

Fine Water: A Blind Taste Test*

Kevin W. Capehart^a and Elena C. Berg^b

Abstract

To test whether consumers can distinguish among different bottled waters and, if so, whether they prefer some to others, we recruited more than 100 subjects to participate in a blind taste test that consisted of four brands of bottled water featured in a restaurant's water menu and a guidebook to fine waters. The tasting involved three successive experiments. First, our subjects tried to distinguish bottled waters in a sensory discrimination test. They were only slightly better than random chance at doing so. Next, they rated bottled waters and tap water on a 14-point scale used at an international water competition. Some subjects preferred the inexpensive tap water to any of the bottled waters, and there was no association or a weak negative association between a bottled water's price and its rating. Finally, our subjects tried to distinguish tap from bottled water while matching the bottled waters to expert descriptions. They were no better than random chance at doing either of those things. Similar results have been found in previous taste tests of beer and wine. Overall, our results suggest consumers do not have strong preferences over different bottled waters to the extent they can even tell a difference. (JEL Classifications: D12, Q25)

Keywords: bottled water, blind tasting, tap water.

I. Introduction

Not all wines are the same. Some undoubtedly taste, look, and smell different. Yet blind taste tests suggest that novice and even expert wine tasters have trouble distinguishing different wines (Ashton, 2017; Storchmann, 2012). Similar results have been found for different beers (Almenberg, Dreber, and Goldstein, 2014). Given those results for other beverages, it seems doubtful that consumers can tell the difference between different bottled waters, which even a "water connoisseur" admits have subtler taste differences than wine, no visual differences among still waters, and no

*The authors thank Karl Storchmann, Jill Hindenach, an anonymous reviewer, and the human subjects research committees that approved our research. We declare that we have no relevant or material financial interests related to the research described in this paper.

^aDepartment of Economics, California State University, Fresno, 5245 N Backer Ave M/S PB20, Fresno, CA 93740; e-mail: kcapehart@csufresno.edu.

^bDepartment of Computer Science, Math, and Environmental Science, The American University of Paris, 5 Boulevard de la Tour-Maubourg, 75007 Paris, France; e-mail: eberg@aup.edu.

odor differences (Mascha, 2006, pp. 31–32). That suspicion is supported by a study in this journal (Capehart, 2015), which used a hedonic pricing approach to show that objective characteristics of the water inside a bottled water—such as the amount of carbonation and various minerals—explained little if any of the price differences among a wide variety of bottled waters. A possible explanation for that finding is that consumers either cannot tell the difference between the water inside different bottled waters or they are indifferent between any discernable differences.

The claim that not all bottled waters are the same has nevertheless been made as the industry continues to expand (Elmhirst, 2016; Spar and Bebenek, 2008) and as bottled waters are increasingly treated with the same sophistication—or perhaps snobbery—as wine. There are now guidebooks to bottled water, akin to wine guides, such as Michael Mascha's (2006) *Fine Waters: A Connoisseur's Guide to the World's Most Distinctive Bottled Waters*. Some restaurants have “water menus” akin to wine menus. For example, Ray's and Stark Bar in Los Angeles, California, has a 45-page water menu with 20 different brands of bottled water selected by the “water sommelier” Martin Riese, who hosts informal water tastings at the restaurant for paying guests (Verive, 2015) and the media (Zagat, 2015). There are also formal water competitions akin to wine competitions. The Berkeley Springs International Water Tasting held annually in the town of Berkeley Springs, West Virginia, crowns the world's best bottled water based on a blind tasting (Fulcher, 2017).

The finding in Capehart (2015) is only suggestive, so to test whether consumers can tell the difference between bottled waters and, if so, whether they prefer some to others, we recruited more than 100 subjects to participate in a blind taste test. Prior to the blind tasting, we gave our subjects a brief training similar to an informal tasting organized by Riese at his restaurant. The brands of bottled water our subjects tasted for their training, as well as the ones they tasted blind, were from that restaurant's water menu and Mascha's (2006) guidebook. The blind tasting involved three successive experiments modeled on informal tastings by Riese, the formal Berkeley Springs tasting, and previous taste tests of water and other beverages. The first experiment was a sensory discrimination test with different bottled waters. In the second test, subjects rated those bottled waters, as well as tap water for comparison, by using the same 14-point scale used at the Berkeley Springs tasting. For the final test, subjects tried to distinguish tap from bottled water while matching expert descriptions to the bottled waters.

We found our trained subjects were better than random chance at discriminating between the bottled waters, but only slightly better; on average they distinguished the bottled waters less than 50% of the time. Our subjects were also no better than chance at either distinguishing tap water from bottled water or matching the expert descriptions to the bottled waters. When asked to rate the waters, the average ratings given to the most and least preferred bottled waters differed by less than 1 point on the 14-point scale with no association or a weak negative association between a bottled water's rating and its price. Some subjects also preferred the tap water to any of the bottled waters, which were orders of magnitude more expensive. Those results are similar to previous findings from taste tests of beer,

wine, and other beverages. Overall, our results suggest consumers do not have strong preferences over bottled waters to the extent they can even tell a difference.

The design and results of our three experiments are discussed in more detail after discussing the subjects we recruited, the training we gave them, the waters we selected for them to taste, and other general procedures that we followed.

II. General Method

A. Subjects

The subjects who participated in water tasting were all undergraduate students at The American University of Paris, a small international university in Paris, France. They were drawn from three general education courses taught by the authors of this article (specifically, an environmental science course in which the water tasting was a lab for their class, as well as principles of micro- and macro-economics courses for which they received extra credit if they participated). They did not receive monetary compensation, they were warned of the risks of drinking excessive amounts of water, and they all signed consent forms, as required by the human subjects research committees that approved our research.

The students' responses to preliminary demographic questions are summarized in [Table 1](#). Our subjects are not necessarily representative of larger populations, but as shown in that table, they were diverse in terms of their nationalities, which is consistent with the university overall. For an open-response question about their nationality, 39% identified at least one of their nationalities as the United States, 16% said the same about France, 4% identified as both French and American, and the rest (49%) identified one or more other nationalities that ranged from Albanian to Zimbabwean. Chinese were the third-largest group at 5%. The participants were young and mostly female, which is again consistent with the university overall. We did not ask about socioeconomic status, but it can be assumed that most are of high status because the university's tuition is high (about €30,000 per academic year) and financial aid is limited.

