding ratings with price. The other two say that price is not considered. However, they are vague about the meaning of “peer group”: if their identification of peers is straightforward, then they should be willing to identify them, but they do not.

The idea of providing ratings relative to a peer group seems reasonable and appealing and convenient—like attempting to capture the relative quality of some modal group of close substitutes among which the consumer may want to choose—but the difficulties arise in the implementation. Indeed, a goal-oriented interpretation of peers, or close substitutes, which allows implicitly for price could boost sales of one or a group of wines considerably.

One can imagine defining peer group so as to distort the apparent quality of lower priced wines. For example, if the peer group is ’02 Champagne Brut Rose’, then all of the peers in that group are expensive, ranging in price per bottle at the time of this writing from $73 to hundreds of dollars using data from the Professional version of www.wine-searcher.com. Thus, if one of these has a 91 rating for $73, then it doubtful that one can say a 91-point NV Champagne Brut Rose’—from the peer group NV Champagne Brut Rose’—priced at $12, the bottom of that group’s range, is about the same quality. However, if one defines the peer group as “Champagne Brut Rose”’, then the NV bottle looks like a considerably better value, even though it could be clearly lower quality than the vintage Champagne.

In this example, one might object that wines from these two groups are unlikely to have comparable quality. However, if the rating allows for value implicitly, then such a comparison would not be unusual (cf. the red Bordeaux example above). One objects because the peer group has been specified, and the distortion is clear: that is rarely the case with the experts.
OTHER ISSUES COMPLICATING THE USE AND INTERPRETATION OF RATINGS

Aside from the questionable value of TNs and the difficulty of interpreting ratings, a third set of concerns about capturing the consumer’s experience with ratings is the share of the content of the expert’s tasting experience that can be transferred to the consumer’s experience. Think of all the advertised circumstances of the expert tasting and the circumstances of one’s actual consumption of a wine—starting with the differences in nose and tongue. Aside from the persistent risk of bottle variation, wine rating emerges from an experience influenced by a dizzying number of factors ranging from observable and measurable differences like ambient temperature, wine temperature, time of day, and glassware to presence of others, mood, and presence of other consumables. Judging that a wine is 92 percent of perfection under one set of circumstances may provide only the most general indication of its relative appeal under different circumstances, especially to a different person.

Fourth, a finding of a positive impact of ratings on price does not tell us that consumers are willing to pay more for wines that they know are better. A significant direct correlation between WTP and ratings could mean that 10 percent of purchasers are buying for investment purposes and assuming that higher ratings yield higher WTP from subsequent buyers; and the other 90 percent are buying for own consumption and essentially disregarding the ratings since they find them worthless in evaluating how much they will enjoy the wine in their circumstances. In order to disentangle the impact of ratings, we would need to know more about motives for purchase and allow for a change of purpose over the life of ownership: for example, some may buy wine for investment and then change to own consumption and vice versa.

The evidence also does not indicate sustained returns over time. Greater WTP at the time of release says nothing about the ultimate return to the rating in subsequent sales. If a rating actually captures the quality of the wine, then the return would accrue only to the initial buyer with all subsequent buyers getting only normal returns—but then again, we do not know the motives of subsequent buyers.

Complicating this is, again, revision of ratings—probably a reasonable development but another indication of the uncertainty of knowing what is in bottle.

WINE GRADES AND STUDENT GRADES

Many consider one of Parker’s more clever innovations in wine evaluation to be the adoption of a grading scheme for wines that resembled the grading schemes that most had encountered as students (McCoy 2005: 63-4, 131-3). He was incisive in realizing that US consumers—especially the relatively educated consumers that were likely to become the core of the US fine wine market—would understand wine “grades” on a 100-point scale.

Perhaps ironically, Parker’s step in that direction can be extended to highlight the often unreliable content or perhaps transitory value of wine ratings. A closer look suggests that wine grades suffer from many of the same difficulties presented by school grades—for example, multidimensional performance usually compressed into a single dimension; selection of evaluation criteria, weighting, and methods based upon the evaluator’s personal preference rather than some well-established set of standards (criteria, weights, and methods); a limited opportunity for evaluation of performance relative to the lifetime of performance of the subject being evaluated; and influences upon the grade from circumstances at the time.

For credibility, both require objectivity, transparency, and thoroughness; but neither reflects or communicates a true and enduring knowledge of what is graded. Perhaps, as with student grades, one gains a better sense of overall quality by collecting numerous evaluations.
from different reliable sources over an extended period (like a Grade Point Average from a university transcript), but this is prohibitively difficult—and one may still be surprised. Moreover, like sorting student quality, it still likely to yield only a handful of quality categories.

THE PHILOSOPHICAL PROBLEM

Our focus has been the question of designating fine wine as an experience good, arising from efforts to identify the determinants of wine prices and, in particular, our ability to represent what consumers know or could know about a fine wine in forming their willingness to pay for it by using expert ratings. Up to now, most of the focus has been upon a critique of the most popular sources of information about wine quality available to consumers aside from extensive personal tasting experience. Based upon the analysis, the case for expert opinions in their various forms providing a good proxy for a consumer’s knowledge of a wine is weak: such information, even if supplemented by personal tasting experience, still leaves most fine wine consumers surprised much of the time.

