Self-ratings of olfactory performance and odor annoyance are associated with the affective impact of odor, but not with smell test results

Short title: Subjective vs. objective olfactory function

Antti Knaapila[1,*], Auri Raittola[1], Mari Sandell[2] and Baoru Yang[1]

[1] Department of Biochemistry, University of Turku, Finland
[2] Functional Foods Forum, University of Turku, Finland

Corresponding author:

Antti Knaapila, Department of Biochemistry, University of Turku, Vatselankatu 2, FI-20500 Turku, Finland.
Email: anjukn@utu.fi

* Present address (2019): Department of Food and Nutrition, University of Helsinki, P.O. Box 66 (Agnes Sjöbergin katu 2), FI-00014 University of Helsinki

Abstract

Our aim was to explore factors potentially associated with subjective (self-rated) and objective (measured using the Sniffin' Sticks Extended test) olfactory performance in the general population without olfactory disorders. We studied associations between olfactory performance and how important odors were in determining liking for new places, things and people (measured using the Affective Impact of Odor scale, AIO) and the average annoyance caused by odors in 117 adults (83 women, 34 men; age 18–69 years, mean age 32 years). In a subset of 44 participants we also studied associations between olfactory performance and spice odor identification task scores (14 odors) and the number of herbs and spices consumed. Self-rated olfactory acuity and odor-related annoyance were associated with the AIO scores, but neither correlated with the smell test results. Instead, the number of spices consumed correlated with spice odor identification score ($r = 0.50$) and the identification (but not threshold nor discrimination) sub-score of the Sniffin' Sticks test ($r = 0.49$). Our results suggest that a tendency to perceive odors in affective terms may be associated with overestimation of olfactory abilities and that recurrent exposure to a large variety of spice odors may improve performance on odor identification.

Keywords: Individual differences, olfaction, olfactory function, Sniffin' Sticks, smell, subjective olfaction

1 Introduction

Self-ratings of olfactory ability (olfactory performance/function/acuity/capacity) are rarely concordant with the results of subsequent smell tests. It has been reported that self-evaluations underestimate the prevalence of olfactory impairment as determined by smell tests (Murphy et al., 2002). This suggests that an individual's olfactory capacity may become impaired without the individual concerned noticing the decline. On average, however, people who have lost their sense of smell (anosmics), or who have significantly decreased olfactory acuity (hyposmics), rate their olfactory abilities lower than do people with an intact sense of smell (normosmics) (Kollndorfer et al., 2015; Murphy et al., 2002; Welge-Luessen, Hummel, Stojan, & Wolfensberger, 2005). This implies that at least some of the people who suffer from significant olfactory impairment are aware of it. In contrast, individuals who have not experienced a dramatic decline in olfactory function appear to be rather unaware of the relative level of their olfactory abilities (i.e. whether their olfactory acuity is average or exceptional). This discrepancy was demonstrated by Nguyen, Nguyen-Thi, and Jankowski (2012) who showed that in patients with olfactory deficits self-rated and measured olfactory function were correlated, whereas in normosmics self-ratings were not reliable.

Several other studies have reported similar results, confirming that among healthy, normosmic adults self-rated olfactory ability is not correlated with the results of subsequent olfactory testing using various methods. Landis, Hummel, Hugentobler, Giger, and Lacroix (2003) used a validated smell test battery, the Sniffin' Sticks Extended test (Hummel, Sekinger, Wolf, Pauli, & Kobal, 1997), Philpott, Wolstenholme, Goodenough, and Murty (2006) used olfactory threshold testing with a standard odorant (phenyl ethyl alcohol or eucalyptol) as the stimulus, and Knaapila et al. (2008) used an odor identification task based on tailored scratch-and-sniff labels.

The lack of correlation between subjective and objective assessment of olfactory function in normosmics is not well understood. Only a few studies have focused on the question. Recently, however, Kollndorfer et al. (2015) reported that vividness of olfactory imagery (measured using the vividness of olfactory imagery questionnaire, VOIQ) predicted self-evaluated olfactory function in hyposmics and healthy controls, but not in anosmics. Landis et al. (2003) discovered that although olfactory acuity as measured using the Sniffin' Sticks Extended test was not correlated with pre-test self-ratings of olfactory function it was correlated with post-test self-ratings ($r = 0.48$, $p = 0.021$).

Knaapila et al. (2008) observed that self-rated olfactory function was moderately correlated ($r = 0.30$) with the average annoyance caused by everyday odors. Odor-related annoyance is obviously related to negative valence (unpleasantness) of odors. Similarly, Knaapila and Tuorila (2014) noted that participants who regarded their sense of smell as exceptionally good (self-reported hyperosmics) reported more negative odor-related experiences than sex- and age-matched normosmic controls. If sensitivity to the affective consequences (annoyance) of unpleasant odors is associated with higher self-ratings of olfactory function, what about pleasant odors? Both pleasant and unpleasant odors can potentially influence individuals' affective response to various objects.

Wrzesniewski, McCauley, and Rozin (1999) developed an eight-item scale, the Affective Impact of Odor (AIO), for measuring the impact of odor on liking for new places, foods, cosmetics, and people. We used the AIO scale in this study and refer to the importance of odor in determining liking for new things as affective impact of odor. People who regard novel objects' odor as an important reason for liking or disliking them (as indicated by a high AIO score) may think that this indicates that they have a particularly acute sense of smell. On his basis, we hypothesized that self-rated olfactory function is associated with the affective impact of odor on the individual concerned (AIO score). If this is the case, then increased sensitivity to the affective impact of odors may contribute to overestimation of one's olfactory abilities.

