The 40-item Monell Extended Sniffin' Sticks Identification Test (MONEX-40)

Jessica Freiherr[a,*], **Amy R. Gordon**[a,d], **Eva C. Alden**[a], **Andrea L. Ponting**[a], **Monica F. Hernandez**[b], **Sanne Boesveldt**[a], and **Johan N. Lundström**[a,c,d]

Jessica Freiherr: jfreiherr@ukaachen.de; Amy R. Gordon: agordon@monell.org; Eva C. Alden: eva.alden@u.northwestern.edu; Andrea L. Ponting: aponting@monell.org; Monica F. Hernandez: monica.hernandez@mail.mcgill.ca; Sanne Boesveldt: sanne.boesveldt@wur.nl; Johan N. Lundström: jlundstrom@monell.org

[a]Monell Chemical Senses Center, 3500 Market St., Philadelphia, PA 19104, USA

[b]Dept. of Clinical Neurosciences, University of Calgary, Calgary, Canada

[c]Dept. of Psychology, University of Pennsylvania, Philadelphia, USA

[d]Dept. of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden

Abstract

Background—Most existing olfactory identification (ID) tests have the primary aim of diagnosing clinical olfactory dysfunction, thereby rendering them sub-optimal for experimental settings where the aim is to detect differences in healthy subjects' odor ID abilities.

Materials and methods—We have developed an extended version of the olfactory ID subtest of the Sniffin' Sticks test battery to better assess the variability in ID scores and thereby olfactory abilities of healthy, adult individuals. Twenty-four odorants, corresponding cue labels, and distractor labels were added to the existing 16-item Sniffin' Sticks ID test to create the 40-item Monell Extended Sniffin' Sticks Identification Test (MONEX-40). The MONEX-40 was administered to 259 healthy young subjects, of which 72 were retested on an average of 212 days (SD 112 days) later.

Results—The added odor items demonstrated good validity, as shown by a significant correlation of the results with the original 16-item ID test. In addition, the MONEX-40 achieved a significant test–retest and split-half reliability.

Conclusions—Taken together, these results suggest that the MONEX-40 is a reliable method for experimental assessment of odor ID ability in healthy, young individuals. Moreover, its use of a wider range of odors allows the experimenter to present subsets of the MONEX-40 within the same experiment while maintaining statistical power.

Keywords

Smell; Olfaction; Reliability; Validity

*Corresponding author at: Department of Diagnostic and Interventional Neuroradiology, RWTH Aachen, Aachen, Germany. Tel.: +49 241 80 35688; fax: +49 241 80 82155.

Conflict of interest

The authors declare that no known conflicts of interest exist. This test will not be patented and the authors have no ambition to gain financially from manufacturing, distributing, or marketing this test.

1. Introduction

Evaluation of olfactory function in humans is commonly performed using tests of olfactory identification (ID). Measures of olfactory ID are, in comparison to olfactory detection threshold tests, broader assessment techniques since they combine detection, discrimination, and odor-related semantic memory tasks (Hornung et al., 1998; Hummel et al., 2007; Jones-Gotman and Zatorre, 1988; Moberg et al., 1999).

Although a plethora of olfactory ID tests already exists (Cain, 1989; Cain et al., 1988; Doty, 2007; Doty et al., 1984a, b; Hummel et al., 1997; Kobal et al., 1996; Kobayashi, 2005; Nordin et al., 1998), the overwhelming majority of these tests was developed to determine olfactory dysfunction in a clinical setting rather than to assess differences in olfactory ID ability between healthy participants in non-clinical research studies. Hence, normosmic subjects commonly perform very well on these tests and often achieve near or at maximum scores meaning that the actual variability in olfactory ID scores among healthy subjects is relatively limited when assessed with clinical tests. Indeed, this seems to be true for the two clinical ID tests most often used in published experimental studies, the UPSIT and Sniffin' Sticks. The median score of young (age 20–39 years), healthy subjects for the UPSIT is 38 out of a total of 40 (Doty et al., 1984a). Similarly, the mean score for young (age 16–35 years), healthy subjects for the Sniffin' Sticks is 14.8 out of a total of 16 (Hummel et al., 2007). These values translate to a 95% and 93% correct response rate, respectively. The need for a research-oriented olfactory ID test capable of assessing the olfactory ID abilities of healthy individuals thus seems warranted. With this in mind, we aimed to expand the original clinical Sniffin' Sticks olfactory ID test to a 40-item version using odors that are more challenging, yet still possible to identify.

As has been clearly demonstrated by others (Haehner et al., 2009; Reden et al., 2006), greater statistical power to test the null hypothesis can arise from a larger allowed variability in the collected data. This has the benefit that smaller differences in olfactory ID ability between two sample populations or conditions, e.g. differences in olfactory ID capacity in different states of satiety, can be detected. Moreover, an extended test also allows for repeated testing within a short period of time (e.g. a day), something that is problematic with a shorter test as subjects can easily remember their answers after the initial administration. An extended test set can also be split into two halves with individual odors and with large enough number of items to generate decent allowed variability in data thus enabling the experimenter to assess short-term effects such as pre- and post-treatment.