From a philosophical perspective, skepticism about the consumer’s ability to know what she is purchasing is on even firmer footing than the preceding discussion suggests. A recent book (Smith 2007a) collects a number of papers which, according to the editor, represent the first “sustained study of the relationship [between philosophy and wine].” In his Introduction (Smith 2007b), Smith wastes no time in questioning what we can know about a fine wine:

- If much of the pleasure of drinking wine is sharing it with others, how can we actually do that if taste is subjective “as we are always told”? Is sharing during tasting truly a shared experience in the sense of having something significant in common and knowing what each other is tasting?
- When we share and reflect upon a bottle of wine, we think we are reflecting upon the same thing—but what are the properties of this so-called shared experience?
- “How accurate or objective is the language we use for describing [the features, qualities, and character of the wines we talk about]?”
- “How much trust should we place in wine connoisseurs or experts” and are they able to communicate effectively with anyone other than perhaps other experts?
- Knowing a wine requires tasting it for oneself, but does our tasting reveal properties of the wine itself or, instead, our subjective responses to the wine?
- Is judgment about wine neither entirely subjective nor objective but perhaps relative—that is, accurate for “a standard or assessment, or set of preferences, “ that is not shared among experts? (pp. xii-xvi)

It is not surprising that none of the answers to these questions are either dismissed or answered definitively. The contributors disagree fundamentally about the existence of objective knowledge of a wine beyond its technical characteristics which, while objective, will make different impressions upon different consumers, depending upon influences such as their own physical characteristics (e.g., genetic differences, “supertasters” vs. “medium tasters”), their backgrounds (e.g., amount of “training”), and the circumstances of the tasting (Goode 2007).

As Smith has stated it: “Here we have a key philosophical question: how subjective are tastes and tasting, or to put it in ontological terms: what are we tasting?” (Smith 2007b: xiii)

CONCLUSION

After considering the definition, one must question whether fine wine is an experience good—not that it overstates the complexity of the consumer’s problem but that it significantly understates it. For example, Nelson’s original examples like home appliances are considerably
easier to know than these wines. If one has no experience with a refrigerator, then one may have a difficult time arriving at a WTP for a given refrigerator. Once one buys a refrigerator, one understands its basic function and most of its competing brands. Very few competing products represent a significant challenge to understanding all refrigerators—and similarly for the kinds of goods Nelson describes. Because of the flood of new consumer technology in recent years, many of us who experienced much of this only as adults can identify with this difficulty: If one has never used a “smart phone”, how does one gauge one’s WTP for it? The idea that one can know a type of wine by buying one of its representatives in the same sense that one can know a type of refrigerator by buying one of its type seems misplaced. At least as important in considering the determination of WTP is the consideration that comprehensive knowledge of substitutes among refrigerators is at least feasible relative to the prospect of gaining comprehensive knowledge of all the choices available as substitutes for a type of wine.

We have considered wine’s status as an experience good primarily through the prospect of the consumer’s vicarious familiarity with a wine through access to others’ experiences as represented by expert ratings and both expert and amateur tasting notes. They may be the best proxy we have for consumer knowledge of a wine; but, as knowledge, they suffer from a variety of flaws. Our review may provide some support for the apparently weak ability of such data to explain wine prices. Philosophy is not always considered a behavioral science, but the questions philosophers raise about our ability to know a wine contribute further to our understanding of the consumer’s challenge in knowing a wine and what it is worth.

The case for fine wine as an experience good is weak. A product that might be a closer match to the good the fine wine consumer purchases is an informed bet on a horse race. By doing one’s research, one can increase one’s likelihood of choosing a winner—or at least a show—but the purchase inevitably includes a large element of chance. It seems unlikely that gamblers and wine aficionados are cut from the same cloth, but they are not unrelated—and those who are sufficiently risk averse tend to avoid purchasing fine wine. We cannot know the extent to which this shrinks the potential market, but we expect that many acknowledged wine buyers are “overwhelmed” and that many more potential buyers usually stay away.

Considering only one’s own consumption, what then is fine wine’s attraction? Part of the appeal is the prospect of winning and the potential returns to careful research. Also important is the sense that, when one finds a “winner”, it is not only a reward for risk taking and an attractive return on investment but also a gift from the winemaker that yields considerable satisfaction.

The philosophical problem presents a challenge to progressing with this. Two promising areas are, first, further exploration of information sources that contribute to the consumer’s ability to recognize a wine. The prospect of learning more, perhaps through neuroscience, about how we register and remember experiences like wine tasting is also attractive.

Bibliography


http://www.jancisrobinson.com/articles/a201303062.html


(accessed 21 April 2014).

Press.


Marketing 68 (January): 1-17.

April): http://wineeconomist.com/2008/04/05/what-are-wine-enthusiasts-looking-for/ (accessed
21 April 2014).

Other websites accessed:
http://users.ox.ac.uk/~bacchus/docs/tasting_sheet.pdf
http://winemag.com/PDFs/2011MediaKit/
http://www.bordoverview.com/?q=Robert-Parker