The Psychophysical Assessment of Odor Valence: Does an Anchor Stimulus Influence the Hedonic Evaluation of Odors?

Marion Clepce, Konrad Neumann, Peter Martus, Marie Nitsch, Jan Wielopolski, Alexander Koch, Johannes Kornhuber, Karin Reich, and Norbert Thuerauf

1Department of Psychiatry and Psychotherapy, Friedrich-Alexander University Erlangen-Nuremberg, Schwabachanlage 6, Erlangen 91054, Germany, 2Department of Biometrics and Clinical Epidemiology, Charité University Hospital Berlin, Hindenburgdamm 30, Berlin 12203, Germany and 3Department of Medical Biometry, Eberhard Karls University Tübingen, Silcherstraße 5, Tübingen 72076, Germany

Correspondence to be sent to: Norbert Thuerauf, Department of Psychiatry and Psychotherapy, Friedrich-Alexander University Erlangen-Nuremberg, Schwabachanlage 6, Erlangen 91054, Germany. e-mail: norbert.thuerauf@uk-erlangen.de

*These authors contributed equally to the work.

Accepted September 23, 2013

Abstract

Olfactory stimuli are experienced primarily in terms of their hedonic tone and the assessment of olfactory hedonic estimates is a prevalent task in scientific and industrial contexts. However, measuring conditions are poorly standardized. Our study aims to fill this gap, focusing on the influence of anchor stimuli on olfactory hedonic evaluations, frequency of anchor presentation, and temporal stability of results. In n = 31 subjects, hedonic estimates for the 16 odors of the Sniffin’ Sticks identification task were assessed on a visual analog rating scale under 4 measuring conditions (nonanchor, pleasant anchor, neutral anchor, unpleasant anchor). To test for stability over time, n = 10 subjects were reassessed 2, 4, and 6 months after original testing. To analyze for possible effects of single versus repeated anchor presentation, n = 15 subjects were retested 2 months after the original session in a multiple anchor presentation format. Statistical analysis revealed significant differences between the 4 anchor conditions, thus highlighting the necessity of specifying assessment methods in scientific research. No significant differences between timepoints were observed, indicating a high temporal stability of olfactory hedonic evaluations, especially from timepoint T2 onward. No overall significant effects of single versus multiple anchor presentation were detected. Findings might help to further standardize testing procedures.

Key words: anchor stimulus, odor valence, olfaction, psychophysics, Sniffin’ Sticks test

Introduction

Odor pleasantness can be considered the primary and most important dimension of olfactory perception (Khan et al. 2007), comparable, for example, to wavelength in vision or frequency in audition. Odor pleasantness was found to correlate with the principal axis of descriptors for physicochemical properties of odorous stimuli (Khan et al. 2007) and the spatial organization of human olfactory epithelium (Lapid et al. 2011). Odorants are involuntarily categorized by their pleasantness (Bensafi et al. 2002) and hedonically toned olfactory stimuli have been shown to possibly influence a variety of human behaviors, such as mate choice (Capparuccini et al. 2010), dream content (Schredl et al. 2009), or nutrition (Stevenson et al. 2007). Therefore, valence ratings of odorants are a prevalent task in various scientific and industrial contexts (Clepce et al. 2010; Cumming et al. 2011; Demarquay et al. 2006; Jiang et al. 2010; Kamath et al. 2011; Lombion et al. 2010; Strauss et al. 2010). However, the psychophysical assessment of olfactory hedonic estimates is still poorly standardized. This may lead to difficulties in interpreting and comparing results of different studies. Among other techniques, a wide variety of Likert scales is employed to obtain pleasantness ratings for stimulus odors. Although some researchers use a 5-point Likert scale ranging from −2 “disgusting,” 0 “neutral” to +2 “pleasant” (Masago et al. 2001), or from −2 (“dislike a lot”) to +2 (“like a lot”) (Cumming et al. 2011), others use a 7-point scale
Materials and methods

Subjects

Olfactory testing was performed in 31 healthy volunteers (mean age: 28.9 years [standard deviation $\{SD\} = 7.6$]; minimum age: 20; maximum age: 50; 13 men, 18 women). Prior to testing, the subjects’ medical history was assessed by a trained clinician focusing on prior or current otorhinolaryngologic complaints. All participants suffering from known anosmia, hyposmia, or reporting any disturbances of the olfactory system were excluded from the study. The remaining subjects clearly perceived all stimulus odors during the test trials.