Odor identification tasks are frequently employed in standard smell test batteries that are used to detect olfactory impairments (ie, anosmia and hyposmia) and they have proved useful in this context. However, odor identification tests designed for screening for olfactory deficits are not necessarily suitable for detecting differences in acuity in the normosmic population (eg, between average and exceptionally high olfactory function), due to ceiling effects. We could reduce ceiling effects by making the identification task more challenging, for example, by adding a larger number and more closely related response options to multiple-choice tasks, which often use four rather dissimilar options per odor (the correct answer and three distractors). However, this may give raise to another problem. The results of odor identification tests depend not only on sensory factors but also on cognitive factors, such as experience (familiarity) with the olfactory stimuli presented in the tests. The average level of familiarity with a given odor often varies from culture to culture (eg, Chrea et al., 2004; Distel et al., 1999), and even within a culture there is variation in odor familiarity from person to person. Some odors, for example odors of a specific food category, may be more familiar to frequent consumers of that food category than to others. We therefore hypothesized that people who habitually consume a larger variety of herbs and spices would score better on a spice odor identification task than people who do not use as many different spices. In formulating this hypothesis we assumed that consumption of a spice could be used as a proxy for familiarity with and exposure to its odor.

The overall objective of the study was to reveal whether selected factors were associated with subjective (self-rated) and objective (measured using the Sniffin' Sticks Extended test) olfactory ability in the general population without olfactory disorders. More specifically, in the first part of the study our aim was to test the hypothesis that self-rated olfactory ability is associated with affective impact of odor (the AIO score). In the second part of the study we tested our hypothesis that the range of spices consumed habitually would be associated with performance on odor identification tasks. The results could be used in future research into the reasons for the lack of correlation between subjective and objective assessments of olfactory performance in normosmics.

2 Methods

2.1 Participants and procedure

The study was split into two parts. Each part involved one visit to a sensory laboratory (designed according to the ISO 8589 standard). At each visit different psychophysical olfactory task (responding to olfactory stimuli) and questionnaires were completed.

In the first part of the study we measured the olfactory performance of 117 individuals (**Table 1**) using the Sniffin' Sticks Extended smell test (Hummel et al., 1997). Participants also completed questionnaires, as detailed below. In the second part of the study a subset of the original sample (*n* = 44, ~38%) (**Table 1**) completed a spice odor identification task and a questionnaire on their consumption of spices (detailed below).

Table 1. Characteristics of the participants and major tasks of the two parts of the study.

	Part I (N = 117)	**Part II (N = 44)**[a]
Gender		
- Women	83 (70.9%)	34 (77.3%)
- Men	34 (29.1%)	10 (22.7%)
Age (years)		
- Range	18–69	18–69
- Mean (SD)	32.0 (10.7)	33.9 (12.5)
Smoking		
- No	102 (87.2%)	42 (95.5%)
- Yes, but not daily	8 (6.8%)	1 (2.3%)
- Yes, daily	7 (6.0%)	1 (2.3%)
Major tasks of the visit[b]		
- Olfactory	Sniffin' Sticks Extended test (Hummel et al., 1997)	Odor identification task for 14 spice odors
- Questionnaire	Affective Impact of Odor (Wrzesniewski et al., 1999)	Consumption of 38 herbs and spices[c]

[a] All participants of the second part of the study had previously completed the first part. Gender distribution and mean age of the participants who completed the both parts did not differ ($p > 0.05$) from those of the participants who completed only the first part ($n = 73$).

[b] In each part of the study a participant visited the research unit once.

[c] Listed in Table 2. Included the 14 spices used as olfactory stimuli in the odor identification task.

Data for the first part of the study were collected between November 2014 and February 2016. Data for the second part were collected in April and May 2016, that is, 2 to 18 months after the first visit.

The participants for the first part were recruited from the population of students and personnel of the University of Turku (via email messages and flyers on bulletin boards) and from the general population (via advertisements in the print and online versions of a free local newspaper and via distribution of flyers in public spaces). The minimum age for participants was 18 years but no other selection criteria were used. The participants for the second part were recruited from the sample used in the first part, no new participants were recruited.

We did not request information about participants' ethnic or cultural background. However, all participants were Finnish residents and we estimated that ~80% of them were of Finnish descent. Nevertheless, in both parts of the study we provided the instructions and questions in English as well as Finnish.

We obtained written, informed consent to participation from all participants. Participation was voluntary and we followed the ethical guidelines of the Declaration of Helsinki. We compensated each participant for his or her contribution with a box of xylitol drops, chocolate bar, or other similar food product.

2.2 Olfactory testing

Sniffin' Sticks Extended test. During the first visit participants' olfactory function was assessed using the Sniffin' Sticks Extended test (Burghart Messtechnik GmbH, Wedel, Germany). The test has been described in detail elsewhere (Hummel et al., 1997). Briefly, the test consists of three parts (sub-tests): (1) measurement of olfactory threshold (detection threshold), (2) odor discrimination task, and (3) odor identification task. The threshold test is available for phenyl ethyl alcohol (PEA, 2-phenylethanol; a rose-like odor) and *n*-butanol. We used the PEA version in this study. The detection threshold is determined using a staircase procedure and scores range from 1 to 16, with higher scores indicating detection at a higher dilution (at lower concentration), that is, a lower detection threshold and better performance. The discrimination sub-test comprises 16 sets of three olfactory stimuli and the participant's task is to indicate the odd one out in each set. The score on this sub-test is the number of odors discriminated correctly (range: 0 to 16). The identification sub-test comprises of 16 target odor stimuli, and the participant's task is to identify each target odor from a set of four different odor names (chance success rate = 1/4). The score on this sub-test is the number of odors identified correctly. Total scores on the Sniffin' Sticks Extended test are referred to TDI scores (the acronym is derived from the initials of the sub-tests). The TDI score is the sum of scores on the three sub-tests and thus ranges from 1 to 48. Normative data based on over 3000 individuals are available for the Sniffin' Sticks Extended smell test (Hummel, Kobal, Gudziol, & Mackay-Sim, 2007). Participants were tested individually by a test administrator.

