



HAL
open science

What do we learn from comparing hedonic scores and willingness-to-pay data?

Emilie Ginon, Pierre Combris, Youenn Loheac, Géraldine Enderli, Sylvie Issanchou

► To cite this version:

Emilie Ginon, Pierre Combris, Youenn Loheac, Géraldine Enderli, Sylvie Issanchou. What do we learn from comparing hedonic scores and willingness-to-pay data?. *Food Quality and Preference*, 2014, 33, pp.54-63. 10.1016/j.foodqual.2013.11.003 . hal-00950490

HAL Id: hal-00950490

<https://hal.science/hal-00950490>

Submitted on 21 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

What do we learn from comparing hedonic scores and willingness-to-pay data?

E. Ginon, P. Combris, Y. Lohéac, G. Enderli, S. Issanchou

PII: S0950-3293(13)00207-3

DOI: <http://dx.doi.org/10.1016/j.foodqual.2013.11.003>

Reference: FQAP 2703

To appear in: *Food Quality and Preference*

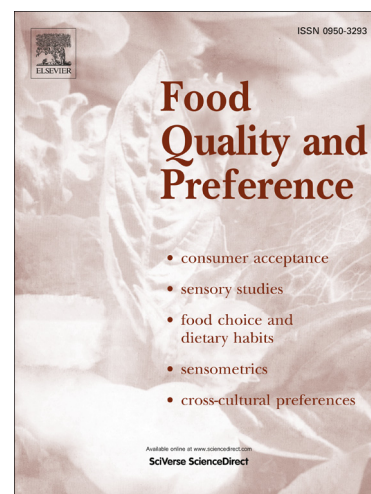
Received Date: 20 June 2012

Revised Date: 30 October 2013

Accepted Date: 13 November 2013

Please cite this article as: Ginon, E., Combris, P., Lohéac, Y., Enderli, G., Issanchou, S., What do we learn from comparing hedonic scores and willingness-to-pay data?, *Food Quality and Preference* (2013), doi: <http://dx.doi.org/10.1016/j.foodqual.2013.11.003>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



1 **What do we learn from comparing hedonic scores and**
2
3 **willingness-to-pay data?**
4

5
6 3
7
8 4 E. Ginon^{abcd}, P. Combris^e, Y. Lohéac^{fg}, G. Enderli^{abce},

9
10 5 S. Issanchou^{abc*}
11
12
13
14
15

16 7 ^a CNRS, UMR6265, Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

17
18 8 ^b UMR1324 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

19
20 9 ^c Université de Bourgogne, Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

21
22 10 ^d LESSAC, Groupe ESC Dijon Bourgogne, F-21000 Dijon, France

23
24 11 ^e INRA, UR1303 ALISS, F-94000 Ivry-sur-Seine, France

25
26 12 ^f France Business School (FBS) – Campus de Brest, F-29000 Brest, France

27
28 13 ^g CREM, UMR6211, CNRS & University Rennes 1, F-35065 Rennes, France
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

58 * Corresponding author. INRA, UMR1324 Centre des Sciences du Goût et de l'Alimentation, 17 rue
59 Sully, BP 86510, 21065 Dijon cedex, France
60 Tel.: 33.3.80.69.30.76. Fax: 33.3.80.69.32.27. E-mail address: Sylvie.Issanchou@dijon.inra.fr
61
62
63
64
65

Abstract

Consumer preferences for different variants of a given food product can be directly obtained with hedonic measurements or revealed with willingness-to-pay measurements. The aim of this paper is to present a comparison of the data collected using these two types of measurements on four data sets collected in our laboratory for different food products (bread, cooked ham, cheese and orange juice). This comparison was conducted at two levels (global and individual) and was based on two criteria: discrimination between variants and consistency in variant ranking. For the four data sets, the hedonic scores and reservation prices were collected for each participant in a 'full information' condition, i.e. in a condition where participants tasted each variant associated with extrinsic information. To reveal consumer willingness-to-pay, the BDM mechanism was used (Becker, DeGroot, & Marschak, 1964), which consists in real sales at a random price. Aggregate results were similar for the two measurements. In addition, in two out of four studies, willingness-to-pay measurements led to slightly higher discrimination between variants than hedonic measurements. At the individual level, more inconsistencies were found. This result is in line with previous studies. Nevertheless, participants were more consistent concerning the most-liked variant than concerning the least-liked variant. Our results also showed that hedonic score distributions did not reveal any cut-off point below which consumers chose the no-purchase option; this cut-off point largely depended on individuals and products.

Keywords

Willingness-to-pay; Hedonic scores; Consumer preferences; Consistency; Discrimination; Food valuation

Highlights

We compare consumer hedonic scores to their willingness-to-pay from four studies.

At the group level, discrimination is similar for the two types of measurements.

At the group level, hedonic scores are consistent with willingness-to-pay.

The two types of measurements generate different individual rankings.

1. Introduction

Hedonic or willingness-to-pay measurements can be used to assess consumer preferences for different variants of a given food product. Aiming both at predicting future choices, these two measurements result from answers to two specific questions: how much do consumers like a given product and which maximum price are they willing to pay for it (reservation price). Being different, these two questions may lead to different answers and to different conclusions. Therefore, given the high stakes linked to the segmentation of food markets and heterogeneity of consumers' behaviours, it is a major issue to better understand the differences between the two methods. Hedonic rating is widely used by sensory scientists, whereas economists usually rely on willingness-to-pay assessment to elicit preferences. This is probably why very few papers use both measurements and even fewer try to compare them.

Comparisons of different evaluation scales to assess preferences have been traditionally conducted by psychologists in the field of judgment and decision making (Hsee, Loewenstein, Blount & Bazerman, 1999). One outstanding example is the "preference reversal" phenomenon, a well-known inconsistency which appears when subjects are asked to rate the attractiveness of different lotteries and also to give a price for each of them. Preference reversal was first reported by psychologists (Lichtenstein & Slovic, 1971). The interest for the phenomenon has spread to experimental economics because it appeared as a major challenge to economic theory (Grether & Plott, 1979). According to basic economic theory, differences in willingness-to-pay should reflect difference in preferences, and using hedonic rating or willingness-to-pay assessment should result in exactly the same ranking of alternatives. Thus, the reservation prices of each individual should be a monotonic transformation of the preferences expressed in their hedonic ratings (Melton, Huffman, Shogren & Fox, 1996). The preference reversal controversy showed that this is not always the case, and that many factors

73 can account for the difference. The main point is that the framing of the evaluation task
74 matters. Joint or separate evaluation, wording of the questions, scales used to measure
75 answers, all these factors focus individuals' attention on different characteristics, or
76 dimensions of the alternatives to evaluate (Nowlis & Simonson, 1997; Tversky, Sattath &
77 Slovic, 1988). Another point raised by economists (Grether & Plott, 1979) is that incentives
78 matter. Indicating hedonic scores on a scale yields no consequence for respondents.
79 Conversely, offering a maximum buying price in a real sale is more involving because
80 participants are actually committed to purchase a product at the end of the experiment.
81 Consequently, hedonic measurement may be subject to the same kind of hypothetical bias
82 which has been largely documented in economic valuation studies. Comparisons between
83 hypothetical and non-hypothetical methods have shown that economic evaluations are largely
84 overstated when elicited in a hypothetical context (List & Gallet, 2001; Murphy, Allen,
85 Stevens & Weatherhead, 2005). To avoid this bias, many applied economists have turn to
86 experimental auctions (Lusk & Shogren, 2007), and more generally towards methods based
87 on mechanisms that motivate participants to reveal as accurately and truthfully as possible
88 their willingness-to-pay (WTP).