To control for possible effects of laterality and uni- and bilateral testing, an additional experiment was conducted in a separate sample of $n = 10$ healthy volunteers (mean age: 28.5 years [SD = 2.014]; minimum age: 25; maximum age: 32; 5 men, 5 women). Subjects were selected according to the screening procedure for the main experiments described above.

All participants provided written informed consent prior to testing. The study was conducted in accordance with the Declaration of Helsinki and was reviewed and approved by the local Ethics Committee (University of Erlangen-Nuremberg).

Olfactory testing

As test stimuli for the assessment of odor hedonics, the 16 standard odors of the Sniffin’ Sticks identification task (Hummel et al. 1997; Kobal et al. 1996) were employed (orange [1], shoe leather [2], cinnamon [3], peppermint [4], banana [5], lemon [6], liquorice [7], turpentine [8], garlic [9], coffee [10], apple [11], clove [12], pineapple [13], rose [14], anise [15], fish [16]). The interval between stimulus presentations was set at 30 s. Both nostrils were tested separately. The contralateral nostril was occluded manually while wearing unscented examination gloves. Subjects evaluated each odorant on a bipolar analog rating scale. The relative scaling for “unpleasantness/pleasantness” ranged from −100 to +100 visual analog rating units, represented by a physical length of 200 mm on paper. The unpleasant end of the scale was marked “−−−−”, the pleasant end of the scale was marked “++ ++.” According to the 4 measuring conditions, the center of the scale was either marked “0” (nonanchor condition), “anchor A,” “anchor B,” or “anchor C” (see paragraphs below). Subjects marked the perceived hedonic valence with a pencil. Hedonic estimates for each odorant were averaged across both nostrils.

In the nonanchor condition, participants were instructed to rate the test odors according to their subjective liking, using “0” for odorants evaluated as neutral. In the 3 anchor conditions, subjects were instructed to rate the test odors in comparison to the presented anchor stimulus.

The three anchor stimuli 1) orange, 2) liquorice, and 3) fish were chosen based on the results of prior studies in the database HeDoS-F (Markovic et al. 2007; Thuerauf et al. 2008), comprising olfactory data of more than 200 healthy subjects assessed in a nonanchor-measuring condition. Accordingly, the odor orange had received highest pleasantness ratings, fish lowest pleasantness ratings, and liquorice neutral ratings.

To test for possible influences of different anchor stimuli on hedonic ratings, 4 measuring conditions were applied during the first week of testing (T1): subjects ($n = 31$) rated the pleasantness of each odor either without a standard odor (0; nonanchor condition) or in comparison with anchor A (orange), anchor B (liquorice), or anchor C (fish). The 4 measuring conditions were presented to the subjects in a random order on 4 separate days within 1 week.

To test for stability over time, a subgroup of subjects ($n = 10$; mean age: 25.8, $SD = 2.35$ years; 5 men, 5 women) repeated the full baseline testing session (measuring conditions 0, A, B, and C) 3 times (T2–T4), within 2 months
(T2), 4 months (T3), and 6 months (T4) after the original assessment.

Another subgroup of subjects \((n = 15); \text{mean age: } 31.20, \text{SD } = 9.45\text{ years}; 5 \text{ men, 10 women},\) was assessed twice, at baseline T1 and T2. For these subjects, the conditions of anchor presentation were changed at T2 to a multiple presentation format, that is, presenting the anchor stimuli not only once per nostril but anew before the evaluation of every single odor (measuring conditions 0, A-M(ultiple), B-M, C-M, for illustration see Figure 1).