Spice odor identification task. During the second research visit participants completed a spice odor identification task designed especially for this study. The task stimuli were 15 samples: 14 spice odors and a blank (odorless). The samples and the order in which they were presented were the same for all participants: (1) *oregano*, (2) *anise*, (3) *rosemary*, (4) *cardamom*, (5) *mint*, (6) *caraway*, (7) *sage*, (8) *blank*, (9) *thyme*, (10) *cinnamon*, (11) *fennel*, (12) *white pepper*, (13) *marjoram*, (14) *garlic*, and (15) *clove*. The olfactory stimuli were commercial preparations of dried, ground spices. The brand used for white pepper stimulus was Santa Maria (produced by Santa Maria Ltd, Mölndal, Sweden); all other spices were from the Salliselta brand (produced by Mauste Sallinen Ltd, Naantali, Finland). Stimuli consisted of a tea-spoonful of spice placed in a 30 ml glass bottle that was capped and wrapped in foil to mask the content. The samples were identified by three-digit random codes for the purposes of the test.

The participants were asked to take the samples, one at a time in the given order, open the cap, sniff, and then choose the most appropriate response option. The same 17 response options were given for all samples (chance success rate = 1/17): the 'no odor' option, the 14 spices used as target stimuli (see above) and two distractors (ginger and chervil); they were listed in alphabetical order. A forced choice procedure was used, that is, participants had to select a single response for each target; there was no 'don't know' option. Having received instructions from the test administrator, each participant provided responses independently at a booth in the sensory laboratory via a computer-aided data collection system using Compusense five software (Compusense Inc., Guelph, Ontario, Canada).

2.3 Questionnaires

The first part of the study. At the start of their first visit to the sensory laboratory (before the Sniffin' Sticks Extended test was administered) participants completed a questionnaire asking about gender, year of birth, and how the participant rated her or his sense of smell (olfactory function) on 0 to 10 scale where 0 was labelled ('My sense of smell is --') 'completely lost'; 1, 'very poor'; 2, 'poor'; 3, 'rather poor'; 4 'moderate'; 5, 'average'; 6, 'rather good'; 7, 'good'; 8, 'very good'; 9, 'exceptionally good'; 10, 'perfect'.

After performing the first part of the Sniffin' Sticks Extended test (the threshold test, which usually took longest) participants completed a second questionnaire. This allowed their olfactory system some recovery time. The questionnaire included questions about smoking habits and annoyance evoked by odors. Participants were asked to rate the average level of annoyance the odors in their standard living environment caused them using a 0 to 5 scale, where 0 indicated 'not at all disturbing'; 1, 'hardly disturbing'; 2, 'a little disturbing'; 3, 'somewhat disturbing'; 4 'fairly disturbing'; and 5, 'very disturbing'. There was also a 'don't know' option, but none of the participants used it.

The questionnaire also included a previously validated measure, the AIO (Wrzesniewski, McCauley, & Rozin, 1999). The AIO measures affective impact of odor, that is, how much odors influence the liking for four categories of *new* stimuli: (1) foods, (2) places, (3) cosmetics and health products, and (4) people. The impact of odor on liking (and not liking) for new stimuli in all four categories was measured with two items (sample items: *'When you like a new food, is it partly because you like the smell?'* and *'When you don't like a new food, is it partly because you don't like the smell?'*). Thus, the AIO comprises of eight items. The original AIO uses the following four response options: 'never', 'rarely', 'sometimes', and 'often', corresponding to the scores 0, 1, 2, and 3, respectively, and AIO score is calculated as the mean of score for the eight items. However, to avoid ceiling effects, we added two response options to the upper end of the scale: ‘almost always’ and ‘always’, corresponding to the score 4 and 5, respectively. This meant that AIO scores on our modified scale ranged from 0 to 5. Although this modification of the scale made it difficult to compare absolute scores with scores on the original scale, the results suggested that additional response options were needed. Depending on the item, 6.0–27.4% and 3.4–12.0% of the respondents chose the options ‘almost always’ and ‘always’, respectively. The internal consistency of the scale as measured by Cronbach’s alpha was 0.79.

The second part of the study. During the second study visit, participants completed a questionnaire asking about their familiarity with and consumption of 38 herbs and spices, listed in **Table 2**. For each herb or spice item, the participants had to choose the most appropriate of the following four options (modified from the ones used by Bäckström, Pirttilä-Backman, & Tuorila, 2004): (1) ‘I haven’t even heard of it’, (2) ‘I know it by name, but I haven’t eaten it’, (3) ‘I know I have eaten it, but I hardly ever use it’, or (4) ‘I use it occasionally or regularly’. The number of items for which a respondent selected the response ‘I use it occasionally or regularly’ was used as the number of consumed herbs and spices in further analyses.

2.4 Data analyses

Although all data were based on categorical scales we assumed that the underlying phenomena were continuous and at least approximately normally distributed and hence that use of parametric statistical tests (Pearson correlation, *t*-test, and ANOVA) was justified. However, to confirm the results, we re-ran the analyses using the corresponding non-parametric tests (Spearman correlation, Mann-Whitney *U*-test, and Kruskal-Wallis one-way ANOVA). The parametric and non-parametric tests yielded consistent findings (the criterion for statistical significance was at $\alpha = 0.05$) in all but one case, which is noted in the ‘Results’.

