Available online at www.sciencedirect.com
SCIENCE DIRECT®
Food Quality and Preference 16 (2005) 13–23
Food Quality and Preference
www.elsevier.com/locate/foodqual

Do trained assessors generalize their knowledge to new stimuli?

Sylvie Chollet [a,*], Dominique Valentin [b], Hervé Abdi [c]

[a] Institut Supérieur d'Agriculture, 41 rue du Port, Université Catholique de Lille, 59046 Lille Cedex, France
[b] ENSBANA—CESG, Université de Bourgogne, France
[c] Program in Cognition and Neuroscience, The University of Texas at Dallas, France

Received 21 April 2003; received in revised form 17 November 2003; accepted 3 December 2003
Available online 22 January 2004

Abstract

Previous work showed that trained assessors are better at discriminating and describing familiar chemico-sensorial stimuli than novices. In this study, we evaluated whether this superiority holds true for new stimuli. We first trained a group of subjects to characterize beer flavors over a two year period. After training was accomplished, we compared the performance of these trained assessors with the performance of novice subjects for discrimination and matching tasks. The tasks were performed using both well-learned and new beers. Trained assessors outperformed novices in the discrimination task for learned beers but not for new beers. But on the matching task, trained assessors outperformed novices for both learned and new beers. These results suggest that perceptual learning is harder to generalize than verbal learning.

Keywords: Training; Generalization; Perceptual learning; Cognitive learning; Expertise; Experts; Novices; Beer

1. Introduction

In the food industry, sensory experts are widely used in product development, product improvement, quality control, and marketing. But, as mentioned by Lawless (1984), training experts assumes that sensory training can develop perceptual abilities. Curiously, only a few studies have explicitly addressed such a critical assumption, and most of these studies focused on *discrimination* performance. Some showed that discrimination does not improve with experience. For example, Roberts and Vickers (1994) and Wolters and Allchurch (1994) showed that training increases the coherence of panelists' judgments but not their discrimination ability. Likewise, Chambers and Smith (1993) showed that experience does not improve product differentiation. In agreement with these results, Bende and Nordin (1997) and Parr, Heatherbell, and White (2002) showed that wine experts and novice subjects have comparable detection thresholds. In contrast, other studies found a positive effect of training on discrimination performance. Clapperton and Piggott (1979) showed that training and experience increase beer flavor discrimination performance as well as assessors' reliability. Cardello et al. (1982) reported an increased textural differentiation of fish types by experts compared to novices. Along the same lines, Peron and Allen (1988) indicated that, with perceptual training, previously non-discriminating assessors were able to detect beer flavors identical to the ones they had learned. Also Rabin (1988) demonstrated that wine experts outperformed novices in an odor discrimination task. So it seems that, at least in some cases, experts tend to discriminate better than novices the stimuli on which they have been trained.

But is expertise only a perceptual phenomenon, or does it also involve a linguistic component? In fact, most descriptive techniques used in sensory evaluations involve the *description* of the organoleptic properties of the stimuli via a lexicon generally acquired during training. To this extent, sensory experts are expected to learn how to recognize, identify, and communicate the sensory properties of a product. Again, only a few studies focused on this issue. For example, Desort and Beauchamps (1974) and Cain (1979) showed that the usual low ability of subjects to identify odors can be improved by practice and feedback. In agreement with

* Corresponding author. Tel.: +33-3-28-38-48-48; fax: +33-3-28-38-48-47.
E-mail address: s.chollet@isa-lille.fr (S. Chollet).

0950-3293/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodqual.2003.12.003

this idea, Lawless (1988) found that trained panelists were more selective and consistent in the use of perfume descriptors that novices. But this improved ability to describe odors does not always imply an increased ability to communicate as illustrated by Lehrer (1975) who reported that wine "connoisseurs" were not better at matching wines with their descriptions than novices. Somewhat in agreement with this last result, Lawless (1984) reported that, although experts match wines to descriptions better than novices, their performance is not very impressive and the difference between the two groups of subjects is rather small. In contrast, Solomon (1990) found that wine experts clearly outperform novices even when novices are provided with experts' descriptions. Solomon suggested that the greater matching ability of experts results from greater perceptual acuity and a more precise descriptive vocabulary. However, the linguistic superiority of experts is not always found. For example, Sauvageot and Fuentes (2000) observed poor performance from both novices and experts matching products to their corresponding expert generated consensual or individual profiles. Along the same line, Gawel (1997) claims that matching tasks can be performed successfully without formal training. And, in fact, he found that untrained-experienced subjects were able to match experts' consensus descriptions well above chance level.

In summary, the general picture that emerges confirms the hypothesis that perceptual and verbal abilities can be developed during sensory training even though this is not always the case. Part of the discrepancies between authors on the effects of sensory training may be due to differences in the length of training and in the definition of expertise. It is quite possible that an improvement in perceptual or verbal abilities requires a longer period of training than that received by some of the experts participating in the previously mentioned studies. According to ASTM standard E-253-1, an *expert* is "a person with *extensive experience* in a product category who performs perceptual evaluations to draw conclusions about the effects of variations in raw materials, processing, storage, aging, etc." In contrast, "a person with a high degree of sensory acuity who has *experience* with the test procedure and an established ability to make consistent and repeatable sensory assessments" should be referred to as a *trained subject*. Nevertheless while on the whole, both experts or trained assessors as well as novices discriminate chemico-sensorial stimuli above what would be expected by chance alone, experts and trained assessors tend to outperform novices. Experts and trained assessors also tend to be better at matching stimuli to their descriptions than novices who are often at chance level. The main problem, with these studies, however, is that most of them compare subjects' performance on stimuli on which the experts or trained assessors have been previously trained. Only a few studies tried to evaluate if the "expert advantage" can be generalized to new stimuli. Rabin (1988) indicates that subjects trained to perform a given odor task are able to generalize their knowledge to both a new task and new odors. Precisely, he showed that subjects trained to name or describe odors are better at performing a discrimination task than novices even when *new* odors were used. On the other hand, Bende and Nordin (1997) report that experts were able to generalize to new odors but not to a new task. Specifically, they found that assessors, trained to discriminate and identify wine aromas performed no better than novices in a detection task whereas, they outperformed them in a discrimination task using unfamiliar odors. Likewise, Chambers and Smith (1993) found that panelists trained to describe cereal products were also able to describe an unfamiliar product such as strawberry jam. These studies suggest that, in some cases, the "expert advantage" observed in discrimination and matching tasks can generalize to new stimuli. What remains unclear is what specific circumstances are needed for this generalization to occur.

