Journal of Experimental Psychology

Vol. 100, No. 2 — October 1973

LONG-TERM MEMORY OF ODORS WITH AND WITHOUT VERBAL DESCRIPTIONS

TRYGG ENGEN[1]
Brown University

BRUCE M. ROSS
Catholic University of America

Three experiments demonstrated the uniqueness of memory for odors. Unlike visual and auditory memories, immediate recognition tests for 20 or more odorants produced numerous errors, but there was little further retention loss for periods up to 3 mo. Auxiliary results showed that neither verbal labeling nor odor familiarity aided memory, while long-term retention held up even when there were no instructions to memorize. Use of similar odors as test distractors impaired recognition significantly, but it remained well above chance after 3 mo. The hypothesis is suggested that odors are coded as unitary perceptual events with little attribute redundancy; this leads to poor immediate retention but great subsequent resistance to distortion of immediately retained odors.

The Proustian view is that odors are not forgotten to the same extent as are other perceptual events. Is there any factual validity for this claim of the artist? Studies pertaining to memory of odors have been reported (Bannister & Zangwill, 1941; Ford, 1940; Gordon, 1925; Jellinek, 1964), but these deal more with qualitative and introspective aspects rather than facts concerning memory under controlled conditions that are comparable to studies of memory of sights and sounds in the Ebbinghaus tradition. Potential differences in memory in different modalities have been noted by Engen, Kuisma, and Eimas (1973). They performed a short-term memory experiment with odors and found that the ability to recognize odors, unlike memory of sights and sounds, seems to be unaffected by the duration of the retention interval between the original smelling of an odorant and a later recognition task. However, the maximum interval in this experiment was only 30 sec. and thus too short to provide a test of what is meant by odor memory in the literature (Moncrieff, 1970).

A corollary to the hypothesis that man's memory for odors is outstanding is that he is also able to identify literally thousands of different odors. Informational analysis of odor judgment does not support this popular assumption (Engen & Pfaffmann, 1959, 1960). Man's ability to identify odors seems to fall well within the range of his ability to identify events in other modalities (Miller, 1956). These informational results are based on experiments in which S must identify odorants presented singly by recalling the correct verbal label in a method described as "absolute judgment" of odorants. However, results supporting the writers who claim excellence for odor memory may be obtained with a method involving discrimination and recognition of odorants presented side by side for comparison. Shepard (1967) devised

[1] Requests for reprints should be sent to Trygg Engen, Walter S. Hunter Laboratory of Psychology, Brown University, Providence, Rhode Island 02912.

such a recognition memory procedure to deal with a similar problem in the memory of sights and sounds.

There were 2 parts to each of the 3 experiments discussed in this study. In general, in the first part of each experiment *S* smelled in turn a number of odorants selected at random from the whole odor library of the laboratory. In the second part of each experiment *S* was presented a smaller set of those *old* odorants smelled in the first part paired with *new* odorants drawn at random from the odor library. Instructions were to judge which member of the pair was the *old* one. The main independent variable was the interval between the first part and the second part of an experiment; the main dependent variable was the average number of correct choices. Additional variables were the effects on the recognition score of the number of alternative odorants inspected in the first part, the familiarity of *S* with the odorant, the nature of the code for the odor, the pleasantness of the odor, and the subjective similarity of the *old* and the *new* odorants.

The *S*s were men and women students and faculty members (N = 179) of Brown University. Although some of them had participated in other psychological experiments, none of them had any special experience with odors or with the details of the present experiments.

A set of 110 diverse odorants representing odors of familiar household products, such as onion and vinegar, as well as more or less familiar chemical compounds and perfumes, such as amyl acetate and Fabergé, were used. Each odorant was presented to *S* on a piece of cotton for sniffing (Engen, 1965). The cotton was attached to a glass rod which was stored in a test tube when not in use and easily removed for repeated sniffing by *S*. The test tube was presented to *S* on an acrylic-plastic test tube rack.

There were 3 separate experiments: the first established a forgetting curve, while the second and third extended the long-term forgetting interval and added new odor identification and recognition test variables.

