

Similarity judgments and recognition memory for some common spices

F. NOWELL JONES, KELYN ROBERTS, and ERIC W. HOLMAN
University of California, Los Angeles, California 90024

A group of subjects rated the similarities of all 55 pairs of 11 spices. Nonmetric multi-dimensional scaling of this judgment matrix indicated a single underlying dimension, interpretable as pleasantness. Short-term recognition memory for the same spices was tested in separate groups of subjects required to remember one, three, and five stimuli, respectively. Performance was best for the group remembering one stimulus, but was not affected by the number of stimuli preceding the correct stimulus in the test. The confusion matrix for the one-stimulus group, symmetrized by removal of a stimulus bias, showed a significant correlation with the judgment matrix.

Judgments of similarity or frequencies of confusion among stimuli are commonly represented by distances in hypothetical psychological spaces. According to these models, stimuli close together in the space should be judged more similar or confused more often than stimuli farther apart. Although many sets of data have been described by various spatial configurations, there has been much less effort to determine whether both judgment and confusion data for a given set of stimuli can be described by the same configuration. Shepard (1974) reports that for English consonants, the configurations that fit the confusion data of Miller and Nicely (1955) are systematically different from those that fit the judgment data of Peters (1963). For Dutch vowels, on the other hand, Van der Kamp and Pols (1971) emphasize the similarity between the configurations that describe their own confusion data and the judgment data of Pols, Van der Kamp, and Plomp (1969).

In a broad sense of the term, the work referred to above is in the tradition of "associationism," and therefore explores the relationships among stored meanings (Titchener, 1909, p. 25). Since it is conceivable that unsuspected verbal labeling may underlie judgments of some stimuli, it is of advantage, if one wishes to generalize results beyond the realm of language, to use stimuli which have a relatively low likelihood of forming, at least readily, verbal associations. Odors fit this description. It has been suggested, for example, that associations between odors and verbal stimuli (Davis, 1975) are difficult to form, and that the retention of odors in recognition memory may behave differently from other kinds of stimuli (Engen, Kuisma, & Eimas, 1973; Engen & Ross, 1973; Lawless & Cain, 1975).

Supported by grants from the Research Committee, University of California, Los Angeles, and by National Institute of Mental Health Grant 1 RO3 MH 29174-04 and National Science Foundation Grant GB 42218 to Eric Holman. The data analyses were performed at the Campus Computing Network.

Recent work by Lawless and Engen (1977) suggests, however, that certain aspects of odor memory may resemble those for verbal memory. We shall be able to comment further on this point later on. That the formation of odor to word associations proceeds rather slowly is further evidenced by attempts to measure channel capacity by the learning of odor names (Engen & Pfaffman, 1960; Jones, Note 1). (Although it is not directly relevant to the purposes of this article, we should like to call attention to a growing literature on the multidimensional analysis of odors. Entry references include Berglund, Engen, & Ekman, 1973; Harper, 1975; Moskowitz & Barbe, 1976; Schiffman, Robinson, & Erickson, 1977; Yoshida, 1976.)

METHOD

Stimuli

Eleven common herbs or seasonings were chosen: basil (*Labiatae Ocimum basilicum*), bay (*Lauineae Umbellularia californica*), celery (*Umbelliferae Apium graveolens*), marjoram (*Labiatae Origanum majorana*), mint (*Labiatae Mentha spicata*), oregano (*Labiatae Origanum vulgare*), parsley (*Umbelliferae Petroselinum arispum*), rosemary (*Labiatae Rosemarinus officinalis*), sage (*Labiatae Salvia officinalis*), tarragon (*Compositae Artemisia dracuncululus*), and thyme (*Labiatae Thymus vulgaris*). All were dried and packaged by Schilling. The stimuli were prepared by placing a teaspoon of the herb in a plastic bottle and adding a small amount of distilled water. Subjects could not see into the bottle and the mouth of the bottle was covered with brass mesh.

Experiment 1

Twenty-two students from the introductory psychology course subject pool served as subjects in order to satisfy a course requirement. The subjects were asked to judge, on a 9-point scale (9 being most similar), the similarity of all the possible 55 pairs of stimuli. The presentation order of the pairs was random for each subject.

Experiment 2

There were four groups of subjects. These were obtained from the same population as Experiment 1, but were not the same individuals. Groups 1, 2, and 3 contained 26 subjects each, and Group 1a contained 18.

