Food Quality and Preference 21 (2010) 562–568

Contents lists available at ScienceDirect

Food Quality and Preference

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

ROC analysis in olive oil tasting: A Signal Detection Theory approach to tasting tasks

Concepción Paredes-Olay *, María M. Moreno-Fernández, Juan M. Rosas, Manuel M. Ramos-Álvarez **

Department of Psychology, University of Jaén, Paraje las Lagunillas s/n, 23071 Jaén, Spain

ARTICLE INFO

Article history:
Received 3 November 2009
Received in revised form 3 March 2010
Accepted 3 March 2010
Available online 7 March 2010

Keywords:
Taste
Olive oil
Signal Detection Theory
Discriminative learning

ABSTRACT

Sensory evaluation of olive oil has been conducted by using the Classical Threshold Theory perspective. This approach does not allow the separation of purely sensory processes from other factors that may affect performance such as tasters' cognitive biases. Two experiments were conducted with the goal of developing a task that, using the logic of Signal Detection Theory (SDT), would allow sensory and decision processes to be separately assessed within an olive oil tasting task, overcoming the shortcomings of the traditional Classical Threshold Theory approach. Experiment 1a used classic psychophysics theory to establish the olive oil concentrations that were used later in Experiment 1b. Experiment 1b presents a taste task based on SDT (by using signal and noise stimuli) that allowed establishing ROC curves and the separate calculation of sensory and decision indexes. Non-parametric detection and sensory indexes, and robust analyses of variance are proposed to overcome the problems associated with the use of a limited number of trials involved. The advantages of the SDT approach in olive oil tasting are discussed.

1. Introduction

Establishing the quality of certain foods requires performing a sensory analysis. Olive oil is in that case. The quality of olive oil depends not only on its physicochemical properties, but also on organoleptic properties such as smell or flavour, that are established by a taster. Accordingly, sensory analysis of food may be considered as an instance of discriminative learning in which the goal is to establish the differences between two similar foods, and to describe the features of a specific food.

Within sensory analysis the use of human senses as measurement instruments is necessary, given that they are more influenced by variations in the intensity of the stimuli being assessed than most laboratory instruments (O'Mahony, 2003). In fact, the human being is the only "machine" able to perform the sensory analysis. This analysis is conducted by using different discriminative tests: Yes–No tasks, Same–Different tasks, Triangular test, Paired test, Duo-Trio test or Forced-Choice tests (for a general description see Macmillan and Creelman (2005); for their application to a tasting task see Alba, Izquierdo, and Gutiérrez (1997)). These tasks are based on Classical Threshold Theory. This approach assumes that sensations directly depend on the intensity of the attributes of the stimulus, so that the stimulus is perceived only when its intensity is above some specific absolute level or threshold. These tasks provide a single index about the sensory-perceptual process, without taking in account possible cognitive biases on the perceiver that may affect participants' performance beyond the differences in the sensory process.

There are a number of studies that suggest that performance within these tasks may be affected by participants' characteristics. For instance, O'Mahony and Rousseau (2003) found that Forced-Choice tasks are useful when the feature that changes can be specified by participants (i.e., in psychophysical studies and studies with expert tasters). However, in situations such as consumer perception studies in which participants cannot specify the feature that leads two stimuli to be considered as different, the use of Same–Different tests is more appropriate, given that their statistical power is greater in this situation than in Forced-Choice tests (O'Mahony, 1995). Another factor that is assumed to affect performance in discriminative tests seems to be the decision rule participants use when confronting a task. For instance, the appropriate rule would be to "compare the distance between stimuli" in Triangular and Duo-Trio tasks, while the rule "draw the line" between sensations would be the proper one in a Forced-Choice task. Given that the use of the adequate rule for each type of task would determine perceiver's performance regardless of the sensory difference among the products, a common strategy is to fix the rule by instructions with the goal of avoiding variability depending on the subjective interpretation of the rule in the specific situation (O'Mahony and Hautus, 2008).

The results summarized above show the role tasters' cognitive biases may play in performance. This problem has led some authors, such as O'Mahony and Hautus (2008) to propose the use of Signal Detection Theory (SDT) and, more specifically, the Receiver Operating Characteristic curve (or ROC) to overcome the

* Corresponding author. Tel.: +34 953 213452; fax: +34 953 211881.
** Corresponding author. Tel.: +34 953 212587; fax: +34 953 211881.
E-mail addresses: cparedes@ujaen.es (C. Paredes-Olay), mramos@ujaen.es (M.M. Ramos-Álvarez).

0950-3293/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodqual.2010.03.003

shortcomings of the use of classical tasks. ROC curves allow for exploring the cognitive strategies participants use in different tests without having to fix them beforehand by instructions, providing a visual tool for examining the tradeoff between the perceiver's sensory performance and perceiver's decision biases.

