Journal of Applied Psychology
Vol. 43, No. 1, 1959

R59-19

# CONTRAST AND CONVERGENCE EFFECTS IN RATINGS OF FOODS [1, 2]

JOE KAMENETZKY

*Quartermaster Food and Container Institute for the Armed Forces*

A major requirement in establishing the acceptability of foods for military use is that of taste-testing. The major purposes of these tests are to evaluate samples submitted by food processors intending to bid on procurement contracts, to determine the effects on preference of certain processing variables, and to assess the degrees of liking for new foods.

In connection with this service function, a considerable amount of criterion and methodological research on affective evaluations has been in progress for several years. For the most part, the rationale for the specific problems investigated has been based on empirical and practical considerations, and the results have proved useful in improving the reliability and interpretability of taste-test data. However, it was felt that increasing emphasis on theory will lead to greater integration of findings and will facilitate applications of methodological research.

A simple model was developed as a starting point. The research strategy was to set forth tentative and perhaps over-simplified assumptions based largely on observations of the testing process and subsequent results, to derive and test hypotheses, and to revise the model accordingly. The broader implications of the assumptions and results will be discussed later.

## *Assumptions*

1. An individual evaluating a given food item in terms of like and dislike bases his judgment on the presence or absence of several characteristics of the food.

2. Characteristics of foods are of two types: negative and positive. For any food, characteristics of both types are likely to be present.

3. (*a*) Noticing the absence of a positive characteristic will result in a lower preference rating for the food.

(*b*) Noticing the presence of a negative characteristic will result in a lower preference rating for the food.

4. (*a*) When the positive characteristics of a good quality food (a "good") predominate, the presence of *some* of the negative characteristics is not noticed or taken into consideration.

(*b*) When the negative characteristics of a poor quality food (a "poor") predominate, the presence of *some* of the positive characteristics is not noticed or taken into consideration.

5. (*a*) Presentation of a "poor" increases an individual's awareness of the presence of some of the same negative characteristics in a "good."

(*b*) Presentation of a "good" increases an individual's awareness of the absence of some of the same positive characteristics in a "poor."

6. (*a*) As successive samples of a "poor" are served, forgetting the absence of some *positive* characteristics takes place.

(*b*) As successive samples of a "good" are served, forgetting the presence of some *negative* characteristics takes place.

## *Experimental Implications*

It can be shown that the above assumptions lead to the following predictions.

1. A "poor" will be rated lower when preceded by a "good" than when it is preceded by another "poor." Consequently, the difference in mean preferences between a "good" and "poor" should be larger when the ratings of the "poor" are obtained after a "good"

[1] This paper reports research undertaken at the Quartermaster Food and Container Institute for the Armed Forces, and has been assigned Number 843 in the series of papers approved for publication. The views or conclusions contained in this report are those of the author. They are not to be construed as necessarily reflecting the views or endorsement of the Department of Defense.

[2] The writer expresses his appreciation to Norman Gutman, Chief, Statistics Branch, and Bettye Johnson, Statistical Clerk, for their valuable assistance. Special acknowledgment is made to Phyllis Whitmer and Audrey Beauvais, Home Economists, Food Acceptance Branch, who performed almost all the laboratory work.

than after another "poor." This effect is called *contrast*.

2. A "good" will be rated lower when preceded by a "poor" than when it is preceded by another "good." Thus, the difference in mean preferences between a "good" and "poor" should be smaller when the ratings of the "good" are obtained after a "poor" than after another "good." This effect is called *convergence*.

3. Preference will increase with successive servings of the same quality, provided no opposite quality intervenes.

## Method

Four independent replications of the experiment were conducted on four separate days between August, 1957, and January, 1958.

### *Foods*

The food tested in the first two replications was a cherry beverage made from a concentrated liquid base. The food tested in the second two replications was a beef broth prepared from a granulated base. On the basis of results from previous taste tests, "good quality" lots of each food were selected. "Poor quality" samples of each food were prepared as follows: For the cherry beverage on the first replication, 44 ml. of vinegar and .55 g. of caffeine were added to the standard ingredients of 570 g. of sugar, 4500 ml. of water, and 95 ml. of beverage base. On the second replication, adulterating ingredients were 13.4 g. "liquid smoke" and 10.8 ml. vinegar. For the beef broth on both replications, the powder was partially burned, thereby producing a definite acrid taste. Since the burning could not be precisely controlled, the tastes of the "poors" on the two replications were not identical. The holding and serving temperatures of the beverage was 50° F., and of the broth, 100° F.