The objectives of this study are to introduce the new Monell Extended Sniffin' Sticks Identification Test (MONEX-40), present split-half and test–retest reliability in a group of healthy, North American subjects, as well as validate the 24 new items against the existing 16 items of the Sniffin' Sticks test. We would like to stress that the MONEX-40 is not designed to serve any clinical function, but rather that this extended version has been developed for use in behavioral or neuroscience research experiments involving healthy subjects.

2. Materials and methods

2.1. Participants

Ability to identify odors using the MONEX-40 was assessed in 269 healthy, young, non-smoking subjects. Anosmic or hyposmic subjects, as determined by an ID score below 11 on the original 16-item Sniffin' Sticks set, and statistical outliers displaying an ID score below or above 3 standard deviations from the mean, were removed from the analysis thus resulting in a total of 259 healthy young subjects (163 females, mean age 24.7 years, SD 4.6

years, range 18–40 years). Removing anosmic and hyposmic subjects from the analyses penalizes certain types of statistics, e.g. the test–retest analyses using Spearman's rank-correlation, since these individuals, by virtue of their impairments, would group naturally together at the end of the distribution in both sessions. However, we opted to include only young healthy adults in our analyses because the overwhelming majority of experimental testing is performed on this population and the main goal of the study was to develop an ID test to be used in an experimental setting. Subjects were not taking any medication known to influence chemosensory perception, had never suffered a major head trauma, and had not been knowingly exposed to toxins affecting chemosensory perception. A subset of 72 randomly selected subjects (40 females, mean age 26.6 years, SD 4.2 years, range 18–36 years) participated in a retest session (average 212 days post initial testing, SD 112 days). There was no statistical difference in age between male and female participants in either the initial testing session (females mean age 24.3 years, SD 4.4 years; males mean age 25.3 years, SD 4.8 years; two-sample t-test, $t(257) = 1.86$, p = ns), or the retest session (females mean age 26.6 years, SD 4.3 years; males mean age 26.5 years, SD 4.1 years; two-sample t-test, $t(70) = .09$, p = ns). A separate subset of 46 randomly selected healthy subjects (23 females, mean age 27.9 years, SD 5.2 years) participated in an additional testing session during which perceptual ratings for the 40 items of the MONEX-40 were acquired. All subjects gave written informed consent, and all experimental procedures were approved by the Institutional Review Board of the University of Pennsylvania.

2.2. Creation of olfactory identification test

The original Sniffin' Sticks olfactory ID subtest (Hummel et al., 1997; Kobal et al., 1996) consists of 16 individual felt-tip pens, each containing a distinct odor. Each odor is paired with four descriptors in a cued, forced-choice testing paradigm, and the subject is instructed to identify the descriptor corresponding to the presented odor. The 40-item olfactory ID test is composed of the original Sniffin' Sticks 16-item olfactory ID subtest and 24 new items.

The 24 new items were selected from a larger set of odors ($n \sim 100$). A pilot group, consisting of olfactory experts, four individuals of North American descent and one of northern European descent living in North America, selected odors whose "odor object" would be generally recognized by the North American population from this set. These nameable odors were then ranked according to familiarity ratings of the actual odor of the object rather than according to familiarity ratings of the object per se. Based on this ranking, 8 high/8 medium/8 low relative familiar odors were selected. Note that all of the 24 newly selected items were on the positive side of the familiarity scale, meaning that the 8 odors labelled as low in familiarity still scored above 5 in familiarity on a 10 grade scale. Intensity of the 24 new items in comparison to the 16 original items was taken into account during the selection process. Additionally, care was taken to include both mono-molecular odorants as well as mixtures. Four milliliters of each compound were injected into empty Sniffin' Sticks (Burghart Instruments, Wedel, Germany). A full list of the compounds and their respective concentrations used to create the 24 new items can be found in Table 1. The ratio of food versus non-food items of the 24 new items (83.3%) was similar to that of the original 16-item test (81.3%). Each of the 24 new items was then paired with four descriptors (Table 2). These descriptors were chosen, by the same pilot group, the following way: one target descriptor corresponds to the odor item, a second descriptor is very similar to the odor item, a third descriptor is of moderate similarity to the odor item, and a fourth descriptor is dissimilar to the odor item. Finally, to verify that all of the odor names used in the ID test, both target and distractor items, were familiar, a pilot group of 20 randomly selected native American-English speakers that were unfamiliar with the odor test indicated whether they were familiar with each odor name used in the ID test by a force-choice dichotomized

question for each item. This pilot group was familiar with 98.8% of all the descriptors used in the 16 original items, compared to 99.5% of the descriptors used for the 24 new items.

