
Matthias Laska and Nina Grimm

Department of Medical Psychology, University of Munich Medical School, Goethestraße 31, D-80336 Munich, Germany

Correspondence to be sent to: Matthias Laska, Department of Medical Psychology, University of Munich Medical School, Goethestraße 31, D-80336 Munich, Germany. e-mail: laska@imp.med.uni-muenchen.de

Abstract

Recently, Olsson and Cain (2000, Chem. Senses, 25: 493) introduced a psychometric method which, for the first time, allows the standardized determination of odor quality discrimination (OQD) thresholds. The method defines a threshold value that is an average fraction by which one odorant has to be substituted with another to reach a criterion level of discrimination. This measure of discrimination is reciprocal in the sense that it is a result of two separate psychometric functions involving two different standards but the same comparison stimuli. Using the same odor stimuli as Olsson and Cain, with six human subjects but adopting a slightly different experimental design, we were able to replicate their finding that the proportion of correct discriminations changes monotonically with the proportion of adulterant in mixtures of eugenol and citral. As the SURE (SUbstitution-REciprocity) method is based on discriminative responses, it should also be applicable with nonhuman species which can be trained to give unequivocal discriminative responses at the behavioral level. Using an olfactory conditioning paradigm, we therefore trained four squirrel monkeys to discriminate between exactly the same pairs of odor stimuli as our human subjects. We found the psychometric functions of the monkeys to be similar to those of the human subjects. Our results show that the SURE method can successfully be employed with nonhuman primates and thus offers a new approach to study the odor spaces of nonhuman species. Future studies should elucidate whether the SURE method allows for direct comparisons of OQD thresholds and of similarities and differences between odor quality perception of different species.

Key words: nonhuman primates, odor psychophysics, odor quality discrimination, psychometric functions

Introduction

The relationship between odor quality and molecular properties of volatile chemical stimuli is arguably one of the central topics in olfactory research (Rossiter, 1996; Yoshii and Hirono, 1996; Chastrette, 1997; Afshar et al., 1998). Although a considerable number of psychophysical studies have tried to reveal, and generally reported, some correlations between odor quality and molecular properties (Pilgrim and Schutz, 1957; Moskowitz and Barbe, 1977; Schiffrin, 1981; Dravnieks, 1985; Jeltema and Southwick, 1986; Cain et al., 1998), most of these studies have failed in providing quantitatively useful data as they usually relied either on enumerative verbal description or on sorting, profiling or scaling procedures. In a thoughtful review of the available methods to quantify odor quality, Wise et al. (Wise et al., 2000) mentioned that such methods, ‘all share a strong subjective component, an unproved ability to discern small differences, and are indeterminate regarding the extent of individual differences’ (Wise et al., 2000). The only way to overcome the problems inherent to the quantification of odor quality is to develop and employ psychophysical methods that are based upon performance rather than on the more common reporting of mental contents. Such methods inevitably tap the discriminative basis of olfactory perception. However, discriminative methods, usually measuring confusability of odors using same-different or oddity procedures (de Wijk and Cain, 1994; Laska and Hübener, 2001; Laska, 2002), have their limitations, too, as they tend to produce ceiling effects, i.e. high rates of correct discriminations, unless qualitatively very similar odors are compared (Laska and Teubner, 1999a,b; Laska et al., 2000). Therefore, differences in performance levels of such confusion tests may be too small to differentiate effectively between dissimilar pairs of odors or between individuals (Wise et al., 2000).

Recently, Olsson and Cain (Olsson and Cain, 2000) introduced a psychometric method which avoids this disadvantage and, for the first time, allows the standardized determination of olfactory quality discrimination (OQD)
thresholds. The method defines a threshold value that is an average fraction by which one odorant has to be substituted with another to reach a criterion level of discrimination. This measure of discrimination is reciprocal in the sense that it is a result of two separate psychometric functions involving two different standards but the same comparison stimuli. An important advantage of the SURE (SUbstitution-REciprocity) method compared to conventional discrimination procedures is that it circumvents one of the major problems of any method based on discriminative responses, i.e. perceived intensity of the stimuli to be compared as an unwanted cue for discrimination (Lawless and Heymann, 1998). This is due to the fact that the method employs both odors to be discriminated as standards and as adulterant stimuli, thereby cancelling out—at least to some degree—possible differences in perceived intensity of the stimuli.