Water is essential to life, so our subjects must consume water on a regular basis. They should nevertheless be considered novice rather than expert water tasters. Indeed, as shown in [Table 1](#), less than 20% of our subjects thought their ability to distinguish different bottled waters would be above average. They were more confident about their ability to distinguish different wines, and even more confident about their ability to distinguish tap from bottled water. Those self-reported abilities are positively correlated, which will be shown as part of [Table 5](#).

B. Training

Given that our subjects should be considered novice water tasters, we gave them a brief training in water tasting before our blind tasting. The goal of the training

Table 1
Participant Information

| <i>Number of Participants</i> | <i>118</i> |
|--|------------|
| Age in years | |
| Average | 21 |
| Standard deviation | 2 |
| Gender | |
| Male | 34% |
| Female | 66% |
| Nationality | |
| United States | 39% |
| France | 16% |
| Other | 49% |
| Self-reported ability to distinguish wines | |
| Very good | 4% |
| Good | 22% |
| Average | 47% |
| Bad | 14% |
| Very bad | 11% |
| Self-reported ability to distinguish tap and bottled water | |
| Very good | 10% |
| Good | 36% |
| Average | 38% |
| Bad | 12% |
| Very bad | 3% |
| Self-reported ability to distinguish bottled waters | |
| Very good | 2% |
| Good | 16% |
| Average | 42% |
| Bad | 26% |
| Very bad | 10% |

Note: This table shows preliminary demographic information about the subjects who participated in our study.

was to dispel any pre-existing belief among participants that all waters taste the same while attuning their palates to the tastes of water. The Berkeley Springs tasting also trains its judges, some of whom are novice tasters (Fulcher, 2017; Lewis-Kraus, 2006). Our training was based on a YouTube video of a non-blind tasting the water sommelier Riese organized for a *Zagat* reporter (Zagat, 2015). In the video, Riese serves the reporter three brands of bottled water with different levels of carbonation and different amounts of “Total Dissolved Solids” (TDS), which include “carbonates, bicarbonates, chlorides, sulfates, phosphates, nitrates, calcium, magnesium, sodium, potassium, iron, manganese, and a few other minerals” (Mascha, 2006, pp. 35–37). He starts by serving the still version of the VOSS brand of bottled water, which has a relatively low TDS level of 44 milligrams per liter (mg/L). The reporter says the water just tastes like water. Next, Riese serves the still version of the Iskilde brand with a higher TDS level of 426 mg/L. The reporter reacts by saying that, to her surprise, she can actually taste the difference. The third water

he serves is the sparkling Vichy Catalan brand with an extremely high TDS level of 3,052 mg/L. She reacts strongly, saying it tastes like Alka-Seltzer. Of note, all three of those bottled waters are included on the water menu at Ray's and Stark Bar, and those waters cover almost the entire range of TDS levels on that menu. Among the waters on the menu, VOSS has the second lowest TDS level, Iskilde has the highest TDS level among the still waters, and Vichy Catalan has the highest TDS level.

The participants in our study were shown that video and served the same three brands of bottled water in a non-blind fashion as they followed along. Anecdotally, their reactions were similar to the reporter's. Some subjects said they could not taste a difference between the VOSS and Iskilde brands, but many said they could, and everyone said the Vichy Catalan brand was noticeably different. The reporter's reactions could have influenced our subjects', but it seems safe to conclude that waters with high enough amounts of carbonation and dissolved minerals can be distinguished from waters with low amounts. Whether subtler differences can be discerned is a question addressed by our study.

Immediately after the non-blind training, a blind taste test was conducted. None of the waters used in the training were used in the blind tasting, and participants were told that beforehand. The brands of bottled water used in the blind tasting were masked from participants throughout and only unmasked afterwards.

C. Waters for Blind Tasting

There are thousands of brands of bottled water (Mascha, 2006, p. 20), but only a few can be served during any tasting. We selected four based on the following criteria. First, to ensure a selection of distinct, high-quality waters, we only considered brands included in both Ray's and Stark Bar's water menu and Mascha's (2006) guide to fine water. Next, we considered only still waters because brands of sparkling water can vary in their level of carbonation in a way that could perhaps be detected visually rather than by taste. After excluding the still waters used in our non-blind training, we picked waters that varied widely in their TDS levels. The water sommelier at Ray's and Stark Bar (see, e.g., Zagat, 2015) and Mascha (2006, p. 35) both emphasize that a water's TDS is the most important dimension to its taste. Finally, we chose waters available at *La Grande Épicerie de Paris*, an upscale grocery store in Paris known for bottled water selection.

Based on those criteria, we selected the Speyside Glenlivet, Acqua Panna, Fiji, and Hildon brands, which vary in their TDS levels from 58 to 312 mg/L. Additional information about those bottled waters—including their sodium, calcium, magnesium, and hardness, where a water's "hardness" is a weighted sum of the amount of calcium and magnesium in the water (see, e.g., Mascha, 2006, p. 38)—is given in Table 2. That information is taken from the water menu at Ray's and Stark Bar.

Table 2
Water Information

| <i>Water</i> | <i>Origin</i> | <i>TDS</i> | <i>Sodium</i> | <i>Calcium</i> | <i>Magnesium</i> | <i>Hardness</i> |
|--------------------|--------------------|------------|---------------|----------------|------------------|-----------------|
| Speyside Glenlivet | Speyside, Scotland | 58 | 4 | 12 | 2 | 38 |
| Acqua Panna | Tuscany, Italy | 188 | 7 | 32 | 7 | 107 |
| Fiji | Viti Levu, Fiji | 224 | 18 | 18 | 15 | 107 |
| Hildon | Hampshire, England | 312 | 8 | 97 | 2 | 249 |
| Tap water | Paris, France | 420 | 10 | 90 | 6 | 250 |

Note: This table shows information about the waters included in our blind taste test. The information on the bottled waters is taken from Ray's and Stark Bar (2015), while the information on the tap water is taken from the public water supply company *Eau de Paris*. The mineral contents are all measured in milligrams per liter.