The olfactory ID testing was performed in a silent, well lit, and well-ventilated room built exclusively for olfactory testing. The MONEX-40 testing procedure runs approximately 15 min.

2.3. Perceptual ratings

As described above, a subset of participants ($n = 46$) was asked to rate the perceptual characteristics of the individual odors on 100-points visual analogue scales (Aitken, 1969) using the E-Prime 2 presentation software (Psychology Software Tools, Inc., Sharpsburg, PA). The perceptual dimensions were intensity (1 = extremely weak, 100 = extremely strong), pleasantness (1 = extremely unpleasant, 100 = extremely pleasant), and familiarity (1 = extremely unfamiliar, extremely familiar).

2.4. Statistical analysis

An initial Kolmogorov–Smirnov test determined that the ID data were not normally distributed; a characteristic previously demonstrated for other olfactory ID tests (Doty et al., 1984a). Therefore, only medians (Mdn), interquartile ranges (IQR) and results from non-parametric tests are reported. We used Wilcoxon signed-rank tests and Spearman's rank-correlation tests, depending on the question at hand, to establish relationships between test scores. Potential sex differences were assessed using a two-sample t-test or Mann–Whitney U-test. SPSS 16 (SPSS Inc., Chicago, IL) was used for all statistical evaluations with the alpha level set to 0.05.

3. Results

The median score on the MONEX-40 was 32 (*IQR* 4). When performance on the two subsets was analyzed separately, subjects' median performance on the original 16-items set was 14 (*IQR* 2) and 18 (*IQR* 3) for the 24 new items. These values correspond to 88% and 75% correct response rate for the 16 original and the 24 new items, respectively. As expected, subjects performed significantly better on the original set than the new set (Wilcoxon signed-rank test, $Z = -13.73$, $p < .001$) and the displayed variability was also greater for the new items than the 16 original items, thus indicating that these items are harder to identify than the 16 original items (Fig. 1). Itemized identification accuracies are shown in Fig. 2.

There was a statistical tendency for women to perform better than men on the MONEX-40 (Mann–Whitney U-test, $U = 6799$, $p = .076$, Table 3). However, when assessing the two ID subsets separately, we found no significant sex-related differences in olfactory ID scores on the original Sniffin' Sticks 16-item ID subtest, whereas the 24 new items indicated a significant difference (Mann–Whitney U-test, $U = 6243$, $p = .006$, Table 3).

One of the more widely used and thoroughly validated olfactory ID tests is the original 16-item clinical Sniffin' Sticks test (Hummel et al., 2007; Kobal et al., 2000). To assess the validity of the additional 24 odors added to form the MONEX-40, we correlated the results of the 16 initial items with those of the 24 new odors. The two subtests of the MONEX-40 show a significant positive correlation (Spearman's rank-correlation tests, $rho(259) = .31$, $p < .001$).

To create two comparable sets, we divided the MONEX-40 into two 20-item halves. The first half, MONEX-40-1, is comprised of the first half of the original Sniffin' Sticks 16-item ID subtest (items 1–8) and the first half of the 24 new items (items 17–28). The second half

(MONEX-40-2) is therefore composed of the second half of the Sniffin' Sticks 16-item test (items 9–16) and the last twelve of the 24 new items (items 29–40). Based on this division, the MONEX-40 demonstrates good split-half reliability as shown by subjects' comparable performance on the first and second halves of the test (MONEX-40-1: *Mdn* 16, *IQR* 2; MONEX-40-2: *Mdn* 16, *IQR* 2). Moreover, subjects' performance on MONEX-40-1 was significantly correlated to performance on MONEX-40-2 (Spearman's rank-correlation tests, $rho(259) = .34$, $p < .001$).

Additionally, the MONEX-40 demonstrates good test–retest reliability. In the retest session, subjects' median MONEX-40 score was 32 (*IQR* 4) [original 16 items: 13 (*IQR* 1), 24 new items: 18 (*IQR* 3)]. A correlation between scores obtained during the first session with those scores obtained in the second session demonstrated a significant test–retest correlation (Spearman's rank-correlation tests, $rho(72) = .68$, $p < .001$, Fig. 3). Furthermore, when assessed separately, there were significant correlations between initial and retest scores for both the 16 original odor items (1–16: Spearman's rank-correlation tests, $rho(72) = .40$, $p < .001$) and the 24 new odor items (17–40: Spearman's rank-correlation tests, $rho(72) = .62$, $p = .001$).

A few response alternatives are repeated in both, the original and the new portion of the test (Table 2). We therefore explored whether these items were overrepresented as answers. However, we observed that in 89 of the original subjects (see Table 2) where itemized responses were collected, repetition of responses do not seem to influence subject's answers.

An overview displaying the intensity, pleasantness, and familiarity ratings for each odor item of the MONEX-40 can be seen in Fig. 4.

