RESEARCH BULLETIN 651 JANUARY, 1958

UNIVERSITY OF MISSOURI COLLEGE OF AGRICULTURE
AGRICULTURAL EXPERIMENT STATION
J. H. LONGWELL, Director

Predicting Consumer Acceptance of Beef Loin Steaks

V. JAMES RHODES, ELMER R. KIEHL,
D. E. BRADY, and H. D. NAUMANN

(Publication authorized January 23, 1958)

COLUMBIA, MISSOURI

TABLE OF CONTENTS

Summary 3
Introduction 5
Review of Literature 6
Grade and Shear as Predictors of Consumer Acceptance 8
- Experimental Procedure of Consumer Test 8
- Acceptability of Loins as Associated with Carcass Grades 12
- Acceptability of Loins as Associated With Shear Values 19
- Other Predictors of Consumer Acceptance 26
- Further Acceptability Results 37
- Evaluation of St. Louis Consumer Panel 43

Appendix 47
- Sampling Procedure 47
- Laboratory Panel Procedure 48
- Comparison of Results of the Two Laboratory Panels 49

References 51

ACKNOWLEDGMENTS

Appreciation is expressed to Mrs. Richie Thornton for editorial assistance and to Dr. Margaret Mangel for guidance. The bulletin is a report on Department of Agricultural Economics Research Project 86, "Meat Preferences," Animal Husbandry Project 5, and Home Economics Project 109.

Portions of this bulletin are taken from a manuscript by Elmer R. Kiehl entitled "Consumer Evaluation of the Product Characteristics of Beef" and being submitted as a Ph.D. thesis to Harvard University.

SUMMARY

This is a report on the eating acceptability of beef loin steaks from 126 carcasses of various shear strengths and federal grades. Eating acceptability was measured by a sample of 266 St. Louis white households in October, 1955.

Both the federal grade of a carcass and its shear measurement were fairly useful in predicting the consumer acceptability of loin steaks. While somewhat related, grade and shear were sufficiently independent that acceptance prediction was materially improved by the use of both as explanatory variables.

Twenty-one loin pairs in each of three comparisons were used. There were significant differences in the acceptability ratings of 17 $Choice_0$-Commercial (now Standard), 14 Good-Prime, and 4 $Choice_1$-$Choice_2$ comparisons. Coefficients of correlation within grades of carcass shear values and acceptability ratings ranged from 0.24 to 0.74.

As indicated by Figure 1, the leaner the grade the poorer its over-all acceptability and the more heterogeneous the acceptability of carcasses within the grade. The acceptability ratings of Prime and Choice were so overlapping that they might have been combined with little loss in eating homogeneity. On the other hand, Good and, especially, Standard carcasses were so heterogeneous that a redefinition into two or more homogeneous grades might be useful. Use of shear measurements in conjunction with grades would improve the homogeneity grouping of these leaner carcasses, but an easier and more efficient grouping method appears desirable.

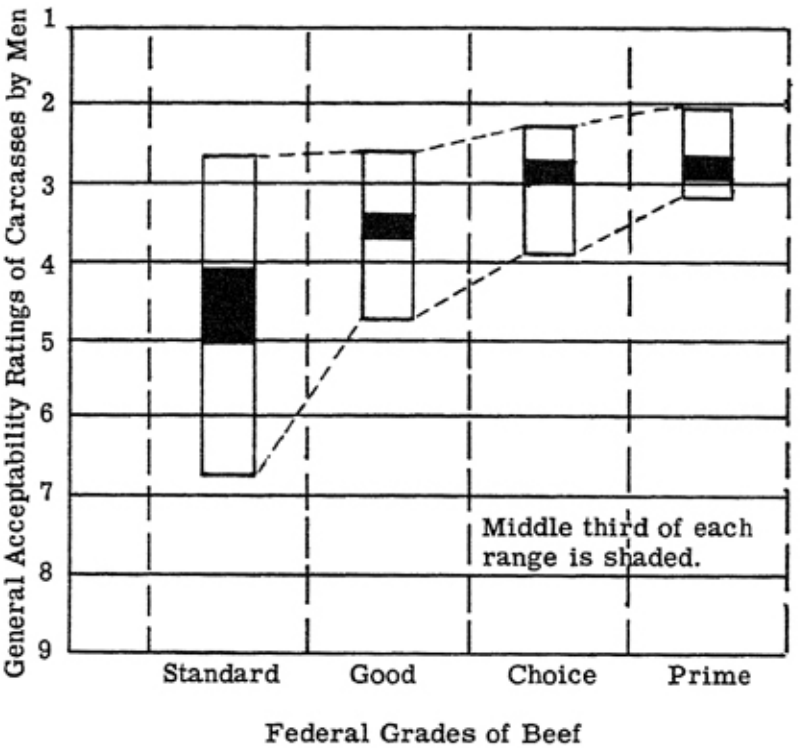
Portions of 84 pairs of these loins were used in trio discrimination tests in the laboratory. There was a positive relationship between the differences found by consumers and by laboratory testers, but it was disappointingly weak.