Table 1

Orders of Presentation and Qualities Presented for Four Treatments

| Treatment | Order of Presentation | | | |
|---|---|---|---|---|
| | First | Second | Third | Fourth |
| 1 | Good | Poor | Good | Poor |
| 2 | Poor | Poor | Poor | Good |
| 3 | Good | Good | Good | Poor |
| 4 | Poor | Good | Poor | Good |

### *Judges*

The judges (*Os*) were randomly drawn from a larger pool of about 700 civilian and military, both

Table 2

Treatment Means for Each Replication[a,b]

| Treatment | Replication | Order of Presentation | | | |
|---|---|---|---|---|---|
| | | First | Second | Third | Fourth |
| 1 | 1 | 7.3 (G) | 3.4 (P) | 7.3 (G) | 3.9 (P) |
| | 2 | 6.8 | 3.0 | 7.0 | 2.5 |
| | 3 | 7.5 | 3.4 | 7.5 | 3.1 |
| | 4 | 7.2 | 3.5 | 6.1 | 4.2 |
| 2 | 1 | 3.5 (P) | 4.2 (P) | 5.1 (P) | 6.9 (G) |
| | 2 | 4.5 | 6.3 | 6.5 | 6.4 |
| | 3 | 4.6 | 5.2 | 5.2 | 7.0 |
| | 4 | 6.7 | 6.1 | 5.9 | 5.5 |
| 3 | 1 | 6.8 (G) | 6.9 (G) | 6.7 (G) | 3.3 (P) |
| | 2 | 7.4 | 6.8 | 6.9 | 1.9 |
| | 3 | 7.9 | 7.4 | 7.5 | 4.1 |
| | 4 | 7.1 | 6.9 | 6.8 | 3.2 |
| 4 | 1 | 3.6 (P) | 6.1 (G) | 3.3 (P) | 6.5 (G) |
| | 2 | 3.7 | 6.7 | 3.8 | 6.8 |
| | 3 | 3.8 | 7.9 | 2.9 | 6.8 |
| | 4 | 6.0 | 6.5 | 4.2 | 6.1 |

[a] A high rating signifies a high preference.
[b] "G" & "P" represent "Good" and "Poor," respectively.



Table 3

Analyses of Variance of Preference Ratings[a]

| Source of Variation | *df* | First Replication (Cherry Beverage) | | Second Replication (Cherry Beverage) | | Third Replication (Beef Broth) | | Fourth Replication (Beef Broth) | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean Square | *p* | Mean Square | *p* | Mean Square | *p* | Mean Square | *p* |
| 1. Poor, 2nd (Treatments 1 vs. 2)[b] | 1 | 3.2000 | | 54.4500 | <.01[c] | 16.2000 | <.05[c] | 33.8000 | <.01[c] |
| 2. Good, 2nd (Treatments 3 vs. 4)[b] | 1 | 3.2000 | | .0500 | | 1.2500 | | .8000 | |
| 3. Poor, 1st & 3rd (Treatments 2 vs. 4)[b] | 1 | 7.2250 | | 30.6250 | <.05[d] | 24.0250 | <.05[d] | 14.4000 | |
| 4. Good, 1st & 3rd (Treatments 1 vs. 3)[b] | 1 | 3.0250 | | .6250 | | .4000 | | .9000 | |
| 5. 1st & 2nd vs. 3rd & 4th | 1 | .9000 | | 8.1000 | <.05[d] | 8.1000 | <.05[d] | 40.0000 | <.01[d] |
| 6. 1st & 3rd vs. 2nd & 4th | 1 | 3.6000 | | 22.5000 | <.01[d] | 2.5000 | | 40.0000 | <.01[d] |
| 7. 1st & 4th vs. 2nd & 3rd | 1 | .9000 | | 32.4000 | <.01[d] | 3.0250 | | 0 | |
| 8. Good vs. poor | 1 | 366.0250 | <.01[c] | 324.9000 | <.01[c] | 462.4000 | <.01[c] | 96.1000 | <.01[c] |
| 9. Good, 4th (Treatments 2 vs. 4)[b] | 1 | .8000 | | .8000 | | .2000 | | 1.8000 | |
| 10. Poor, 4th (Treatments 1 vs. 3)[b] | 1 | 1.8000 | | 1.8000 | | 5.0000 | | 5.000 | |
| 11. Good vs. poor, 1st & 2nd. vs. 3rd & 4th | 1 | .2250 | | 2.5000 | | .0250 | | 1.6000 | |
| 12. Good vs. poor, 1st & 3rd vs. 2nd & 4th | 1 | .6250 | | 8.1000 | <.05[d] | .2250 | | 8.1000 | <.05[d] |
| 13. Good vs. poor, 1st & 4th vs. 2nd & 3rd | 1 | 3.0250 | | 28.9000 | <.01[d] | 0 | | .4000 | |
| 14. Poor, 1st & 3rd; 1st & 4th vs. 2nd & 3rd | 1 | 9.0250 | <.01[c] | 9.0250 | <.01[c] | 5.6250 | <.05[c] | 2.5000 | |
| 15. Good, 1st & 3rd; 1st & 4th vs. 2nd & 3rd | 1 | .0250 | | 1.2250 | | .4000 | | 1.6000 | |
| 16. Judge (within groups) | 36 | 6.1361 | | 4.7903 | | 4.9292 | | 5.1472 | |
| 17. Judge-treatment (within groups) | 108 | 1.4546 | | 1.5106 | | 1.6495 | | 1.8306 | |
| 18. Total | 159 | | | | | | | | |