Data for some variables (such as TDI score, self-rated olfactory function, and olfactory annoyance) were collapsed into fewer categories (three) for use as fixed factors (independent variables) in ANOVA; further details are given below.

Table 2. The 38 herb and spice items involved in the questionnaire in the Part II of the study (in alphabetical order).

Allspice
Anise (aniseed)[a]
Basil
Black pepper
Caraway[a]
Cardamom[a]
Chervil
Chili
Chives
Cilantro (coriander)
Cinnamon[a]
Clove[a]
Curry
Dill
Fennel[a]
Garlic[a]
Ginger
Green pepper
Hyssop
Lavender
Lemon balm
Lovage
Marjoram[a]
Mint[a]
Oregano[a]
Parsley
Rose pepper
Rosemary[a]
Saffron
Sage[a]
Savory
Sour orange (peel)
Star anise(ed)
Tarragon
Thyme[a]
Turmeric
Vanilla
White pepper[a]

[a] Item was also used as an olfactory stimuli in the spice odor identification task.

3 Results

3.1. Olfactory tasks

Sniffin' Sticks Extended smell test. The mean TDI score was 35.40 (SD = 4.50; range: 25.00–43.75). None of our participants was anosmic (criterion for functional anosmia: TDI score < 16, Hummel et al., 2007). Fourteen of the participants (12.0%) could be regarded as hyposmic on the grounds that they scored below the age-adjusted criteria for hyposmia defined by Hummel et al. (2007): 30.3, 27.3, and 19.6 for individuals aged 16–35, 36–55, and > 55 years, respectively. We observed the expected trend towards a negative relationship between TDI score and age (r = -0.15, p = 0.11, N = 117). It is possible that the association was not significant because most of our participants were young adults (only five participants (4.3%) were over 55 years old).

Spice odor identification task. The mean number of correctly identified spice odors (excluding the blank) was 7.8 (SD = 2.2; range: 3–12). As with the TDI score there was a trend towards a negative association between score and age (r = -0.29, p = 0.06, n = 44).

3.2 Self-rated olfactory function

The mean self-rating for olfactory function was 5.7 (SD = 1.3; range: 2–9). The most popular responses were 5 and 6, given by 38 (32.5%) and 37 (31.6%) of the participants, respectively. Self-rated olfactory function was not correlated with the TDI score (r = 0.11, p = 0.23), nor with any of the sub-scores, that is, threshold (r = 0.09, p = 0.32), discrimination (r = 0.13, p = 0.18), or identification (r = -0.01, p = 0.94). Self-rated olfactory acuity also did not correlate with olfactory annoyance ratings (r = 0.14, p = 0.14). In addition, we detected no correlation between self-rated olfactory function and number of spice odors identified correctly (r = 0.04, p = 0.79) in the sub-sample of participants (n = 44) who also completed the spice odor identification task.

3.3 Affective impact of odor

The mean AIO score was 2.70 (SD = 0.72; range: 0.75–4.75). At this point we would like to remind readers that in the original AIO (Wrzesniewski et al., 1999) responses are given using a four-point scale (ranging from 0 to 3), whereas in this study a modified six-point scale (ranging from 0 to 5) including two additional responses at the upper end of the scale was used (see 'Methods' for details). AIO score was correlated with self-rated olfactory function (r = 0.33, p < 0.001) and with odor-related annoyance (r = 0.28, p = 0.002). In contrast, AIO score was not correlated with objective indicators of olfactory function TDI score (r = 0.07, p = 0.44) and spice odor identification score (r = -0.19, p = 0.21).

To examine further the relationships between AIO scores and self-rated and measured olfactory function, we used one-way ANOVA with Tukey's post hoc test to compare the AIO scores of participant groups according to TDI score, self-rated olfactory function, and odor-related annoyance. In order to create a reasonable number of adequately sized comparison groups, the data for each of these independent variables were collapsed into three groups.

The participants were assigned to groups as follows. First, they were classified into three similarly sized groups (ie, tertiles) according to their TDI score (TDI < 33.33, n = 41; TDI 33.33–38.25, n = 38; TDI > 38.25, n = 38). Next, participants were grouped according to their self-rated olfactory function (below average (0–4), n = 13; average (5), n = 38; above average (6–10), n = 66). Finally, participants were collapsed into groups who experienced their everyday olfactory environment as 'not at all' or 'hardly' annoying (0–1, n = 26), 'a little' or 'somewhat' annoying (2–3, n = 64), or 'rather' or 'very' annoying (4–5, n = 27).

Participants who rated their sense of smell as above average had higher AIO scores than those who rated their olfactory performance as average or below average. The latter two groups had similar AIO scores. A similar pattern was observed in AIO scores and odor-related annoyance. However, there was no association between AIO and TDI scores (**Figure 1**).

Three-way ANOVA with TDI category, self-rated olfactory function category, and odor annoyance category as fixed factors (independent variables) revealed no significant interactions between variables. A series of two-way ANOVAs also revealed no interactions between any two of the three factors. This suggests that the main effects of self-rated olfactory performance and odor annoyance on AIO score (observed in the one-way ANOVA) were independent.

3.4 Subjective vs. objective measures of olfaction: under- and overperformers

We wanted to examine not just the relationship between subjective and objective measures of olfactory function, but also the relationship between accuracy of people's self-assessments and other variables. We set the criteria for the accuracy of participant's self-assessments of olfactory performance as follows. About one-third of participants (n = 38, 32.5%), rated their sense of smell as 'average' (5 on the scale 0–10). If all these participants had rated their sense of smell accurately their TDI scores would all have fallen the middle tertile (33.33–38.25). Similarly, participants who rated their sense of smell as less than 5, should have TDI scores in the lowest tertile (< 33.33) and participants who rated their sense of smell as more than 5 should have TDI scores in the highest tertile (> 38.25). Participants who fulfilled these criteria were classified as 'realists'. Participants who overestimated their olfactory acuity were classified as 'underperformers' and those who underestimated their olfactory acuity were classified as 'overperformers'. The classifying criteria and distribution of participants are summarized in **Figure 2**.