The mean ratings of the second replicate of the leaner grades were somewhat better than the first replicate while those of the fatter grades were poorer. Whether this shift was random or was the first part of a trend is obviously an important question, which unfortunately was not answered in this experiment.

Variations in consumer acceptability ratings can be assumed to be largely explained by variations in the test product and in the household environment including the likes of the particular consumers. It is particularly difficult in the testing of conventional meat cuts to separate household and animal variation because of the absence of large quantities of a product which are known, *a priori,* to be homogeneous. Evidence from these experiments is largely based upon means of carcasses which were tested by 12 to 14 households. There is evidence that the variation between carcass means was largely a result of variation in product rather than in consumer likes or preferences.

Over-all ratings were slightly but significantly related to method of cooking and to income. However, there was no relation of particular grade means to either of those variables.

Fig. 1—Distribution of Carcasses (Loins) by Grade, St. Louis Panel. (Rating scale began with 1 as the best. Number of loin pairs in each grade was 21.)



Predicting Consumer Acceptance of Beef Loin Steaks

V. James Rhodes, Elmer R. Kiehl,
D. E. Brady, and H. D. Naumann

INTRODUCTION

The recognition and measurement of quality of agricultural products are important in a market economy. Quality has implications for resource use and production and merchandising policies of firms. It has significance for government programs related to grading and regulation of markets. Educational and service programs to improve quality of product must be concerned about the dimensions of quality and particularly those which are relevant in a market economy.

Over the years there has been a wide diversity of interests among those concerned with product quality. Some have emphasized programs to "educate" the market (consumers) concerning the quality characteristics they have believed important. This approach has its genesis among producers of products and those representing the producer interest. It has been the predominant force in the evolution of standards of quality.

Other interests place more emphasis on nutrition, while a third group approaches product quality from the viewpoint of the consumer. Part of the difficulty lies in the orientation of product quality standards which serve needs at different levels of the market, i.e., the producer, wholesaler, retailer and consumer. The controversy of private versus governmental standards of quality also is involved. This diversity of interests has led to divergent approaches among research groups, each concerned with particular facets of this subject.

Clarification of ideas related to quality would lead to the development of a more logical framework and place the various interests and approaches in proper relationships and perspective to the needs of the market economy.

Recognition of quality variability eventually leads to qualitative standards. Such standards become "the children of trade necessity."**[1] While the case for them in agricultural marketing has been long established, they have many shortcomings in both the manner of development and in their application.

An optimistic view of the progress of standardization was presented in 1928 by Dr. L. S. Tenny, then Chief of the Bureau of Agricultural Economics. He stated:

**Numbers refer to references in the back.

> Real progress has been made during the recent period of agricultural depression when every function of our marketing machinery has been tested with a view to the elimination of waste. The movement has been further expedited by the development of specialized agriculture which has rapidly changed the marketing of farm products from a local or regional problem to one of national and even international significance.[2]

In noting the progress of grading during the two preceding decades, he stated that:

> We have completed the fundamentals of the work and although each commodity has its peculiarities and its problems, we have mastered so many of them that we feel confident we can ultimately formulate standards for all of the factors of all the commodities.[2]

Tenney's views stressing the need for and progress of quality standardization were largely producer or market oriented—that is, they would serve these interests primarily and only incidentally the needs of consumers. This strong bent toward producer-oriented, quality standardization is particularly evident in the case of beef. Adaptability of the beef quality standards to consumer needs has been given scant attention until recent years.

The framework adopted for work reported in this bulletin was that quality must be interpreted in terms having relevance to want-satisfaction. Discovery of these characteristics is the first step. The next steps are identification and measurement of those physical aspects of the product which contribute to the want-satisfaction of important groups of consumers. These steps precede the development of quality standards. The resulting standard has only one main function to perform. That function is to assure that the quality characteristics be fairly homogeneous within each classification and be homogeneous over time. Questions of fitting the products in order along a classification scale in terms of their market value is a task of the market processes and not of the "experts" designing the quality standards. The ideal standard would attain product homogeneity within classifications as measured by factors relevant to consumers. If such a standard can be obtained or approximated, it will clearly be a boon to market and resource efficiency.

In brief, the efficient satisfaction of consumer demand for any product requires that that product be designed to maximize consumer satisfaction at the least possible cost. Effective product designing requires the discovery of which factors can be varied and how these variations affect both production costs and consumer acceptance. The effect of certain variations in beef short loins upon consumer acceptance is reported in this bulletin.