SDT assumes that any relevant stimulus appears confounded with a background stimulation that is constantly activating the sensory organs. The perceiver's task would be to discriminate the Signal (the stimulus feature that participants are required to discriminate) from the Noise (any other stimulus feature to which participants should not respond). In Yes–No tasks there are two types of trials participants are exposed to, Signal + Noise trials and Noise trials. Participants should decide whether the signal is present or not in any given trial. If participants respond that the signal is present when the Signal is present their answer is categorized as a *Hit*. If participants respond that the signal is absent when the Signal is present their answer is categorized as a *Miss*; if participants respond that the signal is present in a Noise trial, their answer would be categorized as a *False alarm*; finally, if participants respond that the signal is absent in Noise trials then their answer will be labeled as a *Correct Rejection*.

According to SDT, participants will respond following their decision rule. When the amount of evidence for the signal is larger than the criterion, participants will respond that the signal is present, and will respond that the signal is absent when the amount of signal it is smaller than their criterion. SDT combines this information to postulate two different indexes, sensitivity index or d' and the decision criterion index or beta (β). Sensitivity (d') may be understood as the distance between the means of the probability distributions associated to the signal and the noise; specifically, it will be estimated as the distance between the z-score associated to Hit (Zh) and False alarm (Zfa) rates. The probability of Hits and False alarms will vary depending on the degree of overlapping between both distributions. Subsequently, the decision criterion (or bias) may be understood as the comparison between the two distributions, and it is estimated as beta (β), or the ratio of the ordinates associated to Hit and False alarm rates. Preferentially, the decision criterion may be estimated by c index that locates the criterion by its distance from the intersection of the two distributions measured in z-score units, or .5 (Zfa + Zh) (Macmillan & Creelman, 1990; Snodgrass & Corwin, 1988).

SDT assumes that sensitivity, as it depends of external factors such as the degree of contrast between the intensity of the signal and the noise, or the taster's physiological state, will be constant for a given participant whenever signal and noise do not vary. However, the decision criterion, which depends on subjective or internal variables such as participants' expectancies or the possible outcome derived from their judgments, may be manipulated through different ways, such as instructions, changing the proportion of signals versus noise in a given situation or using incentives within a payoff matrix (see Macmillan & Creelman, 2005).

The relationship between sensitivity and the decision criterion (or response bias) has been expressed in the ROC graph that we mentioned above. This graph relates the probability of False alarms to Hits, allowing the influence of the different variables that affect psychophysical judgments to be visualized. The ROC curve allows variations in the sensitivity of participants to be detected as well as changes in the decision criterion, revealing whether participants are flexible (with a tendency to say yes more easily, so that they increase the number of Hits, but also the number of False alarms) or conservative (with a tendency to say yes just when they are quite sure about the presence of the signal, decreasing the number of False alarms, but also the number of Hits) (see Swets, 1986b). Thus, ROC's can be useful in olive oil tasting tasks, as they allow separating sensory performance of the taster from his or her hypothetical response biases. SDT has been successfully applied to the study of many psychological processes such as learning, perception or memory, and to a variety of tasks such as medical diagnosis (e.g., Marshall, 2001; Swets, 1996), visual perception (e.g., Zalevski, Henning, & Hill, 2007), weather forecast (e.g., Harvey, Hammond, Lusk, & Mross, 1992) and to the sensory analysis of drinks (e.g., Lawless, 1985; Parr, Heatherbell, & White, 2002) and foods (e.g., Moskowitz, 1988).

The bridge between Thresholds Classic Theory and SDT has not been yet crossed within olive oil tasting. Olive oil tasting is conducted with the type of tasks briefly mentioned above (e.g., Alba et al., 1997). Accordingly, research and practice within the olive oil tasting has not taken in account possible cognitive biases on the perceiver that may affect his or her performance beyond the differences in the sensory process. Thus, the general goal of the present study was to develop an SDT-based task that would allow for a separate evaluation of sensory and decision processes within the field of olive oil tasting. Specifically, we tried to establish a ROC analysis within a discrimination task of similar olive oil concentrations. Sensory evaluation of olive oil is usually conducted by a small group of experts that should try a limited number of samples to avoid overloading of the taste sensory system. Our experiments were designed to approach, as much as possible, the regular tasting situation. This approach included the use of a small number of participants and a small number of samples, so that our results could be generalized to the tasting panels. Aware that this kind of approach makes unlikely to fulfill parametric assumptions such as normality, the second goal of these experiments was to evaluate the use of non-parametric sensory and decision indexes and the use of robust statistical analyses of variance as methodological tools to be applied in the olive oil tasting situations.