Tap water was also used for comparison in some of the experiments discussed later. As with bottled water, there are numerous tap waters. We used the tap water that was the easiest to obtain: the water that could be drawn from a sink in the building in which we conducted our experiments. Additional information about the tap water is given in Table 2; that information is from the public water supply company for the city of Paris.

D. Other General Procedures

The blind tasting, as well as the non-blind training discussed earlier, was conducted in a science teaching lab. Due to the small size of the lab, experiments were run several times in the Spring and Fall 2016 semesters with groups ranging from five to 20. It was not feasible to isolate subjects in that lab, so we did the following to limit their influence on one another. Before subjects arrived, we set up all the experiments. Each subject's setup looked identical, but the waters inside each glass were randomized with the aid of a computer program. We informed everyone of that randomization, emphasizing that different waters would be in different glasses for each subject. Therefore, any opinions about a glass of water expressed by one person should not influence the opinions of others. We kept track of the setups by assigning an identification number to each one. The large number of random setups meant that we—the researchers—were unable to remember during the tasting how the waters were set up. The test was therefore “double blind” in the sense that neither the researchers nor subjects knew which waters were which.

Between the non-blind training and blind tasting, each subject was given a total of 20 glasses of water (specifically, 3 glasses for the training, 12 for the first experiment discussed later, and 5 for the other experiments). The Berkeley Springs tasting asks its judges—which again include trained but still novice tasters—to evaluate 20 glasses of water at the same time and to do so four times in one day (Fulcher, 2017; Lewis-Kraus, 2006, p. 26). Thus, the number of glasses we served our subjects is well within the range of an official tasting event. Neutral water crackers were available as palate cleaners. The Berkeley Springs tasting also makes water crackers available to its tasters (Lewis-Kraus, 2006, p. 26).

According to Mascha (2006, p. 56), the recommended temperature for serving still water is 12°C. Maintaining that cold temperature for all the waters throughout the experiment was infeasible, so we served all of the waters at room temperature. The Berkeley Springs tasting also serves still waters at room temperature (Fulcher, 2017).

III. Study 1: Sensory Discrimination Experiment

A. Method

For the first experiment, we followed similar studies of wines (Ashton, 2014; Weil, 2001, 2005, 2007), beers (Almenberg, Dreber, and Goldstein, 2014), and waters (Dietrich and Gallagher, 2013; Gallagher and Dietrich, 2010) by using so-called “triangle tests” to assess whether participants could tell the difference among bottled waters. In a triangle test, subjects were given three small glasses of water that looked identical. Two of the glasses had the same brand of bottled water in them (the “twin”). The third glass had a different brand of bottled water (the “singleton”). Participants were asked to carefully taste the waters to try to identify the singleton. We emphasized there was no deception involved; one water was different from the other two, even if the differences might be subtle. A subject has a one-in-three chance of correctly identifying the singleton even if they cannot tell a difference.

We gave each subject four sets of triangle tests laid out on a clearly marked template.¹ They were instructed to move sequentially from one triangle test to the next, trying to identify the singleton each time. Each triangle test had a different pair of bottled waters. In the video mentioned earlier (Zagat, 2015) and a similar video (Guardian Food, 2015), Riese moves from waters with lower to higher TDS levels when conducting a tasting. The triangle tests were therefore arranged by their TDS levels from lowest to highest, except for the last triangle test, which involved comparing the waters with the lowest and highest TDS levels. Although participants were not informed of it, for every triangle test except the last, the singleton was the water with the relatively higher TDS level.

In addition to asking our subjects to identify the singletons, we asked them two questions for each triangle test. First, we asked whether they were confident they correctly identified the singleton. They ranked their confidence on a five-item Likert-type scale ranging from “no confidence at all” to “completely confident.” Second, we asked which water they would prefer—the water they identified as the twin or the one they identified as the singleton—as their everyday drinking water. They expressed their preference on a five-item scale with “indifferent between them” as the middle option. We prompted them to consider their preferences in terms of an

¹For each triangle test, the glasses were lined up from left to right. The literature on triangle tests warns that any arrangement that implies an ordering might bias the results (O’Mahony, 1995), but we did not find that the position of a singleton significantly affected its chance of being correctly identified. See our supplementary appendix.

“everyday drinking water” to make the results more comparable to the second experiment we conducted.

B. Results and Discussion

The results of our triangle tests are shown in [Table 3](#). The participants did better than random chance for each test, and statistically significantly so for every test, except the first one, which involved the two waters with the lowest TDS levels.

[Table 4](#) shows the percentage of singletons participants should correctly identify by random chance and the percentage they correctly identified. More participants correctly identified all four singletons than would be expected by chance, and fewer failed to identify any than would be expected. Also, more participants correctly identified two or three than would be expected by chance, and fewer identified only one.

Although our subjects were better than chance at identifying the singletons, the singleton was correctly identified less than half of the time for each triangle test, including the last test with the biggest TDS difference (see [Table 3](#)). Moreover, our subjects correctly identified the singleton in less than half of the triangle tests—about 1.8 out of the four—on average (see [Table 4](#)). Those results suggest our subjects were only slightly better than chance at distinguishing bottled waters.

Similar results have been found for wine and beer tastings. Using triangle tests with various wines and mostly non-expert tasters, Ashton (2014) and Weil (2001, 2005, 2007) found the singleton was usually identified no more than around half of the time. Almenberg, Dreber, and Goldstein (2014) also found similar results for three triangle tests with pale European lagers. Their subjects were not statistically significantly better than chance at distinguishing either Heineken or Czechvar beer from Stella Artois, and they actually did *worse* than would be expected by chance, although not significantly so. For the one test in which they did statistically significantly better than chance, they correctly distinguished Heineken from Czechvar beer only 48% of the time (Almenberg, Dreber, and Goldstein, 2014, pp. 4–5).

Our results can also be compared to previous triangle tests of water. Gallagher and Dietrich (2010) used three undisclosed brands of bottled water with TDS levels of 3, 31, and 524 mg/L, respectively. Out of three triangle tests they conducted with those waters at room temperature and untrained tasters, the singleton was correctly identified as frequently as 66% of the time (when comparing the 31 and 524 mg/L waters) and infrequently as 36% of the time (when comparing the 3 and 31 mg/L waters). Across all three tests, 50% of their subjects correctly identified the singleton (Dietrich and Gallagher, 2013, p. 22).