To address this question, we need to compare the performance of trained assessors and novices on several tasks using stimuli learned by the trained assessors as well as stimuli new to both groups of subjects.

For that, we decided to use expertise in beer. The rational for this choice is that, compared to other products, beers are relatively stable over time and their flavors can easily be manipulated by the addition of artificial aromas. We trained 20 subjects, one-hour a week, over a period of two years, to detect and identify 15 off-flavors [1] in beer (e.g., cardboard, butter, cabbage, lilac) as well as to describe general characteristics of a large number of commercial beers such as sparkleness, taste of hop, or persistence. The method of training was based on active learning: When participants made a mistake, they were not given the answer immediately but had the opportunity to find the right answer by themselves. If after three attempts, they still had not found the correct answer, then the answer was given to them and they had the possibility to retest the beer. To evaluate the efficiency of learning, subjects' identification performance was tested four times over the two years of training. During these evaluation tests, subjects had to identify which aroma(s) were present in a beer among the 15 aromas they learned during training. The aroma concentrations used during these tests were just noticeable so that the performances were very low at the beginning of training (25% correct identification, random level performance was 6.25%, or 1/16—because there were 15 beers plus an additional 16th possible

[1] An off-flavor is a natural beer olfactory component acceptable at a low concentration but considered as a flaw when its concentration reaches a given threshold.

answer as subjects were also allowed to answer that there was nothing in the beer). Subjects' performance improved regularly with training and reached 55% correct identifications after two years (see Chollet et al., 2002). While this level of performance is not very impressive, it shows that subjects acquired some competence during training. To evaluate if this competence could be generalized, we asked our trained assessors and an equivalent number of novices to perform a discrimination and a matching tasks on learned and new beers.

2. Discrimination task

2.1. Material and method

2.1.1. Subjects

Two groups of subjects participated in this study: the "*trained* group" or the "*trained assessors*" received two years of training, and the "*novice* group" or the "*novices*" did not receive any training. The trained group consisted of 19 trained assessors (5 women and 14 men ranging in age from 25 to 65 years, Mean age = 44.16, sdt = 12.62). These subjects were enrolled in a training program designed to produce beer-trained assessors. They were trained one hour per week to detect and identify added flavors [2] in beer and to evaluate the intensity of global beer characteristics. [3] At the time of the experiment, these subjects had already received 72 hours of training and evaluation. The novice group consisted of 20 untrained subjects (10 women and 10 men ranging in age from 24 to 76 years, Mean age = 43.25, sdt = 15.07). They were beer consumers but did not have any formal training or experience in the description of beer flavors.

2.1.2. Stimuli

Eighteen beers were used in this experiment. The first six beers were commercial lagers (noted A to F) which have fairly constant and salient characteristics. Specifically, beer F is bitter, beer D is very light, sweet and not bitter, beers B and C have a higher alcoholic content than the others, and beer E is characterized by a slight banana flavor. Beer A, on the other hand, does not have any distinctive characteristic. The next six beers were created by adding to commercial beer A one out of six different learned aromas (i.e., which were learned during training): honey, caramel, banana, metallic, bread, and lilac. The last six beers were created by adding to commercial beer A one out of six different new aromas (i.e., which were not learned during training): coconut, lemon, orange blossom, thyme, wintergreen, and tar. The aroma concentrations were chosen so as to obtain 60% of correct answers during a pre-test performed with 10 participants who did not have any formal training or experience in the description of beer flavors. All beers were presented in black plastic tumblers and served at 8 °C under red light.

2.1.3. Experimental design

Both group of subjects performed a "same/different" task. Two independent variables were manipulated: the type of beer (learned commercial, learned supplemented, and new supplemented) and the training level (trained assessors vs. novices). The pairs of beers were presented in a random order identical for all subjects.

2.1.4. Procedure

All subjects participated in three sessions in which they received 12 pairs of beers and were asked to indicate if both beers were similar or different. In the first two sessions they tasted the learned beers (commercial and supplemented) and in the last session the new beers. In the first two sessions, 24 pairs were tasted, 12 similar pairs and 12 different pairs: The 12 identical pairs were made of the 6 commercial pairs and the six learned supplemented beers. The 12 different pairs were randomly chosen among the 66 possible pairs ([(6 commercial+6 supplemented)^2 – 2 identical]/2). The resulting 24 pairs were randomly separated in two series, one presented during the first session and the other during the second session. For the third session, 12 pairs of beers were tasted, 6 similar pairs and 6 different pairs. The 6 different pairs were randomly chosen among the 15 possible pairs.

2.2. Results and discussion

The frequency of correct responses (expressed as percentage) was calculated for each subject in each condition. A response was considered correct when two identical beers were judged identical or when two different beers were judged different. Fig. 1 shows trained assessors' and novices' frequency of correct responses for (a) all, (b) learned and (c) new beers.

A series of Student *t*-tests revealed that, for both groups, the proportions of correct responses were significantly above chance level ($\alpha = 0.05$) in all conditions. However, a two-way ANOVA with training level as a between-subject variable and type of beer as a within-subject variable showed a main effect of training level, $F(1,37) = 9.28$, $MSe = 180.40$, $p < 0.01$ and type of beer, $F(1,37) = 4.93$, $MSe = 123.55$, $p < 0.05$. On the whole, trained assessors were more discriminating ($M = 68.57$) than novices ($M = 57.36$); and learned beers ($M = 64.64$) were better discriminated than new

[2] The flavors used to train the experts were: almond, banana, butter, caramel, cabbage, cheese, lilac, metallic, honey, musty, bread, cardboard, phenol, apple, and sulfite.

[3] The global characteristics are bitterness, astringency, sweetness, alcohol, hop, malt, sparklingness, and persistence.