4. Discussion

Here we have described the MONEX-40, an extended version of the Sniffin' Sticks test battery. The MONEX-40 is intended for use as an experimental tool to assess olfactory ID performance in healthy young adults thus more difficult than other existing tests aimed at clinical populations. In addition, due to the relatively large number of odor items, the MONEX-40 can be divided in half and used for repeated testing of odor ID performance within a short time period without learning effects.

The median ID rate for the 24 new items of the MONEX-40 was significantly lower than the median ID rate for the original 16-item ID subtest (Hummel et al., 1997). Additionally, as expected with a harder test, subjects' scores showed greater dispersion for the 24 new items than for the 16 original items. The lower median and greater dispersion in scores for items 17–40 suggest that the MONEX-40 reflects the wider range of ID capabilities among healthy subjects and, therefore, is better able to illuminate differences between non-clinical conditions in within-subjects but also between-subjects designs. Separating the results for the 16 initial items of the MONEX-40, i.e. the original Sniffin' Sticks ID clinical subtest, and the 24 new items enables the experimenter to both screen for clinical olfactory deficiencies and elucidate the effects of experimental manipulation.

One could argue that the greater dispersion for the 24 new items does not originate from fluctuation in ID performance, but rather originates from other irrelevant processes. However, the sex-differentiated result for the new 24-item subtest demonstrating a female superiority, and a lack of sex-dependent effects for the original 16-item test, indicates that the greater variability does indeed reflect variability in ID performance. It has been repeatedly demonstrated that women are, on average, superior to men at identifying odors (Cain, 1982; Doty et al., 1985). Combined, the MONEX-40 detected a statistical tendency toward a significantly higher ID ability in women compared to men.

To date, validation of new olfactory ID tests has been accomplished by comparing to existing clinical olfactory ID tests. Validating an experimental test by comparison to a clinical test means that differences between the two will naturally exist in level of difficulty since a clinical test is designed to be relatively easy for a normosmic individual. Nonetheless, subjects' new 24-item scores and the Sniffin' Sticks original 16-item ID scores were significantly correlated, thereby indicating the statistical validity of the MONEX-40. The obtained correlation coefficient ($rho(259) = .31$) is lower than that of the correlation between, for example, the Sniffin' Sticks 16-item subtest and a 5-item ID test ($r = .61$) (Mueller and Renner, 2006). However, the difference can be partly explained by the fact that the aforementioned five-item test comprised of items from the 16 original items. The increased difficulty of items 17–40, as well the use of more conservative non-parametric testing, may contribute to the significantly weaker correlation in comparison to the correlation between two clinical tests as well as the fact that individuals with olfactory dysfunction and outliers were excluded from these analyses.

The MONEX-40 test–retest correlation coefficient ($rho(72) = .68$) is slightly lower than those of other Sniffin' Sticks-based identification tests [$r = .73$ (Hummel et al., 1997), $r = .83$–$.94$ (Reden et al., 2006), $r = .88$ (Haehner et al., 2009)], the 16-item SOIT [$r = .79$ (Nordin et al., 1998)], and the 40-item UPSIT [$r = .92$ (Doty et al., 1984a, b)]. However, the test–retest intervals employed in these comparisons range from only 4 to 131 days and tend to be toward the shorter end of that spectrum, whereas the MONEX-40 was repeated with a greater interval. It should be kept in mind that the MONEX-40 seems to be a significantly harder olfactory ID test than the aforementioned tests since it is aimed toward testing of healthy young individuals. This will naturally lower test–retest values since a clinical olfactory ID test with a ceiling effect with respect to healthy participant's performance and the inclusion of individuals with anosmia will have, by its design, a very high test–retest result. A more direct and valid comparison would be to assess the difference between test–retest correlation for the well-established 16-items test and the additional 24-items within this population, with the same test–retest time period. The difference in test–retest correlation between the 16 initial items and the additional 24 items were negligible. More importantly, the 24 new items demonstrated a higher test–retest correlation compared to the 16 original items. Those facts let us suggest that the test–retest reliability for the MONEX-40 is comparable to other tests.

One limitation of this work is that during development of the test we did not control for frequency of the response alternatives throughout the test. Some descriptors of the 16 original, as well as of the 24 new items were used more than once. Although the repeated use of descriptors did not lead to an overrepresentation in selection of these response alternatives, and is also seen in some clinical tests, this might be of concern when using this test in other populations. Experimenters choosing to replicate this test in a different subject population should be aware of this and monitor their data for overrepresentation of certain response alternatives.

In conclusion, the 40-item Monell Extended Sniffin' Sticks Identification Test (MONEX-40) is a measure of human olfactory identification ability for use in a population of healthy subjects within the North American culture sphere. Its lack of a clear ceiling effect raises the potential to identify minor differences in identification ability between experimental conditions in healthy subjects. Furthermore, the MONEX-40 can be divided in half to enable comparison of olfactory ID performances within a limited time frame, or before and after experimental manipulation. Based on these data, the 40-item Monell Extended Sniffin' Sticks Identification Test (MONEX-40) provides an evaluation tool of subjects' identification performance in a research setting. This might be of special value as

an evaluation tool of olfactory abilities for functional neuroimaging studies in healthy individuals.