The 'underperformers' had slightly higher AIO scores than the 'overperformers', although the effect was not significant (one-way ANOVA; $F_{2,114}$ = 2.02, p = 0.14) (**Figure 3a**). Concordance between subjective and objective assessments of olfactory performance was not associated with odor annoyance. In the sub-sample who were involved in the second part of the study (n = 44) 'underperformers' reported consumption of fewer spice items than 'overperformers' ($F_{2,41}$ = 4.02, p = 0.026) (**Figure 3b**). However, this result was not confirmed by the corresponding non-parametric test (Kruskal-Wallis one-way ANOVA; p = 0.106). Furthermore, all the accuracy groups had similar spice odor identification scores.

3.5 Relationship between smell test results and consumption of spices

The reported number of spices consumed occasionally or regularly (out of the 38 given, **Table 2**) was correlated with spice odor identification score ($r = 0.50$, $p = 0.001$, $n = 44$) and with identification sub-score of the Sniffin' Sticks Extended test ($r = 0.49$, $p = 0.001$, $n = 44$). In contrast, the number of spices consumed was not correlated with threshold sub-score ($r = 0.07$, $p = 0.63$, $n = 44$) or discrimination sub-score ($r = -0.03$, $p = 0.85$, $n = 44$) (**Figure 4**).

3.6 Comparison of the genders

Women evaluated everyday environmental odors as more annoying than did men (2.6 vs. 2.0; $t_{115} = 2.34$, $p = 0.021$), but there were no gender differences in any other studied variables (all *p*s > 0.05). Women nominally outperformed men in the spice odor identification task (7.9 vs. 7.5) and reported that they consumed a higher number of spice items (19.5 vs. 16.8) although neither of these effects reached statistical significance. There was no clear relationship between gender and TDI score, nor any of the TDI sub-scores.

4 Discussion

The starting point for our study was the often reported lack of correlation between self-rated (subjective) and measured (objective) olfactory performance in the healthy, normosmic general population (Landis et al., 2003; Philpott et al., 2006; Knaapila et al., 2008), a finding that was once again confirmed in this study. We asked why no such correlation exists, that is, why self-evaluations of olfactory function are inaccurate. We approached the question by investigating whether certain factors were associated with self-rated and measured olfactory function. We also compared participants who assessed their sense of smell accurately with those who either over- or underestimated their sense of smell.

Our hypotheses were based on the previous research, especially the findings that self-rated olfactory function was associated with vividness of odor imagery (Kollndorfer et al., 2015) and odor-evoked annoyance (Knaapila et al., 2008). We hypothesized that the degree to which odor influenced affective response to new stimuli (ie, affective impact of odor) could be associated with the self-rated olfactory abilities.

We observed that the participants who rated their sense of smell as above average also had higher AIO scores. In other words they were more strongly influenced by odor in their liking (and not liking) for new stimuli than other participants. Similarly, participants who reported being annoyed by odors also had higher AIO scores. However, AIO score was not associated with TDI score (ie, result of the smell test). These results suggest that people who are easily, often, or strongly affected by smells may, perhaps as a consequence of this, start to think that they have a superior sense of smell. This is consistent with our observation, that a large proportion of the participants overestimated their olfactory ability and with the non-significant trend towards higher AIO scores in participants who overestimated their olfactory function. Interestingly, these olfactory 'underperformers' reported that they consumed a smaller range of herb and spice items than the 'overperformers' (ie, the participants whose TDI scores were better than would be expected from their self-assessment of olfactory acuity). This observation appears to be relevant to our second hypothesis about the connection between spice consumption and identification of their odors. Indeed, we saw a strong correlation between spice odor identification score and the variety of spice items consumed. This was expected, because the questionnaire assessing spice consumption asked about the spices that were used as olfactory stimuli in the spice odor identification task. In addition, however, the number of spices consumed was correlated with the identification sub-score of the Sniffin' Sticks Extended test, although the test included only four of the items in our spice odor identification task and in our consumption questionnaire (anise, cinnamon, clove, and garlic).

In the future understanding of the affective impact of odor and its associations with subjective and objective olfactory abilities might be improved by research into the influence of personality. Recent studies have demonstrated that affective responses to odors can be related to personality traits. For example, Cornell-Kärnekull, Jönsson, Larsson, and Olofsson (2011) reported that neuroticism was associated with negative responses to environmental odors and Seo, Lee, and Cho (2013) also reported associations between personality traits and attitudes towards the sense of smell.

Methodological considerations

We regard it as a strength of our study that subjective data collected via questionnaires were combined with the data from an established, widely used smell test battery, Sniffin' Sticks Extended test. We believe that we are the first group to report results obtained using the extended, three-part version of the test in a Finnish translation. However, the fact that we were unable to persuade all those who took part in the first part of the study to complete the second part reduced the sample size (and hence the statistical power) in the second part.

Conclusions

Our data confirms previous findings of a lack of correlation between self-evaluated and measured olfactory performance in normosmic adults and suggest two potential causes to be further scrutinized: individual differences in affective importance of odor (ie, how important odors are for liking and disliking for various things) and in the variety of odors one sample repeatedly and consciously, including spices and other odorous products.

Acknowledgements
We are grateful to the participants for their time and devotion to the study. We thank Mira Virkki, Li Zhuoran, Jonna-Kaisa Pyykkö, and Huma Ejaz for their contribution to the data collection by administering the Sniffin' Sticks Extended test.

