OXFORD

Original Article

Long-Term Memory for Odors: Influences of Familiarity and Identification Across 64 Days

Stina Cornell Kärnekull[1], Fredrik U. Jönsson[2], Johan Willander[1,3], Sverker Sikström[4] and Maria Larsson[1]

[1]Gösta Ekman Laboratory, Department of Psychology, Stockholm University, Stockholm, Sweden, [2]Department of Psychology, Stockholm University, Stockholm, Sweden, [3]Department of Social Work and Psychology, University College of Gävle, Gävle, Sweden and [4]Department of Psychology, Lund University, Lund, Sweden.

Correspondence to be sent to: Stina Cornell Kärnekull, Gösta Ekman Laboratory at the Department of Psychology, Stockholm University, Stockholm, Sweden. e-mail: stina.cornell.karnekull@psychology.su.se

Abstract

Few studies have investigated long-term odor recognition memory, although some early observations suggested that the forgetting rate of olfactory representations is slower than for other sensory modalities. This study investigated recognition memory across 64 days for high and low familiar odors and faces. Memory was assessed in 83 young participants at 4 occasions; immediate, 4, 16, and 64 days after encoding. The results indicated significant forgetting for odors and faces across the 64 days. The forgetting functions for the 2 modalities were not fundamentally different. Moreover, high familiar odors and faces were better remembered than low familiar ones, indicating an important role of semantic knowledge on recognition proficiency for both modalities. Although odor recognition was significantly better than chance at the 64 days testing, memory for the low familiar odors was relatively poor. Also, the results indicated that odor identification consistency across sessions, irrespective of accuracy, was positively related to successful recognition.

Key words: face recognition, identification consistency, odor recognition, olfactory forgetting

Although empirical evidence is scarce regarding the longevity of olfactory memories, available evidence suggests that the forgetting rate of olfactory representations is rather slow (e.g., Lawless 1978; Murphy et al. 1991; Saive et al. 2014). This study investigated long-term olfactory forgetting as a function of familiarity and identification. Whereas a wealth of evidence indicates a positive influence from semantic factors on episodic retention of verbal and visual information (Larsson 1997; Murphy et al. 1991), the observations are mixed regarding the relationship between odor recognition and semantic factors (e.g., familiarity, identifiability). Early observations in odor recognition over longer time frames (e.g., days, weeks, and 1 year) showed slow forgetting, possibly because of a negligible impact of retroactive interference (Engen and Ross 1973; Lawless and Cain 1975). Based on the rather flat forgetting function, it was assumed that odors are encoded as unitary perceptual events with little attribute redundancy. Typical findings in these early studies

were low initial memory performance, little subsequent forgetting, and no effects of familiarity and identification on subsequent memory performance. Relatedly, research targeting odor memory in the short term (i.e., seconds up to a few minutes), have also shown that memory (e.g., correct same/different discriminations of odor pairs) is relatively unaffected by retention interval (Engen et al. 1973; Jones et al. 1975; Jehl et al. 1994). However, as research on odor memory in the long-term is the primary focus for this study, this literature will be more thoroughly reviewed.

In a seminal study, Engen and Ross (1973) found a virtually flat forgetting curve across a 30-days interval (immediately, 1 day, 7 days, 30 days) in their first experiment. The results showed approximately 70% correct recognition of target odors in a 2-alternative forced choice (2-AFC) task. Memory for familiar and unfamiliar odors, collapsed across retention intervals, did not differ significantly. A subsample of the participants were re-tested after 1 year and performed

significantly better than chance on a recognition task. However, as there is no available information of how this specific subsample was chosen and how the procedure was done, the results should be interpreted with caution. The second and third experiments assessed recognition memory 3 months after encoding and similar performance levels were observed as in their first experiment. Only a weak relationship was found between memory and identification, as indicated by somewhat better memory performance for odors that had been matched with a correct label at encoding as compared with matching it with peers' self-generated names. Relatedly, Lawless and Cain (1975) also used a 2-AFC task and reported of a rather weak relationship between recognition and identification (Ayabe-Kanamura et al. 1997). In this study, performance dropped with approximately 10% in proportion correct recognition over 28 days (10 min, 7 days, 28 days), again in line with the notion of little olfactory forgetting over time. Notably, the study by Engen and Ross (1973) and Lawless and Cain (1975) both lacked an appropriate control task or comparison modality. In contrast, Lawless (1978) compared odor recognition with memory for ambiguous forms and pictures across 4 months (20 min, 7 days, 28 days, 4 months), using a 2-AFC task. The results showed that proportion correct recognition for odors and ambiguous shapes decreased gradually up to 28 days (~15%) but showed no significant decline between 28 days and 4 months. Thus, the slow forgetting rate could also be generalized to equivocal visual information. In comparison, picture recognition was superior to that of odors and ambiguous shapes and decreased with approximately 20%, although with clear ceiling effects during the first month. No effect of perceived familiarity on odor memory was found.

Rabin and Cain (1984) investigated odor recognition across 7 days (10 min, 1 day, 7 days) with a yes/no procedure and found a rather similar memory decrement across that specific time interval as had been observed in previous studies (Engen and Ross 1973; Lawless and Cain 1975; Lawless 1978). However, in contrast to the previous studies on olfactory forgetting, memory appeared to be closely related to both familiarity and identification, with better memory (d') for odors that were perceived as more familiar, identified more accurately, and identified more consistently across encoding and testing. In addition, it was noted that memory for consistently identified odors decreased across time, although not below a d' score of 2.5, which should be considered as a relatively high performance. It was further suggested that verbal codes may facilitate odor memory, although they were not considered as a prerequisite for successful recognition.