Acknowledgments

References

Aitken RC. Measurement of feelings using visual analogue scales. Proc R Soc Med. 1969; 62:989–93. [PubMed: 4899510]

Cain WS. Odor identification by males and females: predictions vs performance. Chem Senses. 1982; 7:129–42.

Cain WS. Testing olfaction in a clinical setting. Ear Nose Throat J. 1989; 68(316):22–8.

Cain WS, Gent JF, Goodspeed RB, Leonard G. Evaluation of olfactory dysfunction in the Connecticut Chemosensory Clinical Research Center. Laryngoscope. 1988; 98:83–8. [PubMed: 3336267]

Doty RL. Office procedures for quantitative assessment of olfactory function. Am J Rhinol. 2007; 21:460–73. [PubMed: 17882917]

Doty RL, Applebaum S, Zusho H, Settle RG. Sex differences in odor identification ability: a cross-cultural analysis. Neuropsychologia. 1985; 23:667–72. [PubMed: 4058710]

Doty RL, Shaman P, Dann M. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. Physiol Behav. 1984a; 32:489–502. [PubMed: 6463130]

Doty RL, Shaman P, Kimmelman CP, Dann MS. University of Pennsylvania Smell Identification Test: a rapid quantitative olfactory function test for the clinic. Laryngoscope. 1984b; 94:176–8. [PubMed: 6694486]

Haehner A, Mayer AM, Landis BN, Pournaras I, Lill K, Gudziol V, et al. High test–retest reliability of the extended version of the Sniffin' Sticks test. Chem Senses. 2009; 34:705–11. [PubMed: 19759361]

Hornung DE, Kurtz DB, Bradshaw CB, Seipel DM, Kent PF, Blair DC, et al. The olfactory loss that accompanies an HIV infection. Physiol Behav. 1998; 64:549–56. [PubMed: 9761231]

Hummel T, Kobal G, Gudziol H, Mackay-Sim A. 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 3,000 subjects. Eur Arch Otorhinolaryngol. 2007; 264:237–43. [PubMed: 17021776]

Hummel T, Sekinger B, Wolf SR, Pauli E, Kobal G. 'Sniffin' sticks': olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. Chem Senses. 1997; 22:39–52. [PubMed: 9056084]

Jones-Gotman M, Zatorre RJ. Olfactory identification deficits in patients with focal cerebral excision. Neuropsychologia. 1988; 26:387–400. [PubMed: 3374800]

Kobal G, Hummel T, Sekinger B, Barz S, Roscher S, Wolf S. Sniffin' sticks: screening of olfactory performance. Rhinology. 1996; 34:222–6. [PubMed: 9050101]

Kobal G, Klimek L, Wolfensberger M, Gudziol H, Temmel A, Owen CM, et al. Multicenter investigation of 1,036 subjects using a standardized method for the assessment of olfactory function combining tests of odor identification, odor discrimination, and olfactory thresholds. Eur Arch Otorhinolaryngol. 2000; 257:205–11. [PubMed: 10867835]

Kobayashi M. The Odor Stick Identification Test for the Japanese (OSIT-J): clinical suitability for patients suffering from olfactory disturbance. Chem Senses. 2005; 30(Suppl 1):i216–7. [PubMed: 15738122]

Moberg PJ, Agrin R, Gur RE, Gur RC, Turetsky BI, Doty RL. Olfactory dysfunction in schizophrenia: a qualitative and quantitative review. Neuropsychopharmacology. 1999; 21:325–40. [PubMed: 10457530]

Mueller C, Renner B. A new procedure for the short screening of olfactory function using five items from the Sniffin' Sticks identification test kit. Am J Rhinol. 2006; 20:113–6. [PubMed: 16539306]

Nordin S, Bramerson A, Liden E, Bende M. The Scandinavian Odor-Identification Test: development, reliability, validity and normative data. Acta Otolaryngol. 1998; 118:226–34. [PubMed: 9583791]