89 Without focusing mainly on the comparison of elicitation methods, some papers have
90 used liking measurement jointly with experimental auctions. For example, Melton, Huffman,
91 Shogren & Fox (1996) found very good correspondence between both systems of preference
92 elicitation with aggregate data. They underlined that discreteness in willingness-to-pay may
93 explain a large part of the inconsistencies with hedonic scores (close scores may command the
94 same WTP, and below some hedonic score threshold all WTPs are equal to zero). In a similar
95 study of quality differentiated meat products, Umberger and Feuz (2004) also found a very
96 significant relation between relative WTPs and relative taste ratings of paired samples of
97 steaks. Roosen, Marette, Blanchemanche & Verger (2007) compared hedonic ratings with

1 98 choices between different quantities of alternative products, a measure actually close to
2 99 relative willingness-to-pay. They found a strong correlation between liking scores and
3
4 100 choices. The significant decrease of this correlation after releasing health information raised
5
6 101 the issue of a possible contradiction between taste preference and purchase preference and
7
8
9 102 highlight extrinsic information as one possible source of difference between hedonic scores
10
11 103 and WTPs. This confirms that hedonic scores and WTPs may diverge in some cases, in
12
13 104 particular when willingness-to-pay accounts for important attributes beyond sensory
14
15 105 characteristics.
16
17

18
19 106 To our knowledge, three papers have further investigated the relationship between
20
21 107 hedonic scores and WTPs: one study concerned champagne (Lange, Martin, Chabanet,
22
23 108 Combris & Issanchou, 2002), one concerned cookies, orange juice and chocolate bars
24
25 109 (Noussair, Robin & Ruffieux, 2004) and the third one concerned spelt (Stefani, Romano &
26
27 110 Cavicchi, 2006). In line with most of the previous papers, these three studies reveal an overall
28
29 111 consistency in variant ranking between hedonic and willingness-to-pay results. Lange et al.
30
31 112 (2002) found that external information (label) and sensory information (taste) have the same
32
33 113 impact on the global product evaluation using hedonic scores or reservation prices. However,
34
35 114 these authors observed a larger inter-individual heterogeneity in the relative weights of
36
37 115 external and sensory information for hedonic scores compared to reservation prices. Stefani et
38
39 116 al. (2006) observed as well more heterogeneity for hedonic scores than for WTPs.
40
41 117 Investigating further the relationships between hedonic scores and WTPs, they finally
42
43 118 suggested that in the presence of symbolic and affective components of the value, both
44
45 119 measurements are not directly comparable because WTPs account for more value components
46
47 120 than hedonic scores. Being the only one to explore the consistency between hedonic rating
48
49 121 and WTP at the individual level, the paper by Noussair et al. (2004) goes deeper in the
50
51 122 discussion of differences between the two measurements. Showing that some differences
52
53
54
55
56
57
58
59
60
61
62
63
64
65

123 actually appear in individual rankings, the authors put forward two main sources of
124 inconsistencies: first, they underlined that liking and purchase intentions are essentially
125 different constructs, and second they suggested that auctions may not reveal the whole range
126 of preferences in particular when the value of the good is low or possibly negative for some
127 participants. However, individual inconsistencies did not prevent both methods to result in the
128 same measure of preference intensity at the aggregate level.

129 The above issues and the potential importance of the two methods in product
130 development are sufficient justification for collecting and comparing hedonic and willingness-
131 to-pay data. Moreover, this analysis is relevant not only when actors in the food chain are
132 interested in consumer acceptance of products based on their sensory characteristics, but also
133 when they want to evaluate how consumers value product characteristics conveyed by a label
134 giving information such as brand, origin, environmental impact, and health effects.

135 To contribute to the comparison of hedonic and willingness-to-pay data, we addressed
136 three main questions. The first group of questions refers to the distributions of values
137 collected with these two measurements: “Are the distributions of hedonic scores and
138 reservation prices similar?”, and more specifically, “What are the relationships between the
139 distributions of hedonic scores and purchase decision?”, i.e., “What are the distributions of
140 hedonic scores for buyers compared to non-buyers?” The second question is “Do the two
141 methods have the same capacity to discriminate different variants of a product?” At the
142 individual level, one might wonder if the individual discrimination for one measure is related
143 to the individual discrimination for the other. Finally, a third group of questions refers to the
144 variant rankings. Consistency between hedonic scores and reservation prices can be examined
145 through the similarity of the product hierarchy or at least the correspondence between the
146 most- and least-liked products. Consistency can be examined at the group or individual level.

147 To answer these three questions, we considered four data sets, each one concerning a
148 different type of food product: bread, cooked ham, cheese and orange juice. The next section
149 presents in detail the selection of participants, the four data sets, the experimental design of
150 each experiment and the data analysis that were performed. Section 3 presents the results
151 which support the answers and discussions to our questions. Last, in section 4, we conclude
152 and have a general discussion.

153 **2. Materials and methods**

154 *2.1. Participants*

155 For each study, a sample of consumers was recruited from the general population using
156 different procedures: random selection in four shops of a bakery chain located in Dijon city
157 and suburbs (bread); random dialling in Dijon city and suburbs (orange juice and cheese);
158 and/or random selection from a panel of volunteer consumers in the PanelSens from the
159 ChemoSens Platform (bread, cooked ham, cheese). Participants were selected if they regularly
160 took part in food purchasing and consumed the products of interest in the study. If they agreed
161 to participate, they were sent a letter that explained the main features of the experiment and
162 provided details about the incentive method. Participants received a fee for their participation.
163 Table 1 summarises the main characteristics of the participants for the four studies whose
164 results were analysed in the present paper.

165 *[Insert Table 1 about here]*

166 Except for the cooked ham study, there was a higher proportion of women in all the
167 studies, particularly in the bread study. Participants in the cheese study were older than those
168 involved in the other ones, while participants in the orange juice study were younger. The
169 monthly per capita income was significantly lower for the bread study participants, possibly

170 because they bought their bread, at least occasionally, in shops of bakery chains that sell
171 standard baguettes at low prices.

172

173 *2.2. Product variants*

174 In all studies, four product variants were tested (Table 2). For the bread study, the
175 selected variants were four industrial French baguettes. Three of them had previously been
176 sold on the market (Standard, Meunière and Cereal). The fourth was a new product named
177 ‘Healthy’ (this bread offers potential health benefits). For the cooked ham study, there were
178 two variants, one regular and one with two nutritional labels (low salt content and ‘natural
179 omega-3’) for two brands (national brand and store brand). For the cheese study, there were
180 two variants, one regular and one with a higher omega-3 content for two types of French
181 cheese (Comté and Cantal). For the orange juice study, there was one pure juice and one
182 nectar version for two brands (store/distributor brand and first price brand). It must be noted
183 that, except for the cheese experiment where we had two subgroups (Comté and Cantal), all
184 variants can be considered as close substitutes.