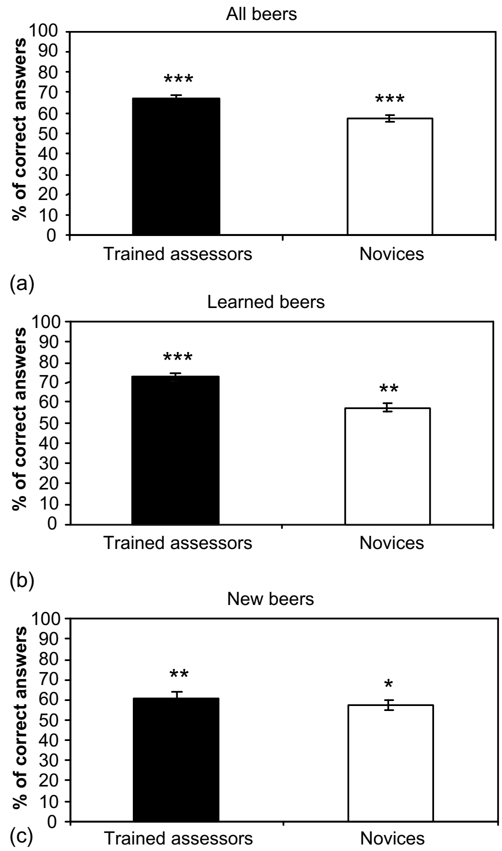

Fig. 1. Percentage of correct discrimination for (a) all, (b) learned, and (c) new beers. The error bars shows the confidence interval at $\alpha = 0.05$, and the stars the student t-test significance level $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

beers ($M = 59.19$). An interaction between training level and type of beer was also apparent, $F(1,37) = 5.32$, $MSe = 123.55$, $p < 0.05$. Whereas novices' correct response rates were similar for the two types of beers ($M = 57.29$ for learned beers and $M = 57.50$ for new beers), trained assessors discriminated the learned beers better ($M = 72.37$) than new beers ($M = 60.96$). Moreover, a series of Student t-tests showed that trained assessors ($M = 72.37$) were better than novices ($M = 57.29$) for learned beers, $t(18) = 4.57$, $p < 0.001$ but not for new beers ($M = 60.96$ vs. $M = 57.50$), $t(18) = 0.77$, ns.

In summary, we found that training improved discrimination performance. However this improvement was observed only for beers on which the trained assessors were trained. Indeed, no significant difference was observed between trained assessors and novices for new beers. The fact that trained assessors were not able to generalize to new aromas in a discrimination task suggests that the familiarity of an aroma plays a part in the discrimination of this aroma in a blending. Rabin (1988) has already reported this effect of familiarity on discrimination in the context of odor perception. Specifically, this author showed that a group of subjects trained to label a set of odors before performing a discrimination task on the same odors performed better than a control group. According to Gibson (1969), verbal learning could act on discrimination performance by both improving odors short-term recognition memory and focusing subjects' attention to salient aspects of the odors.

3. Matching task

3.1. Material and method

3.1.1. Subjects

Two groups of subjects participated in this study. The first group consisted of the same 19 trained assessors used in the discrimination. The second group consisted of 19 novices (5 women and 14 men ranging in age from 21 to 74 years, Mean age = 43.16, sdt = 16.97). They were beer consumers but did not have any formal training or experience in the description of beer flavors. None of them participated in the previous task.

3.1.2. Stimuli

As previously, 18 beers were used in this experiment. The first six beers were the same commercial lagers as those used in the discrimination task. The next six beers were the same learned supplemented beers as in the discrimination task. The last six beers were created from commercial beer A with the addition of one out of six different new aromas: vanilla, kiwi, wax, cut grass, rose, and raw green beans. The aroma concentrations were chosen so as to give a 100% rate of correct identifications during a pre-test realized by 15 participants who did not have any formal training or experience in the description of beer flavors. All beers were presented in black plastic tumblers and served at 8 °C under red light.

3.1.3. Experimental design

Three independent variables were used: the type of beers (learned commercial, learned supplemented, and new supplemented), the level of training (trained assessors vs. novices), and the type of descriptions (written by trained assessors vs. written by novices). The beers were presented in a random order identical for all subjects.

3.1.4. Procedure

Subject's task involved two phases—a description phase and a matching phase – separated by one or two weeks. In the description phase, subjects were asked to describe each of the 18 beers so that another subject could match each of the descriptions with the corresponding beer. They were asked not to describe beer appearance or to cite commercial trademark. Otherwise, they were free to use any descriptors they wanted including hedonic and intensity terms. They were allowed to smell and taste the beers as often as they wished. The beers were presented one at a time along with a sheet of paper on which subjects were asked to write down their descriptions. When they were done with a description, the beer and the sheet of paper were removed and replaced by another beer and another sheet of paper. The matching phase was broken into two sessions separated by a week. In both sessions, the 18 beers were presented one at a time. With each beer, a subject received two descriptions, one corresponding to the presented beer (target) and one corresponding to another beer (distractor). The distractor beer was chosen randomly among the 17 remaining beers and was the same for all subjects. Subjects were asked to determine which description corresponded to the beer under consideration. In the first session, each beer was presented with a pair of descriptions written by another subject of the same group (i.e., trained assessors received descriptions written by another trained assessor and novices received descriptions written by another novice). The same subject wrote the two descriptions in a given pair, but different subjects wrote the pairs of descriptions associated with the 18 beers. As a subject was never provided with his own description, each subject, thus, had to match a description provided by every other subject in the group. The principle was identical for the second session with the exception that the trained assessors were provided with descriptions written by novices and vice versa.