REVIEW OF LITERATURE

The long-time interest of economists in theoretical consumer preferences contrasts with the relatively recent empirical research in that area. In meats, there were a few surveys of consumer attitudes before World War II,[3,4,5] and a

few careful experimental studies have occurred since the war. Postwar studies of visual preferences for grades or other classifications of beef have been reported by researchers at Washington State,[6] Arizona,[7,8] Colorado,[9] Wyoming,[10] Texas,[11] and Missouri.[12,13,14]

There have been a large number of studies concerned with laboratory or panel evaluation of the influence of various variables in the composition of beef or in its processing or cooking upon its acceptability. A few of the more important studies which have attempted to relate acceptability to degree of finish or grade of beef are those by Cover *et al.,* [15] Hankins and Ellis,[16] Wanderstock and Miller,[17] and Husaini, *et al.* [18] While these and other studies have contributed much to the understanding of factors which affect acceptability, there are many relationships which still have not been accurately defined or measured. A major shortcoming of laboratory panel evaluation of acceptability is its uncertain and probably shifting relation to consumer acceptability. There are various appropriate uses for laboratory panels in a preference research program, of course.[19]

The following statement from the recent publication of Cover *et al.* summarizes some of the difficulties in defining important relationships: "At any rate the agreement between fatness and tenderness is low enough so that it is not surprising that a consumer who buys fat loin steaks or even well-marbled loin steaks is sometimes disappointed in their tenderness. Because none of the coefficients is high, it is doubted that fatness by itself is responsible for a marked increase in tenderness and juiciness. It is disconcerting that something which has appeared so obvious to so many for so long should be so extraordinarily difficult to prove in the laboratory."

A large laboratory study of discrimination between and within grades at this station also indicated only a weak relationship between degree of finish and discriminable eating differences.[20]

Only two published reports were found on consumer eating acceptability of beef, although a few stations indicate current research in this area.[12,7] Researchers at the Missouri Station sold rib steaks and/or rib roasts to 92 consumers in 1952. A follow-up after consumption indicated that all 3 Prime, 12 of the 13 Choice, 12 of the 17 Good, and 26 of the 28 Commercial steaks were quite satisfactory. Satisfaction with their purchased roast was expressed by 7 of the 8 purchasers of Prime, all 19 of the purchasers of Choice, 21 of the 23 purchasers of Good, and 20 of the 21 purchasers of Commercial. The few expressed dislikes generally involved toughness in the leaner grades and excessive fat in the fatter grades.

Much the same results were obtained more recently in a larger study at Phoenix. "Of 351 consumers of U. S. Commercial grade beef, 96.1 percent rated their roasts or steaks as 'Excellent,' 'Good,' or 'Satisfactory.' Of 269 consumers of U. S. Choice, 94.8 percent gave the above ratings, while of 234 consumers of U. S. Good beef, 93.2 percent gave the above ratings for their roasts or steaks."[7]

GRADE AND SHEAR AS PREDICTORS OF CONSUMER ACCEPTANCE

Federal grades of beef were developed by the U.S.D.A. in conjunction with cattle producers and meat packers in the 1920's. Grades were developed to facilitate market trading, including the buying and selling without inspection of carcasses and were based upon certain physical descriptions and photographs. In the market, grades of cattle sell at price differentials which are not entirely justified by differences in physical yield to the wholesaler or retailer but rather are mainly justified by differences in retail prices. Therefore, grades must provide meaningful differences in acceptability to consumers. If grades do not perform that function, they promote economic inefficiency. The general belief has been that grades do provide fairly meaningful differences in consumer acceptability.[21,22,23]

The most widely used objective method of measuring tenderness in beef is the Warner-Bratzler mechanical shear. This device measures in pounds the force required to shear a core of cooked or raw meat one inch or one-half inch in diameter. Several researchers have measured the relation between subjective panel measurements of meat tenderness and shear scores and have found a rather large range in the size of the correlation coefficients. Bratzler reports simple *r*'s ranging from 0.299 to 0.986.[24] The typically small number of samples correlated as well as various understandardized variables such as degree of doneness, uniformity of samples, and variability inherent in the machines and in subjective scoring probably contributed to the variability of the simple *r*'s. While the shear, presumably, is not a precise measure of tenderness, its measurements should show some relation to consumer acceptability. [24,25]

Experimental Procedure of Consumer Test.

Resources were available for a sample of approximately 250 households. A larger sample was desired, but this size was considered adequate for the experiment. The sampling was done in the city and county of St. Louis, which had an estimated population of 1,400,000 people in 1955.

Two-stage, cluster sampling procedure was used. Simple random sampling of this metropolitan area would have been much more costly. Only the white population was sampled. This population was further restricted to

(1) Households with annual incomes—after withholding taxes—of $2500 or more.
(2) Households with two adults—but fewer than nine persons—who consumed at least two meals per day at home.
(3) Households in which the person doing most of the meat shopping had 5 or more years of schooling and was under 70 years of age.
(4) Households which had consumed beef steaks or roasts in the previous two weeks.

The purpose of these restrictions was to define a population which actually or potentially represented a market for beef steaks. Further details of the sampling procedure are in the appendix.

Selection of the panel was conducted during the period September 19 to October 7, 1955. A schedule was used to obtain general information concerning consumption and cooking practices and to enable easier establishment of rapport with prospective panel members. In addition, some specific questions were asked to determine eligibility of the household. If the household met the requirements, the interviewer then invited the family to participate in the taste panel. A set of instructions to the cooperating panel members was left at this time. Panel members served without remuneration except for the meat they obtained as a part of the study.