[a] Ordinal values refer to position of sample.
[b] Evaluated against judge (within groups). In all other comparisons, judge-treatment (within groups) interaction is the error term.
[c] One-tailed test.
[d] Two-tailed test.

male and female, employees who regularly participate in taste tests. Departures from randomness occurred when some were absent or were otherwise not available on the days the tests were conducted. Separate selections of 40 *Os* were made for each of the four replications.

### *Procedure*

Each group of 40 *Os* was randomly assigned to one of four treatments that differed in the order in which the different qualities were presented and the number of each quality rated. The nature of the treatments is summarized in Table 1.

Each *O* sat in a semienclosed testing booth. Two-ounce samples in coded cups or glasses were presented one at a time through a turntable in a wall separating the booth from the kitchen. *O* drank as much or as little as he wanted of each sample, rated the product on a nine-point scale described elsewhere (Peryam & Pilgrim, 1957), and rinsed his mouth ad libitum with charcoal-filtered distilled water. The time between the rating of one sample and the presentation of the next was 45 seconds.

On the first replication, the *O*'s, after rating the third and fourth samples, were asked to list the positive and negative characteristics of each sample. Their lack of ability to so verbalize led to the decision to discontinue these questions on the later replications.

## Results

The preference means of each treatment on each replication are given in Table 2. An analysis of variance of the preference ratings was performed for each replication separately, and the results of these analyses are presented in Table 3. In each case the ratings of the "poors" were clearly lower than the ratings of the "goods" (Source of variation No. 8).

### *Contrast Effects*

Inspection of Table 2 reveals that in every replication, the average rating of "poor" in the second position was lower when it was preceded by a "good" than when it was preceded by another "poor" (Source of variation No. 1). Three of the four differences were significant at either the .05 or .01 level. Since the combined probability (Wilkinson, 1951) is less than .001, it is concluded that contrast effects have been demonstrated.

### *Convergence Effects*

Tables 2 and 3 fail to show any consistent or significant (Source of variation No. 2) difference in the ratings of the "goods" regardless of whether a "poor" or "good" preceded. The prediction regarding convergence effects was not confirmed.

### *Effect of Successive Presentations*

It was predicted that as the "poors" are successively presented, the ratings would increase, but would not increase if a "good" intervenes. This means that the algebraic difference between the third and first samples should be greater for those in Treatment 2 than for those in Treatment 4. The comparison of Source of variation No. 14 with the judge-treatment interaction constitutes the appropriate test of this prediction. Significant differences at the .05 or .01 levels were attained for three replications; for the remaining replication, the results were in the expected direction though not significant. The combined probability is less than .001. The hypothesis may be considered confirmed.

It was also predicted that as the "goods" are successively served, the ratings would increase, but would not increase if a "poor" intervenes. Similar to the preceding prediction, the algebraic difference between the third and first samples should be greater for Treatment 3 than for Treatment 1. However, when Source of Variation No. 15 was tested against the judge-treatment interaction, no significant effects emerged. Inspection of Table 2 shows that the differences were not always in the expected direction. Hence, there is no support for the hypothesis.

### *Other Tests of Significance*

The significance of Source of Variation No. 5 shows that the later samples are preferred more than the earlier ones. However, this and other significant sources of variation are of only incidental interest here and will not be further discussed.