3.2. Results and discussion

3.2.1. Frequency of correct matches

The frequency of correct matches was calculated within each experimental condition. In what follows we express, for convenience, the results as the percentage of correct responses. Fig. 2 shows trained assessors' and novices' percentage of correct matches for (a) commercial, (b) learned supplemented, and (c) new supplemented beers. A Student *t*-test showed that trained assessors were significantly better than chance when they matched trained assessors' descriptions for all types of beers [M (learned commercial) $= 64.9\%$, $t(18) = 3.39$, $p < 0.01$; M(learned supplemented) $= 71.0\%$, $t(18) = 5.90$, $p < 0.001$; and M(new) $= 75.4\%$, $t(18) = 5.46$, $p < 0.001$] whereas their ability to match novices' descriptions was at chance level for learned supplemented beers and commercial beers [$M = 53.5\%$, $t(18) = 0.78$, ns; $M = 54.4\%$, $t(18) = 1.05$, ns] but reached significance for new supplemented beers [59.6%, $t(18) = 2.63$, $p < 0.01$]. Novices' correct matching rates for trained assessors' descriptions were better than chance for learned and new supplemented beers [$M = 64.9\%$, $t(18) = 2.85$, $p < 0.01$; $M = 65.8\%$, $t(18) = 4.26$, $p < 0.001$] but not for commercial beers [$M = 54.4\%$, $t(18) = 0.81$, ns]. In contrast, the correct matching rates for novices' descriptions were better than chance for commercial beers [$M = 57.0\%$, $t(18) = 1.80$, $p < 0.05$] but not for learned and new supplemented beers [$M = 57.0\%$, $t(18) = 1.36$, ns; 57%, $t(18) = 1.36$, ns].

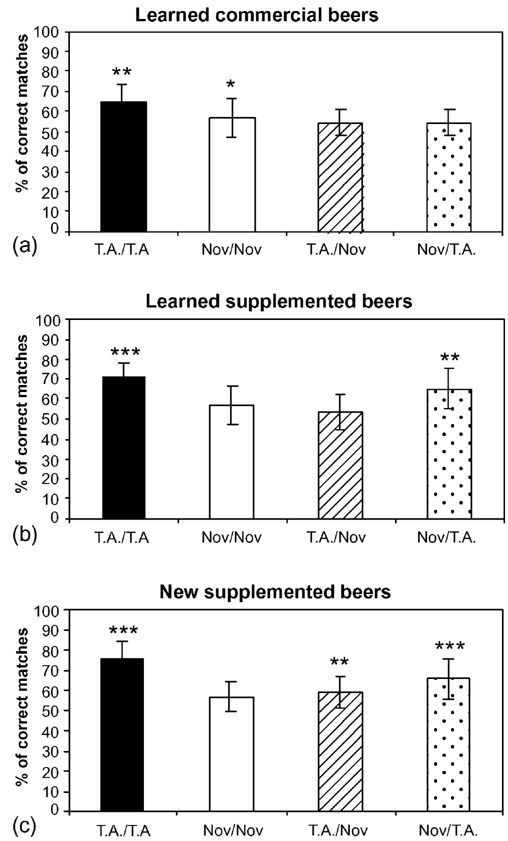

Fig. 2. Percentage of correct matches for (a) commercial, (b) learned supplemented, and (c) new supplemented beers. The horizontal axis represents the four matching conditions: TA/TA (trained assessors matching trained assessors' descriptions), Nov/Nov (novices matching novices' descriptions), TA/Nov (trained assessors matching novices' descriptions), Nov/TA (novices matching trained assessors' descriptions). The error bars shows the confidence interval at $\alpha = 0.05$, and the stars the student *t*-test significance level $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

A three-way ANOVA with level of training as a between-subject variable and type of beers and type of descriptions as within-subject variables revealed a main effect of type of description $F(1, 36) = 13.62$, $MSe = 389.59$, $p < 0.001$. Correct matching rates with trained assessors' descriptions (66.1%) was superior to the rates obtained with novices' descriptions (56.4%) for both trained assessors and novices. No main effect was observed for type of beers or training level, and no interaction between the variables reached significance. However, a Student *t*-test performed on trained assessors' description conditions shows that the correct matching rate of trained assessors (70.5%) was superior to the correct matching rate of novice (61.7%), $t(18) = 2.38$, $p < 0.05$.

3.2.2. Descriptor analysis

The matching performance of both trained assessors and novices do not seem to be linked to the type of beer. Indeed descriptions written by trained assessors lead to higher frequencies of correct matches than descriptions written by novices for both *learned* and *new* beers. These descriptions, on the other hand, lead more often to correct matches when read by trained assessors than by novices. The next questions thus are: What makes trained assessors' descriptions different from novices' descriptions? And are experts using different terms to describe learned and new beers? To answer these questions, we analyzed the terms used by trained assessors and novices to describe the three types of beer. Table 1 shows the terms generated by trained assessors and novices for commercial, learned supplemented, and new supplemented beers. Because of the large number of generated terms, only the terms cited by at least four different subjects are presented in these tables. The efficiency of each term (i.e., the probability that a term leads to a correct match) was evaluated by an index varying between 0 and 1. This index was calculated as follows: When a description led to a correct match, all the terms in this description received a score of 1. When a description did not lead to a correct match, all the terms of the description received a score of 0. An average score was computed for each term: it corresponds to the proportion of correct matches associated with this descriptor. To evaluate the significance of this proportion we used the binomial test assuming equi-probability (i.e., under the null hypothesis, the proportion of correct matches would be equal to 1/2).

To answer the first question, *what makes trained assessors' descriptions different from novices' descriptions?* we separated the terms presented in Table 1 into: common terms (i.e., used by both trained assessors and novices), terms used only by trained assessors, and terms used only by novices. We then analyzed the nature of the terms in each group by categorizing them as (1) specific terms learned by the trained assessors, (2) specific terms not learned by the trained assessors, (3) global terms (e.g., fruity, floral), (4) hedonic terms, and (5) intensity terms. The results are shown Fig. 3.