In planning this phase of the work it was deemed desirable to use beef cuts from carcasses of the grade and weight classes used in the previous study,[20] insofar as the resources permitted. However, carcasses were selected from the middle third rather than the lower third of the grade this time. The tie-in with the laboratory panel through the use of similar grade and weight classes of beef carcasses would provide some evidence of the association of the results of the laboratory panel with those of the household consumer panel.

Resources limited the number of comparisons which could be made with a large consumer panel, so the weight comparisons within a grade were dropped. Two comparisons of the non-adjacent grades—Commercial-Choice$_0$† and Good-Prime—were selected for this phase of the study and an identical comparison of Choice$_1$-Choice$_2$† was used as a control. All carcasses were soft-boned and the Commercial carcasses would now be graded Standard. All carcasses were graded by federal graders after being ribbed down. Hot carcasses weights were permitted to vary from 555 to 605 pounds. The three sets provided for three taste comparisons.

Use of both the left and right loins of the carcass permitted six taste comparisons—that is, each comparison could be repeated once. Servicing a panel of 266 households with six taste comparisons required 21 pairs of loins of each grade or a total of 126 pairs for the six grades.

Short-loins were purchased from commercial packers after selection according to the specifications outlined above. After loins were aged at 38° F for seven to nine days, the tenderloin muscle and ventral, vertebral processes were removed. The steaks were cut as specified below. Then they were wrapped individually in laminated freezer paper, assigned code designations, frozen at -10° F, and stored at 0° F until they were delivered to cold storage facilities in St. Louis. Distribu-

†The subscripts 0, 1, 2 indicated on the Choice grades are used to facilitate discussion. All of the Choice loins in each of the comparisons were of the same weight class and grade and were presumably similar. All carcasses were in the weight class 555-605 pounds.

tion of the meat from cold storage to homes of panel members was handled in double layer cardboard containers. Dry ice was placed in each container to prevent thawing of the steaks during the delivery process.

The general sampling design was influenced by some of the inherent characteristics of the material to be tested. Previous experience had shown that 15 to 18 steaks, ¾ inch thick, can generally be obtained from a beef short loin. A steak 1.5 inches thick from the left loin of each carcass was reserved for shear analysis. In addition, six steaks, ¾ inch thick, were reserved from each of 14 pairs of loins (A-N) for use with a laboratory panel, the objective being to determine to what extent laboratory panel discrimination results could be correlated with large consumer-panel results. Therefore, two replicates of 12 steaks each were obtained from the 14 loin pairs (A-N) and two replicates of 14 steaks each from the 7 loin pairs (P-W).†† The 21 pairs of loins, 14 with 12 steaks each and 7 with 14 steaks each, provided for a taste panel composed of 38 clusters of seven household units each for a total of 266 households. (Table 1).

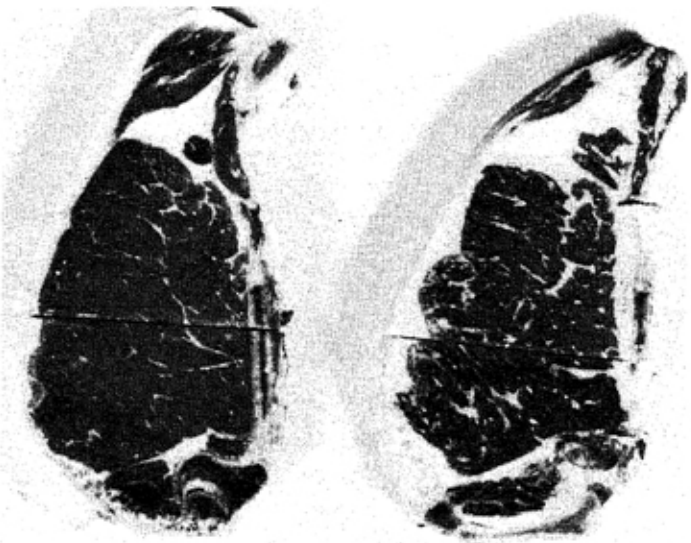
Each steak was cut in half so each of the two adult members of each household could taste each steak. Some means was needed to identify the preparation and cooking of the steaks in the household. Such identification was accomplished by clamping an aluminum ring around the bone on both halves of one of the steaks in the particular comparison. In the replicate taste test of the same comparison, the rings were placed on the other steak. Thus each grade of steaks was identified by the rings an equal number of times in the study. (See photographs, Figure 2.)

The general design showing the assignment of loins and steaks to individual clusters and households is shown in Table 1. For example, clusters 1 through 12 would be tasting from loins A through G. In addition, the steaks from loin A would be tasted by households in 12 different clusters widely separated geographically. The seven households in cluster 1, for example, would be tasting from first position steaks of seven different loins, cluster 2 from the second position steaks of the same set of loins, etc. Loin pairs for each grade were assigned, randomly, the letters A through W (omitting O and Q). A particular pair of loins of one grade was compared with the same lettered pair of the other grade.