Reden J, Mayer A, Hummel T. An extended version of the Sniffin' Sticks. Chem Senses. 2006; 31:A39.
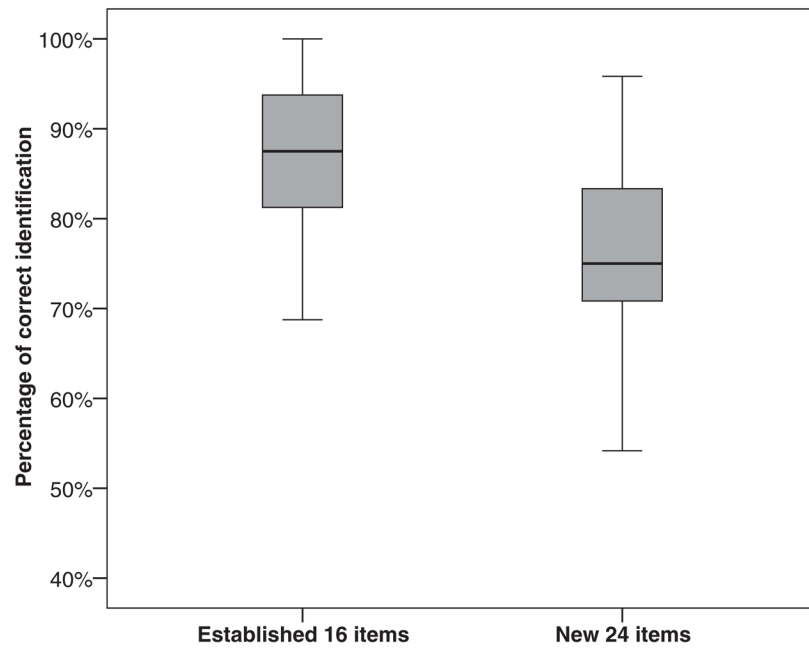
Fig. 1.
Percentage of correct ID for the original 16 and the 24 new items ($N = 259$). Shown are box plots with median, 25th and 75th percentile as well as minimum and maximum values.
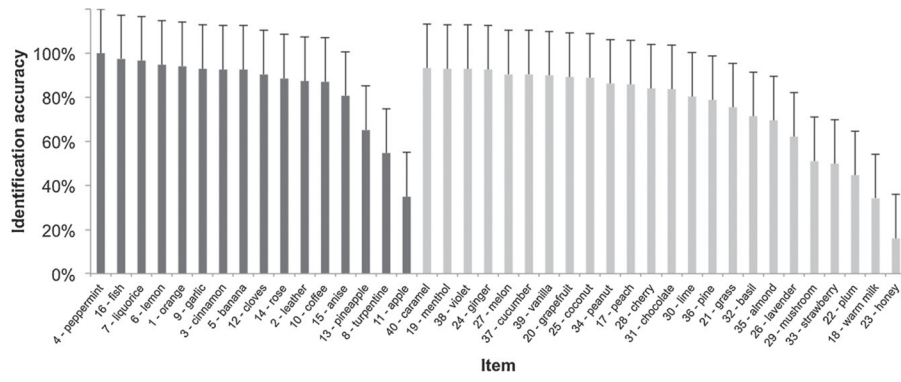
Fig. 2.
Proportional identification accuracy of the 40 items of the MONEX-40 ($N = 259$). The bars represent means and standard deviations.
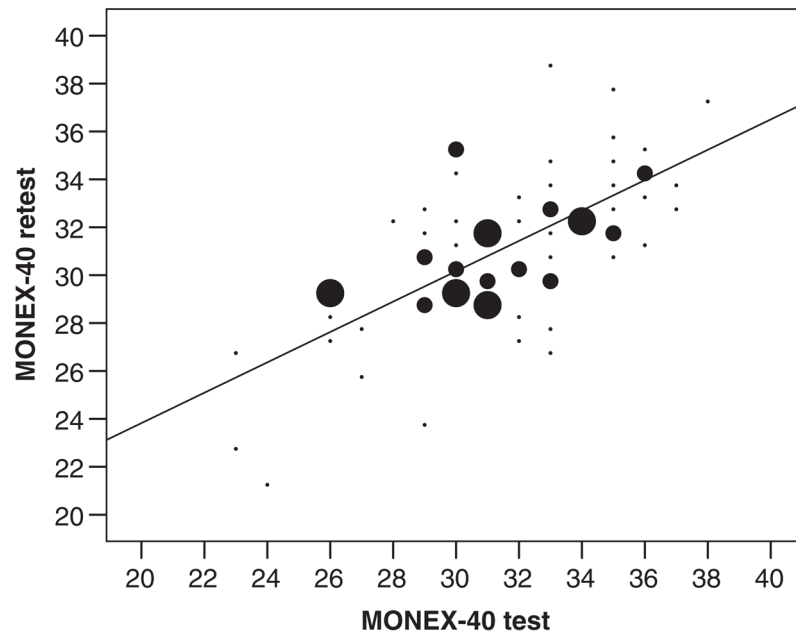
Fig. 3.
Test–retest reliability of the MONEX-40 (size of circle represents number of subjects, smallest size: 1 subject, medium size: 2 subjects, largest size: 3 subjects, $N = 72$).

Fig. 4.
Intensity (light gray, 1 = extremely weak, 100 = extremely strong), pleasantness (medium gray, 1 = extremely unpleasant, 100 = extremely pleasant), and familiarity ratings (dark gray, 1 = extremely unfamiliar, 100 = extremely familiar) of the 40 items of the MONEX-40 (*n* = 46). The bars represent means and standard deviations.