A recognition memory paradigm was used. The subject was given a stimulus or a set of stimuli chosen at random from the 11 to remember. Then he was given the 11 stimuli in a random order and asked to rate the degree to which he felt each stimulus was an old or new item. This 6-point confidence scale ran from -3 to +3. Groups 1, 2, and 3 were given, respectively, one, three, or five stimuli to remember on each of six trials. Group 1a was run later with the same procedure as Group 1 in order to equalize the number of particular stimuli on each trial so that a fairer estimate of the confusion matrix could be obtained. The subjects proceeded at their own pace, with the usual trial lasting around 1 min, with approximately 1 min between trials.

RESULTS

Experiment 1

Multidimensional configurations were fit to the matrix of average similarity judgments using KYST, a descendent of the multidimensional scaling programs of Kruskal (1964a, 1964b) and Young and Torgerson (1967). The analysis assumed the Euclidean metric and a monotonic relation between distances and data. To avoid local minima, both the standard starting configurations in KYST were used, namely the projection of the rotated solution obtained with one more dimension, and the configuration obtained assuming a linear relation between distances and data. The better solutions in one, two, and three dimensions had stress values of 0.150, 0.070, and 0.048, respectively. This pattern of stress values is similar to those obtained by Wagenaar and Padmos (1971) from Monte Carlo data generated by adding random error to a one-dimensional underlying configuration. The horizontal axis of Figure 1 gives the one-dimensional solution for the present data. The dimension can be interpreted as pleasantness, since, in general, subjects disliked the smell of parsley and celery, and liked the smell of mint.

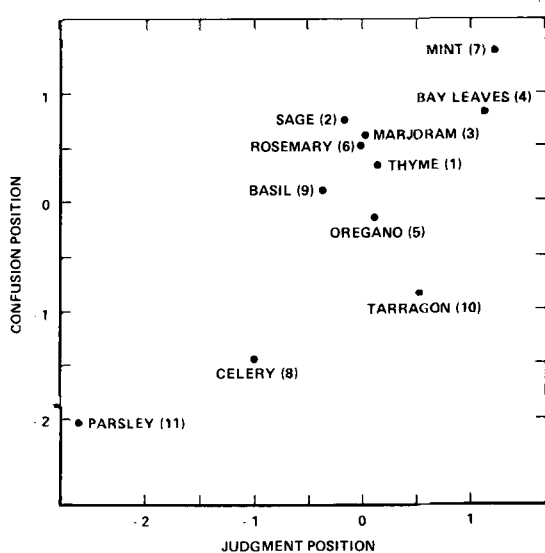


Figure 1. The unidimensional solution for similarity judgments plotted against relative confusion in recognition memory.

Table 1
The Proportion of Correct Responses for Each Spice as a Function of Experimental Group

Spice	Group 1	Group 2	Group 3
Basil	.53	.71	.53
Bay	.67	.49	.51
Celery	.94	.72	.64
Marjoram	.50	.62	.54
Mint	.64	.69	.64
Oregano	.31	.45	.69
Parsley	.88	.85	.68
Rosemary	.38	.50	.49
Sage	.79	.44	.55
Tarragon	.62	.54	.68
Thyme	.85	.78	.82
Overall	.66	.62	.62

Note—Group 1 = one to remember; Group 2 = three to remember; Group 3 = five to remember.

Experiment 2

To estimate the discriminability of stimuli from foils, d' measures were calculated for each subject on the basis of the rating data averaged over the six trials. Group 1 subjects had an average d' of 1.10; Group 2 (three stimuli), 0.62; and Group 3 (five stimuli), 0.45. The differences were reliable [$F(2,75) = 19.93$, $p < .01$]. Post hoc analyses revealed that Group 1 differed reliably from Groups 2 and 3, $p < .05$ and $p < .01$, respectively. When the data were pooled in each group, memory operating characteristic curves (MOC curves) could be plotted. The area under the curve was 0.75 for Group 1, 0.65 for Group 2, and 0.61 for Group 3. Overall percent correct scores (where correct includes yeses for the stimuli correct, nos for the foils) were: 72% for Group 1, 62% for Group 2, and 58% for Group 3.

Table 1 gives the differential recognizability of the individual stimuli according to the number of stimuli to be remembered. The relationship, if any, between the number of stimuli to be remembered and the recognizability of the different stimuli is not obvious.