Experiment 1a was a parametric study conducted with the goal of characterizing performance for a wide range of olive oil concentrations in the context of a simple discrimination task in which participants had to detect the presence of olive oil diluted in sunflower oil. A series of stimuli with different olive oil concentrations were presented to participants. In this experiment we varied the concentration of olive oil from 12.8% to .2% in doubling dilutions following the standards of the International Olive Oil Council (COI/RES-3/75-IV/96) that establishes the tasters' selection through a psychophysical task akin to the constant stimuli method. The concentration of 12.8% was used as the starting point because pilot work in our laboratory had shown that this concentration was easily discriminated by most participants. Participants had to answer whether the presented stimulus was identical to a stimulus that only contained sunflower oil (sample stimulus). Accordingly, a correct performance in this situation implied responding that the target is different from the sample stimulus. This task allows calculating the minimum amount of olive oil needed for participants consider the target stimulus as different from the sample stimulus. We expected that the greater the olive oil concentration the easier it will be for participants to detect that both stimuli are different. This procedure also allows estimating how similar the different stimuli are considered by participants, so that the stimuli subsequently used in Experiment 1b could be selected. Experiment 1b applied the SDT to the olive oil tasting situation. ROC graphs and sensory and decision indexes were calculated for each of the selected olive oil concentrations (1.6%, 3.2% and 12.8%) allowing us to test the idea that the greater the concentration the easier the detection in a situation in which sensory and decision processes can be separately analyzed.

2. Methodology

2.1. Participants

Thirty-two undergraduate students of the University of Jaén participated in the experiment (14 in Experiment 1a and 18 in

Experiment 1b). They were between 19 and 32 years old. Twenty-three were females, and 9 males. All of them were regular consumers of olive oil, using olive oil regularly to season their food, though none of them had previous experience with the task, neither had they been trained in olive oil tasting.

2.2. Stimulus and apparatus

Stimulus samples were prepared by the method of doubling dilutions (1:2) of extra virgin olive oil in sunflower oil, starting on an olive oil concentration of 12.8% and finishing in a concentration of .2%. That is, 12.8%, 6.4%, 3.2%, 1.6%, .8%, .4% and .2% concentrations were used as target stimuli in Experiment 1a. Experiment 1b was conducted with olive oil concentrations of 1.6%, 3.2% and 12.8%. Extra virgin olive oil was provided by Hermejor de la Reina S. L. (Andújar, Jaén, Spain) elaborated from picual and hojiblanca olives. Carrefour S.A. sunflower oil was used for the dilution.

Samples were presented in standard olive oil tasting glasses covered by standard watch-glasses. Water and green apples were used to clean the palate between tastings. A paper table-sheet and a pen were used to record participants' responses.

2.3. Procedure

The experiment was conducted in a laboratory in which participants were placed at different desks separated from each other by a screen. Participants arrived to the laboratory in groups of two and were verbally informed that they were going to participate in an olive oil taste experiment, and that they were going to receive written instructions to taste each of the samples. After participants gave their informed consent to participate in the experiment they were invited to sit by one of the desks.

Tasting glasses were placed on top of the desk, in front of the participant. Glasses were filled with approximately 5 ml of dilution. Instructions were presented in Spanish in a paper sheet that was handed to participants. Instructions read as follows (translated from Spanish): "To conduct the tasting, please take the glass keeping it covered by the watch-glass, tilt it softly, and in this position turn around the glass so that the interior of the glass gets completely wet. Once this operation is finished, please take the watch-glass out, smell the sample, and taste it. Drink the content of the glass trying to distribute the oil all around your mouth, from the front part of the mouth and tongue, to the sides and the back part of the mouth, and to the palate and the throat. After tasting each sample you should bite a small piece of apple and drink some water to clean your palate".

The experimenter was next to the participant while he or she tasted the first sample stimulus, and then the experimenter stepped aside, out of view of the participant. Participants had no time limitations to perform the task, though most of them finished the complete procedure within 45 min after their arrival to the laboratory. No feedback was provided about the accuracy of their responses.

2.3.1. Experiment 1a

Eight glasses were placed on top of the desk. Seven glasses were filled with approximately 5 ml of a dilution of olive oil in sunflower oil. Each of the glasses contained one of the olive oil dilutions used in the experiment (12.8%, 6.4%, 3.2%, 1.6%, .8%, .4% or .2%). Each glass was labeled with a letter from A to G. Which letter contained which specific olive oil concentration was randomly established for each participant in the study. Labeled glasses were aligned in a row (from A to G) in front of the participant. A glass with approximately 40 ml of sunflower oil was placed between the participant and the row of target stimuli. This glass was not labeled and contained the sample stimulus. Participants were instructed to alternatively taste the sample stimulus and each of the stimuli on the line in the order they were aligned (from A to G), beginning always with the sample stimulus. They were instructed to use the response sheet to indicate whether they thought the target stimulus was the same or different from the sample stimulus once he or she had tasted each of pair of stimuli (the sample, and the corresponding olive oil concentration). They were reminded that they should wash their palate after each tasting that is, each time they tasted the sample stimulus and each time they tasted one of the target stimuli.

