Publication Nr 29 of the Cereals Department
of the Central Institute for Nutrition Research T. N. O., Wageningen.

A STUDY IN BREAD FLAVOUR.
SYSTEMS FOR ORGANOLEPTIC EVALUATION OF FLAVOUR DIFFERENCES IN BREAD

by Ir H. De MIRANDA

Cereals Department of the Central Institute for Nutrition Research, Wageningen

Contents of a communication as compiled
for the VIIIth International Congress of Agricultural Industries in Brussels

Publication Nr 29 of the Cereals Department
of the Central Institute for Nutrition Research T. N. O., Wageningen.

A STUDY IN BREAD FLAVOUR. SYSTEMS FOR ORGANOLEPTIC EVALUATION OF FLAVOUR DIFFERENCES IN BREAD

by **Ir H. De MIRANDA**
Cereals Department of the Central Institute for Nutrition Research, Wageningen

Contents of a communication as compiled
for the VIIIth International Congress of Agricultural Industries in Brussels

In the first place I will refer to some of the principal points in the history of bread flavour research.

1. Until recently bread flavour was not a subject of scientific investigation. To obtain tasty bread with attractive flavour was a matter of practical experience, essentially based on organoleptic observation.

2. In the first attempts to attack the problem it was hoped to discover the chemical substances responsible for flavour in bread and to develop analytical methods for their estimation. Particular attention was paid to the by-products of fermentation.

In 1935 VISSER 't HOOFT and DE LEEUW draw attention to acetylmethylcarbinol, a substance which could be oxidised to diacetyl. By analogy with what was known about the flavour of butter it was thought that diacetyl might be a principal component of the bread flavour complex. Suitable methods of analysis have been developed, but it has never been proved that the figures so obtained are indicative of the consumers judgment of different breads. In 1936 MAIDEN proved that the typical bread flavour could not be fortified by addition of acetylmethylcarbinol or diacetyl to the dough.

3. The problem was taken a stage further by the study of crust flavour. In 1939 BAKER and MIZE proved that during the formation of the crust of bread aromatic substances are formed which rapidly diffuse throughout the crumb of the loaf. It is now generally believed that the aromatic substances are by-products of complex reactions between reducing sugars and proteins, known as « browning » or « MAILLARD reactions ». The complexity of these reaction products and their very low concentration make their precise study extremely difficult.

4. Several attempts have thus been made to analyse the bread flavour complex, and some insight onto its nature and origin has been obtained. This information has, at the same time, revealed the essential complexity of the phenomenon and the difficulties of approaching it by the methods of analytical chemistry.

In practice we are still dependent on organoleptic observation if we want to judge bread flavour. The judgment is thus not based on analysis, but on observation of the flavour complex as a whole. The method is accompanied by the known difficulty of separating taste from aroma and by the effects of texture differences upon taste, but it possesses the prime advantage of direct observation of the properties under investigation.

The development of organoleptic techniques is related to the development of statistical methods and especially of ranking methods. To demonstrate the use of ranking methods I shall turn to a practical problem in which our laboratory is involved. That of the judgment of taste and aroma of bread produced from different types of wheat. Samples of these wheats were milled and baked by standardised procedures. Sandwich tins were used in order to minimise texture differences arising from variation in loaf volume. Bread samples were submitted to a tasting panel for assessment of flavour.

Statistically the problem is : A number of objects, bread samples from different wheats, are to be compared by several people. Several working schemes are possible. An obvious scheme is the following. Suppose there are 5 objects and 6 persons. We can give each person all 5 objects together and ask for a judgment of ranking. Each individual then gives to each object a ranking number so that each object gets in all 6 ranking numbers. We then can calculate a mean ranking number for each object and also a mean error of these means. Using all the experimental results it is then possible to apply tests for significance to the found differences.

In judging the taste of bread this scheme is not very suitable. Normally bread has a neutral taste, and differences between breads are not very pronounced. In tasting successively five samples of bread the taste of the first sample has mostly been forgotten when the third or fourth sample is put into the mouth. Five samples are too large a number to compare in one set.