An additional control experiment was conducted to explore possible effects of laterality and uni- versus bilateral testing. A separate sample of \(n = 10\) subjects was assessed twice (at T1 and after 1 week at T2) with anchor condition C. For these subjects in both testing sessions, stimulus odors were presented 1) to the right nostril alone, 2) to the left nostril alone, and 3) to both nostrils. Conditions 1–3 were presented in random order.

**Statistical analysis**

The statistical software package SPSS/Version 18.0 for Windows (SPSS Inc.) and the Statistical Analysis System (SAS Institute Inc.) were used for calculations. All statistical tests were two tailed with a significance level of \(\alpha = 0.05\). We tested for normal distribution of the data employing the Kolmogorov–Smirnov test. Normal distribution could be assumed for all data. Because of the small sample size, a nonparametric approach was chosen for further analysis. Nonparametric analyses for longitudinal data as proposed by Brunner and Puri (2001) were calculated based on rank statistics and a linear mixed model, with hedonic estimates as dependent variable and the different stimulus odors/anchor conditions/timepoints as independent variables. In order to compare results for single versus multiple anchor presentation, a similar nonparametric repeated measures analysis was calculated. For pairwise comparisons between the 4 anchor conditions, Bonferroni correction with factor 6 was applied. To explore possible effects of laterality and uni- and bilateral assessment (control experiment) \(t\)-tests for dependent samples and Pearson correlation coefficients were calculated. For these analyses, the average over all sticks was calculated per subject.

**Results**

**Effects of anchor stimuli on hedonic ratings**

To compare nonanchor and anchor conditions, nonparametric analyses according to Brunner and Puri (2001) were conducted for the first testing session, with hedonic estimates as dependent variable and the 16 stimulus sticks and 4 anchor conditions (0, A, B, C) as independent variables. Analyses showed a significant main effect of anchors \((P < 0.001)\). Pairwise comparisons were significant for anchor 0 versus anchor A and for anchor C versus each of the other anchors \((P < 0.001 \text{ after Bonferroni correction})\). For anchor A versus anchor B and anchor 0 versus anchor B, no significant results were found \((P = 0.19 \text{ and } P = 0.36 \text{ after Bonferroni correction})\).

Furthermore, a similar analysis was conducted to test if the curve shapes for the 4 measuring conditions run parallel. The analyses showed a significant interaction effect of anchor \(\times\) stick \((F(31, 13.1) = 1.93, P = 0.049)\) (for mean hedonic ratings and standard error of the means [SEMs]; see Figure 2).
Thus, the shape of the curves was not parallel. However, none of the 6 pairwise comparisons of the single shapes was significant, the smallest $P$ value was 0.063 between anchor 0 and anchor B (without Bonferroni correction). Thus, results obtained with different anchor stimuli cannot be simply transformed into each other.

Analyzing the different anchors without adjusting for sticks led to overall significant differences ($P < 0.001$). Values were largest for anchor C followed by anchor 0, anchor B, and anchor A. All pairwise comparisons were significant except for anchor 0 and anchor B ($P = 0.231$ without Bonferroni correction) and anchor B with anchor A ($P = 0.009$ without Bonferroni correction, $P = 0.054$ with Bonferroni correction factor 6, “nearly significant”).

**Temporal stability of hedonic estimates**

To test for reproducibility over time, Brunner analyses were calculated to compare hedonic estimates assessed in the non-anchor condition for all 4 testing sessions. No statistically significant differences in hedonic ratings could be observed between the 4 testing sessions T1–T4 ($F(10, 6.01) = 0.66$, $P = 0.65$) (for means and SEMs; see Figure 3). However, separate analyses for the different anchors revealed significant differences over time for anchor C ($P = 0.021$), but not for anchors 0 ($P = 0.57$), A ($P = 0.38$), and B ($P = 0.60$). This significance was due to the comparison of T1 versus T2–T4, differences between T2, T3, and T4 were not significant for anchor C ($P = 0.63$). Moreover, for each of the 4 anchor conditions, the 95% confidence limits of intraindividual variation (Bland–Altman limits of agreement) were approximately ±6 points for the comparison between T1 versus T2, T3, and T4. Thus, subjects’ hedonic evaluations were found to be reproducible from T2 on, but not from T1 and

**Figure 2 and Table 1**. Thus, the shape of the curves was not parallel. However, none of the 6 pairwise comparisons of the single shapes was significant, the smallest $P$ value was 0.063 between anchor 0 and anchor B (without Bonferroni correction). Thus, results obtained with different anchor stimuli cannot be simply transformed into each other.