Murphy et al. (1991, Experiment 2) investigated young and old participants' recognition memory for common odors, faces of presidents, and engineering symbols across 6 months (15 min, 2 weeks, 6 months) by using a yes/no procedure. To the best of our knowledge, this retention interval is the longest applied in a controlled study on olfactory forgetting. Young participants' memory, as calculated by the sensitivity measure A_z (Stanislaw and Todorov 1999), declined progressively for all stimulus types, although slowest for the symbols. Based on the available data, a higher rate of olfactory forgetting was found than in for example the study by Lawless (1978). However, performance stayed above chance over the 6-month period for all stimuli. Hence, this study showed clear olfactory forgetting across time, as opposed to Engen and Ross (1973), although the authors considered the forgetting rates for all stimulus types as relatively slow. The number of identified odors was positively related to hit rates at the 15 minutes and 2 weeks testing, but not after 6 months. The latter may indicate that semantic or verbal factors play a smaller role for the retention of odors across very long time frames. Relatedly, Lehrner

(1993) assessed odor recognition across 21 days (30 min, 1 day, 7 days, 21 days) with a 2-AFC task. Memory performance decreased slowly but significantly from 77% to 69% proportion correct recognition, although neither hit nor false alarm rates changed significantly across the retention intervals. As was indicated in the study by Rabin and Cain (1984), consistently identified odors were better remembered than inconsistently identified odors (Cain and Potts 1996; Lehrner et al. 1999; Frank et al. 2011; Cessna and Frank 2013). These findings suggest that identification consistency across encoding and testing may be more important than identification accuracy (Lehrner et al. 1999; Frank et al. 2011). Cain et al. (1998) showed that inconsistent identification across repeated trial sessions was rather common. Also, odors that were initially correctly identified were also more consistently identified at later trials than incorrectly identified odors.

Although there is still some controversy regarding the role of semantic factors in episodic odor memory (Herz and Engen 1996) a large number of studies, in particular those of later date, underscore the positive influence of familiarity and identification on olfactory retention (Walk and Johns 1984; Cain and Potts 1996; Larsson 1997; Larsson and Bäckman 1997; Olsson et al. 2009). Presumably, these factors are important as they reflect the degree of semantic knowledge and experience with an odor. For example, whereas a low familiarity rating may reflect little experience and knowledge, a high rating may reflect more specific knowledge, such as the odor's category or name (Larsson 2002). Relatedly, the tip-of-the-nose phenomenon (Lawless and Engen 1977; Jönsson et al. 2005) illustrates that a strong feeling of familiarity might be accompanied with some knowledge of the odor, such as the category of objects, but still with an inability to identify the odor. Previous research has shown that odor memory (d') is better for participants instructed to use verbal encoding strategies (i.e., identifying or associating odors to a life episode) than when no specific strategy was promoted (Lyman and McDaniel 1986, 1990), although this finding was not replicated by Zucco (2003). Still, few studies have investigated the relationship between odor memory and semantic factors across longer time frames. Although the studies by Rabin and Cain (1984) and Murphy et al. (1991) provide information on long-term olfactory retention, evidence is based on familiar everyday odors (Lawless 1978). Hence, knowledge regarding memory proficiency for odors with few semantic attributes is yet unknown.

Faces have been described as a unique class of stimuli (Ellis 1975) and research has indicated that face memory is superior to many other visual stimulus materials (Sato and Yoshikawa 2013). Although odors and faces evidently are 2 fundamentally different stimulus types, they share some common features. First, it has been suggested that both modalities encourage holistic encoding (Cain and Gent 1986; Murphy et al. 1991). For example, research indicates that recognition of the whole face is better than of the comprising parts (Tanaka and Farah 1993). In a similar vein, research indicates that odors are perceived unitarily and that its components are difficult to detect or notice (Stevenson and Attuquayefio 2013). Also, both stimulus types are difficult to name (Cohen and Burke 1993; Larsson 1997; Jönsson and Olsson 2003). For these reasons we included famous and non-famous faces as a comparison modality to odors in the present study (Mair et al. 1980; Murphy et al. 1991).

In sum, this work was designed to explore long-term recognition memory for high and low familiar odors and faces across 4 different retention intervals up to 64 days (i.e., immediately, 4, 16, and 64 days after encoding). Of particular interest was to study memory for odors with few semantic attributes that has been neglected in previous work. One important goal was also to determine the importance of identification consistency for successful long-term recognition memory.

Method

Participants

Eighty-four participants were recruited from the Department of Psychology at Stockholm University. As indicated by a questionnaire, all were in good health and reported that they had a normal sense of smell and visual acuity. One participant was excluded due to an interrupted test session, yielding a final sample of 83 participants (43 women, 40 men). The age ranged from 19 to 50 years for the women (M_{age} = 26.3, SD = 5.8) and from 19 to 44 years for the men (M_{age} = 26.1, SD = 5.8). We used a between-groups design, where participants were tested either immediately (10 men, 10 women), 4 days (10 men, 10 women), 16 days (9 men, 10 women), or 64 days (10 men, 14 women) after encoding. They were given course credits or cinema ticket vouchers for their participation.

The study was approved by the Swedish Research Council and it complies with the Declaration of Helsinki for Medical Research involving Human Subjects. The participants provided written consent.

Materials

Odors

A total of 24 odors were used as odor test stimuli, of which half were high in familiarity (n = 12) and the other half low in familiarity (n = 12). As shown in Table 1, the majority of the low familiar odors had been evaluated and defined as unfamiliar in previous studies (Sulmont et al. 2002; Jönsson et al. 2011). Further, Jönsson et al. (2011) reported that these odors were hard to name or *verbalize*.

Familiarity ratings on a seven-point scale (1 = *very unfamiliar*; 7 = *very familiar*) differed significantly between the 2 odor sets, as shown by a paired samples *t*-test, $t(82)$ = 16.92, $P < 0.001$ ($M_{high familiar}$ = 5.86; $SD_{high familiar}$ = 0.73; $M_{low familiar}$ = 4.16; $SD_{low familiar}$ = 1.05). The complete odor set is presented in Table 1. As shown above, the ratings for the low familiar odors were rated approximately in the middle of the scale. However, as we had 2 odor sets that differed in familiarity, the main objectives of the study could be fulfilled. The odors were prepared in 160-mL opaque glass jars by injecting the liquid on a cotton pad that in turn was covered by another pad. For each participant, 12 odors (6 high familiar, 6 low familiar) served as targets at encoding and 12 odors (6 high familiar, 6 low familiar) as distractors at later testing. The selection of target and lure odors was individually randomized for each participant.