A more practicable scheme of comparison is as follows : No more than two objects at once are given for comparison. Out of the five objects ten different pairs can be formed and each person is asked to judge all these pairs separately, indicating in each case which of the two objects he prefers. Treatment of the results is possible in a similar way as for the foregoing scheme. Furthermore it is possible to compare the ability of individuals to judge. For instance, if in comparing samples A, B and C an individual judges : A better than B, B better than C, C better than A, then his judgment is inconsistent. The more inconsistencies he makes, the poorer is his ability to judge.

The statistical basis for the two schemes mentioned here is given by M. G. KENDALL in his book « Rank correlation methods » (1948) under the headings « The problem of m rankings » and « Paired comparisons ». These two schemes are special cases of a more general scheme.

Suppose m persons are required to compare N objects, and it is only possible to compare n objects in one set. We then form groups of n objects in all possible ways out of the total of N. Each person is asked to give ranking numbers to the objects in each group. The sums of the ranking numbers for each object indicate how, on an average, the objects were judged.

The « problem of m rankings » arises from this scheme when $n = N$, and the problem of « paired comparisons » when $n = 2$.

When a large number of persons is available as judges this scheme may be modified as follows. Each person is asked to judge only one group of n objects. The groups are randomly distributed among the persons, each possible group m times. In this form the scheme is suitable for investigations of consumers' judgment.

Another modification of the original scheme concerns the way the judgment is given. It is possible to ask whether the differences found are distinct or not. Five answers are then possible when objects A and B, are compared : distinct preference for object A, or for object B, a slight preference for object A or B, and no preference. When the experiment is dealt with in this way, the treatment of the results is no longer a pure ranking problem.

Examples.

These general schemes may be exemplified by two experimental cases arising from studies in our Institute.

Example 1.

This concerns the comparison of breads, made from very different types of wheat (two American hard wheats, a French wheat, and a Dutch wheat).

A panel, made up of 24 persons, was used for this test. Each was given a packet, containing 3 samples of bread. Following the general scheme mentioned above each packet contained one group of n objects (i. c. $n = 3$). The members were asked to judge the three samples carefully at home, to grade them and to note whether the differences were distinct or slight. The sample which was judged of first preference was rated either + 2, if the preference was distinct or + 1 if it was slight. The second preference was rated zero and the third either — 2 or — 1. The results are summarised in table I.

The treatment of the results was based on application of the F-test. Significant differences were obtained but hardly reached the 1 % level. It may be remarked that even though very different types of wheat were included the flavour differences were slight.

Such results are typical for the organoleptic evaluation of bread. We could sometimes observe that distinct differences in composition had no influence on consumers' judgment. An example follows :

Example 2.

In the course of an investigation upon the influence of the addition of potato starch to flour it was found that additions of 5 % and 10 % caused significant decrease in bread volume and crumb compressibility. The influence upon consumers' judgment was investigated by the method of « paired comparisons ».

In this case the objects were :

0_1 : bread from flour with 0 % potato starch;
0_2 : bread from flour with 5 % potato starch;
0_3 : bread from flour with 10 % potato starch.

Two types of flour were used, type W (78 % extraction) and type A (85 % extraction).

A group of 30 individuals was used and tests were made both at 16 and 40 hours after baking.

At these times every person received a packet containing two pairs of slices of bread (ca 1 cm thick), one pair from type W-flour and one pair from type-A-flour. Each pair contained slices from different objects. The persons were asked to judge the slices carefully as described in example I.

Table II summarises the numbers of persons, giving the following judgments :

+ + : distinct preference for object with lowest number (that is to say : lowest percentage of potato starch).
+ : indistinct preference for object with lowest number.
0 : no preference.
— : indistinct preference for object with highest number.
— — : distinct preference for object with highest number.

It is remarkable that from the 120 answers only 34 were distinct judgments (+ + or — —), and in 47 cases no preference could be given.

Table III shows how the judgments made at the same times by the same persons were correlated. It is quite evident that there is no positive correlation. Even the distinct judgments are not consistent.

This example is given to show an application of the method of paired comparisons. No better method exists for the detection of small differences between samples than their comparison in pairs. We may therefore conclude that in this case the difference in composition between the samples had no significant influence upon the consumers' judgment.