2.3.2. Experiment 1b

Participants had to perform a taste task akin to the one usually performed by professional tasters. They were verbally informed that they were going to taste a series of oil samples and that they had to decide whether each sample contained olive oil. The following paragraph was added to the general instructions reported above (Spanish in the original): "After each tasting you will be requested to give a YES or NO answer to the following question: Do you think the sample you just tasted contained olive oil?" The response sheet for each sample began with this question followed by the sentence "Respond YES or NO, please". With the goal of familiarizing them with the task, participants were exposed to the two stimuli that they were going to taste during the experiment, the sample and the target stimulus containing the appropriate olive oil concentration. In these two practice trials participants received feedback about whether the sample they tasted contained olive oil. The experiment began immediately after practice trials finished. Ten glasses labeled from 1 to 10 were placed aligned in front of the participant. For each participant, five glasses contained the five concentrations of olive oil diluted in sunflower oil (Signal + Noise). The other five glasses contained only the sunflower oil (Noise).The 10 glasses were randomly intermixed.

2.4. Dependent variables and data analyses

2.4.1. Experiment 1a

Given the characteristics of Experiment 1a, response "different" was always correct. Number of participants giving correct responses to each concentration was recorded and transformed to percentages of participants giving a correct response to each of the olive oil concentrations used in the experiment.

2.4.2. Experiment 1b

The rate of Hits (probability of responding that the sample contains olive oil in a Signal + Noise trial), and the rate of False alarms (probability of responding that the sample contains olive oil in a Noise trial) were calculated for each participant. To avoid the problem of extreme rates (i.e., maximum hit rates or null rate of False alarms) a logarithmic-lineal corrective strategy was applied by adding a constant value of .5 to each frequency (Snodgrass & Corwin, 1988; Stanislaw & Todorov, 1999).

2.4.3. ROC analysis

The literature presents two main alternatives to calculate ROC curves: either explicitly manipulating the response criterion or asking for an estimation of participant's confidence in each response (see Macmillan & Creelman, 2005; Swets, 1986a, 1986b; Wickens, 2002). None of these strategies was viable in our experimental situation. Biases of responding were not introduced because our explicit goal in this experiment was to test whether these biases appear in neutral tasting situations. Additionally, the tasting situation involves a short number of trials to avoid saturating participants' senses – as established by the international olive council in COI/RES-3/75-IV/96. This short number of trials does not allow for calculating the three confidence estimations that are at least needed for establishing the ROC curve. These limitations inherent to the tasting tasks led us to estimate *z*ROC curves with a single data point (see Verde, Macmillan, & Rotello, 2006). In *z*ROC curves, the mean False alarms rate is presented as a function of the mean Hits rate, transformed in Zeta scores according to the Normal distribution, following the assumptions of SDT. An isosensitivity contour was estimated from the mean performance in each group. The curve was constructed from a single data point following the standard parametric theoretical model of the theory, projecting a point with a symmetrical bias with respect to the one experimentally obtained (see demonstrations in Macmillan and Creelman (2005)).

2.4.4. Non-parametric indexes

As stated above, the taste task forces the use of a short number of trials to avoid saturation of the senses. Under this limitation it is unlikely to fulfill the implicit assumptions under the use of SDT parametric indexes (d' and β/c): Gaussian distributions of probability associated to Signal and Noise with equivalent variances (Wickens, 2002). Accordingly, we have used a non-parametric computation based on the estimation of the area below the ROC curve from a single data point (see Aaronson and Watts (1987); but see Macmillan and Creelman (1990)). Specifically, non-parametric indexes A' and B' will be used for sensitivity and decision process, respectively. These indexes were computationally developed by Grier (1971) from Hodos (1970) (see Stanislaw & Todorov, 1999). These indexes have been methodologically well developed, and are commonly used in SDT-based research. A' goes from 0 to 1, with .5 reflecting random performance. B' goes from −1 (maximum flexible bias) to +1 (maximum strict bias), with 0 as the neutral value. Following Stanislaw and Todorov (1999), the formulas used to calculate these indexes were (H represents Hits rate and F represents False alarms rate in the formula):

$$A' = .5 + \left[sign(H-F)\frac{(H-F)^2 + |H-F|}{4 \cdot Max(H,F) - 4H \cdot F}\right]$$