185 *[Insert Table 2 about here]*

187 *2.3. Experimental procedure*

188 *2.3.1. Overview of the designs*

189 The four studies were carried out at the National Institute for Agricultural Research in
190 Dijon (France) between spring 2005 and spring 2009. Each study focused on a different
191 product, namely, bread, cheese, cooked ham or orange juice.

192 The main characteristics of each study’s design are presented in Table 3. For each
193 study, hedonic score and willingness-to-pay (WTP) measurements were obtained from the
194 same participants. In each study, four different variants of the product were presented to the

195 participants. The data shown here were collected under a 'full information' condition, in which
196 participants tasted the sample with the product name for bread (standard, Meunière, Cereal,
197 Healthy), and packaging with all mandatory information and nutritional claims for the three
198 other products. For bread, hedonic measurements were collected in a first session, and then
199 willingness-to-pay measurements were obtained two weeks (n=92 participants) or four weeks
200 (n=85 participants) later. For cheese and cooked ham, the two tasks were performed
201 separately within a single session, i.e., the four hedonic scores were collected first, and the
202 four reservation prices were collected second. For orange juice, the two tasks were performed
203 successively for each variant. A sequential monadic presentation was used for hedonic
204 measurements in all studies, as was the case for the willingness-to-pay measurements in the
205 orange juice study for which both measurements were collected in the same variant
206 presentation. For the other studies, the four variants were presented simultaneously while
207 participants were asked to give their reservation prices. For cheese and cooked ham,
208 reservation prices were collected without actual re-tasting but based on tasting memory. In all
209 cases, the presentation order of the four variants followed a Williams Latin square balanced
210 for order and first-order carry-over effects. For a given consumer, this order was different for
211 both measurements in the case of bread but was the same for cheese and cooked ham in order
212 to help consumers to remind of their level of appreciation for each variant. Each session took
213 place in a temperature controlled ($21 \pm 2^{\circ}\text{C}$) sensory room equipped with individual booths.

214 *[Insert Table 3 about here]*

216 2.3.2. Hedonic measurements

217 Participants answered the question “How much do you like this baguette [cooked ham,
218 cheese, orange juice]?” on a linear scale, which was labelled “I don’t like it at all” on the left
219 side and “I like it very much” on the right side. The grades on the hedonic scales were

220 converted into scores from 0 to 10 by measuring the distance between the left side of the scale
221 and the participants' marks.

222

223 *2.3.3. Willingness-to-pay measurement*

224 The experimenter explained the principle of the BDM mechanism to the participants,
225 using concrete examples. Participants were told that one product variant would be sold at a
226 random price at the end of the session and, that individuals with a reservation price higher
227 than the selling price would be committed to paying the random selling price. Then,
228 participants were invited to sign a consent form and commitment to buy according to the
229 principle previously explained. For each variant, participants answered the question "What is
230 the maximum price you are ready to pay for this baguette [cooked ham, cheese, orange
231 juice]?" They could choose a no-purchase option if they did not want to buy the item. In this
232 case, the reservation price was considered as null. For each product variant, participants wrote
233 down their reservation prices on a 'buying form' for orange juice and used a computer for the
234 other studies. Participants were told that at the end of the session, they will be asked to
235 randomly select one variant out of all the products they had evaluated. This procedure was
236 used to maintain the same level of involvement for all variants and to ensure that participants
237 did not have to buy an excess of products, which might have induced them to reduce their
238 reservation prices. Each participant randomly drew one token indicating a selling price.
239 Participants were told that prices on the tokens were distributed according to the selling prices
240 of each product on the Dijon market. However, the market prices were never communicated
241 to the participants. The price written on the token was compared to the reservation price given
242 by the participant. If the participant's reservation price submitted for the selected variant was
243 equal to or higher than the price on the drawn token, the participant had to buy the product
244 variant at the price on the token. If the participant's reservation price was lower than the price

245 on the token, the participant had no opportunity to purchase the item. At the end of the
246 session, participants could ask to examine the bag of price tokens.

248 2.4. Data Analysis

249 All statistical analyses were performed for each study and carried out with SAS/STAT®
250 software, version 9.1 (SAS Institute Inc., Cary, NC, USA, 2002-2003).

251 2.4.1. Comparison of the types of measurements at the global level

252 Firstly, the distributions of hedonic scores and reservation prices were examined.
253 Secondly, in order to compare discrimination between the two types of measurements at the
254 global level, an analysis of variance (ANOVA) was performed on each data set, and F values
255 were compared. The ANOVA procedure was used according to the following model:

256 hedonic (reservation price) = participant + variant + error.

257 When the ANOVA revealed a significant effect ($p < 0.05$), least-square means and the
258 95% confidence intervals were calculated, and t -tests were performed.

259 Finally, consistency between the hedonic scores and reservation prices for the four
260 variants and each study was examined by calculating Kendall correlation coefficients between
261 mean values.

263 2.4.2. Comparison of the types of measurements at the individual level

264 In order to compare discrimination between the two types of measurements at the
265 individual level the coefficients of variation (CV) were calculated for each participant and
266 each type of measurement, then, the Kendall correlation between the two series of CV was
267 calculated for each product.

268 In order to examine consistency at the individual level, several indices were calculated.
269 The same approach as the one used by Noussair et al., 2004) was adopted to declare a

270 participant as consistent. First, for each participant, the Kendall correlation coefficient
271 between the hedonic scores and reservation prices was calculated in order to examine the
272 consistency for all variants. Then, consistency was assessed for the most- and least-liked
273 variants. The first criterion, called ‘strict consistency’, for all variants corresponds to the
274 following situation: for any couple of variants, if the hedonic score [variant A] is higher than
275 the hedonic score [variant B], then the reservation price [variant A] should be higher than the
276 reservation price [variant B]. This equation gives a Kendall correlation equal to 1. The second
277 criterion, ‘Weak consistency’, for all variants, is less severe; it corresponds to the following
278 situation: for any couple of variants, if the hedonic score [variant A] is higher than or equal to
279 the hedonic score [variant B], then the reservation price [variant A] should be higher than or
280 equal to the reservation price [variant B]. In the case of four variants, this equation gives
281 Kendall correlation coefficient above 0.70 and lower than 1. Moreover, consistency on the
282 most-liked and the most-disliked variants were also examined.

283 The impact of several individual characteristics (e.g. sex, age, income) on the Kendall
284 correlation between the coefficients of variation for hedonic scores and the coefficients of
285 variation for reservation prices, as well as on the Kendall correlation between variant rankings
286 was examined. No significant effect appeared and consequently these results are not reported.

288 **3. Results and discussion**

289 The comparison between hedonic scores and reservation prices is presented below at
290 both the global and individual levels. At the global level, a descriptive analysis of the
291 distributions of hedonic scores and reservation prices was performed.