Several points can be noted from this figure. First, a large number of terms are common to trained assessors and novices for both learned beers (41% and 38.5% for commercial and supplemented beers respectively) and to a lesser extent for new beers (28.5%). These common terms include both specific terms learned by the trained assessors (alcohol, bitter, hop, malt, metallic, persistent, sparkling, and sweet) and more general terms including categorical terms (fruity), intensity (light, odorless, soft, strong, and tasteless) and hedonic terms (pleasant, unpleasant) that were not learned by the trained assessors. For supplemented beers (learned and new) the efficiency of these common terms is roughly equivalent for trained assessors and novices. For the commercial beers, the common terms appear to be efficient only for the trained assessors. Second, for all beers, the number of terms used only by the trained assessors is larger than the number of terms used only by the novices (39% vs. 20%; 42% vs. 19.5%; 43% vs. 28.5% for learned supplemented, commercial, and new supplemented beers, respectively). The trained assessors-specific terms consist essentially of learned terms (e.g., astringent, almond, banana, buttery, candy, caramel, cabbage, wax, cheese, lilac) whereas novice-specific terms are mainly global terms (lager, classical, refreshing, aromatized, special) and intensity terms (strong taste, strong smell, no taste). Among the terms used only by the trained assessors, about 30% are efficient for trained assessors and 15% are efficient for novices. In contrast, none of the terms used only by the novices led to successful matches. Finally, on the whole, common terms tend to be more efficient than terms specific to the trained assessors or the novices. Among these common terms, however, specific terms learned by trained assessors (e.g., bitter, sparkling, persistent, sweet) are more efficient than global terms for both trained assessors and novices. We can nevertheless note that some terms such as "unpleasant" or "light" can lead trained assessors and novices to successful matches.

In summary, the main difference between trained assessors' and novices' descriptions is the degree of specificity of the terms. In addition to the set of common words including both specific and global terms, novices use only very general terms whereas trained assessors used only specific terms. The greater efficiency of these terms for trained assessors compared to novices explains the overall trained assessors' advantage in the matching task.

To answer the second question, *are trained assessors using different terms to describe learned and new beers*? we calculated the number of words used by the trained assessors to describe: (1) the three types of beer, (2) two types of beer, and (3) only one type of beer. Our hypothesis was that trained assessors would be able to adapt their descriptions to the type of beer and thus that they should use more words specific to a given type of beer that novices. Fig. 4 shows that this hypothesis is far from being supported by the data. Indeed, the proportion of words common to all types of beer is somewhat larger for trained assessors than for novices (54% vs. 48%) whereas the proportion of words used specifically for a given type of beer is slightly smaller for trained assessors than for novices (8% vs. 10% on the average). Moreover, the terms used specifically by trained assessors to describe the learned supplemented beers do not correspond to the aromas we added in these beers

Table 1
Terms generated by trained assessors and novices for commercial, learned supplemented and new supplemented beers. The 2nd, 3rd, 5th, 6th, 8th, and 9th columns indicate the efficiency frequency (between 0 and 1). The expected terms are marked in bold and the learned terms are marked in italic. The stars show the binomial law significance level $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Learned commercial beers			Learned supplemented beers			New supplemented beers		
Terms	Experts	Novices	Terms	Experts	Novices	Terms	Experts	Novices
Trained assessors								
Bitter	0.68**	0.60	*Bitter*	0.70**	0.70**	*Bitter*	0.79***	0.64*
Pleasant	0.82	0.54	Pleasant	0.80	0.83	Light	0.62	0.60
Persistent	0.79***	0.57	Unpleasant	0.92**	0.50	*Persistent*	0.79**	0.83**
Light	0.83*	0.71	Light	0.71	0.47	*Sparkling*	0.67*	0.67*
Smooth	0.78	0.64	Odorless	0.57	0.78	Unpleasant	1.00**	0.54
Unpleasant	0.28	0.40	Smooth	0.56	0.67	Smooth	0.75*	0.72*
Sparkling	0.72**	0.57	*Sparkling*	0.62	0.69**	Odorless	0.50	0.55
Alcohol	0.66	0.59	*Persistent*	0.75**	0.66	Fruity	0.78	0.40
Insipid	0.40	0.55	*Alcohol*	0.63	0.70*	**Vanilla**	1.00**	0.91*
Fruity	1.00	0.50	*Hop*	0.70	0.71	Insipid	0.78	0.64
Malty	0.63	0.65	Insipid	0.71	0.50	*Alcohol*	0.71	0.65
Strong	0.60	0.50	Fruity	0.60	0.67	*Sweet*	0.76**	0.76**
Hop	0.77**	0.59	***Metallic***	0.80**	0.76**	Floral	0.79*	0.70
Sweet	0.67	0.56	Strong	1.00*	1.00*	*Malty*	0.67	0.63
Astringent	0.74*	0.60	***Bread***	0.80*	0.80**	*Honey*	0.69	0.81*
Banana	0.67	0.61	***Caramel***	0.75*	0.46	*Astringent*	0.67	0.68
Floral	0.62	0.73	*Sweet*	0.70	0.58	*Metallic*	0.86*	0.67
Metallic	0.77	0.53	*Astringent*	0.70	0.76**	*Hop*	0.72	0.59
Caramel	0.67	0.42	***Lilac***	0.69	0.44	Candy	1.00*	0.78
Bread	0.67	0.33	Floral	0.67	0.50	*Cardboard*	0.67	0.47
Sulfite	0.56	0.50	***Honey***	0.55	0.77	*Banana*	0.80	0.38
Cheese	0.86	0.25	***Banana***	0.60	0.62	*Bread*	0.60	0.67
Cabbage	1.00	0.88	*Malty*	0.56	0.74	*Lilac*	1.00**	0.75
Butter	0.75	0.00	*Cardboard*	0.00	0.17	*Caramel*	1.00	0.83
Musty	0.80	0.33	Flat	0.25	1.00	**Wax**	1.00	0.67
			Musty	0.50	0.75	Pharmacy	0.50	0.60
			Sulfite	0.78	0.64	*Phenol*	0.71	0.80
						Almond	0.40	0.71
						Strong	0.57	0.89*
						Cheese	1.00	0.75
Terms	Novices	Experts	Terms	Novices	Experts	Terms	Novices	Experts
Novices								
Bitter	0.62	0.51	*Bitter*	0.55	0.55	*Bitter*	0.59	0.53
Pleasant	0.59	0.56	Pleasant	0.55	0.57	Light	0.58	0.61
Persistent	0.59	0.42	Unpleasant	0.59	0.59	*Persistent*	0.50	0.64
Light	0.56	0.53	Light	0.69	0.62	*Sparkling*	0.67	0.63
Smooth	0.59	0.54	Odorless	0.44	0.28	Unpleasant	0.56	0.65
Unpleasant	0.63	0.50	Smooth	0.36	0.50	Smooth	0.83*	0.80
Sparkling	0.58	0.57	*Sparkling*	0.71*	0.56	Odorless	0.55	0.31
Alcohol	0.31	0.50	*Persistent*	0.65	0.57	Fruity	0.67	0.63
Insipid	0.67	0.50	*Alcohol*	0.60	0.46	**Vanilla**	0.83	1.00
Fruity	0.13	0.86	*Hop*	0.50	0.33	Insipid	0.67	0.80
Malty	0.43	0.57	Insipid	0.60	0.38	*Alcohol*	0.55	0.89*
Strong	0.75	0.63	Fruity	0.29	0.60	*Sweet*	0.83	0.86
Odorless	0.83	0.63	***Metallic***	1.00	1.00	Pleasant	0.56	0.60
Special odor	0.60	0.43	Strong	0.14	0.60	Strong odor	0.75	0.78
Strong taste	0.63	0.33	Strong odor	0.56	0.64	Special odor	0.56	0.56
Pungent	0.57	0.54	Special odor	0.71	0.33	Tasteless	0.64	0.39
Classical	0.43	0.36	Strong taste	0.54	0.60	Special taste	0.83	0.60
Lager	0.74	0.44	Tasteless	0.27	0.11	Strong taste	0.27	0.43
			Transient	0.57	0.50	Transient	0.70	0.63
			Pungent	0.38	0.67	Sour	0.33	0.22
			Aromatic	0.75	0.75	Pungent	0.73	0.50
						Fresh	0.75	0.38
						Aromatic	0.75	0.40
						Lager	0.36	0.54