Experiment I

Method. The *S* was told specifically that this was an experiment concerned with the ability to remember odors, but he was not told the details of the procedure or the names of the odorants. In the first part of the experiment each of 37 *S*s smelled 48 odorants selected at random from the whole set of 110. There was a pause between odorants to counteract fatigue and adaptation. Partly to encourage *S* to pay close attention to each odorant and partly to obtain additional data, he was asked (*a*) to judge whether he was familiar or unfamiliar with each odor, and (*b*) to indicate whether he liked it, felt indifferent about it, or disliked it. The *S* was permitted to smell the odorant again, if he wished. (It should be noted here that while the present experiment is concerned with memory it does not specify the degree to which an odor is originally known, such as might be done in a learning experiment, e.g., by specification of an acquisition criterion.)

In the second part of the experiment, *S* was presented 21 pairs of odorants with the instructions to indicate which of each pair had also been included among the odorants inspected in the first part of the experiment. One member of each pair was an *old* odorant selected at random from the 48 odorants *S* had inspected in the first part, which was paired with a different (*new*) member selected at random from the remaining odorants of the whole set. The order of presentation of the pairs and the placement of the *old* odorant in each pair on the right or the left side of *S* was also determined at random.

This forced-choice recognition part of the experiment was administered to 4 different subgroups of 15 *S*s after 4 different periods of time had elapsed: Group 1, immediately following the inspection part of the experiment; Group 2, after 1 day (24 hr.); Group 3, 1 wk.; and Group 4, 1 mo. Except for 7 *S*s in Group 4, each *S* was tested twice in such a way that for the 3 longer retention intervals half of the *S*s (7 or 8) of each group had no previous recognition test but the other half had been tested for the prior retention interval. Thus, 8 *S*s of Group 1 were also in Group 2, the remaining 7 of Group 2 were also in Group 3, and the remaining 8 of Group 3 were also in Group 4 with 7 new *S*s. The odorants used to test *S*s the second time were selected at random from the odorants remaining in the relevant sets and thus neither *old* nor *new* odorants were presented more than once in the recognition task.

Results. The results are presented in Figure 1. The t test on the recognition scores corresponding to the percentages indicates that all points plotted are significantly different from chance ($p < .001$). The recognition difference between 0 and 30 days is not significant; only the difference between the highest point (1 day) and the lowest (30 days) is significant ($.01 < p < .05$). No reliable differences were observed between *S*s who had served in 1 earlier recognition test and those for whom it was the first test, that is, comparing the 2 halves of *S*s for retention intervals of 1, 7, and 30 days, either for each interval separately or the 3 intervals combined for 30 *S*s. The product-moment correlation coefficient between the scores made by each of the *S*s on the first and second recognition test was not reliable ($r = .06$).

The analysis of the odorants correctly recognized showed that 53% had been described as "familiar" and 47% as "unfamiliar" during the first part of this experiment in which *S*s were asked to inspect them. Although these results, based on all retention intervals, appear to confirm the expectation that previous familiarity may be helpful to *S*, the difference is not reliable. The explanation for this, according to spontaneous comments by many of the *S*s, seems to be that some of the previously unfamiliar odors were recognized in the second part of the experiment because of their novelty as "strange" or "interesting" odors.

A comparison of the percentage of correct vs. incorrect choices in the recognition experiment for odors judged as *like*, *dislike*, or *indifferent* in the first part of the experiment for all retention intervals combined suggested no correlation between preference judgment and recognition accuracy. There is apparently some tendency to categorize unfamiliar odors as *dislike* and *indifferent* rather than *like*, and this indicates that preference judgments were correlated with familiarity judgments. That is, unfamiliar odors were rarely (total of 11%) judged as *like*, but familiar odors were so judged about half of the time (46%). However, these relationships are not directly pertinent to the present problem.

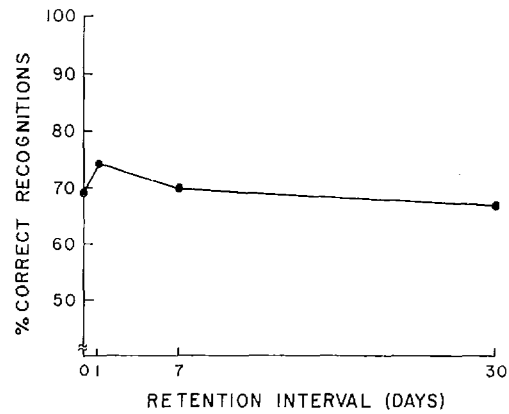

FIG. 1. Mean percent correct recognition of the *old* odorant in test pairs as a function of the duration of the retention interval.