292 *3.1. Global level*

293 *3.1.1. Comparison of the distributions of hedonic scores and reservation prices*

294 When comparing the distributions according to the method, it was observed that each
1
2 295 study yielded a high frequency of null reservation prices (corresponding to the no-purchase
3
4 296 option) but an extremely low frequency of null hedonic scores (Fig. 1). Additionally, the
5
6
7 297 distributions of hedonic scores, but not the reservation prices, for bread and cheese are
8
9
10 298 skewed to the right. These observations are in agreement with previous data (Lange et al.,
11
12 299 2002) on champagne, a product with a high monetary value. These authors suggested that a
13
14 300 greater commitment in the case of willingness-to-pay measurements compared to hedonic
15
16 301 measurements may explain this difference. The present results illustrate that even with a low
17
18
19 302 monetary value product, giving a high hedonic score does not mean to be willing to purchase
20
21
22 303 the product at a high price. It is also important to note that despite the experimental situation
23
24 304 and the fee given to the participants, the reservation prices were rarely lower or higher than
25
26 305 the market prices for the same product category. As demonstrated by Harrison, Harstad &
27
28
29 306 Rutström, (2004), observed reservation prices are censored by market prices. In the case of
30
31 307 orange juice, apart from the null reservation prices, the reservation price distribution and the
32
33
34 308 hedonic score distributions are very flat. Therefore, both measurements revealed great
35
36 309 individual variability in consumer reactions in this study.

37
38
39 310 *[Insert Figure 1 about here]*

40
41 311
42 312 A comparison of the hedonic score distributions for buyers and non-buyers (Fig. 2)
43
44 313 yields no clear cut-off point below which consumers chose the no-purchase option. However,
45
46
47 314 the mean hedonic scores corresponding to no-purchase decisions are not surprisingly lower
48
49
50 315 than for purchase decisions. Nevertheless, the mean scores for no-purchase decisions differ
51
52 316 according to the product. For bread and cheese, the percentage of non-buyers started to
53
54 317 decrease above a hedonic score of 6 while for cooked ham, this percentage started to decrease
55
56
57 318 above a hedonic score of 5, and for orange juice above a hedonic score of 2. Thus, it seems
58
59 319 that consumers are more ready to accept a least-liked variant for orange juice than for the
60
61
62
63
64
65

320 other products, particularly for bread and cheese. The results for orange juice are in agreement
321 with the results obtained by Lange et al. (1999), who examined the role of hedonic scores and
322 prices in purchase behaviour when the consumers were placed in a situation of choice under
323 economical constraint. This study found that consumers could choose their least-liked
324 product. One possible explanation proposed by these authors could also apply here: “non
325 single consumers could have ordered products not only for themselves but for the whole
326 family and thus could have chosen products depending on the family members' preferences”.
327 The results for bread could be explained by bread's important social and cultural role for
328 French consumers (Kaplan, 2002). The effect of socio-cultural values on French consumers
329 can also apply to cheese (Roberts & Micken, 1996). For these two types of products, a higher
330 proportion of consumers must really like the product to be willing to purchase it.

331 *[Insert Figure 2 about here]*

332 3.1.2. Discrimination between variants and consistency between variant rankings

333 The results show a significant variant effect for all products (all p values < 0.0001) and
334 similar F values for both methods (Fig. 3).

335 *[Insert Figure 3 about here]*

336 As shown in Fig. 3, the ranking of the variants, for all products, is the same regardless
337 of the measurement. For orange juice and cooked ham, the same number of groups was
338 obtained regardless of the measurement. For bread and cheese, three groups of variants were
339 obtained for the hedonic scores and four for the reservation prices. Therefore, the willingness-
340 to-pay measurement appears slightly more discriminant than the hedonic measurement. The
341 difference could be related to the different way consumer responses were collected. In our
342 hedonic measurements, consumers expressed their degree of liking on a linear scale, i.e., a

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

345 visual analogic scale. Therefore, they are not aware (except for the anchors) of the exact
346 numerical value deduced from their evaluation. Conversely, for willingness-to-pay
347 measurements, consumers directly gave their reservation price and are thus conscious of the
348 numerical values and more able to express small differences in the reservation prices between
349 variants. Nevertheless, the differences between both measurements in terms of variant
350 discrimination are minor, and we can conclude that there is a relatively high agreement at the
351 global panel level for giving a high hedonic score and a high reservation price.

352 Thus, our results that show a good overall consistency between hedonic scores and
353 reservation prices are in agreement with results from previous studies (Lange et al., 2002;
354 Noussair et al., 2004; Stefani et al., 2006). In fact, all these studies revealed the same ordering
355 of the different variants for both measurements. The main difference is observed with data on
356 champagne (Lange et al., 2002). In the same information condition (bottle and tasting), these
357 authors observed a higher F value for reservation prices than for hedonic scores (data not
358 shown). This difference could be due to champagne's high social value and the importance of
359 brand reputation in particular on monetary value.

360 361 *3.2. Individual level*

362 *3.2.1. Discrimination between variants*

363 The Kendall correlations calculated between individual coefficients of variation (CV)
364 obtained for hedonic scores and reservation prices were significant for all studies except for
365 the bread study (Table 4). This finding means that for three out of the four studies, the level
366 of individual discrimination for hedonic scores is related to the level of individual
367 discrimination for reservation prices. In the case of bread, participants could have liked the
368 variants differently but were not ready to buy them at different prices. This is particularly
369 noticeable as bread was, among the four products studied, the one with the lowest unitary

370 price. However, this behaviour could be related to the fact that this product is bought almost
371 daily and has a stable price, which may have been used as a reference regardless of the
372 perceived quality of the product.

373
374 *[Insert Table 4 about here]*

375 376 3.2.2. Consistency of individual rankings

377 The percentage of participants who satisfied the two consistency criteria is presented for
378 each data set in Table 5. This table shows that there were discrepancies between the hedonic
379 scores and reservation prices of individual participants. Moreover, a significant chi-square is
380 observed ($\chi^2=23.5$, $p<.0001$) for the weak consistency criterion, which reveals that the
381 percentages of participants who satisfied this criterion differ according to product. This
382 significant chi-square is due to the orange juice data, where a higher percentage of weak
383 consistency is observed; the chi-square is no longer significant ($\chi^2=5.1$, $p=0.08$) when the
384 orange juice data set was excluded. This better result for orange juice can be explained by the
385 design; as noted previously, for each variant presented in this experiment, the reservation
386 price was collected immediately after the hedonic score, whereas in the other experiments, the
387 four hedonic scores were collected, then participants gave their four reservation prices. The
388 effect of the experimental conditions is confounded with the product effect, but this
389 hypothesis seems plausible. On average, in the four data sets, only 17.6% of the participants
390 had the same strict ordering of the four variants for both methods. The average consistency
391 between the two measures increases to 50.7% if ex-æquo on one of the methods is permitted.
392 In their study on orange juice, Noussair et al. (2004) obtained a higher level (31.5%) of strict
393 consistency for a similar product and higher levels for both strict and weak consistency than
394 our average values. This discrepancy could be explained by the lower number of variants in

395 their study (three instead of four). Another possible explanation is that in their experiment, the
396 different variants were tasted without any extrinsic information. Extrinsic information, such
397 as brand, could modify the economic value attributed to a variant due to the influence of the
398 market price, possibly because participants do not give the same importance to the sensory
399 and non-sensory properties of each variant when assigning a hedonic score and when stating a
400 reservation price. For example, a participant could like a nectar version more than a pure juice
401 version but does not want to give a higher reservation price for the nectar than for the pure
402 juice if s/he is aware that nectar is made of juice with water and sugar added.