As the SURE method is based on discriminative responses, it should also be applicable with nonhuman species which can be trained to give unequivocal discriminative responses at the behavioral level. This would offer a new approach to study the odor spaces of nonhuman species, and perhaps would allow for direct across-species comparisons of OQD thresholds and of similarities and differences between the odor spaces of different species (Hildebrand and Shepherd, 1997). Therefore, the aims of the present study were twofold: first, to replicate the study of Olsson and Cain (Olsson and Cain, 2000) and, secondly, to determine whether the SURE method can successfully be applied with a nonhuman species. To this end, we used the same odorants but a slightly different experimental design as in the original study with six human subjects, and we trained four squirrel monkeys in an olfactory conditioning paradigm to discriminate between exactly the same pairs of odor stimuli as our human subjects.

Experiment 1: replication study with human subjects

Materials and methods

Human subjects

Six healthy, unpaid volunteers (four females and two males) aged 25–40 years of age participated in the study. All were non-smokers and none had any history of olfactory dysfunction. All subjects had previously served in olfactory tests and were familiar with the basic test procedure. They were informed as to the aim of the experiment and provided written consent. The study was performed in accordance with the Declaration of Helsinki/Hong Kong.

Odorants

Two standards of eugenol and citral were mixed with odorless diethyl phthalate (all from Merck, Darmstadt, Germany) to concentrations of 0.19% (v/v), respectively. These had been determined to be of about equal perceived intensity. Nine liquid phase mixtures of these base concentrations were prepared, ranging from a mixture of 90% eugenol standard and 10% citral standard to a mixture of 10% eugenol and 90% citral (via 80/20, 70/30 and so on). Preliminary tests showed that some of the subjects were able to discriminate the eugenol standard from a 90/10 mixture of eugenol and citral at >75% correct. Therefore, an additional mixture of 99% eugenol and 1% citral was prepared. Altogether, 12 unique stimuli (10 mixtures plus the two standards) were used. With all substances, the highest available degree of purity was used.

Test procedure

A 40 ml aliquot of each odorant was presented in a 250 ml polyethylene squeeze bottle equipped with a flip-up spout which for testing was fitted with a custom-made Teflon nose-piece. Subjects were instructed as to the manner of sampling and at the start of the first session were allowed time to re-familiarize themselves with the bottles and the sampling technique. Care was taken that the nose-piece was only a short distance (1–2 cm) from the nasal septum during sampling of an odorant in order to allow the stimulus to enter both nostrils.

In a three-alternative forced choice (3-AFC) oddity test procedure subjects were asked to compare three bottles and identify the one containing the odd stimulus. Additionally, after each decision subjects were asked whether their choice was predominantly based on perceived differences in odor quality or on perceived differences in odor intensity. Each bottle could be sampled twice with an inter-stimulus interval of at least 3 s. Sampling duration was restricted to 1 s per presentation in order to minimize adaptation effects. The sequence of presenting the stimulus pairs was systematically varied between sessions and individual subjects while taking care to avoid successive presentations of the same combinations and while systematically varying the order in which the stimuli were sampled. The presentation of a given substance as odd or even stimulus was balanced within and between sessions. Approximately 30 s were allowed between trials and no feedback regarding the correctness of the subjects’ choice was given.

Twenty-one different stimulus pairs (11 with the eugenol standard and 10 with the citral standard) were presented once per session so as to give a total of 21 judgements. Testing was repeated in nine more sessions each 1–3 days apart, enabling 10 judgements per stimulus pair and panelist to be collected.

Data analysis

For each stimulus combination, proportions of correct discriminations ($P_c$) were calculated for each of the six subjects separately and then averaged into group results (means ± SE). $P_c$ values for the two standards (eugenol and citral) were plotted as a function of proportion of citral in the mixture.

The OQD (group) threshold value was determined by first...
pooling the two separate functions generated for the two standards. This was done by averaging the two $P_c$ values for each percentage of adulterant in the mixture. Then, a function was fitted to the pooled data from which a combined threshold value $T_c$ could be read for $P_c = 0.75$, i.e. at 75% correct discriminations.