Table 1

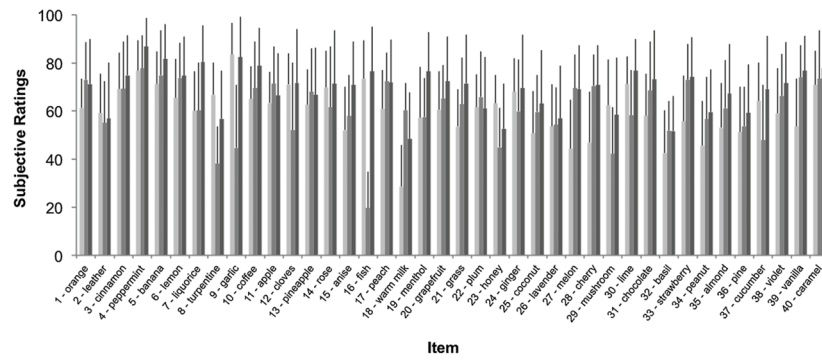Overview of the compounds used to create the 24 new items of the MONEX-40, including the corresponding company name, product number, CAS number, and used concentrations (diluents: DEP, diethyl phthalate; PG, propylene glycol).

No.	Odor descriptor	Chemical compound	Company	Product no.	CAS no.	Concentration
17	Peach	Peach FC	Givaudan	Peach 21059733		100%
18	Warm milk	Warm Milk FC	Givaudan	Warm Milk 21051393		20% in DEP
19	Menthol	L(−)-Menthol	Fisher Scientific	AC125405000	2216-51-5	50% in PG
20	Grapefruit	Grapefruit oil	Givaudan	Grapefruit 5849683		100%
21	Grass	Cis-3-Hexenol	Givaudan	Cis-3-hexenol (grass) 5732451	90045-43-5 8016-20-4	10% in DEP
22	Plum	Plum/Prune FC	Givaudan	Plum 21051123		100%
23	Honey	Honey FC	Givaudan	Miel Blanc 21000513		100%
24	Ginger	Ginger	Givaudan	Ginger CO2 Pure 5845183		100%
25	Coconut	Dihydro-5-pentyl-2(3H)-furanone	Givaudan	Prunolide (coconut) 8487001	104-61-0	100%
26	Lavender	Lavender oil	Givaudan	Lavender Oil 5850703	84776-65-8 8000-28-0	10% in DEP
27	Melon	Melon FC	Givaudan	Melon 21221283		10% in DEP
28	Cherry	Cherry FC	Givaudan	Cherry 21050383		5% in DEP
29	Mushroom	1-Octen-3-ol	Sigma–Aldrich	5284	3391-86-4	50% in DEP
30	Lime	Lime oil	Givaudan	Lime Peru oil 5845273	90063-52-8 8008-26-2	100%
31	Chocolate	Natural Chocolate WONF	Givaudan	Natural Chocolate 96646074		90% in PG
32	Basil	Basil FC	Givaudan	Basil Comores 1463973		10% in DEP
33	Strawberry	Strawberry FC	Givaudan	Strawberry 21051253		20% in DEP
34	Peanut	Natural & Artificial Peanut Favor OS	Takasago International Corp.	TAK-053887		30% in DEP
35	Almond	Benzaldehyde	Givaudan	Benzaldehyde (almond) 1330001		10% in DEP
36	Pine	Pine yarmor oil	Givaudan	Pine oil 8357002		10% in DEP
37	Cucumber	Cucumber FC	Givaudan	Nonedienal (cucumber) 7568001		10% in DEP
38	Violet	Violet FC	Givaudan	Violet 21062503		50% in DEP
39	Vanilla	Vanilla FC	Givaudan	Vanillin 9620001		10% in DEP
40	Caramel	Caramel FC	Givaudan	Caramel 21050343		100%

Table 2

MONEX-40 odor descriptors: items 1–16 originate from the original Sniffin' Sticks ID test and items 17–40 are the 24 newly added odorants (bold font indicates the correct answer, percentages reflect the individual responses of a pilot group of 89 subjects). Descriptors on gray background overlap between 16 original and 24 newly added items.