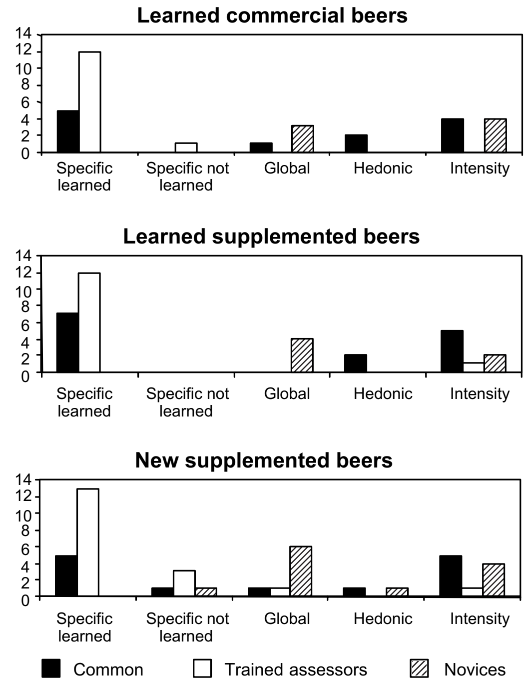

Fig. 3. Percentage of common, trained assessor specific, novice specific terms for the three types of beers.

(honey, banana, caramel, metallic, bread, and lilac). In fact, trained assessors did use these expected terms but they used them to describe the three types of beers. Notice, however, in Fig. 5, that four out of six expected terms (caramel, metallic, bread, and lilac) were more frequently and more efficiently used for the learned supplemented beers than for the other two types of beer. The two other terms are more often and more efficiently used for commercial (banana) or new supplemented (honey) beers. As a comparison, we can note that novices used only one out of six expected terms (metallic). Finally the number of terms used specifically by trained assessors to describe the new supplemented beers is somewhat larger than the number of terms used to de-

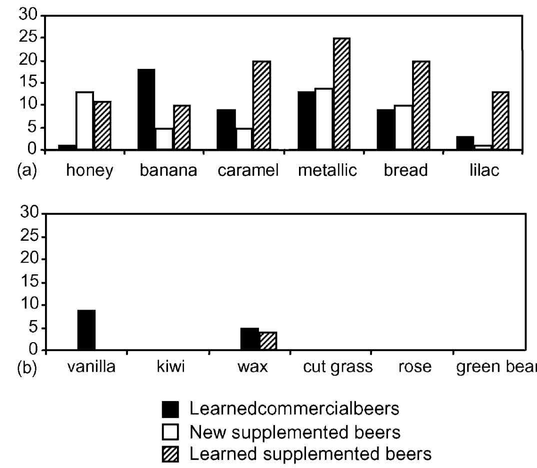

Fig. 5. Distribution of the learned (a) and new (b) expected terms as a function of the type of beers.

scribe learned beers. However, among the six expected terms (vanilla, kiwi, wax, cutgrass, rose, and raw green bean) only two were used (vanilla and wax). The other terms used to describe these beers denote the presence of an unexpected taste for a beer (phenol, pharmacy, and almond) but do not correspond precisely to any of the added aromas.

In summary, trained assessors did not use different terms to describe learned and unlearned beers. Rather they tended to apply the aroma descriptors they have learned during training to all beers even if these terms were not a priori appropriate. This result suggests that while trained assessors are able to provide efficient descriptions for new beers, they are not able to identify new aromas in beer. A similar difficulty to transfer identification abilities to new odorants was reported by Bende and Nordin (1997), who found that while wine experts outperform novices in a familiar odor identification task they are not better than novices at identifying unfamiliar odors. The remaining question is why a

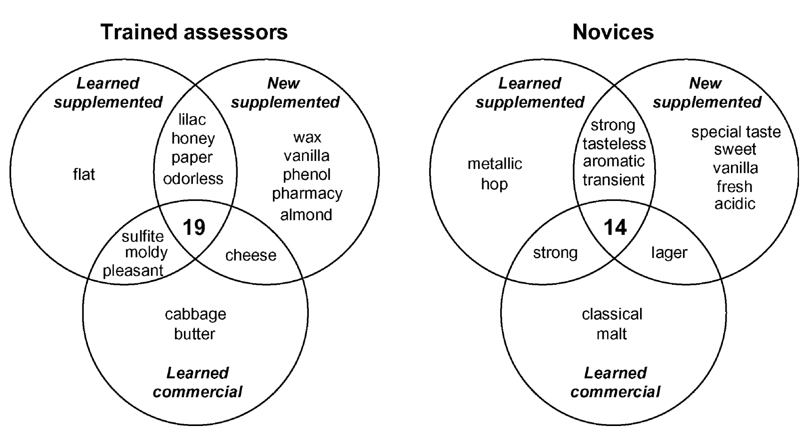

Fig. 4. Distribution of the terms used by trained assessors (left panel) and novices (right panel) to describe the different types of beers. For both groups of subjects a large number of terms is common to all types of beers.

priori not appropriate terms are efficient? Several explanations can be proposed. *First*, it is possible that commercial beers contain these aromas naturally. *Second*, it is possible that trained assessors confuse some aromas added in beer, and are only able to tell whether an aroma has been added or not. *Third*, it is also possible that trained assessors use the learned terms efficiently to describe new supplemented beers because a new aroma is similar to a learned aroma. For example, the aroma "wax" is close to the aroma "honey," thus during the matching session, the term "honey" might be sufficient to lead the rater to decide that the description containing this word corresponds to the "wax" beer.