Funding
This study was funded by the Academy of Finland (grant number 267 698 to AK and 256 176 to MS) and University of Turku.

References

Bäckström, A., Pirttilä-Backman, A.-M., & Tuorila, H. (2004). Willingness to try new foods as predicted by social representations and attitude and trait scales. *Appetite*, **43**, 75–83.

Chrea, C., Valentin, D., Sulmont-Rossé, C., Mai, H.L., Nguyen, D.H., Abdi, H. (2004). Culture and odor categorization: agreement between cultures depends upon the odors. *Food Quality and Preference*, **15**, 669–679.

Cornell-Kärnekull, S., Jönsson, F.U., Larsson, M., & Olofsson, J.K. (2011) Affected by smells? Environmental chemical responsivity predicts odor perception. *Chemical Senses*, **36**, 641–648.

Distel, H., Ayabe-Kanamura, S., Martínez-Gómez, M., Schicker, I., Kobayakawa, T., Saito, S., & Hudson, R. (1999). Perception of everyday odors - correlation between intensity, familiarity and strength of hedonic judgement. *Chemical Senses*, **24**, 191–199.

Hummel, T., Kobal, G., Gudziol, H., & Mackay-Sim, A. (2007). Normative data for the "Sniffin' Sticks" including tests of odor identification, odor discrimination, and olfactory thresholds: an upgrade based on a group of more than 3000 subjects. *European Archives of Otorhinolaryngology*, **264**, 237–243.

Hummel, T., Sekinger, B., Wolf, S. R., Pauli, E., & Kobal, G. (1997). 'Sniffin' sticks': olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. *Chemical Senses*, **22**, 39–52.

Knaapila, A., & Tuorila, H. (2014). Experiences of environmental odors among self-reported hyperosmics: a pilot study. *Journal of Health Psychology*, **19**, 897–906.

Knaapila, A., Tuorila, H., Kyvik, K. O., Wright, M. J., Keskitalo, K., Hansen, J., … Silventoinen, K. (2008). Self-ratings of olfactory function reflect odor annoyance rather than olfactory acuity. *Laryngoscope*, **118**, 2212–2217.

Kollndorfer, K., Kowalczyk, K., Nell, S., Krajnik, J., Mueller, C.A., & Schöpf, V. (2015). The inability to self-evaluate smell performance. How the vividness of mental images outweighs awareness of olfactory performance. *Frontiers in Psychology*, **6**, 627.

Landis, B.N., Hummel, T., Hugentobler, M., Giger, R., & Lacroix, J.S. (2003). Ratings of overall olfactory function. *Chemical Senses*, **28**, 691–694.

Murphy, C., Schubert, C.R., Cruickshanks, K.J., Klein, B.E.K., Klein, R., & Nondahl, D.M. (2002). Prevalence of olfactory impairment in older adults. *JAMA*, **288**, 2307–2312.

Nguyen, D.T., Nguyen-Thi, P.-L., Jankowski, R. (2012). How does measured olfactory function correlate with self-ratings of the sense of smell in patients with nasal polyposis? *Laryngoscope*, **122**, 947–952.

Philpott, C.M., Wolstenholme, C.R., Goodenough, P.C., & Murty, G.E. (2006). Comparison of subjective perception with objective measurement of olfaction. *Otolaryngology - Head and Neck Surgery*, **134**, 488–490.

Seo, H.S., Lee, S., & Cho, S. (2013). Relationships between personality traits and attitudes toward the sense of smell. *Frontiers in Psychology*, **4**, 901.

Welge-Luessen, A., Hummel, T., Stojan, T., & Wolfensberger, M. (2005). What is the correlation between ratings and measures of olfactory function in patinents with olfactory loss? *American Journal of Rhinology*, **19**, 567–571.

Wrzesniewski, A., McCauley, C., & Rozin, P. (1999) Odor and affect: Individual differences in the impact of odor on liking for places, things and people. *Chemical Senses*, **24**: 713–721.