$$B' = sign(H-F)\frac{H(1-H) - F(1-F)}{H(1-H) + F(1-F)}.$$

2.4.5. Statistical analyses

As pointed out above, sensory evaluation of olive oil in natural settings is usually conducted by the small group of tasters that are part of a tasting panel. This characteristic makes it unlikely that the obtained data fulfill the parametric assumptions of normality and homocedasticity. As the ultimate goal of our research is to apply our findings to olive oil taste panels, we have decided to use a descriptive analysis that allow for exploring the assumptions while summarizing the data. We have chosen central tendency and variability indexes within a Box-Plot graph according to robust statistics based on median. Fig. 3 presents Box-Plot graphs for each of the indexes used as dependent variables (A' and B') for each of the olive oil concentrations used in Experiment 1b. The line within each box represents the median, and the size of the box represents variability. Extreme values are presented as points outside the box. Inspection of the figure reveals heteroscedasticity, skewness of distributions, and outliers, so that the assumptions for parametric tests were not met in this experiment. Accordingly, we selected robust analyses of variance using Wilcox computations on medians (Wilcox, 2005). Monte Carlo simulations have shown that percentile boostrap methods H and R for omnibus and contrasts analyses, respectively, are the most appropriated for discrete data with ties (Wilcox, 2006). Within the results section these analyses will be comparatively presented with respect to more conventional parametric analyses showing that none of the conclusions were affected by the use of robust analyses. Finally, for contrast analyses a trend analysis was used in which the coefficients were estimated by the appropriate second grade equations algorithm given that the levels were not equally spaced (Kirk, 1995). Specifically, values [−0.6172, −0.1543, 0.7715] were estimated for the linear order, and [0.5345, −0.8018, 0.2673] for the quadratic order. Analyses were performed using the free-GNU R software, R version 2.9.1 on i386/x64 Windows (R Foundation for Statistical Computing, http://www.r-project.org/, Last update: June, 2009) and Wilcox' Rallfun-v9 libraries (Rand R. Wilcox, Department of Psychology, University of Southern California, Los Angeles; http://www-rcf.usc.edu/~rwilcox/, Last update: May, 2009. Ask for the functions "m1way" and "medpb"). All tests were interpreted with a significance level of .05.

3. Results

3.1. Experiment 1a

Fig. 1 presents percentage of participants with correct responses across the different olive oil concentrations selected in Experiment 1a. Given that this experiment had the goal of obtaining detection curves using the constant stimuli method to establish the stimulus values after which most participants are able to detect the presence of olive oil in the dilution (according to the international olive oil council resolution COI/RES-3/75-IV/96), only descriptive statistics are reported. In general, the percentage of participants with correct responses progressively increased as olive oil concentration increased, reaching 50% at the .8% concentration. An unexpected decrease in performance was found in 6.4% concentration. There is no theoretical reason for such a decrease. Most likely, the decrease is an artifact of an extraneous variable that only affected that specific olive oil concentration, given that subsequent work in our laboratory has found that discrimination of a 6.4% concentration falls within expected levels. At any rate, this result does not compromise the main goal of this experiment as the obtained results allowed us to select the concentrations that

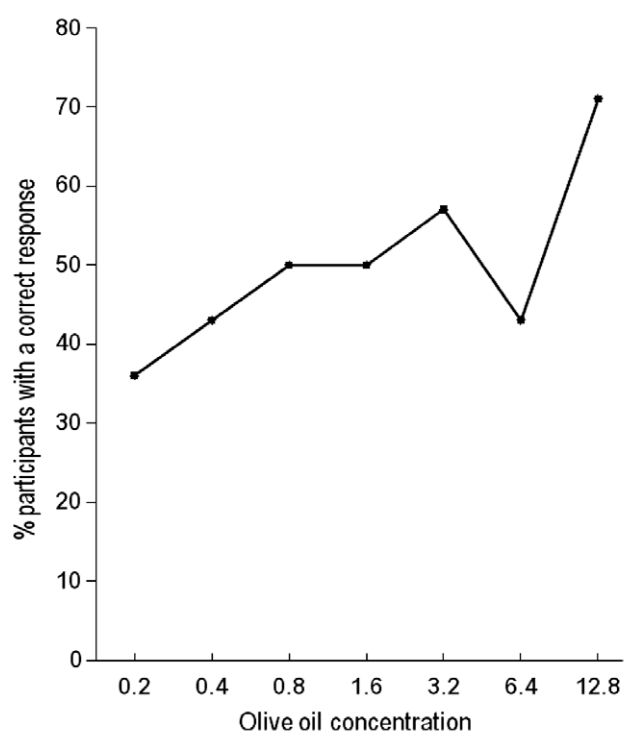

Fig. 1. Percentage of participants with a correct response to each of the olive oil concentrations used in Experiment 1a (.2%, .4%, .8%, 1.6%, 3.2%, 6.4% and 12.8%).

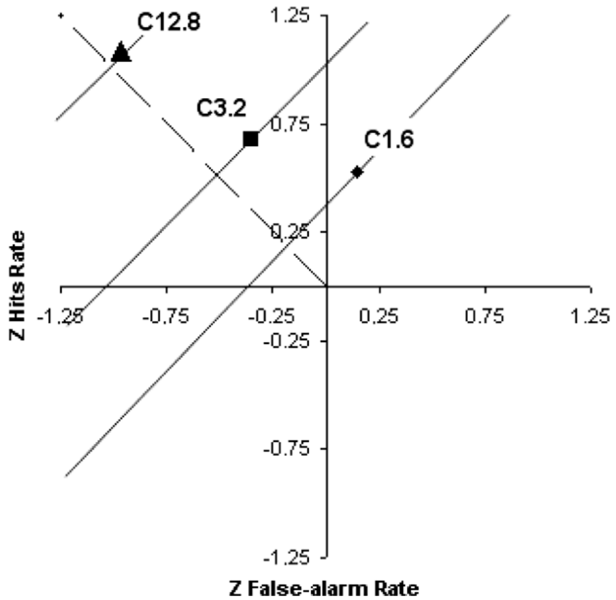

Fig. 2. *z*ROC graph of mean performance in Experiment 1b. Normalized False alarm rates are presented as a function of the normalized Hits rates for each of the three levels of olive oil concentration manipulated in this experiment: 1.6%, 3.2% and 12.8%. The straight lines represent the estimated isosensibility curve for each of the three points in standard scores. The dashed line represents a neutral criterion of performance.