4. General discussion

The main objective of this work was to evaluate if trained assessors are able to transfer their perceptual and verbal knowledge to new material. We found that beer trained assessors outperformed novices in a discrimination task when this task was performed on learned beers but not when it was performed with new beers. In contrast the same trained assessors outperformed novices on a matching task when the task was performed on both learned and new beers. Thus it seems that our trained assessors were able to transfer verbal knowledge but not perceptual knowledge. This interpretation is somewhat in disagreement with previous results by Bende and Nordin (1997). These authors found that wine experts were able to generalize their learning to unfamiliar odors in a discrimination task but not in a detection and an identification task. However, the discrimination task they used was quite different from ours. In their experiment, subjects were presented with pairs of binary mixtures of eugenol and citral and their task was to choose the mixture with an odor of lemon. So their attention was focused on a specific characteristic of the mixtures they smelled. In our case, subjects did not know which characteristics to look for. Moreover, even though wine experts experience only occasionally the odor of clove and lemon in their profession, these two odors are familiar odors, whereas the new beers we used were completely unfamiliar. Thus it is possible that Bende and Nordin's discrimination task reveals a difference in the familiarity with the task (experts are used to looking for a familiar odor in a wine) not a difference in sensitivity. Indeed, they failed to show any significant superiority of experts in an absolute detection task.

Another difference between Bende and Nordin's study and ours is that while their subjects can probably be considered as experts according to the ASTM standard E-253, our subjects are only trained assessors, and as such have less domain-specific experience. The difficulty encountered by our trained assessors to transfer their perceptual ability to new beers is reminiscent of some data reported by Davies, Marley, Ozgen, and Sowden (1996) in the color vision domain. These authors showed that subjects trained to discriminate between color chips were better and faster at discriminating the chips on which they were trained but not new chips. A plausible explanation for this absence of perceptual transfer from one set of stimuli to another one can be cast in terms of perceptual learning. This concept first proposed by Gibson (1969), refers to "an increase in the ability to extract information from the environment, as a result of experience and practice with stimulation coming from it" (p. 3). According to Gibson's view, the perceptual system becomes attuned to its environment and so learns to select and encode the features that are useful to discriminate within a set of stimuli. These features constitute an optimal representation of the set of stimuli from which they have been extracted, but are not useful to distinguish between other stimuli. For example long-term repeated exposure to faces of a given race allows the perceptual system to extract features that enables to accurately distinguish among faces from this race. Faces from other races, however, are not well characterized by these highly specialized features and so we are less accurate at recognizing such faces (O'Toole, Abdi, Deffenbacher, & Valentin, 1995). This well-known phenomenon, referred to as the "other-race effect" in the face processing literature, explains why people are better at recognizing faces from their own race than faces from another race. Likewise, we can speculate that our trained assessors suffered from an "other-beer effect." During training they have been repeatedly exposed to a set of beers and thus extracted a set of features optimal to distinguish between these beers. This optimal set of features allows them to outperform novices on the beers from which they have been extracted but not on new beers. As already mentioned, the verbal labels associated with these optimal features might also have played a role in the trained assessors superiority for the learned beers by either helping to memorize the beers during the discrimination task or helping to focus on the salient information.

The trained assessors limitations observed in the discrimination task were not paralleled in the matching task. The aromas added in this second task were well above threshold and thus the problem subjects were faced with was more of a cognitive nature than of a perceptual one. We may thus think that the trained assessors superiority in that case results from the use of a better strategy than the strategy used by novices. Because trained assessors used the same terms to describe learned and new beers, it seems that when asked to describe a beer, they start by activating the list of terms they have learned during training and then try to apply them to the beers they have to describe. In contrast, novices seem to start by tasting the beers and then generate a list of terms from their sensations. Recent work by Bell and Paton (2000) on odor identification showed that the first strategy ("think and sniff") leads to better odor identification than the second strategy ("sniff and think"). And in agreement with this hypothesis, we found that, indeed, novices improve their performance when provided with trained assessors' descriptions in the matching task. In other words, novices are not able to identify their sensations in a free description task but are able to identify them in a forced choice situation. Trained assessors in the other hand seem to transform the free description task into a forced choice situation. Why does this strategy allow them to describe and match new beers efficiently? In fact our results showed that trained assessors were not able to describe the new beers with the expected labels but rather try to describe their sensations using the terms they have learned during training. This interpretation can be put in perspective with Lawless (1985) and Gawel (1997) suggestion that wine experts store configurations of wine properties in memory and then use these properties to describe wines. When asked to describe a particular wine, let say a red Burgundy wine, a wine expert would start by activating in memory the characteristics of this type of wine and then look for these characteristics. Likewise in a blind tasting situation, the detection of a specific characteristic would trigger the search for characteristics that are generally correlated with this specific characteristic. In agreement with this suggestion, Pangborn, Berg, and Hansen (1963) showed that wine experts, but not novices, judged a solution sweeter after this solution had been colored so as to resemble a rosé/blush wine. More recently, Morrot, Brochet, and Dubourdieu (2001) reported that coloring a white wine in red led wine experts to describe its flavor as that of a red wine. In support to these empirical results, Hughson and Boakes (2002) showed that wine experts recalled better actual wine descriptions (which matched their expectation and knowledge) than random descriptions whereas novices recalled equally well the two types of descriptions. Thus it seems that when sensory experts have to describe some new stimuli, they rely on some kind of semantic networks, or prototypes (Brochet & Dubourdieu, 2001), to infer which characteristics they should be looking for. In support to this idea a recent fMRI study showed that while both sommeliers and novices' scans shows strong activity in the amygdala (an area involved in the processing of emotions), only sommeliers displayed a burst of activity in parts of their frontal cortices (Amadio, Hagberg, & Sturiolo, 2003).