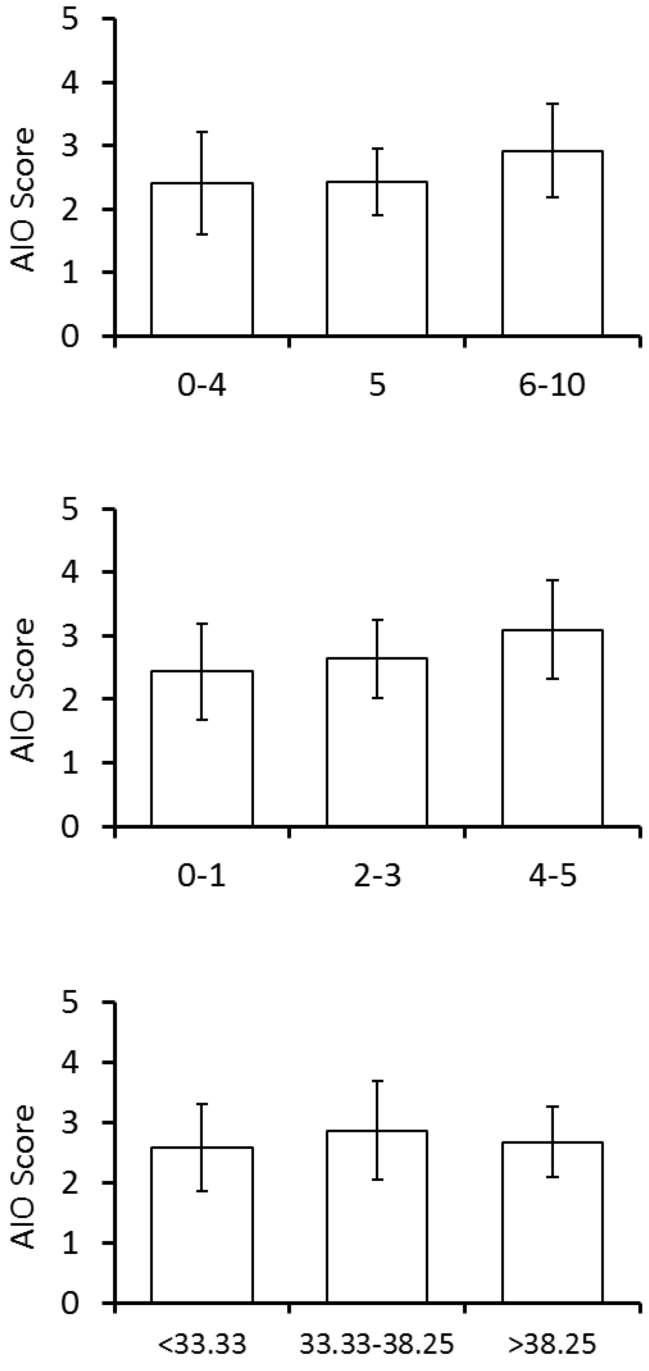

Figure 1. The affective impact of odor (AIO) score was associated with self-rated olfactory performance (one-way ANOVA; $F_{2,114} = 7.42$, $p = 0.001$) and odor annoyance ($F_{2,114} = 6.58$, $p = 0.002$), but not with olfactory performance as measured by the TDI score from the Sniffin' Sticks Extended smell test ($F_{2,114} = 1.58$, $p = 0.21$). Error bars indicate SD. Means of groups marked with different lowercase letters (a, b) differ significantly (Tukey's test, $p < 0.05$).

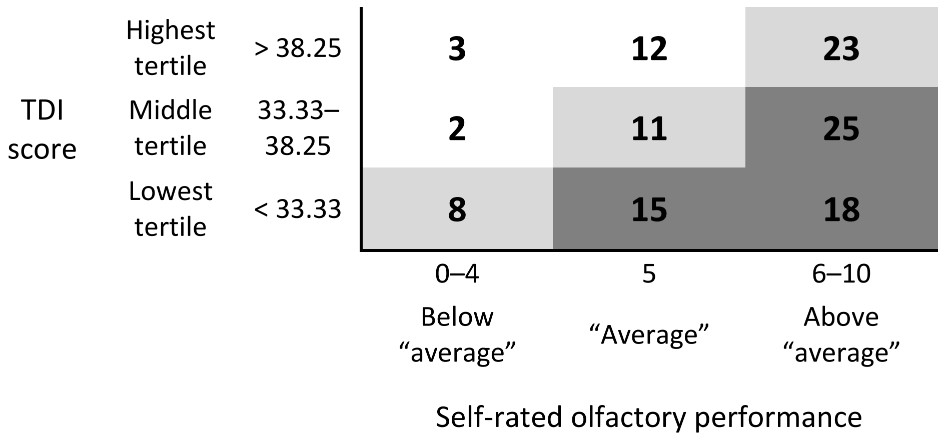

Figure 2. Cross-tabulation showing distribution of participants (N = 117) based on self-rated and measured olfactory performance. The degree of concordance between these two variables was used to classify participants according to the accuracy of their assessment of their sense of smell. Participants whose self-rating corresponded well with the smell test result (TDI score based on the Sniffin' Sticks Extended test) were labeled 'realists' (light gray, n = 42), those who underestimated their olfactory performance were labeled 'overperformers' (white, n = 17), and those who overestimated their olfactory performance were labeled 'underperformers' (dark gray, n = 58).

