[Reprinted from FOOD TECHNOLOGY, 1951, Vol. V., No. 5, Pages 207-210]

Methodology for Sensory Evaluation of Imitation Peppers[a]

DAVID R. PERYAM AND VENONA W. SWARTZ[b]
Quartermaster Food and Container Institute for the Armed Forces, Chicago Quartermaster Depot, Chicago, Illinois

(Manuscript received May 29, 1950)

A method of procedure is described for the evaluation of imitation peppers by comparison with a natural pepper control. Five sensory tests for estimating similarity at successive levels are described and their utility demonstrated. Results of consumer preference tests are reported which call into question the basic importance of pepper to the flavor of foods.

During World War II, when black pepper was not easily available, imitations appeared and found a limited market. Since the war a continued scarcity of natural pepper has kept the interest in such products active. The quality of imitation peppers has been erratic and sales have been limited. However, because of the drive for national self-sufficiency as well as scarcity and high prices, imitation peppers are increasing rather than decreasing in importance and methods of evaluating developments in the field are needed.

The present paper is concerned with the development of sensory evaluation methods only, and not with the physical and chemical aspects. The nature and complexity of the problem militated against the attainment of precise solutions. Techniques adequate for obtaining practical information are described, however, it is recognized that considerable refinement is possible.

Experimentally useful criteria had to be generated from incomplete knowledge of the true role of pepper in food and some decisions were necessarily arbitrary. Since pepper is used primarily not for its own flavor but for its effects in combination with foods, any attempt to judge pepper, as such, does not come to grips with the real problem. It would seem logical to require only that imitations replace natural pepper in foods without lowering acceptability. But experimental difficulties would beset such a program, since pepper may be used in hundreds of different foods or recipes and one would never know in how many an imitation should be tested before approval. This criterion of simple suitability encounters another difficulty. Many imitations have no proven effect on the acceptability of foods, yet have failed on the consumer market. This happened because the consumer did not judge them with regard to their possible use in foods. Imitations were rejected because under optimal sensory conditions they did not match his idea of pepper. This suggests another extreme criterion of a practical sensory identity with genuine pepper. Obviously, any imitation which met this criterion would be highly satisfactory but at the present time it is impractical in that it seems impossible of attainment. The present investigation adopted a plan of evaluation in successive stages, using criteria between these two extremes.

EXPERIMENTAL

Flavor Carriers. Pepper and other spices have been judged by the method of direct stimulation by mouth *(1)*. However, pepper is a strong stimulant and direct exposure to it creates a psychological disturbance that prejudices any skilled act. Pepper at full strength arouses sensations of both pain and heat as well as those of true flavor, whereas in normal usage with extreme dilution the former are reduced below their conscious thresholds even though they may contribute to the total effect. In the present investigation the observers' strong negative attitudes when subjected to pepper at high intensity constituted a controlling factor that required the use of flavor carriers.

A number of foods, including tomato juice, hamburger, soup stock, white sauce, scrambled eggs, and mashed potatoes were evaluated for suitability as flavor carriers on the basis of the ease with which differences caused by variation in quality and quantity of pepper could be detected in them and also for the stability of observers' quality judgments in replicate testing of a group of imitation peppers. None of the foods showed any definite superiority and tomato juice was selected since it permitted better physical control during testing.

The best discrimination in taste tests was obtained using distilled water as a diluent. Water infusions were prepared by a standard series of operations representing extraction conditions more thorough than those encountered in the mouth during eating, but certainly not as severe as those which pepper undergoes during preparation of most cooked foods. One gram of pepper or imitation was added to one liter of distilled water at 36.7° C. (98° F.). This mixture was allowed to stand for one minute and was then filtered rapidly. Lesser concentrations were obtained by dilution. Infusions were made fresh each day. Dry pepper was used for odor tests, one gram of the ground material being spread out on the bottom of a 300 ml. Erlenmeyer flask fitted with a ground glass stopper. Anticipating an accusation of inconsistency in that this represents testing without dilution, it is pointed out that in all odor testing air is the effective diluent. The dilution is great; unfortunately it is also variable. Pepper was added to the tomato juice in the proportion of 0.04 g. per 100 ml. The level was arbitrarily selected since pepper is customarily added "to taste." At this level the flavor of genuine pepper was definitely perceptible to most people yet few considered it too strong.

Observers. Observers of two different types were used. Preference testing was done by people who were as representative of the unselected consumer as it was possible to obtain. These observers were called randomly from among a large available group, excluding anyone with either knowledge of the present problem or technical knowledge of spice quality, and all persons on the difference testing panel.