5. Conclusion

During training, sensory experts acquire two types of expertise: a perceptual expertise and a cognitive expertise. Our study showed that the cognitive expertise can generalize to new material but that the perceptual expertise does not. Why is that? We may speculate that the cognitive ability involved in chemico-sensorial perception relies on the same mechanisms as other types of expertise such as chess or mathematical expertise (but see Abdi, 2002). Because these mechanisms are common to several domains it would make sense that they should be flexible. In contrast, perceptual expertise is domain specific—learning music does not help discriminating between odors—thus the need for flexibility might be less crucial. One can wonder, however, if a longer period of training might not lead to some perceptual generalization abilities.

We also found that the sensory superiority of the trained assessors, compared to the novices, was moderate. Is this effect modality specific or more general? It is difficult to generalize at this point, but this effect seems to surface in other domains of expertise. For example, a recent series of papers, (Bigand, Madurell, Tillmann, & Pineau, 1999, 2001; Tillmann & Bigand, 1998) describes how musical experts performed at the same level (or almost) as novices in several musical tasks (but see also Dowling, 1986, for a diverging point of view). These recent results, along with the results reported here, suggest that the effects of expertise need to be evaluated in relation with the length of training, the domain of expertise as well as the tasks performed by the participants.

Acknowledgements

This work was supported by the Association Nationale de la Recherche Technique, Qualtech, Kronenbourg, and Danone.

Thanks are due to W. Jay Dowling, and two anonymous reviewers for comments on a previous version of this paper.

References

Abdi, H. (2002). What can cognitive psychology and sensory evaluation learn from each other? *Food Quality and Preference, 13*, 445–451.

Amadio, L., Hagberg, G., & Sturiolo, A. (2003). I sensi diVini—Il segreto del cervello nella degustazione. In *37th meeting of the Italian Wine Academy od Roma*, Sienna, Italia.

Bigand, E., Madurell, F., Tillmann, B., & Pineau, M. (1999). Effect global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 184–197.

Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition, 81*, 11–20.

Bell, G. A., & Paton, J. E. (2000). Verbal-cognitive strategy and influence odor judgment. *The Aroma-Chology Review, 9*, 3–9.

Bende, M., & Nordin, S. (1997). Perceptual learning in olfaction: Professional wine tasters versus controls. *Physiology & Behavior, 62*, 1065–1070.

Brochet, F., & Dubourdieu, D. (2001). Wine descriptive language supports cognitive specificity of chemical senses. *Brain Language, 77*, 187–196.

Cardello, A. V., Maller, O., Kapsallis, J. G., Segars, R. A., Sawyer, F. M., Murphy, C., & Moskowitz, H. R. (1982). Perception of texture by trained and consumer panels. *Journal of Food Science, 47*, 1186–1197.

Cain, W. S. (1979). To know with the nose: Keys to odor identification. *Science, 203*, 467–470.

Chambers, E., IV, & Smith, E. A. (1993). Effects of testing experience on performance of trained sensory panelists. *Journal of Sensory Studies, 8*, 155–166.

Chollet, S., Valentin, D., Meyer, A., Leboeuf, L., Rogeaux, M., Flayeux, R., & Sauvageot, F. (2002). Comparaison entre deux méthodes d'entraînement à l'identification de composés dans la bière: «groupe» contre «individuel». *Sciences de l'Aliment, 22*, 401–424.

Clapperton, J. F., & Piggott, J. R. (1979). Flavor characterization by trained and untrained assessors. *Journal of Institute of Brewing, 85*, 275–277.

Davies, I., Marley, L., Ozgen, E., & Sowden, P. (1996). Perceptual learning and same-different judgment of color. In *19th European Conference on Visual Perception*. Strasbourg, France.

Desort, J. A., & Beauchamps, G. K. (1974). The human capacity to transmit olfactory information. *Perception & Psychophysic, 16*(3), 551–556.

Dowling, J. (1986). Context effects on melody recognition: Scale-steps versus interval representation. *Music Perception, 3*, 281–296.

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

Gibson, E. J. (1969). *Principles of perceptual learning and development*. Englewood Cliffs, NJ: Prentice Hall.

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

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

Lawless, H. T. (1985). Psychological perspectives on winetasting and recognition of volatile flavors. In G. G. Birch (Ed.), *Alcoholic beverages* (pp. 97–113). London: Elsevier Applied Science Publishers.

Lawless, H. T. (1988). Odor description and odor classification revisited. In D. M. H. Thomson (Ed.), *Food acceptability* (pp. 27–40). London: Elsevier Applied Science Publishers.

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

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

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., & Valentin, D. (1995). A perceptual learning theory of the information in faces. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition* (pp. 159–182). London: Routledge.

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

Parr, W. V., Heatherbell, D., & White, K. G. (2002). Demystifying wine expertise: Olfactory threshold, perceptual skill and semantic memory in expert and novice wine judges. *Chemical Senses, 27*, 747–755.

Peron, R. M., & Allen, G. L. (1988). Attempts to train novices for beer flavor discrimination: A matter of taste. *The Journal of General Psychology, 115*, 403–418.

Rabin, M. D. (1988). Experience facilitates olfactory quality discrimination. *Perception & Psychophysic, 44*, 532–540.

Roberts, A. K., & Vickers, Z. M. (1994). A comparison of trained and untrained judges' evaluation of sensory attribute intensities and liking of cheddar cheeses. *Journal of Sensory Studies, 9*, 1–20.

Sauvageot, F., & Fuentes, P. (2000). Une approche pour valider la technique du profil sensoriel: la technique de l'appariement. *Sciences de l'Aliment, 20*, 467–489.

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

Tillmann, B., & Bigand, M. (1998). Influence of global structure on musical target detection and recognition. *International Journal of Psychology, 33*, 107–122.

Wolters, C. J., & Allchurch, E. M. (1994). Effect of training procedure on the performance of descriptive panels. *Food Quality and Preference, 5*, 203–214.