**Results**

Figure 1 shows the psychometric functions for quality discrimination of eugenol, citral and mixtures of them for the group of six human subjects. The results indicate that the proportion of correct discriminations changes monotonically with the proportion of adulterant in the mixture and that it took proportionally less citral to adulterate eugenol than vice versa. The psychometric functions of the individual subjects (not shown here) revealed that this was true for all six participants.

Figure 2 shows the psychometric function for the combined functions of Figure 1.

Using the threshold criterion adopted by Olsson and Cain (Olsson and Cain, 2000), i.e. $P_c = 0.75$ (corresponding to 75% correct discriminations), a combined OQD threshold value $T_c = 0.32$ could be read from the function fitted to the pooled data. The $T_c$ values of the individual subjects (not shown here) ranged from 0.26 to 0.36 and thus OQD thresholds showed only little interindividual variability.

Using a second threshold criterion that lies halfway between chance and perfect performance, i.e. $P_c = 0.67$ (corresponding to 67% correct discriminations), a combined OQD threshold value $T_c = 0.23$ could be read from the function fitted to the pooled data.

With all 21 odor pairs <12% of decisions were reported to be based upon perceived differences in odor intensity rather than odor quality (cf. Test procedure). Altogether, 93.0% of decisions were reported to be based upon perceived differences in odor quality. Odor pairs that were readily discriminated yielded lower percentages of perceived intensity as the choice criterion compared to odor pairs that presented some difficulties to the subjects. Accordingly, a negative correlation between discriminability of the odor pairs and the frequency of perceived differences in odor intensity as the choice criterion was found (Spearman, $r_s = -0.70$).

**Experiment 2: extension study with squirrel monkeys**

**Materials and methods**

**Animals**

Testing was carried out using four adult male squirrel monkeys (*Saimiri sciureus*), maintained as part of an established breeding colony. All animals had served as subjects in previous olfactory experiments and were completely familiar with the basic test procedure (Laska and Hudson, 1993a,b, 1995; Laska *et al.*, 1996, 1999a,b, 2000a,b; Laska and Freyer, 1997; Laska and Teubner, 1998; Laska and Seibt, 2002a,b). The colony was housed in a double enclosure comprising a 23 m$^3$ home cage joined to a 7 m$^3$ test cage by two tunnels which could be closed by sliding doors to allow the temporary separation of animals for individual testing. Animals were provided with marmo-
set pellets (Ssniff®, Soest, Germany), fresh fruit, vegetables and water ad libitum.

The experiments reported here comply with the Guide for the Care and Use of Laboratory Animals (National Institutes of Health Publication No. 86-23, revised 1985) and also with current German laws.

**Odorants**

Exactly the same set of odorants as for the human subjects was used.

**Behavioral test**

In a task designed to simulate olfactory-guided foraging, opaque 1.5 ml Eppendorf® flip-top reagent cups were fitted with absorbent paper strips (35 × 7 mm; Sugi, Kettenbach, Germany) impregnated with 10 µl of an odorant signaling either that they contained a peanut food reward (S+) or that they did not (S–). The odor strips were attached to the vials by cutting a slit in each strip and slipping it over the flip-up lid which was connected to the vial by a narrow band. Eighteen such cups, nine positive and nine negative, were inserted in pseudorandom order in holes along the horizontal bars of a climbing frame in such a way that some effort was required for the animals to remove them. The frame was mounted to one of the enclosure walls and consisted of a 2.5 m vertical pole (40 mm diameter) fitted with 7 cross-bars (20 mm diameter) 30 cm apart, the middle three of which extended 50 cm to either side and were equipped with conically bored holes to hold the cups (Laska and Hudson, 1993a).

In each test trial each monkey was allowed 1 min to harvest as many baited cups from the frame as possible, a time too short to allow all of them to be inspected and removed. Five such trials were conducted per animal per session and usually two sessions were conducted per day. Cups were used only once and the odorized strips were prepared fresh at the start of each session.

In a first series of experiments, eugenol was assigned as the rewarded stimulus (S+) and all mixtures as well as the citral standard were used as S–. In a second series of experiments, citral was assigned as the S+ and all mixtures as well as the eugenol standard were used as S–.

To allow an animal to build a robust association between food reward and a given S+, each of the two experimental series started with conducting four sessions using anethole as S–. This substance does not share any apparent structural or qualitative similarities with any of the S+ tested. After these four sessions, all animals reliably scored >90% correct choices, implying that they had learned to correctly assign the reward value of the S+.