A selected panel of observers was used for all difference tests. Starting with a group of 88 persons, selection of 25 persons of superior skill in detecting quality and intensity differences between peppers was accomplished through elimination

[a] Presented at the Tenth Annual Meeting of the IFT, Chicago, Illinois, May 24, 1950. This paper reports research undertaken at the Quartermaster Food and Container Institute for the Armed Forces, and has been assigned number 316 in the series of papers approved for publication. The views or conclusions contained in this report are those of the authors. They are not to be construed as necessarily reflecting the views or endorsement of the Department of the Army.

[b] The authors gratefully acknowledge helpful advice and criticism from Dr. H. L. Fevold, Quartermaster Food and Container Institute for the Armed Forces, throughout the course of the work here reported.

by successive series of the difference tests which are described later. Thus selection was on the basis of consistently demonstrated skill in the type of tests that were used.

Genuine Pepper Control. A good typical black pepper supplied by a major spice company served as a constant reference point in all tests. A 50-lb. lot of this freshly ground pepper was packed in air-tight 1-lb. cans for storage. New cans were opened as needed and after opening the pepper was kept in tightly closed glass jars. Difference tests showed that there was no loss in strength throughout the period of the experiment.

Test Methods. The tests included the five approaches which are briefly described below. Of these, the dual standard odor test and the triangle test have been described in detail in a previous article *(2)*.

1. **Dual standard odor test.** Two pairs of samples, each consisting of a control pepper and a variant being tested are presented in Erlenmeyer flasks covered with paper in order to hide possible differences in appearance. The observer first studies a pair which is plainly identified, being allowed to sniff one and then the other until he feels that he has formed a concept of the difference. Then he attempts to identify the samples in the unmarked second pair. Each observer gives two judgments at a session.

2. **Imitation odor identification test.** The observer is presented with a single pair of unidentified samples in the covered flasks, one containing control pepper and the other the variant, and is instructed: "One of these is genuine pepper; the other an imitation which will be more or less like the genuine pepper. Identify the imitation." He is required to guess if he is not sure. Two judgments are given at an experimental session.

3. **Triangle test in water infusion.** The standard water infusions are diluted using 3 parts of the stock infusion to 7 parts of distilled water. Thresholds for true pepper flavor were established for the panel members and this concentration was selected as being well above the threshold of the least sensitive panel member yet not high enough to cause distress. Samples of approximately 20 ml. are presented at room temperature in 1-oz. glasses. Following a "warm-up," which is a sample of the control, and a subsequent distilled water rinse, three coded samples, two controls and a variant, are presented simultaneously and the observer is asked to identify the variant. The observer gives a second judgment on a second set of samples after a water rinse.

4. **Triangle test in tomato juice.** Samples are prepared by adding control pepper and variant to unseasoned tomato juice at the concentration of 0.04 g. per 100 ml. Any difference in appearance is cancelled by adding an excess of neutral pepper-base to both lots until they appear equal. Temperature, size of sample, and test procedure are the same as for the water infusion test.

5. **Paired comparison preference test.** Samples of tomato juice are prepared as for the triangle test; they are 30 ml. in quantity and are presented at room temperature. Following a "warm-up" and rinse, two coded samples comprising the pair are presented simultaneously. The question is asked "Which do you prefer?" A second pair is presented after a rinse.

RESULTS AND DISCUSSION

The methods described here have been used regularly to evaluate imitation peppers and ingredients proposed for use in their formulation. Not all samples are subjected to all five tests since in many cases sufficent information can be obtained by use of only one or two of them.

Data are presented in Table 1 for nine samples that were subjected to the full battery of tests. In the triangle tests, 33⅓ percent correct responses would be expected by chance alone if there were no difference between control and sample; 50 percent correct responses or preference choices for the sample would be expected on the other three tests. Observed percentages deviating from these chance percentages with significance at the 1, 5, and 10 percent levels are shown. In reporting results of the odor identification tests on genuine peppers, "correctness" is defined as selection of the variant.

The testing of samples A, B, and C, brands of genuine pepper selected at random, was a necessary rationalization of the procedure which used a single lot of pepper as a control. It defined the relationship between various natural black peppers. All three samples were significantly different from the control pepper on the dual standard odor test. In the odor identification test, which reflects interpretation as well as simple detection of difference, discrimination was less precise as shown by the smaller percentages found correct. Sample A was interpreted as being "imitation" significantly more often than the control. Sample C showed no difference, and sample B was called "imitation" even less often than the control. Results on both triangle tests show no variation from the chance pattern and therefore the preference test was not tried.