In another study, Dietrich and Gallagher (2013) took an undisclosed brand of bottled water with a TDS level of 524 mg/L and diluted it with filtered water to obtain waters with various TDS levels. Out of nine triangle tests they conducted

Table 3
Results of Triangle Tests

| Triangle Test | | | Confident Discerners (%) | | | | | |
|---------------|--------------------|--------------------|--------------------------|----------------------|-------------|-------------|------------------|---------------------------|
| # | Twin | Singleton | Correct (%) | Strongly Prefer Twin | Prefer Twin | Indifferent | Prefer Singleton | Strongly Prefer Singleton |
| 1 | Speyside Glenlivet | Acqua Panna | 38 | 8 | 23 | 33 | 33 | 3 |
| 2 | Acqua Panna | Fiji | 48*** | 6 | 45 | 29 | 18 | 2 |
| 3 | Fiji | Hildon | 48*** | 16 | 27 | 20 | 27 | 10 |
| 4 | Hildon | Speyside Glenlivet | 42** | 7 | 38 | 24 | 24 | 7 |

Note: This table reports the results of four triangle tests between different bottled waters. Participants should correctly identify the singleton one-third of the time by chance alone, so we indicate if the percentage who identified the singleton is statistically significantly greater than that based on the adjusted Chi-squared test suggested by Heymann and Lawless (1999, pp. 131–132).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4
Frequency of Identified Singletons

| <i>Count</i> | <i>Expected (%)</i> | <i>Actual (%)</i> |
|--------------|---------------------|-------------------|
| 0 | 20 | 10** |
| 1 | 40 | 32 |
| 2 | 30 | 32 |
| 3 | 10 | 21*** |
| 4 | 1 | 4** |

Note: This table shows, for all four triangle tests, the percentage of subjects who would be expected by chance alone to correctly identify a given number of singletons, the percentage who actually did, and whether the difference between those percentages is statistically significant based on a Chi-squared test. The percentages were rounded.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

with those waters at room temperature and untrained tasters, the singleton was correctly identified as frequently as 65% of the time when comparing the undiluted bottled water to a diluted water with a TDS level of 31 mg/L. The singleton was identified as infrequently as 39% of the time when comparing waters with a smaller, although not the smallest, TDS difference. Across all nine tests, 50% of their subjects correctly identified the singleton (Dietrich and Gallagher, 2013, p. 259).

It is noteworthy that Dietrich and Gallagher (2013) found a close but not monotonic relationship between the difference in TDS levels and the percentage of subjects who correctly identified the singleton. They found that result even though they were diluting the same brand of bottled water (rather than comparing different brands) and even though different subjects (rather than the same ones who might get fatigued) participated in each triangle test. The fact that subjects in our study did not perform the best on the last triangle test with the highest TDS difference is therefore not necessarily a sign of fatigue. Even if it is a sign of fatigue, that raises questions about the ability of Berkeley Springs judges to evaluate upwards of 80 glasses of water in one day.

Thus, our subjects performed about as well as those in previous triangle tests of wines, beers, and waters. Some of our subjects performed better than others, yet those observed abilities had either no correlation or a slight negative correlation with their self-reported abilities. Table 5 shows the rank correlation between each subject's self-reported ability to distinguish different wines or waters and the number of singletons they correctly identified in the triangle tests. Curiously, there is a small but statistically significantly negative correlation between their self-reported ability to distinguish bottled waters and their observed ability ($\rho_s = -0.19$, p -value = 0.05). Less confident subjects were perhaps more careful during their tasting.

As previously noted, in addition to asking subjects to try to identify the singleton in a triangle test, we asked if they were confident they correctly identified the singleton and whether they preferred the singleton they identified or the twin. Part of Table 3 shows, for each triangle test, the preferences expressed by subjects who

Table 5
Correlations between Self-Reported and Observed Tasting Abilities

| | Self-report Ability to Distinguish... | | | Observed Ability |
|--------------------------|---------------------------------------|-----------------------|--------------------------|------------------|
| | Different Wines | Tap and Bottled Water | Different Bottled Waters | |
| Different wines | 1 | 0.08 | 0.25*** | -0.09 |
| Tap and bottled water | | 1 | 0.46*** | -0.03 |
| Different bottled waters | | | 1 | -0.19** |
| Observed ability | | | | 1 |

Note: This table shows the Spearman rank-correlation coefficient for 113 subjects' self-reported ability to distinguish between different wines, tap and bottled water, and different bottled waters, as well as their observed ability to distinguish different bottled waters based on the number of singletons they correctly identified in the triangle tests. Five subjects did not respond to one or more of the preliminary questions about their tasting abilities.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

correctly identified the singleton and expressed more than no confidence in their choice. We restrict our attention to those confident discerners because preferences expressed by subjects who misidentified the singleton or randomly guessed correctly are less interesting.

Many of the confident discerners expressed little or no preference between the bottled waters they tasted. At least 20% and as many as 33% said they were indifferent between the waters in each triangle test. No more than 16% expressed a "strong" preference over any of the waters. That said, among those who expressed any preference, some waters were more popular than others. In addition to most of the confident discerners preferring Acqua Panna to Fiji, most preferred Fiji to Hildon, and most also preferred Hildon to Speyside Glenlivet in their respective triangle tests. Similar preferences were expressed in the next experiment we conducted, which involved comparing all the waters on a simultaneous, rather than pairwise, basis.

IV. Study 2: Preference Rating Experiment

A. Method

The second experiment was conducted immediately after the first one with the same participants. For that experiment, we gave each subject five glasses of water on another clearly marked template. Four of the glasses contained the four brands of bottled water used in the triangle tests. A fifth glass contained tap water. We informed our subjects that each glass had a different water, but did not inform them that one water was tap and the others were the same bottled waters from the triangle tests. As before, the glasses of water looked identical, they were laid out ahead of time and kept at the same temperature, the position of each type of water was randomized, and subjects were informed of that randomization.

The subjects were asked to rate each of those five waters on the 14-point scale used as part of the Berkeley Springs tasting. On that scale, a water should be assigned the highest rating of 14 points if you can say, “This water tastes really good. I would be very happy to have it for my everyday drinking water.” A water should be assigned the lowest rating of 1 point if it makes you say, “This water has a terrible, strong taste. I can’t stand it in my mouth.” The full scale can be found in Lewis-Kraus (2006).