Table 1. Test sets of odors high and low in familiarity

High familiar odors	Low familiar odors
Anise[a]	2-Phenylethyl ethyl ether (PEE)[b,c]
Bitter Almond[a]	2-Phenylethyl pentyl ether (PPE)[b,c]
Clove[a]	3.7-Dimethyloctanenitrile (DON)[b,c]
Lemon[a]	Bornyl acetate (BOR)[b,c]
Lily of the valley	Citrowanil (2-ethenyl-2-methyl benzene-propanal) (CIW)[b,c]
Mushroom	Dec-9-en-1-ol (DEO)[b,c]
Peppermint	Leather[a]
Petrol	Menthyl acetat (19)[b,c]
Soap	Methyl benzoate (MBE)[b,c]
Tar	Malodorous composition[a]
Vanilla[a]	Tridec-2-enenitrile (TDN)[b,c]
Violet[a]	Ylang-Ylang[a]

[a]Purchased from Essencefabriken, Stockholm.

[b]Donated by the Department of Organic Chemistry at Stockholm University.

[c]Odors low in familiarity were sampled from Sulmont et al. (2002).

Faces

In a pilot experiment, 20 subjects were presented with 200 faces. Half of them were photos of famous Swedish persons (e.g., politicians, actors, and news presenters), and the other half photos of nonfamous Swedish persons. All photos were edited to the same size and converted to black-and-white photos. All faces were displayed on a white background and under the face a black collar was inserted in order to conceal variation in clothing. The pictures were presented on a 22-inch LCD monitor controlled by an E-prime 2.0 script (Psychological Software Tools). For each face, participants rated perceived familiarity on a seven-point scale (1 = *very unfamiliar*; 7 = *very familiar*). Since we noticed high performance and risks of ceiling effects in face recognition in previous research (Murphy et al. 1991), a larger set size of pictures than that of odors was selected. The 80 most familiar and the 80 least familiar faces were selected for the main study, with 40 female and 40 male faces per familiarity category. Familiarity ratings on a seven-point scale differed significantly between the 2 face sets, as shown by a paired samples *t*-test, $t(82)$ = 25.70, $P < 0.001$ ($M_{high familiar}$ = 5.13; $SD_{high familiar}$ = 1.22; $M_{low familiar}$ = 1.58; $SD_{low familiar}$ = 0.53). For each participant, 80 faces (40 high familiar, 40 low familiar) served as targets at encoding and 80 faces as distractors (40 high familiar, 40 low familiar) at testing. The target and lure selection was individually randomized and the participant registered the responses with a keyboard.

Procedure

All participants attended one encoding session and one test session. They were tested individually in a well-ventilated room by the same experimenter. First, the participant provided informed consent. Next, a questionnaire was distributed regarding educational background, age, health status, sensory (visual, auditory, olfactory) aptitude, and smoking habits. At both encoding and testing, odors and faces were presented in 2 separate blocks and the order of modality was counterbalanced across participants. For a given participant the modality order was the same in the encoding and the test session. At encoding, the participant was informed that the stimuli would be memorized for a later memory test.

Odor encoding and testing

The participant was blindfolded throughout the odor presentation and the experimenter told when a new odor was about to be presented. The participant was instructed to breathe normally when an odor was presented under the nose. Each odor was presented for approximately 5 s, after which the participant first rated the familiarity of the odor on a seven-point scale (1 = *very unfamiliar*; 7 = *very familiar*), followed by an attempt to identify it by name as precisely as possible (e.g., banana rather than fruit). To minimize effects of adaptation, the inter-stimulus presentation interval was 30 s. The experimenter recorded all answers.

At testing all odors from the encoding session was presented along with the same number of distractor odors, that is, in total 24 odors (12 high familiar, 12 low familiar). The tasks were to first decide whether each of the presented odors had been presented at the encoding session (yes/no) and then to identify it.

Identification consistency and accuracy for odors

Three categories of identification performance, regardless of accuracy across encoding and testing were analyzed: (i) consistent identification at encoding and testing (i.e., identifying identically or similarly), (ii) inconsistent identification (i.e., different identifications at encoding and testing, identification at encoding only, or

identification at testing only), and (iii) no identification at encoding and testing (i.e., omissions). Regardless of accuracy, an odor was scored as consistently identified when the same or nearly the same label was used across sessions (e.g., lemon–lemon candy).

Moreover, identification accuracy was analyzed. The generated names were dichotomously scored as correct or incorrect. Both exact matches as well as names that were close to the veridical name (e.g., lemon candy for lemon) were scored as correct. Incorrect names and omissions were scored as incorrect identifications. Inter-rater reliability for the scoring of odor names was assessed between 2 of the authors, and showed high agreement between the scorings ($r = 0.95$, $P < 0.001$).

Face encoding and testing

Each face was shown for 4 s, after which the participant rated its perceived familiarity on a seven-point scale (for a maximum of 5 s; 1 = *very unfamiliar*; 7 = *very familiar*), followed by an attempt to name the person shown (for a maximum of 8 s). When the time limits were reached the program continued automatically.

At test, 80 target faces along with the remaining 80 distractor faces were presented, that is, in total 160 faces (80 high familiar, 80 low familiar). For each face, a yes/no recognition task and an identification task followed. The same time frames were used at test as at encoding. The generated names were dichotomously scored as correct or incorrect, where a correct answer represented a correct first name, surname, or both. Incorrect answers included inaccurate names and omissions.

For the purpose of this study, only the familiarity ratings and recognition data for the faces were used.