The household consumption of the loin steaks began October 10 and continued for three weeks. Deliveries were made twice weekly. A brief preference and acceptance schedule was picked up on the following delivery. Acceptance ratings and a preference choice were obtained on each pair of steaks from each of two designated adult consumers in each household. Information on cooking methods and degree of doneness was also obtained.

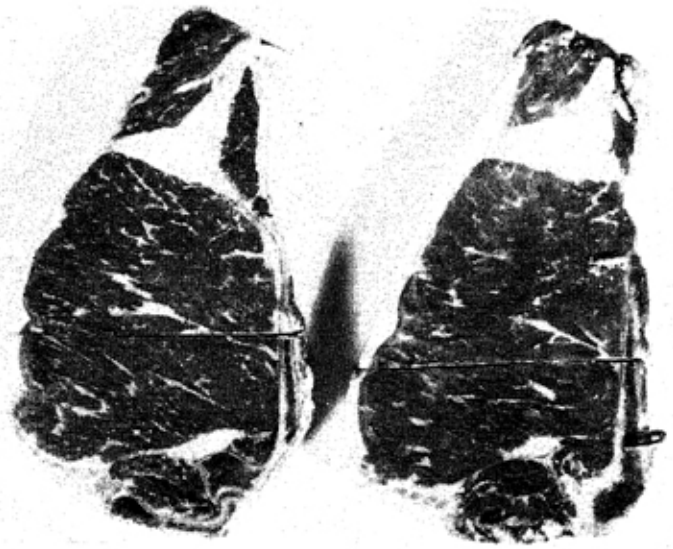
††Letters O and Q were removed in the assigned lettering sequence because of the danger of confusing one for the other.

Fig. 2—Samples of the Three Comparisons.



Good, left, compared with Prime.

Standard, left, compared with $Choice_0$.



$Choice_2$, left, compared with $Choice_1$.

TABLE 1--GENERAL DESIGN SHOWING ASSIGNMENT OF LOINS AND STEAKS TO INDIVIDUAL HOUSEHOLDS WITHIN CLUSTERS

Cluster number	Loin and steak position code[a]						
1	A-1	B-1	C-1	D-1	E-1	F-1	G-1
2	A-2	B-2	C-2	D-2	E-2	F-2	G-2
3	A-3	B-3	C-3	D-3	E-3	F-3	G-3
4	A-4	B-4	C-4	D-4	E-4	F-4	G-4
5	A-5	B-5	C-5	D-5	E-5	F-5	G-5
6	A-6	B-6	C-6	D-6	E-6	F-6	G-6
7	A-7	B-7	C-7	D-7	E-7	F-7	G-7
8	A-8	B-8	C-8	D-8	E-8	F-8	G-8
9	A-9	B-9	C-9	D-9	E-9	F-9	G-9
10	A-10	B-10	C-10	D-10	E-10	F-10	G-10
11	A-11	B-11	C-11	D-11	E-11	F-11	G-11
12	A-12	B-12	C-12	D-12	E-12	F-12	G-12
13	H-1	I-1	J-1	K-1	L-1	M-1	N-1
14	H-2	I-2	J-2	K-2	L-2	M-2	N-2
15	H-3	I-3	J-3	K-3	L-3	M-3	N-3
16	H-4	I-4	J-4	K-4	L-4	M-4	N-4
17	H-5	I-5	J-5	K-5	L-5	M-5	N-5
18	H-6	I-6	J-6	K-6	L-6	M-6	N-6
19	H-7	I-7	J-7	K-7	L-7	M-7	N-7
20	H-8	I-8	J-8	K-8	L-8	M-8	N-8
21	H-9	I-9	J-9	K-9	L-9	M-9	N-9
22	H-10	I-10	J-10	K-10	L-10	M-10	N-10
23	H-11	I-11	J-11	K-11	L-11	M-11	N-11
24	H-12	I-12	J-12	K-12	L-12	M-12	N-12
25	P-1	R-1	S-1	T-1	U-1	V-1	W-1
26	P-2	R-2	S-2	T-2	U-2	V-2	W-2
27	P-3	R-3	S-3	T-3	U-3	V-3	W-3
28	P-4	R-4	S-4	T-4	U-4	V-4	W-4
29	P-5	R-5	S-5	T-5	U-5	V-5	W-5
30	P-6	R-6	S-6	T-6	U-6	V-6	W-6
31	P-7	R-7	S-7	T-7	U-7	V-7	W-7
32	P-8	R-8	S-8	T-8	U-8	V-8	W-8
33	P-9	R-9	S-9	T-9	U-9	V-9	W-9
34	P-10	R-10	S-10	T-10	U-10	V-10	W-10
35	P-11	R-11	S-11	T-11	U-11	V-11	W-11
36	P-12	R-12	S-12	T-12	U-12	V-12	W-12
37	P-13	R-13	S-13	T-13	U-13	V-13	W-13
38	P-14	R-14	S-14	T-14	U-14	V-14	W-14

[a]The loin and steak position code also served to identify households in the panel.