[Insert Table 5 about here]

3.2.3. Consistency on the most-liked and the most-disliked variants

407 We further explored the relationships between the two measurements by looking at the
408 agreement level for the most-liked (most-valued) and the least-liked (least-valued) variants.
409 Concerning the ranking of the four variants, we defined strict and weak consistency criteria.
410 The strict preference consistency corresponds to the case in which participants gave the
411 highest score and the highest reservation price to the same variant. The weak preference
412 consistency corresponds to the case where there were ex-æquo values for one method.
413 The percentage of participants who satisfied these two consistency criteria is presented for
414 each data set in Table 6.

[Insert Table 6 about here]

418 The strict consistency criteria between hedonic scores and reservation prices are higher
419 for the most-liked variant than for the least-liked variant. Concerning the weak consistency

420 criterion, this result is only true for bread and orange juice studies. The greater difference
421 between strict consistency and weak consistency for the least-liked variants than for the most-
422 liked variants can be explained by a tendency of participants to give a zero reservation price
423 not just for the least-liked variant. In fact, for the orange juice, cheese and cooked ham studies
424 where the difference between strict consistency and weak consistency for the least-liked
425 variants was particularly high, we observed 39.5, 30.8 and 21.4% of participants, respectively,
426 who did not want to purchase variants that they had ranked at the fourth and third hedonic
427 positions.

428

4. General discussion and conclusion

430 In accordance with previous studies, aggregate results are similar for the two
431 measurements. In addition, our willingness-to-pay measurements led to slightly higher
432 discrimination between variants in two out of four studies. One reason for that difference is
433 the fact that hedonic measurements use an unstructured analogic scale (the position on a bar),
434 whereas willingness-to-pay is directly measured in Euros, which may enable a higher
435 precision of the latter. This point does not seem to have received much attention yet. It can be
436 pointed out that when two variants of a product were significantly different for willingness-to-
437 pay measurements but not for hedonic measurements, the ‘healthy’ variants (B4 for bread and
438 C2 for cheese) were less valued than the standard variants.

439 At the individual level, we found more inconsistencies, which again is in line with
440 previous studies. The two types of measurements generate different individual rankings, and
441 there is no way to identify whether this discrepancy is due to the type of measurement or to a
442 change in participant preferences. In fact, several authors have shown that, even when using
443 the same type of measurement, participant preferences can change even within a session. For
444 example, when participants were asked to indicate their preferred variant among two out of

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

445 five different orange juices, which they had just previously rated on a linear hedonic scale,
446 only 66% were consistent (Cordelle, Lange & Schlich, 2004). Lévy and Köster (1999)
447 obtained similar results (58% consistency) when participants chose their preferred variants
448 among three soft drinks previously rated on a 9-point hedonic scale. This volatility of
449 preferences is a well-acknowledged fact among sensory scientists (Köster, 1991; Köster,
450 2003; Köster, Couronne, Léon, Levy & Marcelino, 2003). This low level of consistency
451 observed could also be explained by the fact that, except for the cheese experiment where we
452 had two subgroups (Comté and Cantal), our variants can be considered as close substitutes.
453 Nevertheless, we did not observe a higher consistency for cheese compared to the three other
454 products. However, despite these individual inconsistencies, between hedonic scores and
455 reservation prices, on the rankings of all variants, it appears that participants were more
456 consistent on the most-liked variant than on the least-liked.

As mentioned in the introduction the two methods refer to different constructs (i.e.,
hedonic value and economic value). Hedonic scores are supposed to reflect private values.
However, when hedonic scores are collected in an informed condition (i.e., tasting with
external information) they could be influenced by a desirability bias due to the hypothetical
situation, in particular when information is related to nutritional characteristics with potential
health benefits. This possibility of such a bias has been previously underlined by several
authors (e.g., Dailliant-Spinnler & Issanchou, 1995; Lundgren, 1981). On the contrary,
reservation prices are supposed to be less influenced by a desirability bias due to the non-
hypothetical situation. This could explain the greater discrimination between standard and
healthy variants for bread and cheese. Nevertheless, different factors could impact reservation
prices. Firstly, as nicely demonstrated by Muller and Ruffieux (2011) reservation prices do
not exclusively correspond to private values but are partly influenced by common value
(market price). Secondly, budget constraints could influence reservation prices. Nevertheless,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

470 in the present experiment individual consistency, measured by Kendall correlation, did not
471 differ according income per capita (data not shown). Thirdly, inconsistency could be due to
472 the fact that hedonic scores are not supposed to be influenced by consumer's need of the
473 product whereas it could be the case for reservation prices.

474 In spite of inconsistencies at the individual level, both methods lead to very similar
475 conclusions when aggregated data are considered. Our results also revealed that there was no
476 clear cut-off point below which consumers chose the no-purchase option; this cut-off largely
477 depends on individuals and on products. Further research that focuses on specific points of the
478 protocol (numeric scales used for both methods, simultaneous or sequential evaluation of the
479 different variants, presence of external information such as a brand, a nutritional label, use of
480 a common market value as a reference price) may contribute to a deeper understanding of the
481 inconsistencies at the level of individual participants.

482 In conclusion, as previously underlined by Lange et al. (2002), willingness-to-pay
483 measurement seems to be a relevant approach to reveal the true individual value of a product
484 when external information is provided. This seems particularly important when information is
485 about product healthiness as in such a case hedonic measurement may suffer of a desirability
486 bias.

488 **Acknowledgements**

489 The research reported in this article was financially supported by INRA through the PRA
490 (Programme Recherche Alimentation) programme 'Cafet'n'lab' and by the Regional Council
491 of Burgundy. The studies on cooked ham and cheese were conducted within the Truefood
492 "Traditional United Europe Food", an Integrated Project financed by the European
493 Commission under the 6th Framework Programme for RTD (Contract n. FOOD-CT-2006-

494 016264); the information in this document reflects the authors' views only, and the
1
2 495 Community is not liable for any possible use of the information contained therein. The
3
4 496 authors would like to thank Fabienne Bouillot and Christophe Martin (ChemoSens) for their
5
6
7 497 critical logistical help during the experiments.
8
9

10 498

11 499 **References**

12
13
14 500 Cordelle, S., Lange, C., & Schlich, P. (2004). On the consistency of liking scores: Insights
15
16 501 from a study including 917 consumers from 10 to 80 years old. *Food Quality and Preference*,
17
18 502 *15*, 831-841.

19
20
21 503 Daillant-Spinnler, B., & Issanchou, S. (1995). Influence of label and location of testing on
22
23 504 acceptability of cream cheese varying in fat content. *Appetite*, *24*, 101-106.

24
25 505 Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal
26
27 506 phenomenon. *The American Economic Review*, *69*, 623-638.

28
29
30 507 Harrison, G. W., Harstad, R. M., & Rutström, E. E. (2004). Experimental methods and
31
32 508 elicitation of values. *Experimental Economics*, *7*, 123-140.