Results

For the analyses of odor and face recognition memory, we computed hits, false alarms, and d' scores (Stanislaw and Todorov 1999). The mean proportions were analyzed with separate 4 (retention interval: immediate, 4, 16, 64 days) × 2 (modality: odor, face) × 2 (familiarity: high familiar, low familiar) mixed ANOVAs, with repeated measures on the last 2 factors. Only significant main and interaction effects are reported. The alpha level was set at 0.05. Initially we also included gender as a factor in the ANOVA, as previous research has found that women are better in identifying odors than men (Doty and Cameron 2009). Because no reliable main or interaction effects involving gender appeared, we excluded this factor in the analyses reported below.

d' Scores

To obtain a recognition discrimination index, d' was calculated—an unbiased sensitivity measure commonly applied in recognition memory research. In the signal detection theory model, d' is defined as the z-transformed difference between proportions hits (H) and false alarms (F); [$d' = z(H) - z(F)$] (Macmillan and Creelman 2005). Hit and false alarm proportions of 1 and 0 were adjusted according to Macmillan and Creelman's (2005) suggestions, namely to $1 - 1/(2N)$ and $1/(2N)$, respectively. As illustrated in Figure 1 and confirmed by the ANOVA, d' scores were significantly higher for faces ($M = 2.22$; $SD = 0.80$) than for odors ($M = 1.38$; $SD = 0.74$) [$F(1, 79) = 179.40$, $P < 0.001$, $\eta_p^2 = 0.69$]. Performance was significantly better for high familiar ($M = 2.05$; $SD = 0.72$) than low familiar stimuli ($M = 1.59$; $SD = 0.57$) [$F(1, 79) = 53.47$, $P < 0.001$, $\eta_p^2 = 0.40$]. Further, d' decreased significantly across retention intervals [$F(3, 79) = 52.39$, $p < 0.001$, $\eta_p^2 = 0.67$], and Tukey's HSD post hoc test confirmed that all comparisons except that between immediate and 4 days were statistically significant ($Ps < 0.01$). However, the decrease in performance across time differed between odors and faces, as reflected in the significant interaction between modality and retention interval [$F(3, 79) = 5.79$, $P = 0.001$, $\eta_p^2 = 0.18$]. To disentangle this interaction, 2 separate one-way ANOVAs were conducted for odors and faces with retention interval as independent variable. Tukey's HSD post hoc test confirmed that the differences in odor recognition performance were significant between all retention intervals ($Ps < 0.05$) except between 4 and 16 days [$F(3, 79) = 15.02$, $P < 0.001$]. In contrast, face recognition memory decreased significantly across all retention intervals except between immediate and 4 days testing ($Ps < 0.01$) [$F(3, 79) = 39.65$, $P < 0.001$]. Furthermore, the main ANOVA showed that retention interval interacted with familiarity [$F(3, 79) = 3.18$, $p = 0.028$, $\eta_p^2 = 0.11$], indicating that the difference between high familiar and low familiar stimuli decreased significantly across retention intervals. This finding was followed up with paired t-tests, with a Bonferroni corrected alpha level, which showed that although the familiarity effect was significant at immediate, 4,
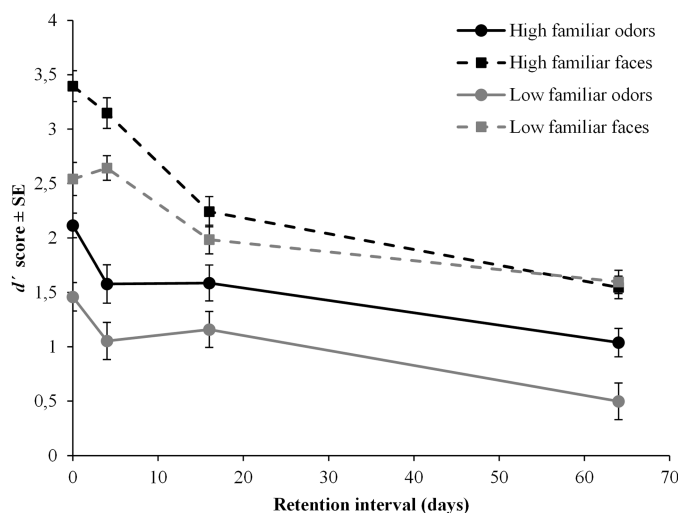

Figure 1. d' Scores (±SE) as a function of modality, familiarity, and retention interval.

and 16 days testing ($Ps < 0.01$), d' for high familiar and low familiar stimuli did not differ significantly at the 64 days testing.

As noted above, it was of particular interest to explore the forgetting function for olfactory information with less semantic attributes. As illustrated in Figure 1, immediate recognition performance for low familiar odors was 1.46 ($SD = 0.58$) and declined to 0.50 ($SD = 0.83$) over the 64 days interval. However, although the performance drop was substantial, recognition memory performance at the 64 days testing was significantly better than chance (score of zero), as indicated by a one-sample t-test [$t(23) = 2.95$, $P = 0.007$].

Hit rates

The ANOVA revealed significant differences in hit rates between odors ($M = 0.72$; $SD = 0.17$) and faces ($M = 0.80$; $SD = 0.13$), [$F(1, 79) = 17.43$, $P < 0.001$, $\eta_p^2 = 0.18$], and between high familiar ($M = 0.85$; $SD = 0.12$) and low familiar ($M = 0.67$; $SD = 0.16$) stimuli [$F(1, 79) = 132.04$, $P < 0.001$, $\eta_p^2 = 0.63$]. Further, there was significant decrement across time [$F(3, 79) = 17.64$, $P < 0.001$, $\eta_p^2 = 0.40$] and a significant interaction between modality and retention interval [$F(3, 79) = 3.07$, $P = 0.03$, $\eta_p^2 = 0.10$]. In order to disentangle this interaction, 2 one-way ANOVAs were performed for odors and faces, respectively, with retention interval as independent variable. For odors, Tukey's HSD post hoc test [$F(3, 79) = 6.29$, $P = 0.001$] confirmed that the drop in hit rates after immediate testing was significant ($Ps < 0.05$ vs. later recognition performances), but remained constant from 4 to 64 days (Figure 2). Hit rates for faces, however, decreased significantly across all time intervals ($Ps < 0.02$) except between immediate and 4 days testing [$F(3,79) = 22.86$, $P < 0.001$].