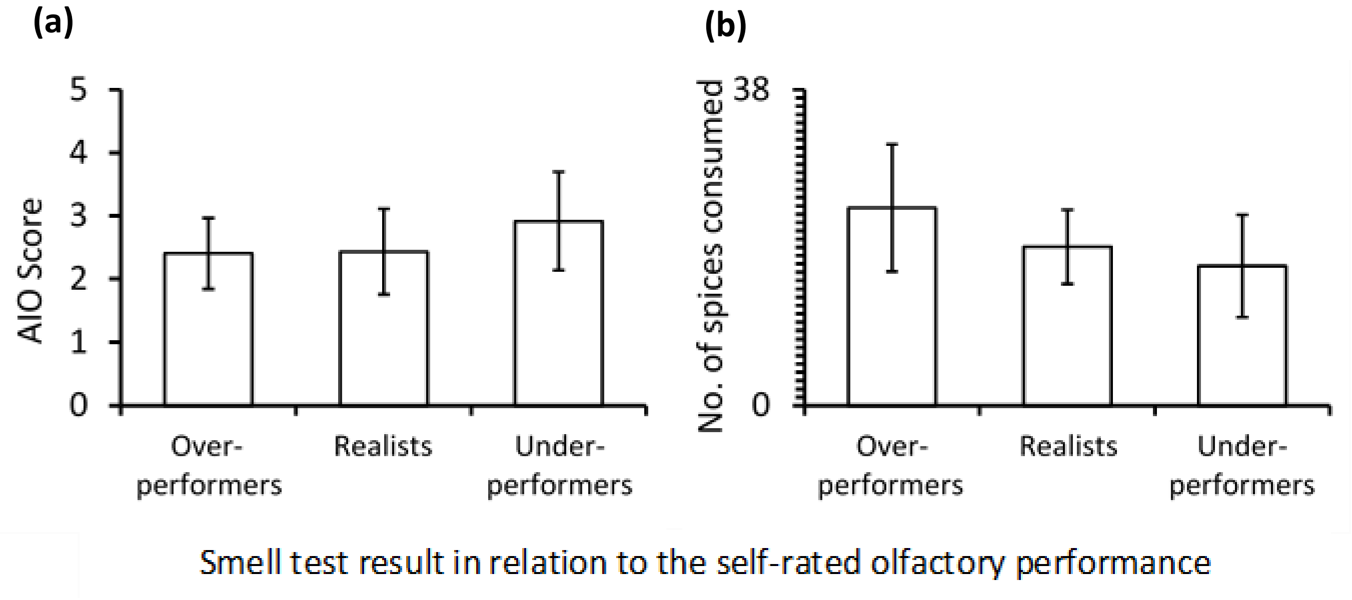

Figure 3. Comparison of the participants whose smell test result (TDI score) was better ('overperformers'), in line with ('realists'), or worse ('underperformers') than would be expected from their self-rated olfactory performance. (a) The affective impact of odor (AIO) scores of the groups were similar ($N = 117$, $p > 0.05$). (b) There were group differences in the number of spice items consumed (data only available for those who completed the second part of the study, $n = 44$, $F_{2,41} = 4.02$; $p = 0.026$). Means of groups marked with different lowercase letters (a, b) differ (Tukey's test, $p < 0.05$). Error bars indicate SD.

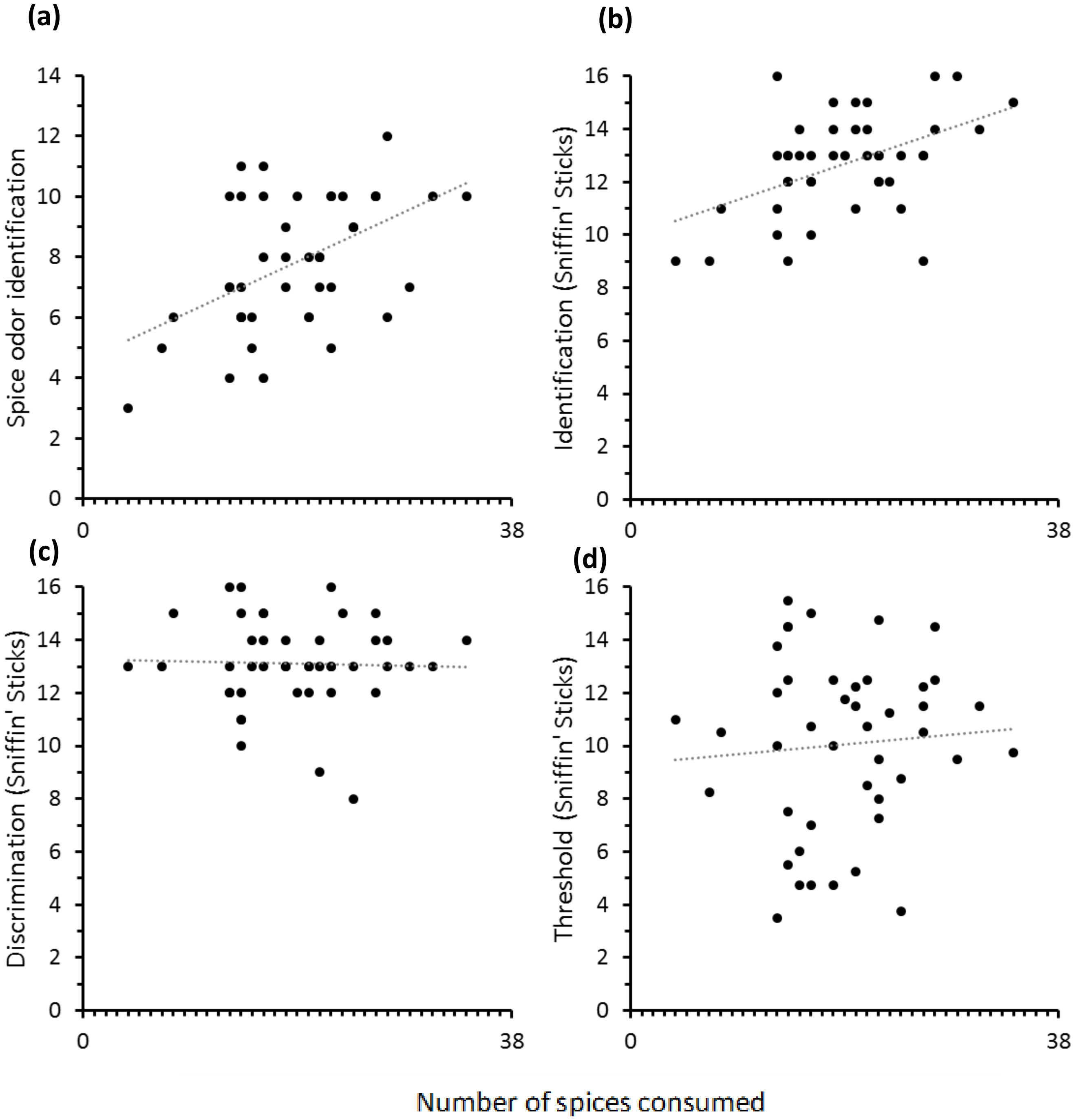

Figure 4. Scatter plots showing relationships between the number spices (out of a list of 38, see Table 2) consumed and a measure of olfactory function, either (a) number of spice odors identified correctly, or a sub-score of the Sniffin' Sticks Extended smell test: (b) identification, (c) discrimination, or (d) threshold (for phenyl ethyl alcohol). *r*, Pearson's correlation coefficient (**, $p < 0.01$).