No.	Descriptors			
1	**Orange (87.6%)**	Strawberry (2.2%)	Blackberry (5.6%)	Pineapple (4.5%)
2	Smoke (3.4%)	**Leather (85.4%)**	Glue (6.7%)	Grass (4.5%)
3	Honey (5.6%)	Chocolate (0.0%)	Vanilla (3.4%)	**Cinnamon (91.0%)**
4	Chives (0.0%)	Onion (0.0%)	Wood (0.0%)	**Peppermint (100.0%)**
5	Coconut (1.1%)	Walnut (0.0%)	**Banana (93.3%)**	Cherry (5.6%)
6	Peach (1.1%)	**Lemon (98.9%)**	Apple (0.0%)	Grapefruit (0.0%)
7	**Liquorice (95.5%)**	Mint (1.1%)	Cherry (1.1%)	Cracker (2.2%)
8	Mustard (2.2%)	Menthol (39.3%)	Rubber (1.1%)	**Turpentine (57.3%)**
9	Onion (2.2%)	**Garlic (96.6%)**	Sauerkraut (1.1%)	Carrot (0.0%)
10	Cigarette (1.1%)	Wine (0.0%)	**Coffee (93.3%)**	Smoke (5.6%)
11	Melon (28.1%)	Orange (16.9%)	Peach (22.5%)	**Apple (32.6%)**
12	Cinnamon (2.2%)	**Cloves (89.9%)**	Pepper (7.9%)	Mustard (0.0%)
13	Pear (10.1%)	Peach (10.1%)	Plum (12.4%)	**Pineapple (67.4%)**
14	**Rose (88.8%)**	Chamomile (10.1%)	Raspberry (1.1%)	Cherry (0.0%)
15	**Anise (85.4%)**	Honey (3.4%)	Rum (7.9%)	Wood (3.4%)
16	Bread (0.0%)	Cheese (1.1%)	**Fish (98.9%)**	Ham (0.0%)
17	Blackberry (5.6%)	**Peach (86.5%)**	Cherry (3.4%)	Orange (4.5%)
18	**Warm milk (36.0%)**	Coca cola (32.6%)	Wood chips (16.9%)	Cream cheese (14.6%)
19	Bacon (1.1%)	**Menthol (95.5%)**	Chocolate (3.4%)	Peach (0.0%)
20	Peach (2.2%)	**Grapefruit (87.6%)**	Grape (5.6%)	Strawberry (4.5%)
21	Rose (4.5%)	Cabbage (18.0%)	Carrot (3.4%)	**Grass (74.2%)**
22	Apple (13.5%)	Melon (38.2%)	**Plum (43.8%)**	Liquorice (4.5%)
23	**Honey (37.1%)**	Almond (27.0%)	Liquorice (15.7%)	Rum (20.2%)
24	Chili (2.2%)	Cloves (3.4%)	**Ginger (94.4%)**	Pepper (0.0%)
25	Cinnamon (4.5%)	Chocolate (2.2%)	Peanut (3.4%)	**Coconut (89.9%)**
26	Grass (0.0%)	Pine (34.8%)	**Lavender (64.0%)**	Rose (1.1%)
27	Lemon (3.4%)	Cabbage (1.1%)	Banana (9.0%)	**Melon (86.5%)**
28	Peach (2.2%)	**Cherry (77.5%)**	Apple (2.2%)	Strawberry (18.0%)
29	Garlic (1.1%)	**Mushroom (57.3%)**	Ham (4.5%)	Wood (37.1%)
30	Lime (73.0%)	Leather (0.0%)	Cedar (24.7%)	Jasmine (2.2%)
31	Bread (9.0%)	Rubber (1.1%)	Peach (0.0%)	**Chocolate (89.9%)**
32	Pepper (14.6%)	Walnut (11.2%)	**Basil (62.9%)**	Mustard (11.2%)
33	**Strawberry (44.9%)**	Blueberry (21.3%)	Raspberry (18.0%)	Cranberry (15.7%)
34	Ham (7.9%)	Cream (12.4%)	Mustard (5.6%)	**Peanut (74.2%)**
35	Honey (6.7%)	**Almond (68.5%)**	Orange (6.7%)	Maple (18.0%)
36	Grass (5.6%)	Cumin (14.6%)	Chili (5.6%)	**Pine (74.2%)**

No.	Descriptors			
37	**Cucumber** (**86.5%**)	Cabbage (7.9%)	Rose (1.1%)	Raisin (4.5%)
38	Cinnamon (2.2%)	Lemon (2.2%)	**Violet** (**94.4%**)	Melon (1.1%)
39	**Vanilla** (**86.5%**)	Almond (0.0%)	Anise (0.0%)	Sugar (13.5%)
40	Raisin (3.4%)	Carrot (1.1%)	Chocolate (3.4%)	**Caramel** (**92.1%**)

Table 3

Descriptive statistics (Mdn: median; IQR: interquartile range) of the 16 original Sniffin' Sticks items, the 24 new items, and the 40-item Monell Extended Sniffin' Sticks Identification subtest (MONEX-40) as well as results of the independent statistical tests investigating sex differences.

	All subjects	Female subjects		Male subjects
n	259	163		96
16-Item original identification test				
Mdn	14	14		14
IQR	2	2		1.75
p			0.810	
24 new items				
Mdn	18	19		17.5
IQR	3	3		3.75
p			0.006	
40-items Monell Extended Sniffin' Sticks Identification Test				
Mdn	32	32		32
IQR	4	4		5
p			0.076	