False alarm rates

New items incorrectly recognized as old were classified as false alarms. The ANOVA yielded significant differences in false alarms between odors ($M = 0.26$; $SD = 0.15$) and faces ($M = 0.13$; $SD = 0.10$) [$F(1, 79) = 73.12$, $P < 0.001$, $\eta_p^2 = 0.48$]. Also, significantly more false alarms were generated for high familiar ($M = 0.21$; $SD = 0.14$) than low familiar stimuli ($M = 0.17$; $SD = 0.11$) [$F(1, 79) = 5.64$, $P = 0.02$, $\eta_p^2 = 0.07$]. False alarms increased significantly across retention intervals [$F(3, 79) = 14.67$, $P < 0.001$, $\eta_p^2 = 0.36$]. Also, a significant interaction between modality and retention interval [$F(3,$

79) $= 3.49$, $P = 0.02$, $\eta_p^2 = 0.12$] reflected that the increment of false alarms across retention intervals differed between modalities (Figure 3). Two separate one-way ANOVAs were performed for the modalities, to further investigate this interaction. Tukey's HSD post hoc test confirmed that the false alarm rate for odors was stable up to 16 days, but increased reliably after 64 days compared with the previous test occasions ($Ps < 0.01$) [$F(3, 79) = 9.71$, $P < 0.001$]. False alarm rates for faces were significantly higher at 64 days than at immediate and 4 days ($Ps < 0.001$), and higher at 16 days than 4 days testing ($P < 0.05$), although no other comparisons were significant [$F(3, 79) = 11.40$, $P < 0.001$]. Furthermore, the main ANOVA revealed a significant interaction between familiarity and retention interval [$F(3, 79) = 6.68$, $P < 0.001$, $\eta_p^2 = 0.20$], indicating that the general effect of familiarity differed across time. The interaction was followed up with paired t-tests, with a Bonferroni corrected alpha level, showing that the false alarm rate was significantly higher for low familiar than high familiar stimuli at immediate testing ($P < 0.01$), whereas the opposite pattern was found at 64 days ($Ps < 0.05$). False alarm rates did not differ between the odor sets at the 4 and 16 days testing. Moreover, familiarity interacted significantly with modality [$F(1, 79) = 12.0$, $P = 0.001$, $\eta_p^2 = 0.13$], reflecting significantly higher false alarm rates for high familiar than low familiar faces ($P < 0.001$), whereas no difference was observed between familiar and low familiar odors.

Identification consistency, identification accuracy, and hit rate for odors

The aim was to investigate the relationship between identification consistency (regardless of accuracy) and recognition. Because only target odors could be analyzed for consistency we focused on hit rate performance. As already noted in the method section, 3 categories of consistency performance were analyzed: (i) consistent identification, (ii) inconsistent identification, and (iii) no identification at encoding and testing (i.e., omissions).

As shown in Table 2, consistently identified odors had higher mean hit rate performance than both inconsistently identified odors and odors that had not been identified at either of the test sessions. The proportion of consistently identified odors decreased significantly across retention intervals [$F(3, 79) = 9.61$, $P < 0.001$]. Furthermore, individual proportions of consistent identification
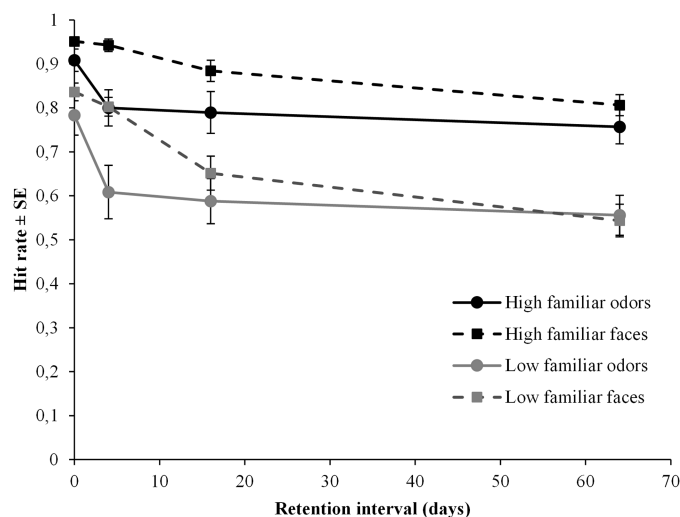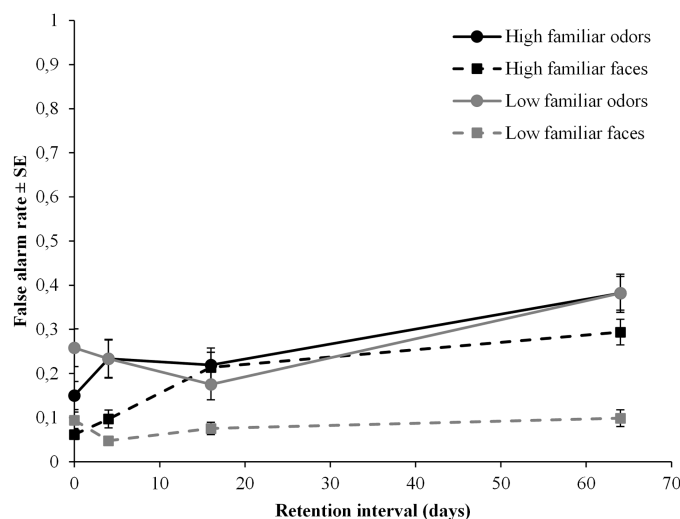

Figure 2. Proportion hit rate (±SE) as a function of modality, familiarity, and retention interval.

Figure 3. Proportion false alarm rate (±SE) as a function of modality, familiarity, and retention interval.

Table 2. Proportion hit rate as a function of identification (ID) consistency between encoding and testing for high and low familiar odors

	Retention interval				
	Immediate	4 days	16 days	64 days	Grand mean
Consistent ID at encoding and testing					
High familiar odors	0.98 (87)	0.99 (73)	0.97 (67)	0.95 (59)	0.97 (286)
Low familiar odors	0.97 (35)	0.87 (23)	0.82 (17)	0.93 (15)	0.91 (90)
Inconsistent ID at encoding and testing					
High familiar odors	0.76 (25)	0.55 (44)	0.55 (38)	0.61 (76)	0.60 (183)
Low familiar odors	0.69 (55)	0.51 (76)	0.60 (65)	0.48 (94)	0.56 (290)
No ID at encoding and testing					
High familiar odors	0.63 (8)	0.00 (3)	0.44 (9)	0.78 (9)	0.55 (29)
Low familiar odors	0.73 (30)	0.67 (21)	0.44 (32)	0.60 (35)	0.60 (118)

Data is presented across retention intervals. The total number of test items (*N*) is presented in the parentheses.