Analyzing the different anchors without adjusting for sticks led to overall significant differences ($P < 0.001$). Values were largest for anchor C followed by anchor 0, anchor B, and anchor A. All pairwise comparisons were significant except for anchor 0 and anchor B ($P = 0.231$ without Bonferroni correction) and anchor B with anchor A ($P = 0.009$ without Bonferroni correction, $P = 0.054$ with Bonferroni correction factor 6, “nearly significant”).

**Figure 3** Mean hedonic estimates for 16 odors in the nonanchor condition for T1–T4. Error bars are representing SEMs.
Table 2. Means and SEMs (in brackets) for 16 stimulus odors in all 4 anchor conditions at testing sessions T1–T4 (n = 10).

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonanchor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>3.00 (0.67)</td>
<td>-1.74 (0.76)</td>
<td>1.88 (1.07)</td>
<td>2.64 (0.91)</td>
</tr>
<tr>
<td>T2</td>
<td>4.16 (0.53)</td>
<td>-0.86 (0.73)</td>
<td>1.83 (1.21)</td>
<td>3.60 (1.25)</td>
</tr>
<tr>
<td>T3</td>
<td>3.88 (0.61)</td>
<td>-0.81 (0.59)</td>
<td>1.93 (0.94)</td>
<td>3.41 (0.95)</td>
</tr>
<tr>
<td>T4</td>
<td>3.48 (0.74)</td>
<td>-0.58 (0.68)</td>
<td>2.47 (0.98)</td>
<td>3.39 (1.07)</td>
</tr>
<tr>
<td>Anchor A</td>
<td>-3.18 (0.66)</td>
<td>-0.44 (1.25)</td>
<td>0.24 (0.86)</td>
<td>1.28 (0.91)</td>
</tr>
<tr>
<td>T2</td>
<td>-3.38 (0.69)</td>
<td>-1.38 (0.73)</td>
<td>-0.27 (1.10)</td>
<td>2.07 (0.62)</td>
</tr>
<tr>
<td>T3</td>
<td>-2.86 (0.55)</td>
<td>-1.29 (0.78)</td>
<td>-0.30 (1.35)</td>
<td>1.81 (0.97)</td>
</tr>
<tr>
<td>T4</td>
<td>-3.53 (0.58)</td>
<td>-2.01 (0.68)</td>
<td>-1.27 (1.19)</td>
<td>1.68 (1.05)</td>
</tr>
<tr>
<td>Anchor B</td>
<td>1.05 (0.87)</td>
<td>-2.93 (0.71)</td>
<td>0.15 (1.32)</td>
<td>1.54 (0.84)</td>
</tr>
<tr>
<td>T2</td>
<td>3.03 (0.67)</td>
<td>-3.08 (0.66)</td>
<td>-0.88 (0.92)</td>
<td>1.47 (0.85)</td>
</tr>
<tr>
<td>T3</td>
<td>2.59 (0.99)</td>
<td>-2.48 (0.57)</td>
<td>-1.31 (0.79)</td>
<td>0.84 (0.90)</td>
</tr>
<tr>
<td>T4</td>
<td>2.74 (1.04)</td>
<td>-1.69 (0.58)</td>
<td>-0.22 (0.68)</td>
<td>1.46 (0.86)</td>
</tr>
<tr>
<td>Anchor C</td>
<td>5.91 (0.84)</td>
<td>2.91 (0.75)</td>
<td>4.26 (0.82)</td>
<td>5.80 (0.67)</td>
</tr>
<tr>
<td>T2</td>
<td>6.65 (0.80)</td>
<td>3.62 (0.62)</td>
<td>4.84 (0.74)</td>
<td>6.41 (0.75)</td>
</tr>
<tr>
<td>T3</td>
<td>6.90 (0.63)</td>
<td>3.97 (0.74)</td>
<td>5.31 (0.66)</td>
<td>6.82 (0.74)</td>
</tr>
<tr>
<td>T4</td>
<td>6.93 (0.76)</td>
<td>4.13 (0.71)</td>
<td>5.49 (0.56)</td>
<td>6.79 (0.60)</td>
</tr>
</tbody>
</table>