We also asked participants to use their ratings to rank the waters from their most to least preferred. They were strongly encouraged to break any ties if they gave the same rating to more than one water.

Asking our subjects to evaluate five glasses of water at the same time is somewhat demanding, but again the Berkeley Springs tasting asks its judges to simultaneously evaluate 20 glasses of water. Olkin et al. (2015, p. 24) also suggest that, when it comes to wine, five is a small enough number for most people to simultaneously evaluate.

B. Results and Discussion

Table 6 shows the distribution of rankings assigned to the five waters. Specifically, it shows the percentage of subjects who ranked a water as their most preferred, their second-most preferred, and so forth. That table reflects the rankings of only 110 of our 118 subjects. We ignored the rankings by eight participants who gave the same ranking to two or more waters because those ties may reflect “lazy tasting” (to borrow a phrase from Quandt, 2006, p. 9).

The Fiji brand of bottled water was given the most first-place rankings with a quarter of participants ranking it as their most preferred. Tap water was given the most last-place rankings with 29% ranking it as their least preferred. However, a quarter of participants also said that Fiji was their least preferred, and almost 20% said that tap water was their most preferred. Thus, there is no clear consensus about which waters are preferable to others.

One way to try to find consensus is by examining the rank-sums for each water (i.e., the sum of the ranks that subjects assigned to each water) shown in Table 6. If a subject assigns a ranking lower than first place, then that can be thought of as a vote against the water, as discussed by Quandt (2006). The water that received the fewest votes against it—and was the most preferred in that sense—was Acqua Panna. The waters that received the most votes against them were Speyside Glenlivet and tap, which were the waters with the lowest and highest TDS levels, respectively.

In terms of the ratings underlying the rankings, histograms of the ratings given to each of the five waters are shown in Figure 1. The subjects made full use of the scale with ratings as low as 1 and as high as 14, but the average ratings were similar. The average (standard deviation) of the ratings assigned to Acqua Panna, Hildon, Fiji, Speyside Glenlivet, and tap water were 10.5 (2.5), 10.2 (2.6), 9.8 (3.2), 9.6 (2.8),

Table 6
Distribution of Rankings

| Rank | Rankings (%) | | | | |
|-------------------|--------------------|-------------|------|--------|-----------|
| | Speyside Glenlivet | Acqua Panna | Fiji | Hildon | Tap Water |
| 1st | 15 | 23 | 25 | 18 | 19 |
| 2nd | 17 | 27 | 17 | 24 | 15 |
| 3rd | 17 | 20 | 16 | 30 | 16 |
| 4th | 29 | 18 | 16 | 15 | 21 |
| 5th | 22 | 12 | 25 | 13 | 29 |
| Rank-sum (Number) | 359 | 296 | 327 | 309 | 359 |

Note: Our subjects ranked five different waters. This table shows the percentage of subjects who gave a water a given ranking, as well as the sum of the rankings for each water. Eight sets of rankings with one or more ties were ignored. The rank-sum for the tap water is about 1.1 times higher than the average rank-sum of the bottled waters, which is statistically significantly higher based on Quandt's (2007) rank-sum test (p -value ≈ 0.03 based on 10,000 Monte Carlo simulations).

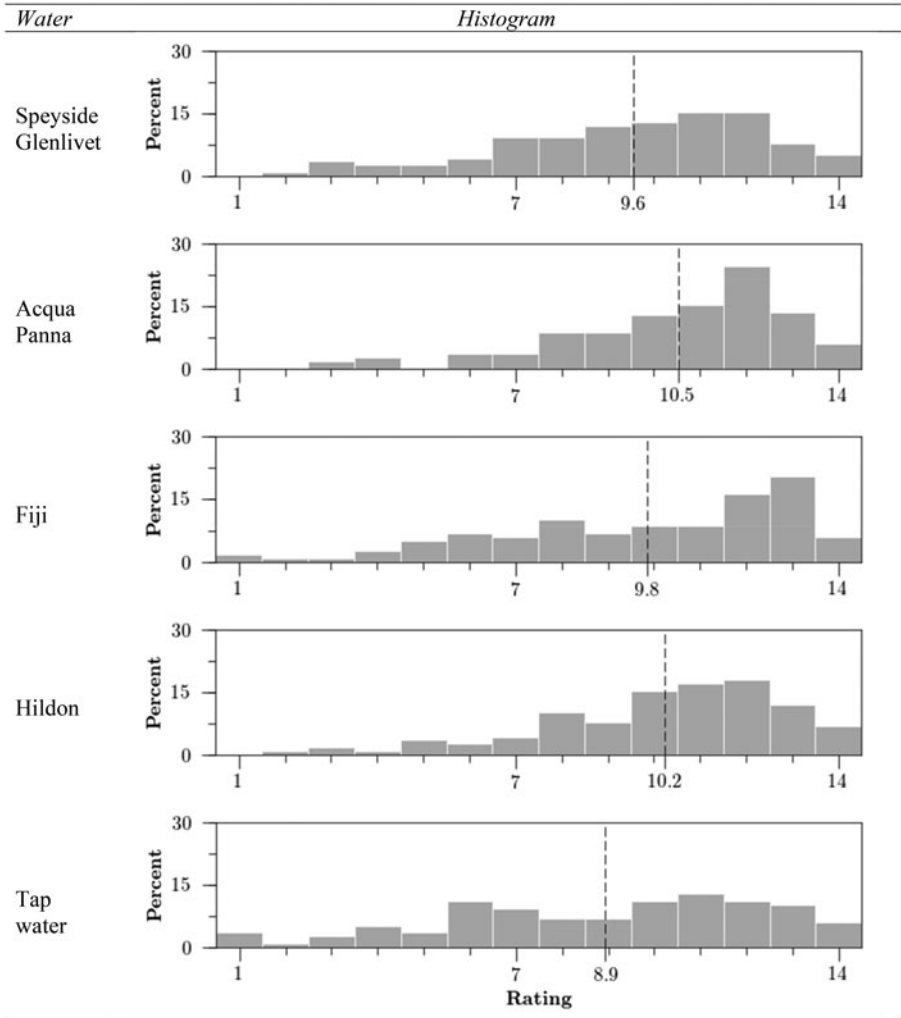
and 8.9 (3.4), respectively. Note the bottled water with the highest average rating, Acqua Panna, was rated less than 1 point higher on the 14-point scale than the bottled water with the lowest average rating, Speyside Glenlivet.