## Discussion

Both predictions concerning changes in preference for the "poors" were substantiated, and both predictions involving changes in preference for the "goods" were not. In fact, the ratings of the "goods" remained almost invariant regardless of the nature or number of samples preceding them. It is possible that the "goods" were so good that the negative characteristics, present to a marked degree in the "poors," were completely absent or below threshold. Indeed, on the first replication, *Os* were unable to specify *anything* negative about the "good." Absence of negative characteristics makes Assumptions 3(*a*), 5(*a*), 6(*b*), and part of Assumption 2 inapplicable.

Obviously, definitive results could be obtained only with the accompaniment of an independent assessment of the presence and absence of both positive and negative characteristics. However, present methods for determining these characteristics are not satisfactory, primarily because judges are unable reliably and independently to describe their introspective experiences. Currently, research is being considered on pychometric multivariate methods for inferring these characteristics without resort to verbalizations by the judges.[3]

Because the positive and negative characteristics were not independently established, this experiment cannot be considered to be a crucial test of the validity of the assumptions. Apart from the consideration that development of methods for assessing these characteristics will enable a rigorous test of the assumptions, tentative retention of the assumptions set forth here should prove useful.

First, it is advocated that the following additional assumption be added: Presentation of a "poor" increases an individual's awareness of the presence of positive characteristics in a "good," i.e., the individual doesn't appreciate the excellence of a "good" until the absence of the positive characteristics in the "poor" makes him cognizant of their presence in a "good."

When there is no independent determination of the individual's perception of the presence of positive and negative characteristics, inclusion of this assumption would preclude derivation of certain predictions, such as the ones regarding convergence effects in the present experiment. On the other hand, even when there is no independent determination, predictions of more complex phenomena can be made.

Consider, for example, four types of an orange soda pop: carbonated, slight off-flavor; carbonated, marked off-flavor; noncarbonated, slight off-flavor; noncarbonated, marked off-flavor. If it is independently demonstrated that most people prefer the carbonated beverage over the noncarbonated one and that the slight off-flavor beverage is preferred to the marked off-flavor one, it can be shown that at least seven predictions are deducible. For example, it would be predicted that a carbonated, slight off-flavor sample will tend further to depress the ratings of a carbonated, marked off-flavor sample that follows it; and a noncarbonated, marked off-flavor sample should have the opposite effect. Testing of these predictions is contemplated.

Another reason for at least tentatively retaining the previous assumptions is that they focus attention and may provide answers to problems facing manufacturers of consumer goods. Consider the case, for example of a manufacturer of a hi-fi component which is well liked by its users. Suppose, also, that he is considering adding to his line an improved component, which, because of its extra cost, is not expected to be bought by as many people. The question arises: Will the new component cause a decreased preference level for the older one with a consequent reduction in sales for it, or will the effect be mainly one of increased liking for the new and no change in liking for the old?

Similarly, when those who have had pleasant experiences with such optional automobile equipment as automatic transmissions and power brakes are in the market for a new car, their preference level for autos without these accessories may decline, while their preference levels for autos with these remain constant. If level of preference is related to willingness to buy and if such a decline in preference occurs, then those who are able to afford just the basic auto might rather forego its purchase or buy a used one with these extras.

Thus, in many cases where contrast effects appear, it is important to determine whether preference for the "good" rises or whether preference for the "poor" declines. To the extent the model is able to predict which—the "good" or the "poor"—is responsible for the increased or decreased differences in preference, its bearing on marketing problems increases.

[3] Impetus for this line of psychometric research was given by results of certain preference tests wherein departures from the postulates of order (transitivity and asymmetry) were evident among different samples of the same product.

## Summary

A set of assumptions was made that led to the hypothesis that preference ratings for poor quality food will be lower when preceded by a good quality food than when preceded by another poor quality item (contrast effects). It was also hypothesized that preference for a good quality food will be higher when preceded by another good quality item than when preceded by a poor quality product (convergence effects). The other predictions were that preference will increase with successive presentations of the same quality item, provided no opposite quality intervenes. The predictions concerning preference for the poor quality foods were clearly confirmed, but those involving the good quality foods were not substantiated. Experimental and practical implications of the assumptions and results are discussed.


*Received May 1, 1958.*

## References


Peryam, D. R., & Pilgrim, F. J. Hedonic scale method of measuring food preferences. *Food Technol.*, 1957, 11(9), Supplement, 9–14.

Wilkinson, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, 48, 156–158.