Control experiment: effects of laterality and un-/versus bilateral presentation

Possible lateralization effects were analyzed by calculating Pearson correlation coefficients for mean hedonic ratings over all stimuli. The left nostril, the left nostril and both nostrils were strongly correlated both for testing session T1 (r = 0.895).

Single versus repeated presentation of anchor stimuli

Possible effects of single versus repeated anchor presentation were analyzed in Figure 4. See Table 3 for comparison of single and repeated presentation format. It is most prominent for anchor C. See Table 4 for mean hedonic ratings for all anchor conditions at T1–T4.
Additionally, \( t \)-tests were calculated for pairwise comparisons of mean hedonic evaluations over all stimulus odors for uni- and bilateral presentation. We found no statistically significant differences between 1) right and left nostril (T1: \( t(9) = -0.663, P = 0.524; \) T2: \( t(9) = -0.612, P = 0.556 \)), 2) right nostril and bilateral presentation (T1: \( t(9) = -0.620, P = 0.551; \) T2: \( t(9) = -0.413, P = 0.689 \)), and 3) left nostril and bilateral presentation (T1: \( t(9) = -1.693, P = 0.125; \) T2: \( t(9) = -0.809, P = 0.440 \)). For means and SEMs, see Table 4.

**Table 3** Means and SEMs (in brackets) for 16 stimulus odors in the single and repeated anchor presentation format \((n = 15)\).

<table>
<thead>
<tr>
<th>Anchor</th>
<th>Right</th>
<th>Left</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchor A</td>
<td>3.45 (0.74)</td>
<td>-0.83 (0.75)</td>
<td>3.13 (0.82)</td>
</tr>
<tr>
<td>Anchor A-M</td>
<td>3.31 (0.78)</td>
<td>-0.94 (0.79)</td>
<td>3.03 (0.80)</td>
</tr>
<tr>
<td>Anchor B</td>
<td>3.21 (0.81)</td>
<td>-2.56 (0.82)</td>
<td>2.49 (0.80)</td>
</tr>
<tr>
<td>Anchor B-M</td>
<td>3.07 (0.83)</td>
<td>-2.57 (0.84)</td>
<td>2.49 (0.82)</td>
</tr>
<tr>
<td>Anchor C</td>
<td>6.90 (0.34)</td>
<td>2.57 (0.67)</td>
<td>4.97 (0.79)</td>
</tr>
<tr>
<td>Anchor C-M</td>
<td>6.89 (0.34)</td>
<td>2.57 (0.67)</td>
<td>4.97 (0.79)</td>
</tr>
</tbody>
</table>

**Table 4** Control experiment: means and SEMs (in brackets) over 15 stimulus odors for uni- and bilateral stimulation at testing sessions T1 and T2 \((n = 10)\).

<table>
<thead>
<tr>
<th></th>
<th>Right</th>
<th>Left</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>2.74 (0.36)</td>
<td>2.86 (0.38)</td>
<td>2.67 (0.36)</td>
</tr>
<tr>
<td>T2</td>
<td>2.54 (0.48)</td>
<td>2.62 (0.53)</td>
<td>2.45 (0.43)</td>
</tr>
</tbody>
</table>