Figure 2 gives the percent correctly identified as a function of "lag" from when the stimulus was presented until the correct choice was presented. Here "lag" is measured in terms of number of intervening stimuli and foils. The reader is reminded that both choice and order of stimuli and the order of the foils were randomly determined. There may be a slight downward trend in Group 1, but the tendency is slight and not borne out by the other groups.

Finally, two average confusion matrices were derived, respectively, from the errors and the confidence ratings made by Groups 1 and 1a combined. Since the results of further analyses were very similar for both matrices, only the latter will be discussed here. To denote the entries of this matrix, let $p(xy)$ be the average rating of foil y when stimulus x was presented. The matrix was asymmetric in the sense that $p(xy)$ was usually not the same as $p(yx)$, although the distances used to represent

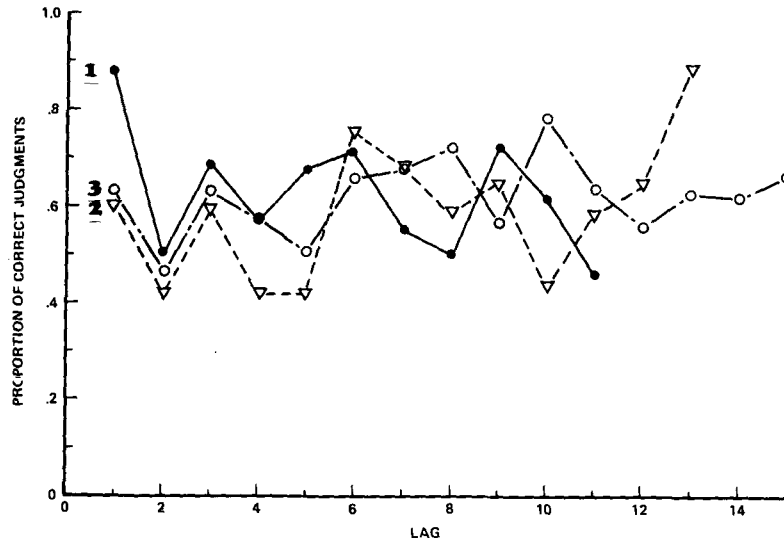


Figure 2. Correct responses as a function of lag. Correct responses include both hits and correct rejections.

such matrices are symmetric. Several methods can be used to derive symmetric from asymmetric matrices. The most common approach is to average the two halves of the matrix: if $s(xy) = [p(xy) + p(yx)]/2$, then the matrix s is symmetric. Alternatively, several models attribute asymmetry to bias. If the bias depends upon stimuli, then such models entail a bias vector, b , and a symmetric matrix, sr , such that:

$$b(x) > b(y) \text{ and}$$

$$sr(wx) > sr(yz) \text{ imply } p(wx) > p(yz). \quad (1)$$

If the bias depends upon foils, then there will be a different symmetric matrix, sc , such that:

$$b(x) < b(z) \text{ and}$$

$$sc(wx) > sc(yz) \text{ imply } p(wx) > p(yz). \quad (2)$$

Luce's (1959) choice model satisfies (2); Tversky's (1977) feature models satisfy (1) if stimulus features attract more attention and (2) if foil features attract more attention. The present analyses compare the three symmetric matrices s , sr , and sc . The latter two matrices were inferred by a computer program that minimizes the violations of (1) or (2), using an iterative algorithm similar to the one described by Holman (1978). The minimum proportions of violations of (1) and (2) were .052 and .062, respectively, indicating an adequate fit for either model.

The three matrices were scaled using KYST as in Experiment 1. The stress values in one, two, and three dimensions were, respectively, 0.165, 0.096, and 0.051 for s , 0.222, 1.130, and 0.079 for sr , and 0.264, 0.129, and 0.065 for sc . Comparison with the Monte Carlo

data of Wagenaar and Padmos (1971) again suggests one-dimensional underlying configurations. The higher stress values for the matrices derived by extracting stimulus or foil bias suggest slightly more error in these matrices than in the average confusion matrix or in the data of Experiment 1.