To try to understand why some waters were preferred to others, we ran the simple ordinary least squares (OLS) reported in Table 7. The dependent variable in each regression is the rating a subject gave a water on the 14-point scale. Almenberg and Dreber (2011), Ashenfelter and Jones (2013), and Goldstein et al. (2008) ran similar OLS regressions for wine ratings. Fixed effects for each subject were used in case some subjects tended to rate all the waters higher or lower on the 14-point scale. Dummy variables for the position of a water were also included to see whether the way in which the waters were arranged—from left to right in a row—affected their rating. None of the positional dummies were statistically significant (either individually or jointly) in any of our regressions, which suggests the arrangement of the waters did not bias the ratings.

For the first regression in Table 7, dummy variables for the different waters were included. The omitted dummy was tap water. We see that Acqua Panna and Hildon tended to be rated more than 1 point higher than tap water on the 14-point scale, Fiji was rated somewhat higher than tap, and Speyside Glenlivet was rated similarly to tap. We also see again that Acqua Panna tended to be rated less than 1 point higher than Speyside Glenlivet, even after controlling for subject-specific fixed effects.

The waters differ in terms of their TDS levels, so the second regression in Table 7 includes both the TDS level of a water and the square of its TDS level to capture a nonlinearity. The regression implies that, other things being equal, a water with TDS level of about 210 mg/L would be rated the highest. Waters with higher or lower TDS levels would be rated lower. That finding is consistent with Teillet et al. (2010) who suggest waters with intermediate levels of mineralization are the most preferred.

Figure 1
Distribution of Ratings
Distribution of Ratings



Note: This figure shows, for each water subjects rated on the 14-point scale, a histogram of the ratings given to those waters. The dashed line indicates the mean rating.

The third regression in Table 7 includes some of the minerals that make up a water’s TDS level. Specifically, the regression includes the water’s sodium level and hardness. The regression implies that (again, other things being equal) a water with a sodium level of about 13 mg/L would be rated the lowest. A water with a hardness of about 182 mg/L would be rated the highest.

Table 7
OLS Results for Ratings

| Dependent Variable: Rating Independent Variables | (I) | | (II) | | (III) | | (IV) | |
|---|---------|--------------|----------|--------------|----------|--------------|---------|--------------|
| | Coef. | (Std. error) | Coef. | (Std. error) | Coef. | (Std. error) | Coef. | (Std. error) |
| Brand | | | | | | | | |
| Speyside Glenlivet | 0.66 | (0.48) | | | | | | |
| Acqua Panna | 1.52*** | (0.43) | | | | | | |
| Fiji | 0.86* | (0.47) | | | | | | |
| Hildon | 1.27*** | (0.41) | | | | | | |
| Water-related variables | | | | | | | | |
| TDS | | | 0.01** | (0.00) | | | | |
| TDS squared | | | -3E-5*** | (0.00) | | | | |
| Sodium | | | | | -1.75*** | (0.61) | | |
| Sodium squared | | | | | 0.07*** | (0.02) | | |
| Hardness | | | | | 0.09*** | (0.02) | | |
| Hardness squared | | | | | -3E-4*** | (0.00) | | |
| Log of price per liter | | | | | | | 0.16*** | (0.01) |
| Glass position | | | | | | | | |
| First position | -0.00 | (0.41) | 0.11 | (0.42) | -0.00 | (0.41) | 0.05 | (0.41) |
| Second position | 0.13 | (0.40) | 0.18 | (0.40) | 0.13 | (0.40) | 0.10 | (0.39) |
| Fourth position | 0.59 | (0.45) | 0.65 | (0.45) | 0.59 | (0.45) | 0.71 | (0.44) |
| Last position | 0.11 | (0.47) | 0.22 | (0.46) | 0.11 | (0.47) | 0.18 | (0.47) |
| Subject fixed effects | Yes | | Yes | | Yes | | Yes | |
| Observations | 590 | | 590 | | 590 | | 590 | |
| Degrees of freedom | 465 | | 467 | | 465 | | 468 | |
| R-squared | 0.28 | | 0.28 | | 0.28 | | 0.27 | |

Note: This table reports the results of OLS regressions for the rating a subject gave a water on the 14-point scale. Fixed effects were used for each subject, and standard errors were clustered by subject. The number of observations corresponds to 118 subjects, all of whom rated each water. The four positional dummies are not jointly significant in the first ($F = 0.83$, p -value = 0.50), second ($F = 0.65$, p -value = 0.63), third ($F = 0.63$, p -value = 0.64), or fourth ($F = 0.91$, p -value = 0.46) regression.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The price of a water can be included in a regression like the ones in [Table 7](#), although the price of tap water will obviously be an outlier. The prices we paid at *La Grande Épicerie de Paris* for the Acqua Panna, Hildon, Speyside Glenlivet, and Fiji bottled waters were about 2.93, 3.00, 3.07, and 3.75 euros per liter, respectively. By comparison, the tap water was free to us and, although someone was ultimately paying for it, the cost of tap water in France has been estimated to be a mere 0.0038 euros per liter (Brei, 2018, p. 2).

The final regression in [Table 7](#) follows Goldstein et al. (2008) by including the log of the price of the waters in euros per liter. Goldstein et al. (2008) found that, in double-blind tastings with a variety of wines, non-experts tended to rate cheaper wines as better, even after controlling for individual fixed effects. We find that more expensive waters were rated as better, although the statistically significant effect is arguably not substantively significant. Our regression implies the most expensive bottled water we served our subjects would be rated only about 1.1 points higher on the 14-point scale than the inexpensive tap water.

Moreover, if we re-ran the last regression in [Table 7](#) with only bottled waters, then the estimated coefficient on the log of the price of a bottled water would be about -1.50 (s.e. = 1.73). That estimate would not be significantly different than zero (p -value = 0.39), but it suggests there is no correlation or perhaps even a negative correlation between the price of a bottled water and its rating.