were going to be used in Experiment 1b. With the goal of exploring different levels of performance in Experiment 1b, we selected the highest concentration discriminated by 50% of participants (1.6%) as the lowest concentration used in Experiment 1b, increasing the level of concentration of olive oil from there on. However, as performance with the concentration of 6.4% in this experiment did not seem representative of actual performance with such a concentration in other situations, Experiment 1b was conducted only with 1.6, 3.2 and 12.8 olive oil concentrations.

3.2. Experiment 1b

Fig. 2 presents *z*ROC curves corresponding to each Hits–False alarm rate pair for 1.6, 3.2, and 12.8 olive oil concentrations of Experiment 1b. Each of the three isosensitivity lines allows the changes in the decision criterion for a given sensitivity value to be quantified. Displacement of the points within each line shows participants' decision criterion. Specifically, if a point is displaced to the right participants would be showing a flexible criterion (many Hits, but also many False alarms). The contrary would be true if the point is displaced to the left. The diagonal that goes from (0, 0) to (1.25, 1.25) (dashed line in the figure) shows a neutral performance within the decision process. As may be seen in Fig. 2, the three points from which the lines have been estimated varied according to what would be expected from changes only in sensory processes. The three points increase along the diagonal, from the starting point to the left top vortex as olive oil concentration increased, showing greater sensibility as the salience of olive oil within the mixture is greater. Additionally, the three points are close to the neutral diagonal, except the one corresponding to 1.6 olive oil concentration that is biased towards values that reflect a flexible criterion.

Let us turn now to the statistical analyses of psychophysical indexes that quantify sensory (A') and decision (B') processes, respectively. The left section of Fig. 3 presents a Box-Plot graph in which sensory performance (A') for each olive oil concentration is presented. Fig. 3 shows a progressive increase in detection as concentration increased: Median of .50, .77 and .90 for 1.6, 3.2, and 12.8 concentrations, respectively. The graph also reveals that variability decreased as concentration increased across groups, a feature that goes against homocedasticity assumption. Additionally, the normality assumption is unlikely to be filled given that the function for the 12.8 concentration is mostly uniform. The viability of parametric statistics is further compromised as, at least for the 12.8 olive oil concentration, there is an outlier as the dot shows in Fig. 3. As stated in the statistical analyses section above, these features recommended the use of a robust perspective for the analysis of results. The Robust omnibus Wilcox analysis for A' index revealed significant differences (medians: .50, .77, and .90; $H_W = .028$, $p = .008$). Trend analyses found that only the linear trend was significant ($R_{W,Linear} = .270$, p-value = .016 and critical $p = .025$). The quadratic trend was not significant, ($R_{W,Quadratic} = -.110$, p-value = .166 and critical $p = .050$). Conventional analysis of variance confirmed the conclusions of the robust analysis. There was an effect of concentration on A' [means: .53, .78, .91; $F(2,15) = 8.495$; $p = .003$; MSe = .027; $\eta_p^2 = .531$]. The linear trend was also