EXPERIMENT II

Method. The purpose of this experiment was (*a*) to supplement the findings of Experiment I, and (*b*) to observe the effect of the similarity of the *old* and *new* odorants on the recognition scores. Sixty-eight *S*s participated in this experiment as part of their laboratory requirement for an introductory psychology course. In contrast to *S*s in Experiment I, these *S*s were given no information about the later recognition, but were presented the first part of the experiment as an investigation of the ability of people to identify odors. For this purpose they were presented a diverse set of 20 odorants, 1 at a time in a different irregular order for each *S*, with instructions to make an identification by a short written description or, preferably, by a specific label. At the conclusion of the session the correct answers were provided as feedback information, and *S* did smell the odorants again at that time.

The second part of the experiment was presented 3 mo. later (at the end of the semester). The procedure was the same as in Experiment I with 20 pairs presented to *S* in an irregular order with instructions to indicate which member of each pair was one of the odors smelled in the first part of the experiment. For half of the *S*s ($n = 34$) the *new* member had been chosen at random from the remaining odorants in the large set. For the other half, the *new* member had been selected by *E* to be similar to the *old* in terms of the use or the kind of odorant. For example, bourbon was matched to Scotch whisky, amyl acetate to butyl acetate, and onion to garlic. At the end of the recognition task, *S*s were asked to judge the similarity of the members of each pair on the numerical scale from 0 (completely different) to 100 (identical) in order to supplement the judgment made by *E*.

Results. When the number of alternative odorants inspected in the first part of the present experiment was reduced from 48 (Experiment I) to 20, the recognition score increased from about 67% to 77% for the comparable group tested with *different* distractors. The t test was made after arc sine transformation of the percentage of correct choices for individual *S*s indicated that this difference is reliable ($p < .001$). In other words, *S*s who inspected a smaller number of alternative odorants in the first part of this experiment did better in the recognition task even though they had a much longer retention interval than *S*s in the previous experiment (1 vs. 3 mo.).

The main purpose of Experiment II was to observe the effect of the distractor vis-à-vis the similarity of the *old* and *new* members of the pair of odors. The mean number of correct judgments for the groups with the *similar* distractors was 12.8 (64%) of a possible 20 compared with 15.4 (77%) for those with *different* distractors. The t test indicates that this is a significant difference ($p < .001$). The subsequent judgments of these pairs by *S*s on the similarity scale from 0–100 resulted in an average score of 37.8 for the *different* pairs and 58.6 for the *similar* pairs, an increase of 56% in rated subjective similarity. This difference is also significant ($p < .001$) and verifies the judgment of *E*. The product-moment correlation between the number of *S*s whose recognition judgment of a certain pair was incorrect and the mean similarity score for that pair was .50 for all *S*s in both groups ($n = 68$).

Finally, it may be of interest to consider these results in light of the ability of *S*s to identify the odorants. In the first part of the experiment the number of correct free identifications for the average *S* was only 6.3 (31.5%) of a maximum of 20 with a range 1–12 and a standard deviation of 2.1 for the 68 *S*s. If one accepts good associations, for example calling amyl acetate "airplane glue" or alcohol "disinfectant," the average score increased to 10.9 (54.5%). These results seem consistent with the data on familiarity in Experiment I, which together imply that the ability to recognize odors exceeds the ability to label single odorants.

Experiment III

Method. The purpose of this experiment was to obtain information about the effect of coding or verbal labeling on odor memory. Seventy-four *S*s participated in this experiment as part of the laboratory requirement for an introductory psychology course. The stated purpose of the experiment was to assess *S*'s ability to identify odors, and no information was given about the later recognition experiment.

Again, there were 2 parts to the experiment. In the first part of this experiment *S* was required to match each of the 20 odorants (same as in Experiment II), presented in an irregular order, to 1 of 20 labels provided on a list. Each of the odorants was numbered and *S*'s task was to write its number next to the appropriate label. As in the other experiments, there were no time restrictions on the task. There were 2 lists of labels; half of the *S*s received one list, and the other half received the other list. The *S*s were randomly assigned. One list consisted of the common names for the odorants, and the other frequent (as judged informally by *E*) associations made to the same odorants by *S*s in Experiment I in addition to or instead of these common names. For example, antiseptic, Life Savers, and root beer were the labels used on this second list instead of alcohol, wintergreen, and licorice on the first list. Feedback information about the accuracy of the matching was again provided.

The second part of this experiment took place after an interval of 3 mo., as in Experiment II. In this case, the *new* odorant was again selected at random from the large set.