Acceptability of Loins as Associated with Carcass Grades.

Three comparisons were made of steak grades by the household panel—Standard-Choice$_0$, Good-Prime, and Choice$_1$-Choice$_2$. Ratings of steaks of each grade in these comparisons were made twice by each panel member—that is, there were two replicates. Since there were 266 panel households each with two panel members, there was a possible total of 6,384 individual ratings. Due to illness or other family reasons, two households withdrew from the study. In addi-

tion, some panel members failed to give ratings or gave ratings inconsistent with preferences, resulting in a loss of 1.6 percent of ratings possible.‡ Inasmuch as the panel members were involved in the study over a three-week period, the total loss of data was surprisingly low.

The descriptive hedonic rating scale used to obtain ratings of each steak sample in the comparison by each panel member in the household was assigned a numerical rating of 1 to 9 to facilitate analysis. The smaller number denoted a superior rating while the higher number indicated an inferior rating on the hedonic scale. The mid-point in the scale, "Neither Like nor Dislike," was equivalent to 5, numerically (Table 2).

TABLE 2--FREQUENCY OF RATINGS OF BEEF STEAKS ACCORDING TO SCALE, ALL COMPARISONS

Hedonic scale rating	Numerical rating	Men		Women	
		Number	Percent	Number	Percent
No rating[a]	--	48	1.5	54	1.7
Like extremely	1	342	10.8	341	10.7
Like very much	2	869	27.3	966	30.4
Like moderately	3	813	25.6	770	24.2
Like slightly	4	464	14.6	439	13.8
Neither like nor dislike	5	242	7.6	224	7.0
Dislike slightly	6	170	5.3	158	5.0
Dislike moderately	7	101	3.2	96	3.0
Dislike very much	8	98	3.1	89	2.8
Dislike extremely	9	33	1.0	43	1.4
		3180	100.0	3180	100.0

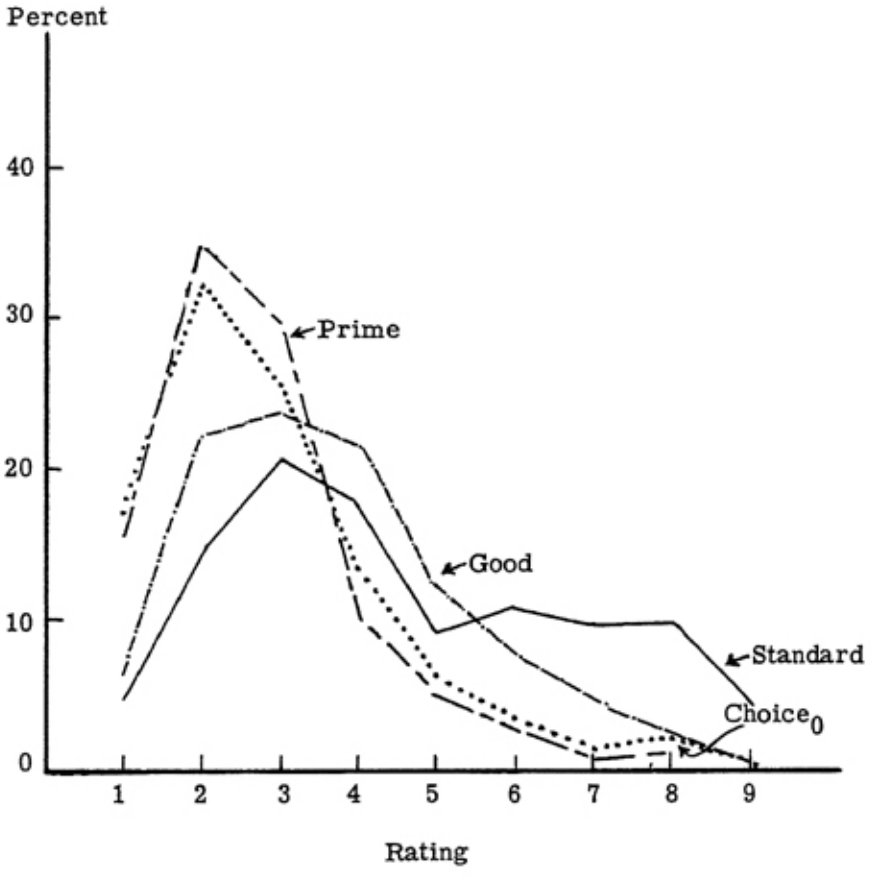
[a]Includes ratings contradictory to expressed preference.

The modal rating given by panel members was 2, i.e. "like very much." More than 50 percent of the total ratings were 2 and 3. Nearly one-third of the ratings by women were rated 2—"Like Very Much." One might assume that the modal rating should be near the mid-point of the scale. However, the fact that loin steaks were used would suggest higher modal ratings, since loin steaks are among the most generally preferred portions of the beef carcass.