(from a total of 12 target odors) were used for a number of Pearson product-moment correlations. First, we noted that proportion consistent identification was highly correlated with proportion inconsistent identification, $r(81) = -.69$, $P < 0.001$. More importantly, and as suggested by the descriptive data (Table 2), there was a positive correlation between the proportion of consistently identified odors and hit rate, $r(81) = 0.50$, $P < 0.001$. Moreover, 2 separate correlation analyses for high and low familiar odors both indicated positive relationships between consistent identification and hit rate (high familiar: $r(81) = 0.57$, low familiar: $r(81) = 0.39$, Ps < 0.001). It should also be noted that a much higher proportion consistent identification was observed for high familiar ($M = 0.57$, $SD = 0.26$) than for low familiar odors ($M = 0.18$, $SD = 0.18$). Although the proportion of consistent identification was higher for correctly ($M = 0.50$) than incorrectly ($M = 0.07$) identified high familiar odors, the hit rate was of similar magnitude ($M = 0.94$ vs. $M = 0.98$, respectively).

Four separate correlation analyses were performed for each retention interval (including all target odors), indicating positive correlations between consistent identification and hit rate at all occasions, although only statistically significant at the 4 days testing, $r(18) = 0.60$, $P < 0.01$. The correlations at immediate ($r(18) = 0.21$, $P = 0.37$), 16 days ($r(17) = 0.37$, $P = 0.12$), and 64 days ($r(22) = 0.26$, $P = 0.23$) were moderate but nonsignificant. A follow-up hierarchical regression analysis was performed to investigate whether the correlations differed significantly from each other. Specifically, proportion consistent identification, retention interval, and an interaction term (proportion consistent identification by retention interval) were entered as independent variables in this order with hit rate as the criterion variable. Proportion consistent identification was the only factor that significantly predicted hit rate (b = 0.54, $t(79) = 4.70$, $P < 0.001$). The observation that the interaction term did not account for any significant variation in hit rate ($P = 0.15$) suggests that the effect of consistent identification on hit rate was of equal magnitude across the retention intervals.

Moreover, the relationship between identification accuracy and memory performance was investigated for high familiar odors, as only these odors had expected labels that could be scored for accuracy. Descriptive data across items indicated that the hit rate was highest for correctly identified stimuli at encoding and testing ($M = 0.97$) and lower when identification was only correct at one of the 2 sessions ($M = 0.54$) or not identified at any of the sessions ($M = 0.55$). Thus, these observations corroborated the findings from the identification consistency analyses.

Follow-up analysis on olfactory forgetting (*d′*) for low familiar odors

The low familiar odors were specifically selected to be impossible to identify (i.e., they had no correct labels) to minimize semantic

influence on recognition memory. Nevertheless, and as reported above, a rather small proportion of the low familiar odors were still named consistently. In a follow-up analysis, consistently named odors were removed for each participant and new hit rates were calculated for the remaining (low familiar) target odors. For the majority of participants (88%) the number of remaining odors ranged between 4 and 6 items. New d' scores for the low familiar odors were calculated for each participant, based on the new hit rates and the already existing false alarms rates.

The d' scores were entered into an ANOVA as the dependent variable and retention interval (immediate, 4, 16, 64 days) as independent variable. The results indicated that d' scores differed significantly between retention interval, $[F(3, 79) = 4.75, P = 0.004, \eta_p^2 = 0.15]$, and Tukey's HSD post hoc test confirmed that performance at the 64 days testing differed significantly from the immediate ($P < 0.01$) and 16 days testing ($P < 0.05$). Performance did not differ significantly between any of the other retention intervals. Mean d' scores (SD) at immediate, 4 days, 16 days, and 64 days testing were 1.19 (0.70), 0.92 (0.79), 1.06 (0.73), and 0.38 (0.85). Memory at the 64 days testing was followed up with a one-sample t-test $[t(23) = 2.19, P = 0.039]$, which indicated that performance was significantly better than chance (score of zero).

Analysis of response bias

Response bias was calculated in order to investigate whether there were any differences in response styles across the retention intervals. In signal detection theory, c (criterion) is a measure of response bias and defined as $c = -\frac{1}{2} [z(H) + z(F)]$ (Macmillan and Creelman 2005). A score of zero reflects a neutral response style. A repeated measures ANOVA was performed with c as dependent variable, and modality and retention interval as independent variables. The results indicated that response bias differed between modalities $[F(1, 79) = 10.16, P = 0.002]$, such that the response bias for faces was moderately conservative ($M = 0.17, SD = 0.32$) and virtually neutral for odors ($M = 0.02, SD = 0.38$). Retention interval did not influence bias ($P = 0.29$) and the interaction between retention interval and modality was not reliable ($P = 0.10$). Descriptive data on bias for odors across retention interval did not reveal any clear pattern $[M_{immediate} = -0.10 (0.37), M_{4\ days} = 0.09 (0.45), M_{16\ days} = 0.18 (0.31), M_{64\ days} = -0.06 (0.35)]$, whereas bias for faces showed a small nominal increment across retention interval $[M_{immediate} = 0.11 (0.30), M_{4\ days} = 0.19 (0.27), M_{16\ days} = 0.16 (0.34), M_{64\ days} = 0.22 (0.37)]$.

Discussion

The aim of the study was to investigate long-term odor recognition memory and forgetting as a function of familiarity and identification. The results showed significant forgetting for both odors and faces, as indicated by the decrement in d' scores across 64 days. Although a significant interaction between modality and retention on odor memory was observed, the forgetting functions for the 2 modalities were not fundamentally different. Previous research has suggested that the forgetting rate for odors may be similar to simple figures (Lawless 1978), symbols and faces (Murphy et al. 1991), but slower than forgetting for pictures (Lawless 1978). In the present work, the forgetting rate was comparable to some of the previous reports (Lawless 1978; Rabin and Cain 1984), although methodological differences between studies should be acknowledged. Furthermore, the hit rates for odors were stable across the 4, 16, and 64 days testing, whereas false alarm rates increased significantly between 16 and 64 days testing. These findings differ from the observed face

recognition performance, where hit rates decreased and false alarm rates increased with longer retention intervals; the so called mirror effect (Glanzer and Adams 1985). Also, the results indicated a relatively conservative response style for faces whereas it was virtually neutral for odors. Although not of primary interest, we also note that face memory in general was superior to odor memory. This is noteworthy as the set size of faces was substantially larger than for odors and that odors were perceived as more familiar than the faces. This outcome concurs with the notion of an exceptional human ability to recognize faces (Yin 1969; Sato and Yoshikawa 2013). In congruence with a large body of evidence, recognition memory (d') was better for high familiar than low familiar stimuli. Although the relative difference in perceived familiarity was smaller for odors than for faces, the familiarity effect on recognition was true for both modalities (Klatzky and Forrest 1984; Rabin and Cain 1984; Murphy et al. 1991; Trinkler et al. 2009).