Results. The results are summarized in Table 1. When *S*s were provided a list of the correct labels and asked to match the odor to 1 of the labels on the list, the mean number of correct judgments was 13.9. This is significantly better ($p < .001$) than the average results of 7.9 obtained by the group of *S*s who were asked to match the odors to a list consisting of the correct associations made by *S*s in Experiment II. It should be emphasized that this mean of 7.9 from matching is better than the mean free-identification score of 6.3 of *S*s in Experiment II. Matching is clearly easier than free identification, and the nature of the labels or the code used in the matching task is also an important factor.

The main purpose of Experiment III was to observe the effect of such coding in the first part of the experiment on the subsequent recognition task. The mean recognition score for *S*s asked to match in terms of the correct label was 15.3 (76%) compared with 14.1 (70%) for those asked to match in terms of the associations. The *t* test indicates that this difference is reliable ($.01 < p < .05$). Note that this recognition performance is very similar to that of the group in Experiment II with *different* distractors (77% correct) who initially were asked to give free identifications of the same set of 20 odorants.

TABLE 1

PERCENT CORRECT IDENTIFICATIONS AND LATER RECOGNITIONS OF ODORANTS MATCHED TO CODES BASED ON ASSOCIATIONS MADE TO THE ODORANTS BY OTHER *S*s OR ON USE OF THEIR COMMON NAMES

Response	Code	
	Association (e.g., "bubble-gum")	Name of substance (e.g., "raspberry flavor")
Identification	39%	69%
Recognition	70%	76%

DISCUSSION

The present study clearly supports the view that odor memory is less influenced by the passage of time than are auditory and visual memories. Shepard (1967), for example, showed in the same kind of experiment that *recognition* of *old* vs. *new* pictures started near 100% but then gradually decreased and approached the 50% chance level after a delay of 120 days. By contrast, the present data show a relatively constant level near 70% for the same kind of task. It should be noted further that memory of odors was "incidental" in Experiments II and III in that *S*s were given no instructions to remember odors. When due allowance is made for experimental design differences, the retention percentages found in these experiments appear to be in line with the percentages found for intentional memory in Experiment I, but they are strikingly higher than the results one would expect for long-term *incidental* memory in the major sense modalities.

The retention curve in Experiment I and the retention results in Experiments II and III would predict that long-term memory of odors was not at all close to termination after 3 mo. A follow-up study strongly supported this prediction. It was possible to test 15 *S*s from Experiment I a third time 1 yr. after exposure with 6 odorants that had not been used in either of the 2 previous recognition tests as distractors. The average recognition score was 64.6%, which is reliably above the 50% chance level ($.005 < p < .01$).

In the first experiment there is a suggestion of an effect resembling the so-called "reminiscence effect," because there is an improvement in the recognition score between the immediate test and the test after 1 day has passed following inspection. However, the difference is not statistically reliable and, as in the case of a similar effect in short-term memory for odors (Engen et al., 1973), the explanation for the initial increment in the score is probably associated with fatigue from the long initial testing session or with some sensory adaptation due to long exposure to moderately strong odors.

In general, there is no essential difference between the present long-term and the previously reported short-term odor memory results. A nearly flat retention function is maintained across all time intervals. The relatively wide and usually random sampling of odorants made for each individual *S* would indicate that the present results are not limited to a special sample of odors. The failure of recognition accuracy to change over time also cannot be attributed to lack of sensitivity of the measure of memory employed, as measurement was shown to be sensitive to the customarily effective variables of distractor similarity and number of odorants to be remembered. Nevertheless, these variables had to be strongly applied in order to produce significant changes in recognition accuracy. Thus a 56% increase in rated subjective similarity produced only a 13% decrease in recognition accuracy, and an increase from 20 to 48 odorants produced a 10% decrease in accuracy. (This latter comparison is inexact since the comparison test intervals differed.) Results from both variables suggest that long-term recognition of odors is a "robust" phenomenon that can maintain itself even under difficult test conditions.

Another important finding, in line with our previous results for short-term odor memory, was the relatively poor maximum recognition score of 70% for either short- or long-term memory in the present study compared to near 100% accuracy for short-term visual recognition (Shepard, 1967). The explanation for this difference in modalities will require a better understanding of modality coding. One obvious consideration is that pictures and most other visual configurations possess multiple perceptual attributes for coding, while odors produce a much more unitary perceptual experience. Therefore, initially, it is likely that for the task of recognition a redundant number of visual attributes are retained by everyone. On the other hand, the all-or-none coding of odors encourages immediate errors but by the same token renders them quite resistant to subsequent confusions.