Of course this single experiment does not throw full light on the disadvantage of using flour with the additions mentioned, but it is beyond the scope of this paper to go further into this aspect of the subject.

REFERENCES

VISSER 't HOOFT, F. and F. J. G. De LEEUW. — The occurrence of acetylmethylcarbinol in bread and its relation to bread flavor. Cereal Chem., **12** (1935), 213.

MAIDEN, A. M. — A system of judging flavour in bread. Confectionery and Baking Craft, 1936, 78.

BAKER, J. G. and M. D. MIZE. — Some observations regarding the flavour of bread. Cereal Chem., **16** (1939), 295.

KENDALL, M. G. — Rank correlation methods. London : Griffin & Co., Ltd., 1948.

TABLE I

Objects : O_1 — O_4

Persons : P1 — P 24

Group 1

	0_2	0_3	0_4
P 1	— 1	0	+ 2
P 2	0	0	0
P 3	0	— 1	+ 1
P 4	— 2	+ 1	0
P 5	— 2	0	+ 1
P 6	— 2	0	+ 1
	— 7	0	+ 5

Group 2

	0_1	0_3	0_4
P 7	— 2	0	+ 2
P 8	— 2	0	+ 1
P 9	— 1	+ 1	0
P 10	— 1	0	0
P 11	0	— 1	+ 1
P 12	0	+ 1	— 1
	— 6	+ 1	+ 3

Group 3

	0_1	0_2	0_4
P 13	0	+ 1	— 1
P 14	— 1	0	+ 1
P 15	0	+ 1	— 1
P 16	— 1	+ 1	0
P 17	0	+ 1	— 1
P 18	0	— 1	+ 1
	— 2	+ 3	— 1

Group 4

	0_1	0_2	0_3
P 19	— 1	0	0
P 20	— 1	0	+ 1
P 21	0	— 2	+ 1
P 22	0	0	0
P 23	0	+ 2	0
P 24	+ 1	— 1	+ 1
	— 1	— 1	+ 3

	0_1	0_2	0_3	0_4
Group 1		— 7	0	+ 5
Group 2	— 6		+ 1	+ 3
Group 3	— 2	+ 3		— 1
Group 4	— 1	— 1	+ 3	
Totals ..	— 9	— 5	+ 4	+ 7

Mean error of individual ratings within the 12 colums :

$S = 0{,}87$

Mean error of totals per object :

$S_1 = 0{,}87 \sqrt{18} = 3{,}7$ (df = 60)

« Mean error », calculated from variation between totals

$S_2 = 7{,}5$ (df = 3)

$F = (S_2/S_1)^2 = 4{,}11^x$ (P = 0,01 for F = 4.13)

TABLE II

	FLOUR TYPE W					FLOUR TYPE A				
	+ +	+	0	—	— —	+ +	+	0	—	— —
16 hours after baking										
$0_1 — 0_2$	2	1	5	1	1	1	3	4	0	2
$0_2 — 0_3$	2	1	3	3	1	1	2	7	0	0
$0_1 — 0_3$	4	0	4	1	1	2	3	3	2	0
40 hours after baking										
$0_1 — 0_2$	1	0	7	1	1	0	3	4	2	1
$0_2 — 0_3$	3	1	1	3	2	2	2	2	2	2
$0_1 — 0_3$	2	3	3	1	1	1	1	4	3	1
Total A										
$0_1 — 0_2$	3	1	12	2	2	1	6	8	2	3
$0_2 — 0_3$	5	2	4	6	3	3	4	9	2	2
$0_1 — 0_3$	6	3	7	2	2	3	4	7	5	1
Total B	14	6	23	10	7	7	14	24	9	6

Total C						Total number of answers : 120
$0_1 — 0_2$	4	7	20	4	5	
$0_2 — 0_3$	8	6	13	8	5	
$0_1 — 0_3$	9	7	14	7	3	
Total D	21	20	47	19	13	

TABLE III

Flour type A

Flour type W	+ +	+	0	—	— —
+ +	2	0	6	2	3
+	1	2	2	2	0
0	1	5	12	4	1
—	0	5	4	1	0
— —	2	2	1	0	2