Perhaps a more important comparison, however, is in the consistency between each of the confusion matrices and the judgment data of Experiment 1. One measure of such consistency is the correlation between corresponding entries in the confusion and judgment matrices. Since (1) and (2) determine only the rank order of entries in sr and sc , the Goodman-Kruskal (1954) gamma was used as a measure of rank correlation. Values of gamma between the judgment data and s , sr , and sc were, respectively, 0.30, 0.39, and 0.17; the first two of these are significantly greater than 0 ($p < .01$), and the last is not. Another measure of consistency is the correlation between the positions of the stimuli in the one-dimensional scaling configurations. Since these positions are (practically) unique up to linear transformations, product-moment correlations were used. Correlations between positions derived from the judgment data and from s , sr , and sc were, respectively, 0.69, 0.84 and 0.49; again, the first two are significantly greater than 0 ($p < .05$ and $p < .01$, respectively), and the last is not. Both for the original scores and for the derived configurations, therefore, removing a stimulus bias from the confusion data produced the greatest consistency between experiments, removing a response bias produced the least consistency, and averaging was in between.

The vertical axis of Figure 1 gives the one-dimensional solution for sr , the confusion matrix with stimulus bias removed. As before, the dimension seems to reflect pleasantness. The integers in parentheses

after the stimulus names are the rank-order values of the stimulus bias. No relation is evident between bias and position in the configuration. Perhaps the observed bias, in which thyme was most prominent, is compatible with the opinion expressed by one herb specialist that "thyme is essential to every kitchen because of its abilities to enhance combinations of herbs" (Hooker, 1971, p. 21).

Correlations between confusions and judgments also provided evidence on the appropriate dimensionality for the configurations. When the judgment data were predicted by the configurations in one, two, and three dimensions derived from s_r , gamma was respectively 0.46, 0.40, and 0.40. When s_r was predicted by the judgment configurations in one, two, and three dimensions, gamma was, respectively, 0.37, 0.37, and 0.35. For predicting across experiments, therefore, the one-dimensional configurations were as accurate as higher-dimensional configurations, and in fact were as accurate as the data themselves.

DISCUSSION

Our basic results both from the multidimensional scaling procedures and from the recognition memory experiment correspond to other work in the literature. A common result of multidimensional analysis with odors is a very large pleasantness-unpleasantness dimension (cf. Schiffman, 1974). No doubt the relatively homogeneous set of odors which we used as stimuli contributed to the amount of variance assignable to this dimension. It was also not unexpected that memory for odors was revealed as rather poor and not much affected by lag (Engen et al., 1973). More of interest in the present context, however, are two other outcomes. First, the correlation between judgments of similarity and confusions in recognition, and, second, the asymmetry of the confusion matrix.

The use of a homogeneous set of odors reduces the chance of finding a significant correlation between our two approaches. That a significant relationship was indeed found increases our confidence in the usefulness of the multidimensional scaling approach. Not only does the magnitude of the relationship indicate that our data are reliable, but also our results can be regarded as being a demonstration of validity in the general sense that converging operations tend toward the same results. The subjective structure of spice quality operated to help determine more than one kind of response.

At first glance, this conclusion may seem to be at variance with Davis' (1977) finding that the degree of similarity among odors to be recognized had little effect on accuracy of recognition. However, the outcomes of the two experiments are not contradictory. Davis used three "target" sets of odors and presented all three sets in the recognition situation. Although the three target sets differed greatly in within-set similarity,

they differed relatively slightly in average similarity between the target items and the distractor items in the other two sets. The present experiment found that the latter similarity was related to recognition, and this is consistent with Davis' finding of insignificant differences in recognition among the three target sets.

The necessity of dealing with asymmetries in the confusion data is of basic theoretical interest. It is obvious that a simple distance model is insufficient. Moreover, introducing a response or foil bias, as in Luce's (1959) choice model, practically destroys the correlation with the judgment data. Tversky's (1977) feature models are consistent with the observed stimulus bias, but are not obviously compatible with the one-dimensional representation that gave an acceptable account of the symmetric part of the data. Models that incorporate stimulus biases and distances may be worth examining. This bias could well be due to the strategies subjects employ in encoding the stimulus odors.

We would conclude, then, that our odors have behaved not too differently from semantic material, and that, at least so far as the one-stimulus case is concerned, there is a stable subjective structure underlying judgments of similarity and also determining confusions. We cannot, however, specify whether this structure necessarily involves distances, features, or both.