Samples D and E were preparations available commercially, consisting of pepper extractives adsorbed on a salt base, and should be considered as pepper extenders. The dual standard odor test revealed differences from the control of about the same order as with the genuine peppers. In the identification test, however, the preparations were consistently judged as

TABLE 1

Genuine and Imitation Peppers Compared to Genuine Pepper Control

	Sample	Dual Standard Odor Test			Imitation Odor Identification Test			Triangle Test in Water Infusion			Triangle Test in Tomato Juice			Preference Test in Tomato Juice		
		N	Percent Correct [d]	Sig.[c]	N	Percent Correct [df]	Sig.[c]	N	Percent Correct [e]	Sig.[c]	N	Percent Correct [e]	Sig.[c]	N	Percent Preferring Sample [d]	Sig.[c]
Genuine Peppers	A	20	90.0	***	36	69.4	**	40	37.5	0	40	37.5	0	....	....	....
	B	40	75.0	***	40	35.0	**	40	32.5	0	40	40.0	0	....	....	....
	C	40	72.5	***	40	55.0	0	40	42.5	0	40	37.5	0	....	....	....
Pepper Extenders	D	40	80.0	***	30	76.7	***	40	40.0	0	38	31.6	0	50	52.5	0
	E	38	86.8	***	40	87.5	***	68	38.2	0	40	30.0	0	80	55.0	0
Commercial Imitation Peppers	F	20	100.0	***	40	87.5	***	40	70.0	***	40	70.0	***	40	45.0	0
	G	20	100.0	***	40	80.0	***	40	57.5	***	40	42.5	0	40	52.5	0
Experimental Imitation Peppers	H	40	90.0	***	20	90.0	***	40	67.5	***	40	55.0	***	111	56.7	*
	I	20	95.0	***	21	95.2	***	40	67.5	***	50	50.0	**	75	54.7	0

[c] Code: *** significant at the 1% level
** significant at the 5% level
* significant at the 10% level
0 not significant

[d] 50% correct expected by chance
[e] 33⅓% correct expected by chance
[f] Identification of the non-control pepper as the imitation called "correct"

"imitation." Since no difference was shown by either triangle test the "no difference" result on the preference test would be expected.

Samples F and G were commercial imitations which appeared on the consumer market during the war. Odor differences were marked, with perfect scores on the dual standard test and nearly as unanimous agreement on identification. Sample F showed the greatest difference on the taste tests of any of the samples, while Sample G dropped to the chance level on the tomato juice test. Note, however, that even Sample F was not inferior to the control on the basis of preference.

Samples H and I were two laboratory formulations which originally had seemed promising. They exhibited extreme difference from the control on both odor tests. The percentages found correct dropped in the water infusion triangle test and still more in the tomato juice test, even though all differences remained significant at either the 1 or 5 percent levels. In spite of such definite qualitative differences, there was but little discrimination on preference.

In Table 2 the results on all nine samples are brought together for purposes of comparison with only the significance of each test being shown. There is a definite order of sensitivity within the group of tests, ranging from the dual standard which showed all nine samples to be different from the control, to the preference test which found only one slight difference; also it is possible to place the samples in a definite order of similarity to the control pepper. Table 2 demonstrates that the two series can be presented simultaneously without a single

TABLE 2

Significance of Test Results: Imitation and Genuine Peppers vs. Genuine Pepper Control [g]

Test	Experimental Imitation Peppers		Commercial Imitation Peppers		Pepper Extenders		Genuine Peppers		
	H	I	F	G	D	E	A	B	C
Dual Standard Odor	***	***	***	***	***	***	***	***	***
Imitation Odor Identification	***	***	***	***	***	***	**	**	0
Triangle in Water Infusion	***	***	***	***	0	0	0	0	0
Triangle in Tomato Juice	***	**	***	0	0	0	0	0	0
Preference in Tomato Juice	*	0	0	0	0	0	....	....	....

[g] Codes as in Table 1.

inversion of either order, which means that the results on each sample agreed perfectly with the general trend of the data. An orderly system of measurement is thus implied, although the agreement depended also on there being an orderly series of flavor properties in the samples. That it actually so happened was, in large part, chance. Even so, the orderly nature of these results implies validity of the system of evaluation.

Considering the present stage of development in the field, it can hardly be required that all work on imitation peppers continue until identity with natural pepper is attained. For this reason a test which will predict the utility of an imitation that only partially meets the objective of identity is valuable. This was one consideration in the selection and development of the tests.

The odor tests on the dry pepper represent the situation in which a buyer of pepper is most likely to judge the imitation, i.e., by direct odor examination. The dual standard test sets optimum conditions and is too sensitive to represent practical evaluation. The odor identification test, at the next level of sensitivity, would give better prediction of whether an imitation were reasonably close to the customer's concept of natural pepper. The triangle test in water infusion has no analogue in practical use. It is designed to exhibit, under conditions made optimum for discrimination, those flavor properties which are effective when pepper is eaten in food. An imitation pepper which passed this test could be substituted for genuine pepper in any food without changing its flavor. The practicability of this test is shown in that it passed without question all three genuine peppers and also the two commercial preparations containing the pepper extractives. The inference is that it would not reject any imitation with a true pepper flavor.