The introspective Proustian explanation as to why odor memory in one's personal life is so excellent is that odors are associated with significant emotional experiences. In fact, it is also a common idea that odors have no names of their own but are known through association with other things. As Nabokov (1970) puts it, "as we know, memory can restore to life everything except smells, although nothing revives the past so completely as a smell that was once associated with it [p. 60]." The present odors were of no known relevance to *S*s and could only have been so by chance. Having *S*s rate a randomly selected sample of odors as to degree of liking or familiarity does not of course indicate how odors are learned or become meaningful to the individual. If it were possible to take advantage of personal relevance or to establish it through conditioning in the present type of experiment, recognition scores might indeed be made to approach 100% accuracy and remain close to that level, unaffected by the passage of time.

The present study exhibits other interesting aspects of the problem of coding. The *S*s who were provided with the correct labels did better than those provided only the associations made to the odors by a peer group ($p < .05$), but they did no better than those given no code. In other words, rather than concluding that the list of correct labels helps *S* recognize odors, it appears that providing him with a poor code—in this case somebody else's associations to the odors—can interfere with his own code and his consequent ability to make accurate recognitions.

Another coding comparison showed, as expected, that recognition of *old* vs. *new* odors presented pair-wise was much easier than free identification of single odors by specific labels. The present results indicate that the average *S* could retain up to 15 bits of information in making decisions in a recognition task compared with only 4 bits obtained previously (Engen & Pfaffmann, 1960) in a labeling task, even when labels were *S*'s own associations. However, a score of 15 bits is still far below the level required to retain information about thousands of odors as has been claimed by many writers (e.g., Moncrieff, 1970).

All results taken together demonstrate conclusively that time decay is only a minor factor in retention of odors. As most of the odorants would not be encountered in everyday living, the related theory of forgetting due to "item disuse" is also rejected. The Proustian insight is vindicated! As retention explanations in terms of interference can only be invoked where there is some performance deficit, they too have little pertinence. Nor can a meaningful distinction be made between short-term and long-term memory stages. These conclusions are strengthened by the fact that rehearsal could have played only a minimal role in retention maintenance. The sizable number of odorants administered in each experiment guarded against using imagery for rehearsal, although it is doubtful that unpracticed *S*s could make effective use of odor imagery anyway. Lack of rehearsal, or at the least ineffectiveness, is also supported by the finding that *S*s did not perform more accurately either with more familiar or more preferred odorants. As cited above, other results showed that verbal labeling provided no rehearsal help, and use of preordained verbal associations handicapped retention. Thus, as our discussion has stressed, a considerable degree of uniqueness must be attributed to the odor-coding process since it is less accurate but far more enduring than coding of the major sense modalities.

REFERENCES

Bannister, H., & Zangwill, O. L. Experimentally induced olfactory paramnesia. *British Journal of Psychology*, 1941, **32**, 155–175.

Engen, T. Psychophysical analysis of the odor intensity of homologous alcohols. *Journal of Experimental Psychology*, 1965, **70**, 611–616.

Engen, T., Kuisma, J. E., & Eimas, P. Short-term memory of odors. *Journal of Experimental Psychology*, 1973, **99**, 222–225.

ENGEN, T., & PFAFFMANN, C. Absolute judgments of odor intensity. *Journal of Experimental Psychology*, 1959, **58**, 23–26.

ENGEN, T., & PFAFFMANN, C. Absolute judgments of odor quality. *Journal of Experimental Psychology*, 1960, **59**, 214–219.

FORD, F. A. The nose in the chemistry laboratory. *Journal of Chemical Education*, 1940, **17**, 17–19.

GORDON, K. The recollection of odors. *Journal of Experimental Psychology*, 1925, **8**, 225–239.

JELLINEK, J. S. Introduction to and critical review of modern methods of sensory analysis (odour, taste and flavour evaluation) with special emphasis on descriptive sensory analysis (flavour profile method). *Journal of Nutrition and Dietetics*, 1964, **1**, 219–260.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

MONCRIEFF, R. W. *Odours*. London: William Heinemann Medical Books, 1970.

NABOKOV, V. *Mary*. (Trans. by Michael Glenny) New York: McGraw-Hill, 1970.

SHEPARD, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156–163.

(Received November 15, 1972)