REFERENCE NOTE

1. Jones, F. N. Information content of olfactory quality. In N. Tanyolac (Ed.), *Theories of odor and odor measurement*. NATO Summer School, Istanbul, 1968.

REFERENCES

- BERGLUND, B., BERGLUND, U., ENGEN, T., & EKMAN, G. Multidimensional analysis of twenty-one odors. *Scandinavian Journal of Psychology*, 1973, **14**, 131-137.
- DAVIS, R. G. Acquisition of verbal associations to olfactory stimuli of varying familiarity and to abstract visual stimuli. *Journal of Experimental Psychology*, 1975, **104**, 134-142.
- DAVIS, R. G. Acquisition and retention of verbal associations to olfactory and abstract visual stimuli of varying similarity. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3**, 37-51.
- ENGEN, T., KUISMA, J. C., & EIMAS, P. R. Short-term memory for odors. *Journal of Experimental Psychology*, 1973, **99**, 222-225.
- ENGEN, T., & PFAFFMAN, C. Absolute judgments of odor quality. *Journal of Experimental Psychology*, 1960, **59**, 214-219.
- ENGEN, T., & ROSS, B. Long term memory of odors with and without verbal descriptions. *Journal of Experimental Psychology*, 1973, **100**, 221-227.
- GOODMAN, L. A., & KRUSKAL, W. H. Measures of association for cross-classifications. *Journal of the American Statistical Association*, 1954, **49**, 732-754.
- HARPER, R. Some chemicals representing particular odor qualities. *Chemical Senses and Flavor*, 1975, **1**, 353-357.
- HOLMAN, E. W. Completely nonmetric multidimensional scaling. *Journal of Mathematical Psychology*, 1978, in press.
- HOOKE, A. *Herb cookery*. San Francisco: 101 Productions, 1971.
- KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964, **29**, 1-27. (a)

- KRUSKAL, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964, **29**, 115-129. (b)
- LAWLESS, H. T., & CAIN, W. S. Recognition memory for odors. *Chemical Senses and Flavor*, 1975, **1**, 331-337.
- LAWLESS, H., & ENGEN, T. Associations to odors: Interference, mnemonics, and verbal labeling. *Journal of Experimental Psychology: Human Learning & Memory*, 1977, **3**, 52-59.
- LUCE, R. D. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959.
- MILLER, G. A., & NICELY, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, **27**, 338-352.
- MOSKOWITZ, H. R., & BARBE, C. D. Psychometric analysis of food aromas by profiling and multidimensional scaling. *Journal of Food Science*, 1976, **41**, 567-571.
- PETERS, R. W. Dimensions of perception of consonants. *Journal of the Acoustical Society of America*, 1963, **35**, 1985-1989.
- POLS, L. C. W., VAN DER KAMP, L. J. T., & FLOMP, R. Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America*, 1969, **46**, 458-467.
- SCHIFFMAN, S. Physiochemical correlates of olfactory quality. *Science*, 1974, **185**, 112-117.
- SCHIFFMAN, S., ROBINSON, D. E., & ERICKSON, R. P. Multi-dimensional scaling of odorants: Examination of psychological and physiochemical dimensions. *Chemical Senses and Flavor*, 1977, **2**, 375-390.
- SHEPARD, R. N. Representation of structure in similarity data: Problems and prospects. *Psychometrika*, 1974, **39**, 373-421.
- TITCHENER, E. B. *Experimental psychology of the thought processes*. New York: Macmillan, 1909.
- TVERSKY, A. Features of similarity. *Psychological Review*, 1977, **84**, 327-352.
- VAN DER KAMP, L. J. T., & POLS, L. C. W. Perceptual analysis from confusions between vowels. *Acta Psychologica*, 1971, **35**, 64-77.
- WAGENAAR, W. A., & PADMOS, P. Quantitative interpretation of stress in Kruskal's multidimensional scaling technique. *British Journal of Mathematical and Statistical Psychology*, 1971, **24**, 101-110.
- YOSHIDA, M. Individual differences among perceived similarity of odors. *Bulletin of the Faculty of Science and Engineering*, Chuo University, 1976, **19**, 221-232.
- YOUNG, F. W., & TORGERSON, W. S. TORSCA: A Fortran IV program for Shepard-Kruskal multidimensional scaling analysis. *Behavioral Science*, 1967, **12**, 498.

(Received for publication March 9, 1978;
revision accepted May 26, 1978.)