33
34
35 509 Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals
36
37 510 between joint and separate evaluations of options: A review and theoretical analysis.
38
39 511 *Psychological Bulletin*, *125*, 576-590.

40
41 512 Kaplan, S. L. (2002). *Le retour du bon pain. Une histoire contemporaine du pain, de ses*
42
43 513 *techniques et de ses hommes*. Paris (FRA): Editions Perrin.

44
45
46 514 Köster, E. P. (1991). The dynamics of consumer preferences and aversions. In *Food*
47
48 515 *Ingredients Asia Conference Proceedings 1991*, Expoconsult, Maarssen, The Netherlands.

49
50
51 516 Köster, E. P. (2003). The psychology of food choice: some often encountered fallacies. *Food*
52
53 517 *Quality and Preference*, *14*, 359-373.

54
55
56 518 Köster, E. P., Couronne, T., Léon, F., Levy, C., & Marcelino, A. S. (2003). Repeatability in
57
58 519 hedonic sensory measurement: a conceptual exploration. *Food Quality and Preference*, *14*,
59
60 520 165-176.

- 521 Lange, C., Martin, C., Chabanet, C., Combris, P., & Issanchou, S. (2002). Impact of the
522 information provided to consumers on their willingness to pay for Champagne: comparison
523 with hedonic scores. *Food Quality and Preference*, *13*, 597-608.
- 524 Lange, C., Rousseau, F., & Issanchou, S. (1999). Expectation, liking and purchase behaviour
525 under economical constraint. *Food Quality and Preference*, *10*, 31-39.
- 526 Lévy, C. M., & Köster, E. P. (1999). The relevance of initial hedonic judgements in the
527 prediction of subtle food choices. *Food Quality and Preference*, *10*, 185-200.
- 528 Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in
529 gambling decisions. *Journal of Experimental Psychology*, *89*, 46-55.
- 530 List, J. A., & Gallet, C. A. (2001). What experimental protocol influence disparities between
531 actual and hypothetical stated values? *Environmental and Resource Economics*, *20*, 241-254.
- 532 Lundgren, B. (1981). Effect of nutritional information on consumer responses. In J. Solms, &
533 R. L. Hall (Eds), *Criteria of food acceptance*, (pp. 27-33). Zurich: Forster-Verlag AG.
- 534 Lusk, J. L., & Shogren, J. F. (2007). *Experimental auctions. Methods and applications in*
535 *economic and marketing research*. Cambridge: Cambridge University Press.
- 536 Melton, B. E., Huffman, W. E., Shogren, J. F., & Fox, J. A. (1996). Consumer preferences for
537 fresh food items with multiple quality attributes: Evidence from an experimental auction of
538 pork chops. *American Journal of Agricultural Economics*, *78*, 916-923.
- 539 Muller, L., & Ruffieux, B. (2011). Do price-tags influence consumers' willingness to pay? On
540 the external validity of using auctions for measuring value. *Experimental Economics*, *14*, 181-
541 202.
- 542 Murphy, J. J., Allen, P. G., Stevens, T. H., & Weatherhead, D. (2005). A meta-analysis of
543 hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, *30*,
544 313-325.
- 545 Noussair, C., Robin, S., & Ruffieux, B. (2004). A comparison of hedonic rating and demand-
546 revealing auctions. *Food Quality and Preference*, *15*, 393-402.

547 Nowlis, S. M., & Simonson, I. (1997). Attribute-Task Compatibility as a Determinant of
1 548 Consumer Preference Reversals. *Journal of Marketing Research*, 34, 205-218.

549 Roberts, S. D., & Micken, K. S. (1996). Le fromage as life: French attitudes and behavior
5 550 toward cheese. *Advances in Consumer Research Volume*, 23, 111-119.

551 Roosen, J., Marette, S., Blanchemanche, S., & Verger, P. (2007). The effect of product health
9 552 information on liking and choice. *Food Quality and Preference*, 18, 759-770.

553 Stefani, G., Romano, D., & Cavicchi, A. (2006). Consumer expectations, liking and
14 554 willingness to pay for specialty foods: Do sensory characteristics tell the whole story? *Food*
17 555 *Quality and Preference*, 17, 53-62.

556 Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice.
20 557 *Psychological Review*, 95, 371-384.

558 Umberger, W. J., & Feuz, D. M. (2004). The usefulness of experimental auctions in
25 559 determining consumers' willingness-to-pay for quality-differentiated products. *Review of*
28 560 *Agricultural Economics*, 26, 170-185.

561

562

563 Tables

564 Table 1

565 Characteristics of the participants in each study

566

		Study			
		Bread	Cheese	Cooked	Orange
		177	195	ham	juice
		n		126	86
Gender	Male	58	84	66	31
	Female	119	111	60	55
Age	Mean	42.2	50.0	44.8	38.0
	Min	19.0	21.0	25.0	16.0
	Max	81.0	81.0	66.0	81.0
	SD	16.2	14.4	12.8	14.7
Income ^a	Mean	1003.8	1256.9	1293.3	1216.6
	Min	0	330.0	514.3	0.0
	Max	2520.0	3480.0	4020.0	9500.0
	SD	575.9	572.3	596.5	1167.6
Household size	1	33	29	18	18
	2	74	86	44	27
	3-4	51	66	52	32
	>4	16	14	3	9

^a Monthly income per capita (in €)

568

569 Table 2

570 Characteristics of the product variants in each study

Study	Variant	Brand	Label	Market price (€)
Cooked Ham	H1	National	/	2.67
	H2	National	Low salt content and 'natural ω -3'	2.78
	H3	Store	/	2.14
	H4	Store	Low salt content and 'natural ω -3'	3.10
Cheese	C1	AOP	Comté	2.00
	C2	AOP	Comté 'Higher ω -3 content'	2.25
	C3	AOP	Cantal	Test product ^a
	C4	AOP	Cantal 'Higher ω -3 content'	Test product ^a
Bread	B1	National	Standard	0.64
	B2	National	Meunière	0.87
	B3	National	Cereal	1.06
	B4	Experimental	Healthy	0.95
Orange juice	J1	Lowest-price	Nectar	0.31
	J2	Store	Nectar	1.00
	J3	Lowest-price	Pure juice	0.63
	J4	Store	Pure juice	1.51

571 ^aTest product not sold on the market, produced by INRA – URH Theix

572

573 Table 3

574 Main characteristics of each study's design

Study	(Measure * Variant) order		Variant order ^a for hedonic scores
	Step 1	Step 2	(Hedo) and willingness-to-pay (WTP)
Bread	Hedo * 4 variants	WTP * 4 variants	Different per participant
Cheese	Hedo * 4 variants	WTP * 4 variants	Same per participant
Cooked ham	Hedo * 4 variants	WTP * 4 variants	Same per participant
Orange juice	Step 1 & 2 combined: (Hedo + WTP) * 4 variants		Same per participant

575 ^a All orders were balanced (Williams Latin squares)

576

577 Table 4

1
2 578 Kendall correlations between individual coefficients of variation for hedonic scores and for
3
4
5 579 reservation prices in each study

Study	Mean value of Kendall correlation	<i>p</i> value	Missing values ^a
Bread	0.05	0.37	4
Cheese	0.28	<0.0001	2
Cooked ham	0.31	<0.0001	0
Orange juice	0.28	0.0002	3