V. Study 3: Description Matching Experiment

A. Method

In our final experiment, we gave participants five written descriptions corresponding to each of the five waters they had rated in the previous experiment. We asked them to try to match each water to its corresponding description. Note that a randomly guessing subject would have a one-in-five chance of correctly matching a given water to its description.

The description we gave participants for the tap water was simply that it was tap water. The descriptions of the bottled waters, which were taken from the water menu at Ray's and Stark Bar, are reproduced as part of [Table 8](#). Those are expert descriptions in the sense that they were curated by the water sommelier who designed that menu.

B. Results and Discussion

The results of that experiment are reported in [Table 8](#). For each water, the participants were not significantly better than chance at matching the water to its description, except for Acqua Panna, but even in that case, less than 30% of participants were able to correctly match it to its description.

Table 8
Results of Description-Matching Test

| Description | Correct (%) |
|--|-------------|
| Good water makes good whisky, and the still version of this slightly alkaline, low TDS water is the perfect match for your single-malt whisky. The spring, located in Speyside, Scotland, away from habitation, industry, and agriculture, is locally known as Slochd. Speyside Glenlivet produces more single-malt whisky than any other part of the world. Quartzite at the source keeps high levels of minerals from mixing with the water. | 17 |
| The Acqua Panna source is nestled among beech wood and chestnut forests on the slopes of Mount Gazzaro, 3,700 feet high in the pristine Apennines Mountains of Tuscany. This serene, vast, unspoiled natural reserve brings forth a luminous water with neither the slightest effervescence nor odor. The first sip has a pleasant, fresh taste due to the low acid and low mineral salt content. Not without structure, it imparts a taste that is as light as a feather, pleasantly soft, and velvety. | 28** |
| Fiji water begins as rain falling on the tropical island of Viti Levu's north coast and is drawn from an artesian aquifer that lies hundreds of feet below the edges of a primitive rainforest. As it filters down through layer after layer of ancient volcanic rock, the natural filtration process introduces the mineral silica which contributes to Fiji's distinctly soft, smooth "mouth feel" (a professional tasters' term that refers to a liquid's taste and texture). That distance and isolation is part of what makes Fiji water so pure and rich in taste. | 23 |
| The distinctive terroir of Hildon natural mineral water comes from deep beneath the chalk hills of the beautiful Hampshire, England, countryside. This chalk environment filters the water to exceptional purity and embodies it with a specific composition of minerals characteristic of the geological strata it flows through. This "fingerprint" gives this water a taste and composition which is uniquely Hildon. | 18 |
| Tap water | 24 |

Note: This table shows the percentage of subjects who correctly matched a water to its description and whether that percentage is statistically significantly greater than 20% based on a Chi-square test. The descriptions of the bottled waters were taken from Ray's and Stark Bar (2015), except brand names and geographic locations that could reveal the brands were redacted from the descriptions given to our subjects.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The fact that only 24% of participants correctly identified the tap water also means that 76% mistook a bottled water for tap water. Speyside Glenlivet, Hildon, Acqua Panna, and Fiji were mistaken for tap 28%, 27%, 24%, and 18% of the time, respectively. Those percentages are all close enough to 25% that we should perhaps not over-analyze them, but it is interesting that the two bottled waters most likely to be mistaken for tap were very different in their TDS levels yet similar in that they received lower ratings on average. Bad water is tap water, our subjects might have presumed.

As further evidence that participants were no better than chance at matching the descriptions, Table 9 shows the percentage of descriptions that participants would be expected to correctly match by random chance and the percentage they correctly matched. Those percentages are not significantly different.

Table 9
Frequency of Matched Descriptions

| <i>Count</i> | <i>Expected (%)</i> | <i>Actual (%)</i> |
|--------------|---------------------|-------------------|
| 0 | 37 | 36 |
| 1 | 38 | 35 |
| 2 | 17 | 19 |
| 3 | 8 | 8 |
| 4 | 0 | 0 |
| 5 | 1 | 2 |

Note: This table shows the percentage of subjects who would be expected by chance alone to correctly match a given number of waters to their description and the percentage who actually did. The expected percentages were calculated as in the Montmort matching problem. The differences between those percentages are neither jointly nor individually significant at the 10% level based on Chi-square tests.

The expert descriptions of the bottled waters given in [Table 8](#) are admittedly not very descriptive of their tastes. More detailed descriptions, like the ones given to wines, could perhaps be easier to match. Yet previous studies suggest that wine tasters, especially novice ones, are no better than chance at matching wines to descriptions (Storchmann, 2012, pp. 24–25). Better descriptions would not necessarily improve the performance of our subjects, whose performance is consistent with previous findings for wine.

VI. Conclusions

Despite the fact that all water is simply H₂O at the molecular level, it is true that waters are not identical. Indeed, the difference between potable and polluted water can be the difference between life and death. Waters with higher levels of carbonation or dissolved minerals can also be distinguished by taste or visual inspection from waters with lower amounts, as discussed earlier. Thus, the entire bottled water industry cannot be dismissed as selling the exact same water in different bottles.

That said, our “Judgement of Paris” for bottled water raises questions about whether consumers have strong preferences over different bottled waters or can even reliably distinguish them. We found our trained subjects were only slightly better than random chance at distinguishing bottled waters, even though we served them ones that a water sommelier selected for his restaurant’s water menu and that a water connoisseur included in his guidebook to “the world’s most distinctive bottled waters.” We also found our subjects were no better than chance at either distinguishing tap from bottled water or matching expert descriptions to the bottled waters.

Even when our subjects could tell a difference, they did not express strong preferences. When asked to rate the waters we served them on the same scale used at an international water competition, some were rated higher on average, yet the average ratings of the most and least preferred bottled waters differed by less than 1 point on the 14-point scale. Most of our subjects exhibited a stronger preference

for bottled over tap water, but the average rating for tap was less than 2 points lower than the highest rated bottled water, and about 20% of subjects rated tap higher than any of the bottled waters.