**Discussion**

Our results clearly demonstrate the influence of anchor stimuli on olfactory hedonic estimates. In the field of cognitive psychology, it is a robust finding that specific anchors may affect cognitive judgments. Tversky and Kahneman (1974)
defined the so-called anchoring effect as follows: “… different starting points yield different estimates, which are biased toward the initial values. We call this phenomenon anchoring.” Context or anchoring effects have been reported for various research areas such as time estimation (König 2005; Thomas and Handley 2008), decision processes (Englisch et al. 2006; Jacowitz and Kahneman 1995; McNicol and Pennington 1973; Northcraft and Neale 1987), visual discrimination (Ware 1980), or taste perception (Schifferstein 1994; Schifferstein and Frijters 1992). Context effects were demonstrated for intensity ratings in the field of olfaction (Hulshoff Pol et al. 1998). Odors were evaluated as more intense after the presentation of low-intensity stimuli, whereas no changes in intensity ratings occurred after the presentation of high-intensity odors. As an explanation, Hulshoff Pol and colleagues (1998) refer to Helson’s adaption level theory (Helson 1948), suggesting that sensory perception of a test stimulus is influenced by prior stimuli. According to this, the presentation of strong stimuli should result in an underestimation of the test stimulus and the presentation of weak stimuli in an overestimation of the test stimulus. Only a few studies exist concerning context influences on the hedonic perception of chemical sensory stimuli. In a cross-modality experiment, subjects expressed lower palatability ratings for a drink after positive odor exposure (Stevenson et al. 2007). According to our results, it can be assumed that anchor stimuli do influence olfactory hedonic judgments. However, the curves for the 4 anchor conditions were not significantly superimposable. Thus, assessments of hedonic ratings obtained with different anchor stimuli cannot be simply transformed into each other. For future research on odor hedonics, our findings highlight the necessity of clearly specifying assessment methods until common standards are established.

Our study also demonstrates high temporal stability of hedonic ratings assessed via a VAS, especially from time-point T2 onward. This result is in line with prior findings confirming a high test–retest reliability for VAS ratings in various research areas, such as pain (Carlsson 1983) or mood estimates (Steinacher et al. 1998). Thus, VAS can be considered a reliable technique for the assessment of olfactory hedonic judgments.

The anchor presentation format (single vs. multiple) had no overall significant effect on olfactory hedonic judgments. Although Cain and Johnson (1978) found that prolonged exposure to an odorant influences subsequent olfactory hedonic ratings, even a single presentation of the anchor stimulus had a significant effect on hedonic ratings in our study. However, a repeated presentation format seemed to influence hedonic ratings differently depending on the anchor stimulus. This finding again highlights the necessity of clearly specifying assessment methods in future research on olfactory hedonics.

The results of our control study concerning possible effects of laterality are in line with prior findings. For testing in a nonanchor condition, we previously showed that odor intensity but not the hedonic estimation is lateralized (Thuerauf et al. 2008). Differences in odor detection, discrimination, and identification did not reach a statistically significant level, but for all these parameters the scores of the right nostril were slightly higher compared with the left nostril. For odor identification; however, a statistical tendency was observed. Zatorre and Jones-Gotman found a right nostril advantage for odor discrimination (Zatorre and Jones-Gotman 1990). In the present study, we found no differences in the hedonic evaluation of odors between the 2 nostrils and an uni- versus bilateral presentation format even for an anchor condition.

Limitations

The relatively small sample size has to be considered a limiting factor of the present study, especially concerning the analysis of temporal stability and the comparison between single and multiple anchor presentation formats. Therefore, future confirmatory studies on a larger population are needed. It would also be interesting to test if the presented findings can be replicated for different stimulus odors and different anchors, especially nonfood anchors.

Conclusion

The present study for the first time demonstrates the influence of anchor stimuli on olfactory hedonic estimates, and therefore the importance—for industrial and scientific applications—to weigh anchor for the psychophysical assessment of odor valence.

Funding

This study is part of the Neurotrition Project, which was supported by the FAU (Friedrich-Alexander University Erlangen-Nuremberg) Emerging Fields Initiative.

Acknowledgement

We would like to thank Diplom-Psychologin Katharina Glaser and Diplom-Psychologe Christoph Stein for their assistance in conducting the control experiment.

References