The tomato juice triangle represents use in a typical food. It gives an estimate of the probability of the existence of a flavor variation that could carry through when blended with food flavors. A test comparable in purpose to this one could be set up using any one of many foods. Although our preliminary work on flavor carriers indicated that there would be little difference in sensitivity of the test with any of the foods tried, the list of flavor carriers tested is not all-inclusive. It seems certain that foods could be selected which would give either greater or lesser sensitivity.

The preference test in tomato juice should be the most meaningful of the whole series since it evaluates any flavor variations that may exist by that final practical criterion, consumer preference. Here the preference test failed to discriminate between samples, since even those imitations which were definitely different from the control pepper were preferentially equal to it. However, it is possible that an imitation may be considered at some time which could not meet this criterion; certainly the test must be used whenever a difference can be shown.

Is Pepper Essential to Flavor?

The question of the true role of pepper was a natural corollary to these results. If an imitiation pepper, quite different from the genuine material, is preferentially equal to it, how important is pepper itself to the flavor of foods?

To answer this question consumer preference tests were run in which foods seasoned normally with pepper were compared directly with the same foods without pepper. Seasoning levels of these foods were determined by usual recipes or by military specifications. The basic data were gathered by the same paired-comparison method used for the tomato juice preference test. The results obtained are given in Table 3. More data were obtained on some foods than on others since any test was extended when first results indicated a trend in either direction.

Only in the case of vegetable soup was it definitely established (1 percent level) that pepper improved

TABLE 3

Paired-Comparison Preference Tests of Foods With and Without Pepper

Food	Percent Pepper by Weight	N	Percent of Observers Preferring		Significance [i]
			Sample without Pepper	Sample with Pepper	
Cream of Tomato Soup	0.015	80	60.0	40.0	**
Tomato Juice	0.040	60	60.0	40.0	*
Corned Beef Hash [h]	0.038	80	57.5	42.5	0
Beef Stew [h]	0.021	39	53.8	46.2	0
Hamburger	0.220	60	48.3	51.7	0
Mashed Potato	0.040	40	47.5	52.5	0
Beef and Gravy [h]	0.045	80	45.0	55.0	0
Soup Stock	0.020	74	41.9	58.1	*
Pork Sausage	0.130	122	41.8	58.2	**
Vegetable Soup	0.010	80	33.8	66.2	***

[h] Canned ration items.
[i] Codes as in Table 1.

flavor. There were two trends for the food containing pepper, also two trends against it, but in all four cases the variation from the chance percentage is not large. To test the hypothesis that this may have been due to the test method, which employed a number of small samples presented at one session, results on four of the foods were checked by a second method. The observer ate 2-ounce portions of each food but only one sample at a session, and rated each on a scale whose nine points ranged in successive categories from "dislike extremely" to "like extremely." This preference scaling device is used regularly in our laboratory and has been found to give reproducible results. Three canned ration items and pork sausage were tested using consumer observers. Mean ratings on the scale were almost identical for the canned meats with and without pepper. With pork sausage there was a very slight trend toward preference for the sample containing pepper, which agrees with results of the paired-comparison test.

These results are not conclusive but they strongly suggest that the traditional role of pepper has been overestimated and that many people would be content to use much less of it or none at all. There are two factors which qualify this tentative conclusion: (1) pepper was tested in only a few of the many ways in which it is used, and (2) testing was confined to one group of consumer-observers who can not be considered representative of all potential pepper-users. However, even if the conclusion can not claim universality, it constitutes a serious questioning of the importance of natural pepper to flavor. This is significant to the development of imitiation peppers because, if pepper itself is relatively unimportant in foods, then the task of developing suitable substitutes should be less imposing.

SUMMARY AND CONCLUSIONS

A battery of five tests was developed for the purpose of evaluating imitation peppers by comparison with a natural pepper control. They fall into a consistent order of sensitivity. The most sensitive, the dual-standard odor test with the dry peppers, revealed definite differences even between natural peppers and the control. The least sensitive, preference in tomato juice, showed but little difference between imitation and the control pepper. All tests with the exception of the triangle difference test in water infusion are analogues of practical evaluation situations. This series of tests will show relative values of imitation peppers; single tests can be selected for specific purposes.

Consumer preference tests showed that pepper does not improve the flavor of many foods in which pepper is normally used.

LITERATURE CITED

1. Cartwright, L. C., and Nanz, R. A. Comparative evaluation of spices. *Food Technol.*, **2**, 330 (1948).
2. Cartwright, L. C., and Nanz, R. A. Flavor improved, sales boosted through organoleptic tests. *Food Inds.*, **20**, 1608 (1948).
3. Peryam, David R., and Swartz, VeNona W. Measurement of sensory differences. *Food Technol.*, **4**, 390 (1950).