While the over-all modal rating was 2, the modal rating for Standard and Good grades of steak was 3 and for $Choice_0$ and Prime it was 2. In the case of $Choice_0$ and Prime, one-third of the ratings were 2, while less than one-fourth of the Standard and Good steaks were rated either 2 or 3. One-third of the ratings of Standard were below 5 (poorer than 5). The distributions are generally more skewed proceeding in grade-ranked order from Standard through Prime (Figure 3). One might infer from the distributions that the lower grades might be more heterogeneous in eating characteristics than the higher grades. It must be remembered that these are sensory elevations of people with varying experience in eating beef; their preferences are based in part on that experience. Yet

‡Schedules from those panel members rating one steak superior and then showing a "preference" for the other steak were omitted from the rating and preference analysis.

Fig. 3—Distribution of Ratings of Steaks by Grade, Both Replicates. (Rating scale began with 1 as the best.)



the tentative inference is plausible, recognizing the weaknesses in methodology concerning the obtaining of the ratings, the limited number of comparisons, and the size of the sample. Further evidence of heterogeneity will be presented later.

The mean rating of men for all samples was 3.31 while that of women was 3.25. Although women rated the samples slightly higher, the difference was not significant. The mean ratings of grades for both men and women were in grade-ranked order—that is, Prime received the lowest numerical mean rating (superior) while Commercial received the highest numerical mean rating. Good and Choice grades were rated in grade-ranked order between Standard and Prime. The Choice grade ratings in the control comparison ($Choice_1$-$Choice_2$) were lower in hedonic scale sense than the ratings of $Choice_0$ in the comparison with Standard (Table 3).

The difference in the ratings of Choice in the control comparison as contrasted with the non-adjacent grade comparisons cannot be fully explained. Part of this difference might be inherent in non-adjacent grade comparisons because panel members were in this instance rating the Choice in relation to Standard grade. In the control comparison the ratings were in relation to grades presumed

TABLE 3 -- MEAN RATING OF LOIN BEEF STEAKS BY GRADE

Grade	Mean Ratings[a] Men	Women
Standard	4.54	4.63
$Choice_0$	2.84	2.75
Good	3.65	3.52
Prime	2.71	2.57
$Choice_1$	3.12	3.09
$Choice_2$	3.14	3.05
Combined Mean	3.31	3.25

[a]Weighted Mean

to be identical. The mean shears of $Choice_1$ and $Choice_2$ were about 2 pounds higher than mean shears of $Choice_0$. This matter is further explored in later sections.

Two major sources of variation in consumer acceptance ratings are (1) differences in attitudes of consumers toward an identical product, and (2) differences in sensory characteristics of products. The second source of variation can largely be controlled in many manufactured or processed products in which considerable homogeneity can be obtained by careful mixing. However, homogeneity of beef steaks cannot be assumed to extend beyond the groups of steaks from a single animal—and small differences may even exist there.

Therefore, differences in mean ratings of groups of steaks from the various carcasses may be caused by sensory differences in the carcasses and/or differences in acceptance attitudes of the consumers rating those steaks. To minimize the influence of particular consumer attitudes upon carcass ratings, each carcass was randomly distributed among 12 or 14 households in as many different neighborhoods. Evidence will be presented later to indicate that most of the differences in mean carcass ratings should be attributed to sensory differences in carcasses rather than to mean differences among groups of consumers.

Sensory differences among carcasses within the same grade were expected. While only beef loins were used in this study, the term "carcass" is employed here as comprising both the left and right loins. Table 4 gives mean ratings, by men, of both left and right loins of the 21 carcasses in the sample for each grade. The range in mean ratings of carcasses was greatest for standard (2.65-6.74) and smallest for Prime (2.07-3.17), see Figure 1.

There were significant differences in the acceptability ratings of 17 of the 21 Standard-$Choice_0$ comparisons, 14 of the 21 Good-Prime comparisons, and 4 of the 21 $Choice_1$-$Choice_2$ comparisons.

Analysis in terms of the mean ratings of eating characteristics given to the 21 carcasses (loin pairs) within each grade suggests wide differences in characteristics for the lower grades, Standard and Good. The variation within a grade resulting from the differences in the carcass mean ratings was not significant for $Choice_0$, $Choice_1$, and Prime. This suggests more homogeneity among carcasses of the higher grades than was true of the lower grades (Table 5).