In order to prevent serial order or training effects from confounding the results, the tasks were not presented according to a systematic change in the proportions of the mixture components but according to a pseudorandomized sequence, with the exception that the very first task of an experimental series involved the discrimination of the pure standards, and the second task involved a mixture (as S–) that was maximally different and thus presumably easy to discriminate from the reference odor (as S+), and the very last task of a series involved a mixture that was maximally similar and thus presumably difficult to distinguish from the reference odor. Therefore, the order of task presentation was:

- citral (as S+) versus eugenol—10/90, 50/50, 30/70, 80/20, 40/60, 70/30, 20/80, 60/40, 90/10 mixtures of citral and eugenol (as S–);
- eugenol (as S+) versus citral—90/10, 50/50, 70/30, 20/80, 60/40, 30/70, 80/20, 40/60, 19/99 mixtures of citral and eugenol (as S–).

All experimental conditions were conducted for a total of four sessions (i.e. 20 × 1 min trials performed on two consecutive days) and a new task was only begun when the previous task was completed. In order to prevent the more challenging conditions leading to extinction or to a decline in the animals' motivation and in order to obtain a baseline value allowing the evaluation of the discrimination performance of the squirrel monkeys in the critical tasks, two sessions of an easy control task were interspersed between the two experimental series. This consisted of the discrimination between the S+ and anethole.

**Data analysis**

In assessing performance of the squirrel monkeys, only cups inspected by the animals were scored. For each individual, the percentage of correct choices from the best two sessions, i.e. from 10 trials of 1 min comprising a total of at least 60 decisions, was calculated. Correct choices consisted both in animals correctly rejecting negative cups by failing to open or remove them and in identifying positive cups by removing and opening them to obtain the food reward. Conversely, errors consisted in animals opening or removing negative cups, or failing to remove and open positive cups.

Further analysis of data followed exactly the same procedure as described in experiment 1.

**Results**

Figure 3 shows the psychometric functions for quality discrimination of eugenol, citral and mixtures of them for the group of four squirrel monkeys. In agreement with the human data, the results indicate that the proportion of correct discriminations changes monotonically with the proportion of adulterant in the mixture and that it took proportionally less citral to adulterate eugenol than vice versa. The psychometric functions of the individual animals (not shown here) revealed that this was true for all four monkeys.

Figure 4 shows the psychometric function for the combined functions of Figure 3.

Using the threshold criterion adopted by Olsson and Cain (Olsson and Cain, 2000), i.e. \( P_c = 0.75 \) (corresponding to
75% correct discriminations), a combined OQD threshold value $T_c = 0.24$ could be read from the function fitted to the pooled data.

**Discussion**

The results of the present study demonstrate (i) that we were able to replicate the finding of Olsson and Cain (Olsson and Cain, 2000) that the proportion of correct discriminations changes monotonically with the proportion of adulterant in mixtures of eugenol and citral and (ii) that the SURE method can successfully be employed with nonhuman primates.

Although it is near at hand to compare the OQD threshold value reported for human subjects in the original study ($T_c = 0.34$) with that obtained in the present study for human subjects ($T_c = 0.32$) and for the squirrel monkeys ($T_c = 0.24$), two caveats should be considered.

First, we used a slightly higher concentration for one of the stimuli (0.19% eugenol standard rather than 0.09%) with both human subjects and squirrel monkeys than in the original study. The use of a slightly higher concentration of eugenol was recommended to us by one of the authors of the original study (M. Olsson, personal communication) as their choice of concentrations for eugenol and citral turned out to be less than optimal with regard to the intended intensity-matching of the two standards. Our finding that 93.0% of decisions were reported to be based upon perceived differences in odor quality rather than intensity confirms that the slight adjustment of the eugenol concentration indeed led to an excellent—and probably better than in the original study—matching of perceived stimulus intensities. Although the possibility that this change might have contributed to the difference in the human OQD threshold of the two studies cannot be ruled out completely, it seems rather unlikely as our increase in the concentration of eugenol would—at least theoretically—have been expected to shift $T_c$ to a higher rather than to a lower value.