Although people in general have a rather poor ability to identify odors (Larsson 1997; Jönsson and Olsson 2003), identification proficiency is positively related to memory performance (Rabin and Cain 1984; Frank et al. 2011). Our findings extend and replicate this observation by showing that identification consistency may be a stronger predictor for successful recognition than correct identification (Frank et al. 2011). First, irrespective of accuracy, high familiar consistently identified odors were associated with the highest hit rates. However, it should be noted that identification consistency was markedly lower for incorrectly identified odors (Cain et al. 1998). Second, the hit rates did not differ for odors identified correctly at one session only or for odors that were not identified at any of the sessions. The influence of consistent identification on memory was also examined at an individual level and indicated that proportion consistent identification across encoding and testing was positively correlated to hit rates irrespective of odor familiarity. Moreover, a follow-up analysis where consistently identified items were removed from the low familiar odor set, revealed an even lower memory performance across all retention intervals. Although memory was still significantly better than chance at the 64 days testing, the performance level was relatively low.

Overall, the pattern of findings suggests that olfactory memory is dependent on odor familiarity and consistent identification. Consistent identification is related to familiarity and correct identification, as indicated by a higher frequency of consistent identifications for high than low familiar odors as well as for correctly than incorrectly identified (high familiar) odors. The positive effects of familiarity and identification on memory could be due to several factors. For example, it has been suggested that identification influences the perception of the odor or makes it more meaningful that ultimately result in better memory performance (de Wijk et al. 1995). Also, it is possible that the odor names generated at encoding may serve as retrieval cues at recognition (Crowder and Schab 1995).

Recently, Cessna and Frank (2013) suggested that the consistency effect varies with regard to the temporal order of the identification and recognition tasks (Frank et al. 2011). Specifically, their results suggested that the consistency effect was largest when the identification task preceded the recognition task, but markedly reduced for the reversed order. This implied both better memory performance for consistent identifications as well as worse performance for inconsistent identifications. Intriguingly, the effect was not reliable when identification was assessed after the recognition test. Thus, these findings suggest that the relative impact of odor name generation on odor recognition depends on whether such strategies are promoted in the specific task setting (Lehrner 1993). Although the participants

in our study were instructed to identify the odor after the recognition judgment, it is possible that covert naming or related semantic associations occurred and facilitated recognition (Larsson and Bäckman 1993).

Although gender effects were not the primary aim of the present study, it is of interest to note that men and women did not differ in recognition memory for neither odors nor faces. Previous research presents a mixed pattern of findings regarding potential gender differences in olfactory functions (Brand and Millot 2001; Doty and Cameron 2009), although evidence appears to be most consistent for a female superiority effect in odor identification (Doty et al. 1985; Fusari and Ballesteros 2008) and odor recognition (Oberg et al. 2002; Choudhury et al. 2003). However, the fact that the manifestation of the gender effects in odor memory vary largely between studies and that some studies have failed to find gender effects (Lawless and Cain 1975; Larsson et al. 2006) indicates that further research is warranted.

In conclusion, we reported of substantial long-term forgetting for odors and faces across 64 days. Memory for faces was better than for odors, but no fundamental differences in forgetting functions between modalities were observed. Familiarity was related to memory for both modalities, and for odors consistent identification exerted a significant influence on performance.

Funding

Acknowledgments

References

Ayabe-Kanamura S, Kikuchi T, Saito S. 1997. Effect of verbal cues on recognition memory and pleasantness evaluation of unfamiliar odors. *Percept Mot Skills*. 85(1):275–285.

Brand G, Millot JL. 2001. Sex differences in human olfaction: between evidence and enigma. *Q J Exp Psychol B*. 54(3):259–270.

Cain WS, de Wijk R, Lulejian C, Schiet F, See LC. 1998. Odor identification: perceptual and semantic dimensions. *Chem Senses*. 23(3):309–326.

Cain WS, Gent JF. 1986. Use of odor identification in clinical testing of olfaction. In: Meiselman HL, Rivlin RS, editors. *Clinical measurement of taste and smell*. New York (NY): MacMillan Publishing Company. p. 170–186.

Cain WS, Potts BC. 1996. Switch and bait: probing the discriminative basis of odor identification via recognition memory. *Chem Senses*. 21(1):35–44.

Cessna TC, Frank RA. 2013. Does odor knowledge or an odor naming strategy mediate the relationship between odor naming and recognition memory? *Chem Percept*. 6(1): 36–44.

Choudhury ES, Moberg P, Doty RL. 2003. Influences of age and sex on a microencapsulated odor memory test. *Chem Senses*. 28(9):799–805.

Cohen G, Burke DM. 1993. Memory for proper names: a review. *Memory*. 1(4):249–263.

Crowder RG, Schab FR. 1995. Odor recognition memory. In: Schab FR, Crowder RG, editors. *Memory for odors*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers. p. 9–20.

de Wijk RA, Schab FR, Cain WS. 1995. Odor identification. In: Schab FR, Crowder RG, editors. *Memory for odors*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers. p. 21–37.

Doty RL, Applebaum S, Zusho H, Settle RG. 1985. Sex differences in odor identification ability: a cross-cultural analysis. *Neuropsychologia*. 23(5):667–672.

Doty RL, Cameron EL. 2009. Sex differences and reproductive hormone influences on human odor perception. *Physiol Behav*. 97(2):213–228.

Ellis HD. 1975. Recognizing faces. *Br J Psychol*. 66(4):409–426.

Engen T, Kuisma JE, Eimas PD. 1973. Short-term memory of odors. *J Exp Psychol*. 99(2):222–225.

Engen T, Ross BM. 1973. Long-term memory of odors with and without verbal descriptions. *J Exp Psychol*. 100(2):221–227.