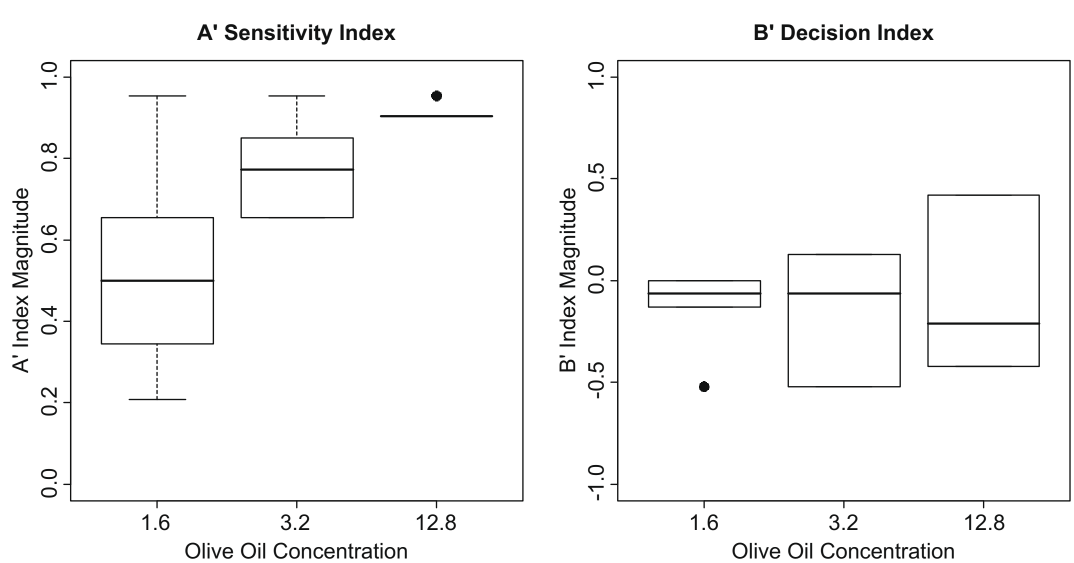

Fig. 3. Sensitivity (left) and decision (right) indexes as a function of the olive oil concentrations used in Experiment 1b. The straight lines represent the median, boxes represent variability, and outliers are represented by dots outside the boxes.

significant, [$F(1,15) = 14.888$; $p = .001$; MSe = .027; $\eta_p^2 = .498$], but the quadratic trend was not significant [$F(1,15) = 2.101$; $p = .168$; MSe = .027; $\eta_p^2 = .123$]. In summary, sensitivity measured through A' shows a linear increase as a function of the increase in the olive oil concentration.

The right part of Fig. 3 presents a Box-Plot graph in which decision index (B') for each olive oil concentration is presented. Performance was neutral (not biased) as the median was near zero in all groups. Additionally, the graph uncovers the same problems for B' index already shown with A' with respect to the statistical assumptions. Exploratory analysis suggested heterocedasticity patterns, uniform distributions, and outliers. Robust omnibus Wilcox analysis for B' index yielded no significant differences (median: −.07, −.07, and −.21; $H_W = .000$, $p = .963$). The same was true when standard ANOVA analyses were conducted [means: −.13, −.15, −.07; $F(2,15) = 2.460$; $p = .138$; MSe = .101; $\eta_p^2 = .014$]. In summary, no variations in the decision process were found across increasing concentrations of olive oil. In fact, the criterion was neutral, given that the values were close to zero.

4. Discussion

The main goal of this experimental series was to develop an olive oil tasting task under the logic of Signal Detection Theory so that researches could measure separate indexes for sensory and decision processes. Experiment 1a established the optimal levels of concentration needed for untrained participants to detect the presence of olive oil diluted in sunflower oil by using a method similar to the method of constant stimuli in classical psychophysics. Participants were exposed to different olive oil concentrations and they were required to estimate whether they were different from a sample that only contained sunflower oil. This method allowed us to establish that from a concentration of .8% up, at least 50% of participants detected differences between the target stimuli and the sample (with the exception of the 6.4% concentration; see Fig. 1). However, this classical approach to olive oil tasting does not allow for detecting participants' cognitive biases that may affect performance, and that play an important role in these kinds of tasks.

As stated above, standard olive oil tasting procedures have the limitation of not allowing for a separate evaluation of sensory and decision processes affecting performance. Experiment 1b presented a task adapted to the STD methodology allowing these shortcomings to be overcome. A detection task was used in which participants had to detect whether olive oil was present within the samples they tasted. Three different groups of participants received three different olive oil concentrations (1.6%, 3.2% and 12.8%). Within each group, half of the samples contained olive oil diluted in sunflower oil, and the other half contained only sunflower oil. This allowed for recording both Hits and False alarms, essential elements for calculation of the ROC curve. As expected, ROC curves showed an increase in sensitivity (detection) as the concentration of olive oil increased. This curves equally allowed observing that the decision criterion of our participants was basically neutral. This neutrality of the decision criterion was further confirmed when sensitivity and decision indexes were calculated. The increase in olive oil concentration increased detection, improving performance, but did not affect the decision index that was basically zero across different oil concentrations. This result is not surprising given that this experiment did not manipulate factors that are known to affect the decision criterion such as instructions, signal probability, or payoffs matrices (Macmillan & Creelman, 2005).

In summary, Experiment 1b shows the usefulness of SDT for sensory analysis of olive oil. Experiment 1b provides a task that it is adapted to the peculiarities of the olive oil tasting procedures in which both Hits and False alarms are recorded and represented. Both of these dependent variables are necessary for calculating sensory and decision indexes. Similar to other activities in which discrimination is involved, olive oil tasting presents the psychological problem of being potentially affected by cognitive biases, as it has been shown within the study of sensory analysis of other types of food (O'Mahony and Hautus, 2008; O'Mahony and Rousseau, 2003). The advantage of this procedure based on SDT is that allows separating the pure sensory process (the ability of the taster to detect the feature) from the decision criterion (possible cognitive biases that may affect taster's performance). To fully understand and evaluate the utility of SDT theory on its application of olive oil tasting, it will be needed to manipulate factors that affect decision criterion leaving the sensory process intact. Our research group already has produced some data that show that both instructions and incentives may affect decision criterion in a task similar to this one (e.g., Ramos-Álvarez, Moreno-Fernández, Paredes-Olay, & Rosas, submitted for publication). However, this research goes well beyond the scope of this paper, in which the main goal is presenting a useful task to evaluate separately sensory and decision processes, a goal that seem to be achieved with the results obtained in Experiment 1b.