The bottled waters are orders of magnitude more expensive than tap water—even without trying to account for any negative externalities generated by the production, distribution, and disposal of bottled water—so our subjects might not be willing to pay much more for waters they liked at best only slightly more than tap. Examining willingness to pay for waters in a blind setting—like others have done for alcoholic and non-alcoholic beverages (Combris, Lange, and Issanchou, 2006, 2007; Tozer et al., 2015)—is one possible direction for future research.

Overall, however, the results of our blind water tasting support and extend the conclusions of Capehart's (2015) hedonic analysis of the price of bottled waters. Consumers seem to be largely indifferent between the water inside bottled waters, suggesting that taste cannot be a major reason why consumers purchase and pay more for some bottled waters than others or tap water. Just as there is more to a wine than the look, smell, or taste of what is inside its bottle, there must be more to bottled waters than what is inside, especially since there are no visual differences among still waters, no odor differences, and subtle or non-existent taste differences. Consumers' willingness to pay for an expensive bottled water must be rooted in other aspects besides the taste of the water inside it.

Supplementary Material

For supplementary material accompanying this paper visit <https://doi.org/10.1017/jwe.2017.50>.

References

- Almenberg, J., and Dreber, A. (2011). When does the price affect the taste? Results from a wine experiment. *Journal of Wine Economics*, 6(1), 111–121.
- Almenberg, J., Dreber, A., and Goldstein, R. (2014). Hide the label, hide the difference? American Association of Wine Economists, Working Paper No. 165, August. Available at http://www.wine-economics.org/aaawe/wp-content/uploads/2014/08/Aawe_WP165.pdf.
- Ashenfelter, O., and Jones, G. V. (2013). The demand for expert opinion: Bordeaux wine. *Journal of Wine Economics*, 8(3), 285–293.
- Ashton, R. H. (2014). 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. (2017). Dimensions of expertise in wine evaluation. *Journal of Wine Economics*, 12(1), 59–83.
- Brei, V. A. (2018). How is a bottled water market created? *Wiley Interdisciplinary Reviews: Water*, 5(1), 1–14.
- Capehart, K. W. (2015). Fine water: A hedonic pricing approach. *Journal of Wine Economics*, 10(2), 129–150.

- Combris, P., Lange, C., and Issanchou, S. (2006). Assessing the effect of information on the reservation price for champagne: What are consumers actually paying for? *Journal of Wine Economics*, 1(1), 75–88.
- Combris, P., Lange, C., and Issanchou, S. (2007). Product information, hedonic evaluation, and purchase decision: An experimental study of orange juice. *Journal of Wine Economics*, 2(1), 40–54.
- Dietrich, A. M., and Gallagher, C.D. (2013). Consumer ability to detect the taste of total dissolved solids. *Journal of the American Water Works Association*, 105(5), 255–263.
- Elmhirst, Sophie. (2016). Liquid assets: How the business of bottled water went mad. *The Guardian*, October 6. Available at <https://www.theguardian.com/business/2016/oct/06/liquid-assets-how-business-bottled-water-went-mad>.
- Fulcher, J. (2017). Judging water at Berkeley Springs international water tasting competition. *Water Environment Federation Highlights*, April 7. Available at <http://news.wef.org/judging-water-at-berkeley-springs-international-water-tasting-competition/> (accessed 7 April 2017).
- Gallagher, C. D., and Dietrich, A. M. (2010). TDS and temperature affect consumer taste preferences. *OpFlow*, 36(12), 20–22.
- Goldstein, R., Almenberg, J., Dreber, A., Emerson, J. W., Herschkowitsch, A., and Katz, J. (2008). Do more expensive wines taste better? Evidence from a large sample of blind tastings. *Journal of Wine Economics*, 3(1), 1–9.
- Guardian Food*. (2015). Testing the waters: What we learned from a “water sommelier.” you. [be/yyrH0M5buQk](https://www.youtube.com/watch?v=be/yyrH0M5buQk) (accessed 19 February 2015).
- Heymann, H., and Lawless, H. (1999). *Sensory Evaluation of Food: Principles and Practices*. New York: Springer Science.
- Lewis-Kraus, G. (2006). The water rush: The pursuit of perfect drinking water ends in Berkeley Springs, West Virginia. *The Oxford American*, 53, 22–28.
- Mascha, M. (2006). *Fine waters: A Connoisseur's Guide to the World's Most Distinctive Bottled Waters*. Philadelphia: Quirk Books.
- Olkin, I., Lou, Y., Stokes, L., and Cao, J. (2015). Analyses of wine-tasting data: A tutorial. *Journal of Wine Economics*, 10(1), 4–30.
- O'Mahony, M. (1995). Who told you the triangle test was simple? *Food Quality and Preference*, 6(4), 227–238.
- Quandt, R. (2006). Measurement and inference in wine tasting. *Journal of Wine Economics*, 1(1), 7–30.
- Quandt, R. (2007). A note on a test for the sum of ranksums. *Journal of Wine Economics*, 2(1), 98–102.
- Ray's and Stark Bar. (2015). Water menu. Available at patinagroup.com/restaurants/133/menus/Rays-and-Stark-Bar-Water-Menu.pdf (15 accessed May 2015).
- Spar, D., and Bebenek, K. (2008). Profitable springs: The rise, sources, and structure of the bottled water business. *Entreprises et Histoire*, 50(1), 100–118.
- Storchmann, K. (2012). Wine economics. *Journal of Wine Economics*, 7(1), 1–33.
- Teillet, E., Urbano, C., Cordelle, S., and Schlich, P. (2010). Consumer perception and preference of bottled and tap water. *Journal of Sensory Studies*, 25(3), 463–480.
- Tozer, P., Galinato, S., Ross, C., Miles, C., and McCluskey, J. (2015). Sensory analysis and willingness to pay for craft cider. *Journal of Wine Economics*, 10(3), 314–328.
- Verive, J. (2015). “I went to a \$50 water-tasting, and now I'm thirsty for more.” *Paste*, February 12. Available at pastemagazine.com/articles/2015/02/i-went-to-a-50-water-tasting-and-now-im-thirsty-fo.html.
- Weil, R. L. (2001). Parker v. Prial: The death of the vintage chart. *Chance*, 14(4), 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(3), 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.
- Zagat (2015). What we learned from a water sommelier. Available at zagat.com/v/los-angeles/what-we-learned-from-a-water-sommelier (accessed 16 September 2015).