17 580 ^a Missing values due to no purchase for all variants

18
19
20 581
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

582 Table 5

583 Percentages of participants satisfying the strict and weak consistency criteria in each study

Study	Strict consistency (%)	Weak consistency (%)
Bread	20.9	40.7
Cheese	13.8	52.3
Cooked ham	17.5	47.6
Orange juice	19.8	72.1
Pooled data	17.6	50.7

584

585 Table 6

1
2 586 Percentages of participants^a satisfying the strict and weak consistency criteria for the most-
3
4 587 liked and most-disliked variants in each study

Study	Most-liked variant		Most-disliked variant	
	Strict consistency	Weak consistency	Strict consistency	Weak consistency
Bread	64.4%	74.6%	48.0%	62.1%
Cheese	59.0%	70.3%	30.3%	69.7%
Cooked ham	57.1%	67.5%	34.9%	69.8%
Orange juice	68.6%	81.4%	27.9%	72.1%

19 588 ^a Excluding participants with to no or only one purchase

20
21 589
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

590 List of figures and captions

1
2 591

3 592 Fig. 1. Distributions of hedonic scores and reservation prices in each study. The vertical lines
4 indicate the extreme values of price on the market. The arrow indicates the average market
5 593 price
6
7 594

8
9 595
10
11 596 Fig. 2. Distributions of hedonic scores for the buyers and non-buyers in each study. Light grey
12 = non-buyers; dark grey = buyers.
13 597

14 598
15
16 599 Fig. 3. Averages of the hedonic scores and reservation prices for each product variant.
17

18 600

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

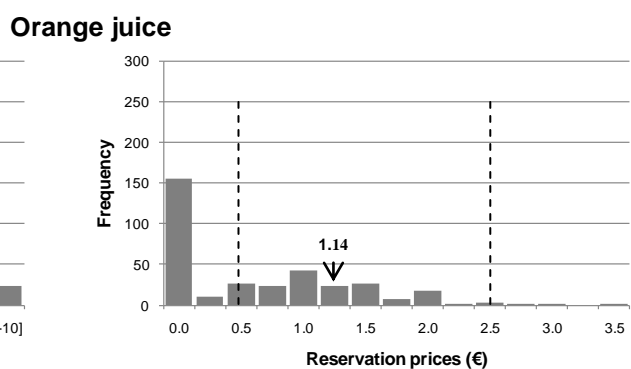
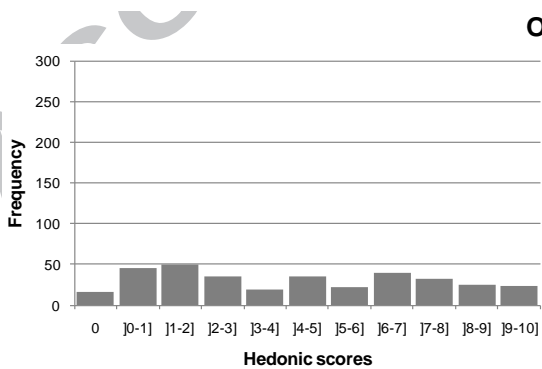
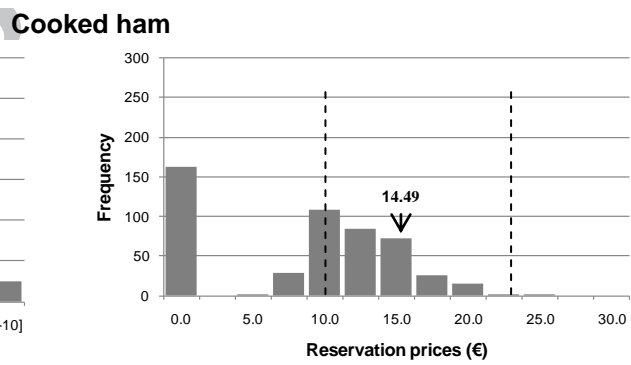
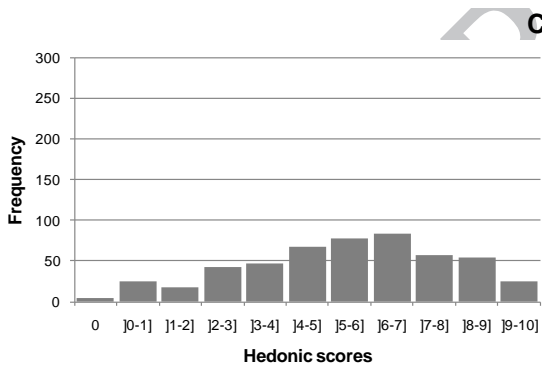
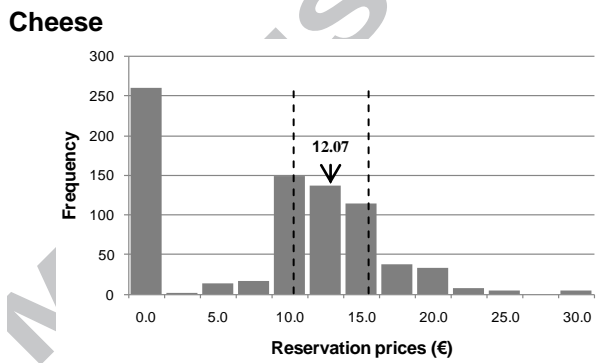
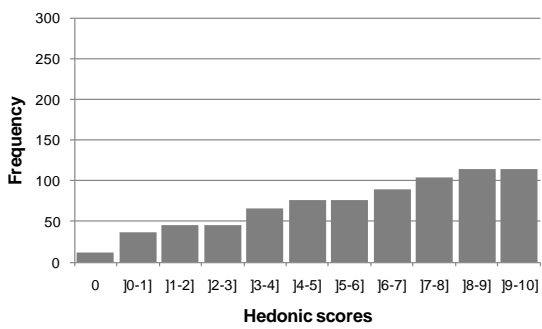
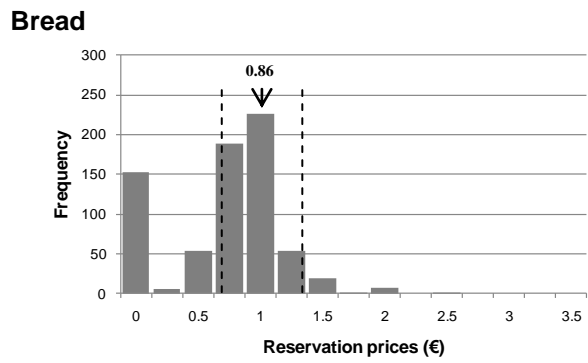
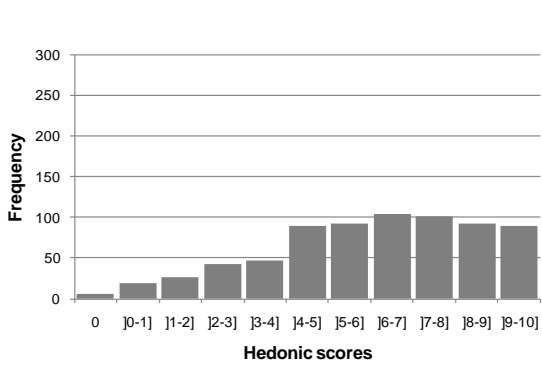
63