From a more practical point of view, the experiments conducted here have allowed us to explore and adjust the appropriated statistical indexes for the specific features of the tasting situation. As it has been pointed above, the need of working with a limited (and small) number of trials makes it difficult to fulfill the assumptions of parametric statistics. We have shown that robust statistics are suitable as a tool to be applied in this kind of situation. Accordingly, we can conclude that our procedure is a tool that overcomes the limits of classical psychophysics allowing for detecting factors that affect sensory and decision processes within the olive oil tasting situation, identifying how these processes affect performance.

Acknowledgements

This research was funded by research grant HUM642 and P06-HUM-01391, from the Consejería de Innovación, Ciencia y Empresa of Junta de Andalucía, and by research grant ACS/PP2005 of the University of Jaén.

We would like to thank J.B. Nelson for his assistance on the preparation of the final version of this manuscript.

M.M. Moreno-Fernández participation was funded by a grant from the FPU program of the Ministerio de Educación de España (AP2007-01332).

References

Aaronson, D., & Watts, B. (1987). Extensions of Grier's computational formulas for A′ and B″ to below-chance performance. *Psychological Bulletin, 102*, 439–442.

Alba, J., Izquierdo, J. R., & Gutiérrez, F. (1997). *Aceite de oliva virgen. Análisis sensorial*. Madrid: Editorial Agrícola Española.

COI/RES-3/75-IV/96. *Método revisado para la valoración organoléptica del aceite de oliva*. Madrid: Consejo Oleícola Internacional.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*, 424–429.

Harvey, L. O., Jr., Hammond, K. R., Lusk, C. M., & Mross, E. F. (1992). The application of signal detection theory to weather forecasting behavior. *Monthly Weather Review, 120*, 863–883.

Hodos, W. (1970). Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin, 74*, 351–354.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole Publishing Co.

Lawless, H. (1985). Psychological perspectives of wine tasting and recognition of volatile flavours. In G. Birch & M. Lindley (Eds.), *Alcoholic beverages*. London: Elsevier Applied Science.

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin, 107*, 401–413.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Marshall, N. W. (2001). The practical application of signal detection theory to image quality assessment in X-ray image intensifier-TV fluoroscopy. *Physics in Medicine and Biology, 46*, 1631–1649.

Moskowitz, H. (1988). *Applied sensory analysis of foods*. Boca Raton, Florida, USA: CRC Press.
O'Mahony, M. (1995). Sensory measurement in food science. Fitting methods to goals. *Food Technology, 49*, 72–82.
O'Mahony, M. (2003). Who said sensory measurement was simple? *Journal of Cosmetic Science, 54*, 92–93.
O'Mahony, M., & Hautus, M. J. (2008). The signal detection theory ROC curve: Some applications in food sensory science. *Journal of Sensory Studies, 23*, 186–204.
O'Mahony, M., & Rousseau, B. (2003). Discrimination testing: A few ideas, old and new. *Food Quality and Preference, 14*, 157–164.
Parr, W. V., Heatherbell, D., & White, G. (2002). Demystifying wine expertise: Olfactory threshold, perceptual skill and semantic memory in expert and novice wine judges. *Chemical Senses, 27*, 747–755.
Ramos-Álvarez, M. M., Moreno-Fernández, M. M., Paredes-Olay, C., & Rosas, J. M. (submitted for publication). A psychophysical approach within the ecological context of olive oil tasting: Evidence about the dissociation between perception and decision processes through a double dissociation additive test.
Snodgrass, G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34–50.
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers, 31*, 137–149.
Swets, J. A. (1986a). Form of empirical ROC's in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99*, 181–198.
Swets, J. A. (1986b). Indices of discrimination or diagnostic accuracy: Their ROC's and implied models. *Psychological Bulletin, 99*, 100–117.
Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers (Scientific Psychology Series)*. Mahwah, NJ: Lawrence Erlbaum Associates.
Verde, M. F., Macmillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision and robustness of d′, Az, and A′. *Perception and Psychophysics, 68*, 643–654.
Wickens, T. D. (2002). *Elementary signal detection*. Oxford; New York: Oxford University Press.
Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). San Diego, California: Academic Press.
Wilcox, R. R. (2006). Comparing medians. *Computational Statistics and Data Analysis, 51*, 1934–1943.
Zalevski, A. M., Henning, G. B., & Hill, N. J. (2007). Cue combination and the effect of horizontal disparity and perspective on steroacuity. *Spatial Vision, 20*, 107–138.