Frank RA, Rybalsky K, Brearton M, Mannea E. 2011. Odor recognition memory as a function of odor-naming performance. *Chem Senses*. 36(1):29–41.

Fusari A, Ballesteros S. 2008. Identification of odors of edible and nonedible stimuli as affected by age and gender. *Behav Res Methods*. 40(3):752–759.

Glanzer M, Adams JK. 1985. The mirror effect in recognition memory. *Mem Cognit*. 13(1):8–20.

Herz RS, Engen T. 1996. Odor memory: review and analysis. *Psychon Bull Rev*. 3(3):300–313.

Jehl C, Royet JP, Holley A. 1994. Very short term recognition memory for odors. *Percept Psychophys*. 56(6):658–668.

Jones BP, Moskowitz RH, Butters N. 1975. Olfactory discrimination in alcoholic Korsakoff patients. *Neuropsychologia*. 13(2):173–179.

Jönsson FU, Møller P, Olsson MJ. 2011. Olfactory working memory: effects of verbalization on the 2-back task. *Mem Cognit*. 39(6):1023–1032.

Jönsson FU, Olsson MJ. 2003. Olfactory metacognition. *Chem Senses*. 28(7):651–658.

Jönsson FU, Tchekhova A, Lönner P, Olsson MJ. 2005. A metamemory perspective on odor naming and identification. *Chem Senses*. 30(4):353–365.

Klatzky RL, Forrest FH. 1984. Recognizing familiar and unfamiliar faces. *Mem Cognit*. 12(1):60–70.

Larsson M. 1997. Semantic factors in episodic recognition of common odors in early and late adulthood: a review. *Chem Senses*. 22(6):623–633.

Larsson M. 2002. Odor memory: a memory systems approach. In: Rouby C, Schaal B, Dubois D, Gervais R, Holley A, editors. *Olfaction, taste, and cognition*. Cambridge: Cambridge University Press. p. 231–245.

Larsson M, Bäckman L. 1993. Semantic activation and episodic odor recognition in young and older adults. *Psychol Aging*. 8(4):582–588.

Larsson M, Bäckman L. 1997. Age-related differences in episodic odour recognition: the role of access to specific odour names. *Memory*. 5(3):361–378.

Larsson M, Oberg C, Bäckman L. 2006. Recollective experience in odor recognition: influences of adult age and familiarity. *Psychol Res*. 70(1):68–75.

Lawless HT. 1978. Recognition of common odors, pictures, and simple shapes. *Percept Psychophys*. 24(6):493–495.

Lawless HT, Cain WS. 1975. Recognition memory of odors. *Chem Sens Flav*. 1(3):331–337.

Lawless H, Engen T. 1977. Associations to odors: interference, mnemonics, and verbal labeling. *J Exp Psychol Hum Learn*. 3(1):52–59.

Lehrner JP. 1993. Gender differences in long-term odor recognition memory: verbal versus sensory influences and the consistency of label use. *Chem Senses*. 18(1): 17–26.

Lehrner JP, Glück J, Laska M. 1999. Odor identification, consistency of label use, olfactory threshold and their relationships to odor memory over the human lifespan. *Chem Senses*. 24(3):337–346.

Lyman BJ, McDaniel MA. 1986. Effects of encoding strategy on long-term memory for odours. *Q J Exp Psychol*. 38A(4): 753–765.

Lyman BJ, McDaniel MA. 1990. Memory for odors and odor names: modalities of elaboration and imagery. *J Exp Psychol Learn*. 16(4): 656–664.

Macmillan NA, Creelman CD. 2005. *Detection theory: a user's guide*. 2nd ed. Mahwah (NJ): Erlbaum.

Mair R, Capra C, McEntee WJ, Engen T. 1980. Odor discrimination and memory in Korsakoff's psychosis. *J Exp Psychol Hum*. 6(3):445–458.

Murphy C, Cain WS, Gilmore MM, Skinner RB. 1991. Sensory and semantic factors in recognition memory for odors and graphic stimuli: elderly versus young persons. *Am J Psychol*. 104(2):161–192.

Oberg C, Larsson M, Bäckman L. 2002. Differential sex effects in olfactory functioning: the role of verbal processing. *J Int Neuropsychol Soc*. 8(5):691–698.

Olsson MJ, Lundgren EB, Soares SC, Johansson M. 2009. Odor memory performance and memory awareness: a comparison to word memory across orienting tasks and retention intervals. *Chem Percept*. 2(3): 161–171.

Rabin MD, Cain WS. 1984. Odor recognition: familiarity, identifiability, and encoding consistency. *J Exp Psychol*. 10(2):316–325.

Saive AL, Royet JP, Plailly J. 2014. A review on the neural bases of episodic odor memory: from laboratory-based to autobiographical approaches. *Front Behav Neurosci*. 8:240.

Sato W, Yoshikawa S. 2013. Recognition memory for faces and scenes. *J Gen Psychol*. 140(1):1–15.

Stanislaw H, Todorov N. 1999. Calculation of signal detection theory measures. *Behav Res Methods*. 31(1):137–149.

Stevenson RJ, Attuquayefio T. 2013. Human olfactory consciousness and cognition: its unusual features may not result from unusual functions but from limited neocortical processing resources. *Front Psychol*. 4:819.

Sulmont C, Issanchou S, Köster EP. 2002. Selection of odorants for memory tests on the basis of familiarity, perceived complexity, pleasantness, similarity and identification. *Chem Senses*. 27(4):307–317.

Tanaka JW, Farah MJ. 1993. Parts and wholes in face recognition. *Q J Exp Psychol A*. 46(2):225–245.

Trinkler I, King JA, Doeller CF, Rugg MD, Burgess N. 2009. Neural bases of autobiographical support for episodic recollection of faces. *Hippocampus*. 19(8):718–730.

Walk HA, Johns EE. 1984. Interference and facilitation in short-term memory for odors. *Percept Psychophys*. 36(6):508–514.

Yin RK. 1969. Looking at upside-down faces. *J Exp Psychol*. 81(1):141–145.

Zucco GM. 2003. Anomalies in cognition: olfactory memory. *Eur Psychol*. 8(2):77–86.