64

65

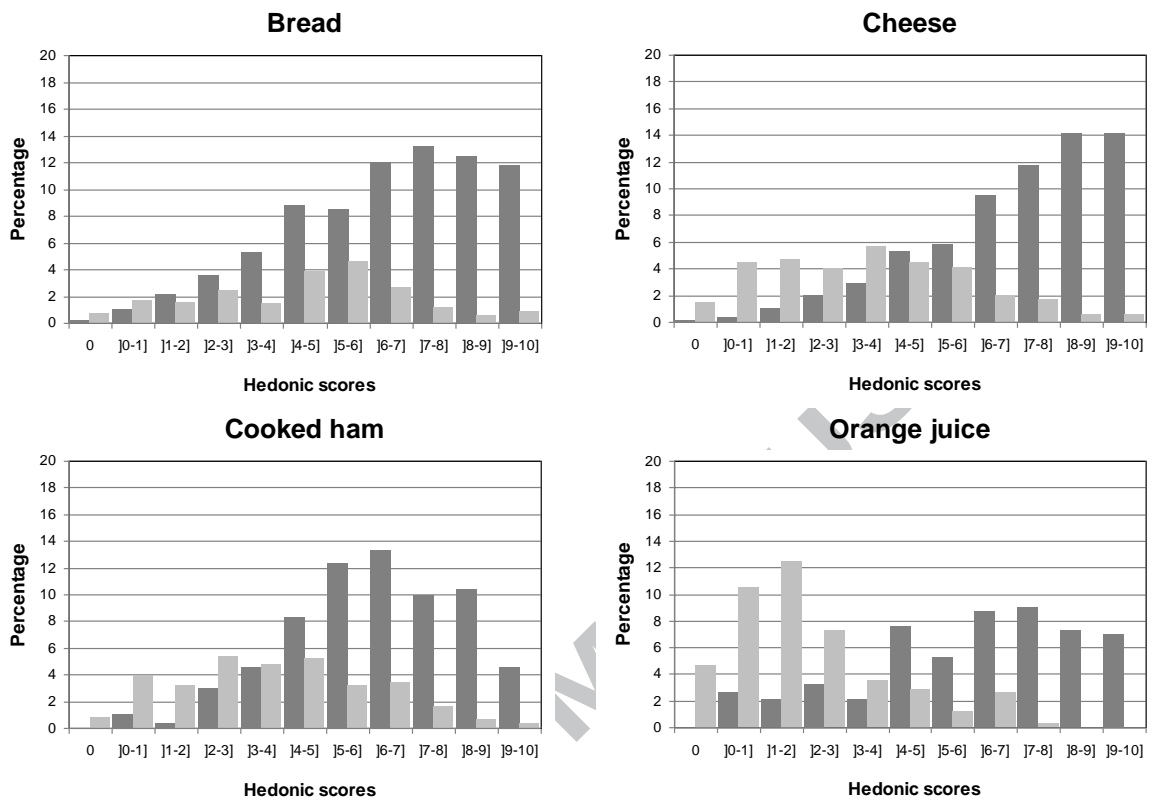
601 Fig. 1

602
603



604

605 Fig. 2

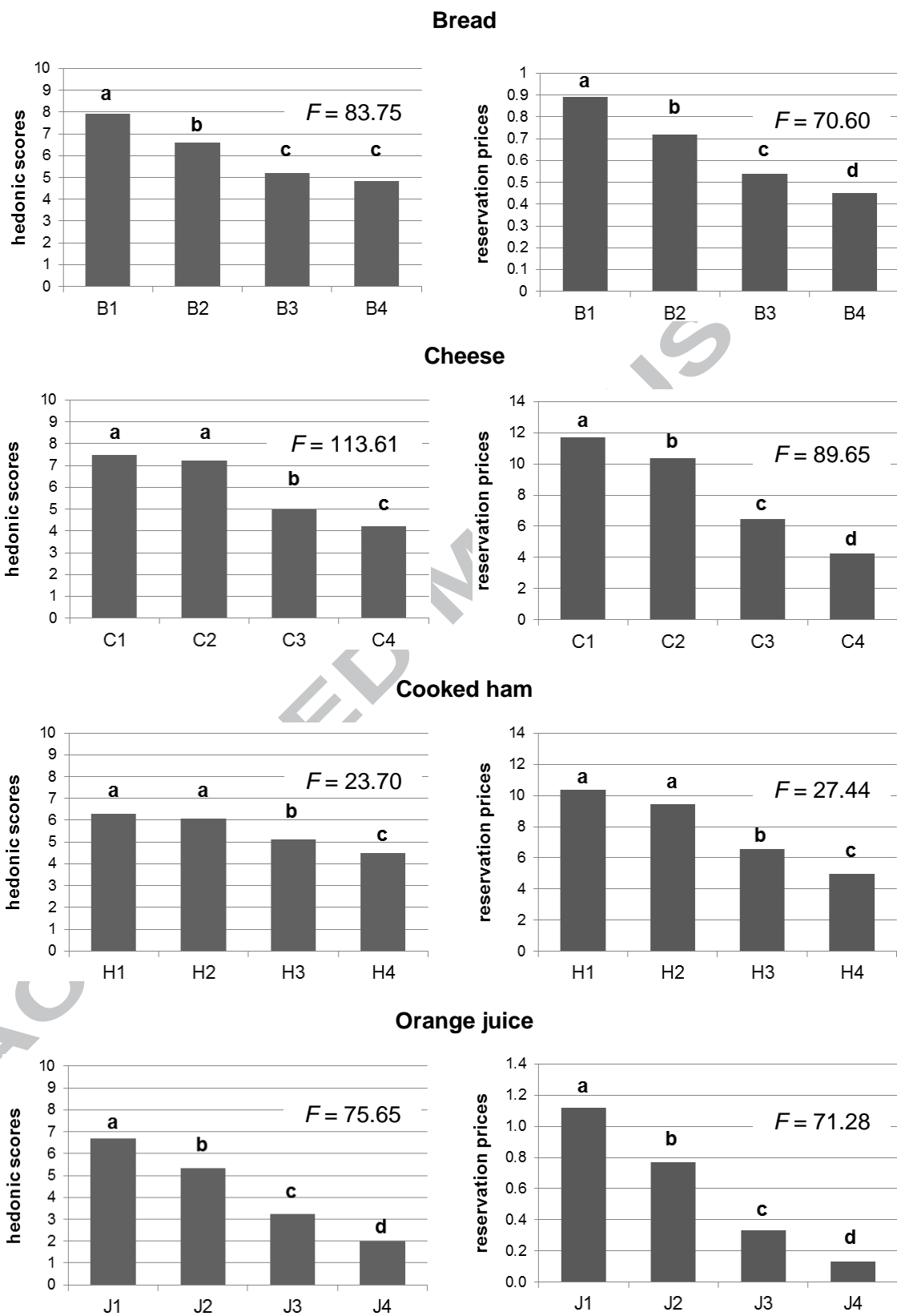


608
609

ACCEPTED

610 Fig. 3.

611



612