Secondly and perhaps most importantly, we used the 3-AFC oddity method rather than the rarely employed ABX-design for discrimination of odors by the human subjects. This led to different chance levels of performance in the two studies. Whereas the probability for a subject to pick the correct bottle in a given trial is 0.33 with the 3-AFC oddity method, it is 0.5 with the ABX-design. (In the 3-AFC oddity situation, a subject has to decide which of the three stimuli is different from the other two; in the ABX-design the subject is first presented with two different stimuli and then with a third stimulus, and has to decide whether the third stimulus is identical to the first or the second one.) This, in turn, implies that the threshold criterion of 75% correct decisions used in both studies is halfway between chance and perfect performance with the latter method, but not with the former method. One possibility to avoid such a mismatch of threshold criteria between studies could be to adopt the criterion of ‘halfway between chance and perfect performance'.

**Figure 3**

Psychometric functions for odor quality discrimination of eugenol, citral and mixtures of them shown for the group of four squirrel monkeys (means ± SE). Open circles indicate proportions of correct discrimination between a standard of pure eugenol and a mixture where citral was substituted for eugenol. Filled circles indicate the reversed case where citral was the standard and eugenol the substitute. Theoretical data points ($P_c = 0.5$) were added for the two theoretical cases where the standard would have been compared with itself. The dotted line indicates the chance level of performance.

**Figure 4**

The psychometric function for the combined functions of Figure 3. The combined OQD threshold $T_c$ at 67 and 75% correct discriminations for the group of four squirrel monkeys is given. A theoretical data point was added for the theoretical case where the standard would have been compared with itself ($P_c = 0.5$). The dotted line indicates the chance level of performance.
performance’ to our data. In doing this, a combined OQD threshold value $T_c = 0.23$ could then be read for $P_c = 0.67$, i.e. at 67% correct discriminations (which is halfway between the chance level of 0.33 and perfect performance at 1.00), from the function fitted to our pooled human data (cf. Figure 2).

A comparison of OQD threshold values using the criterion of 75% correct discriminations would suggest excellent agreement between the two studies (with $T_c$ values of 0.34 and 0.32, respectively), whereas a comparison of $T_c$ values using the criterion ‘halfway between chance and perfect’ would suggest a considerable difference in the outcome of the two studies (with OQD threshold values of 0.34 and 0.23, respectively).

However, it is arguable whether meaningful comparisons—beyond the fact that in both studies human discrimination performance changes monotonically with the proportion of adulterant in mixtures of eugenol and citral—and between studies using different experimental designs are possible at all. Therefore, no further conclusions as to reproducibility of results should be drawn here. Future studies should address this point by directly and systematically comparing the performance of the same set of subjects tested on the same discrimination tasks but with different methods.

In the second part of the present study, we were able to show that the SURE method can successfully be employed with nonhuman primates. In line with the results obtained with the human subjects, we found that the proportion of correct discriminations performed by the squirrel monkeys with nonhuman primates for this particular pair of odor stimuli are quite similar (cf. Figures 1 and 3).

A comparison of OQD threshold values of human subjects and squirrel monkeys using the criterion of 75% correct discriminations would suggest a superior performance of the nonhuman primates (with $T_c$ values of 0.52 and 0.24, respectively), whereas a comparison of $T_c$ values using the criterion ‘halfway between chance and perfect’ would suggest almost identical performance of the two species (with OQD threshold values of 0.23 and 0.24, respectively). Note that a comparison between squirrel monkeys and the human subjects of the original study would also suggest a superior performance of the nonhuman primates. However, here too, comparisons between studies using different experimental designs should be considered with caution for the reasons mentioned above and no further conclusions as to possible differences or similarities in performance between species should be drawn at this point.

Nevertheless, our finding that the SURE method can successfully be employed with squirrel monkeys offers a new approach to study the odor spaces of nonhuman species. Future studies should elucidate whether this method allows for direct or at least relative comparisons of OQD thresholds and of similarities and differences between odor quality perception of different species.

Acknowledgements

We thank our panelists for their willingness to participate in this study, Angela Meckl for expert maintenance of the animals and the Deutsche Forschungsgemeinschaft for financial support (La 635/10-2).

References


Laska, M. and Teubner, P. (1999b) Olfactory discrimination ability of...


Accepted 28 November 2002