TABLE 4--MEAN RATINGS BY MEN OF CARCASSES BY GRADES IN BOTH REPLICATES[a]

	Grade												
	Standard		Choice$_0$		Good		Prime		Choice$_1$		Choice$_2$		Group
Carcass	Rank	Rating	Rating	Rank	Rank	Rating	Rating	Rank	Rank	Rating	Rating	Rank	Mean
A	5	3.71	3.33	19	10	3.58	2.08	2	7	2.88	2.75	5	3.05
B	1	2.65	3.65	20	8	3.38	2.67	7	20	3.71	2.83	7	3.15
C	6	4.00	2.45	4	2	2.64	2.95	17	12	3.14	3.00	11	3.03
D	20	5.79	2.62	7	14	3.67	2.83	12	14	3.17	4.54	21	3.77
E	12	4.65	3.22	18	20	4.77	2.95	17	14	3.17	3.71	19	3.75
F	2	3.08	3.12	17	9	3.54	3.00	19	9	2.92	2.91	10	3.10
G	16	5.04	2.46	5	4	3.17	2.79	11	6	2.86	3.55	18	3.31
H	8	4.14	2.76	10	6	3.35	2.35	4	8	2.91	2.91	10	3.07
I	7	4.12	3.00	15	13	3.63	2.79	11	18	3.50	3.29	14	3.39
J	9	4.39	2.91	12	16	3.96	3.13	19	2	2.52	3.13	13	3.34
K	18	5.52	2.74	9	17	4.04	2.50	5	16	3.38	2.54	2	3.45
L	19	5.61	3.00	15	8	3.38	2.79	11	1	2.27	4.18	20	3.54
M	12	4.65	2.43	2	21	4.79	3.17	21	3	2.57	3.39	16	3.50
N	15	4.96	3.08	16	12	3.61	2.17	3	10	3.00	3.48	17	3.38
P	10	4.64	2.96	13	5	3.33	3.00	19	17	3.39	2.64	4	3.33
R	4	3.48	3.89	21	3	3.04	2.92	15	5	2.81	2.59	3	3.12
S	13	4.89	2.86	11	18	4.08	2.68	8	4	2.75	2.89	8	3.36
T	21	6.74	2.44	3	1	2.59	2.07	1	21	3.74	2.52	1	3.35
U	17	5.33	2.15	1	19	4.22	2.89	14	19	3.54	3.35	15	3.58
V	3	3.11	2.50	6	11	3.60	2.56	6	15	3.35	2.77	6	2.98
W	15	4.96	2.74	9	15	3.93	2.85	13	11	3.11	3.07	12	3.44
Mean Grade Rating		4.54	2.84			3.65	2.71			3.12	3.14		3.31
Range		2.65-6.74	2.15-3.89			2.59-4.79	2.07-3.17			2.27-3.74	2.52-4.54		2.98-3.77

[a]Average of the mean ratings of left and right loins. Since the ratings of husbands and wives were highly correlated, the ratings of only the men are used in much of the analysis.

TABLE 5 -- SOURCE OF VARIATION IN MEAN RATINGS BY MEN OF CARCASSES BY GRADES

Source of Variation	Degrees of Freedom	Observed F Ratio Grade					
		Standard	Choice$_0$	Good	Prime	Choice$_1$	Choice$_2$
Between Carcasses Within a Grade	20	8.48**	1.02	4.31**	1.62	1.00	3.73**
Within Carcass Means of Two Replicates	21						

**Significant at .01 level.

When the individual loin score ratings of men were analyzed for significant sources of variation among carcasses within grades, it was found that all 21 Prime grade loin pairs were rated quite uniformly (Table 6). There was significant disagreement in the ratings of Commercial and only slightly less in the case of Good. The three groups of the Choice grade were rated quite uniformly but less so than the Prime grade. A significant source of variation in individual ratings appeared in three of six rating opportunities for the Choice grade.

TABLE 6 -- SOURCE OF VARIATION IN INDIVIDUAL RATINGS BY MEN AMONG LOINS BY GRADES

Source of Variation	Grade					
	Standard	Choice$_0$	Good	Prime	Choice$_1$	Choice$_2$
Replicate 1						
Degrees of Freedom						
Between Loins	20	20	20	20	20	20
Within Loins	229	229	229	229	224	224
Observed F Ratio	3.49**	1.91*	2.39**	.82	2.11**	.91
Replicate 2						
Degrees of Freedom						
Between Loins	20	20	20	20	20	20
Within Loins	226	226	223	223	220	220
Observed F Ratio	3.08**	1.14	1.94*	1.39	.92	2.26**

*Significant at .05 level.
**Significant at .01 level.

Graphic representation of the distribution of the mean carcass ratings for each grade is shown in Figure 4. Although there were but 21 carcasses in the sample for each grade, an indication of the nature of the distributions can be obtained. The leaner grades exhibit a wider dispersion of mean carcass ratings. On the other hand, the range of mean carcass ratings of Choice and Prime is less. In terms of rating evaluations, Standard is the most heterogeneous and Prime the most homogeneous of the four grades. Yet, some Standard carcasses were rated as high as Prime carcasses—that is, the distributions overlap considerably. The distributions of the carcass ratings of all grades overlap and in the case of Choice$_0$ and Prime in the first replicate there is considerable similarity in the distributions. The distribution of carcass ratings shifted to the right from the first to the second replicate for the two fatter grades and to the left for the two leaner grades.

Fig. 4—Distribution of Mean Loin Ratings by Grade. (Rating by men, rounded to nearest quarter.)

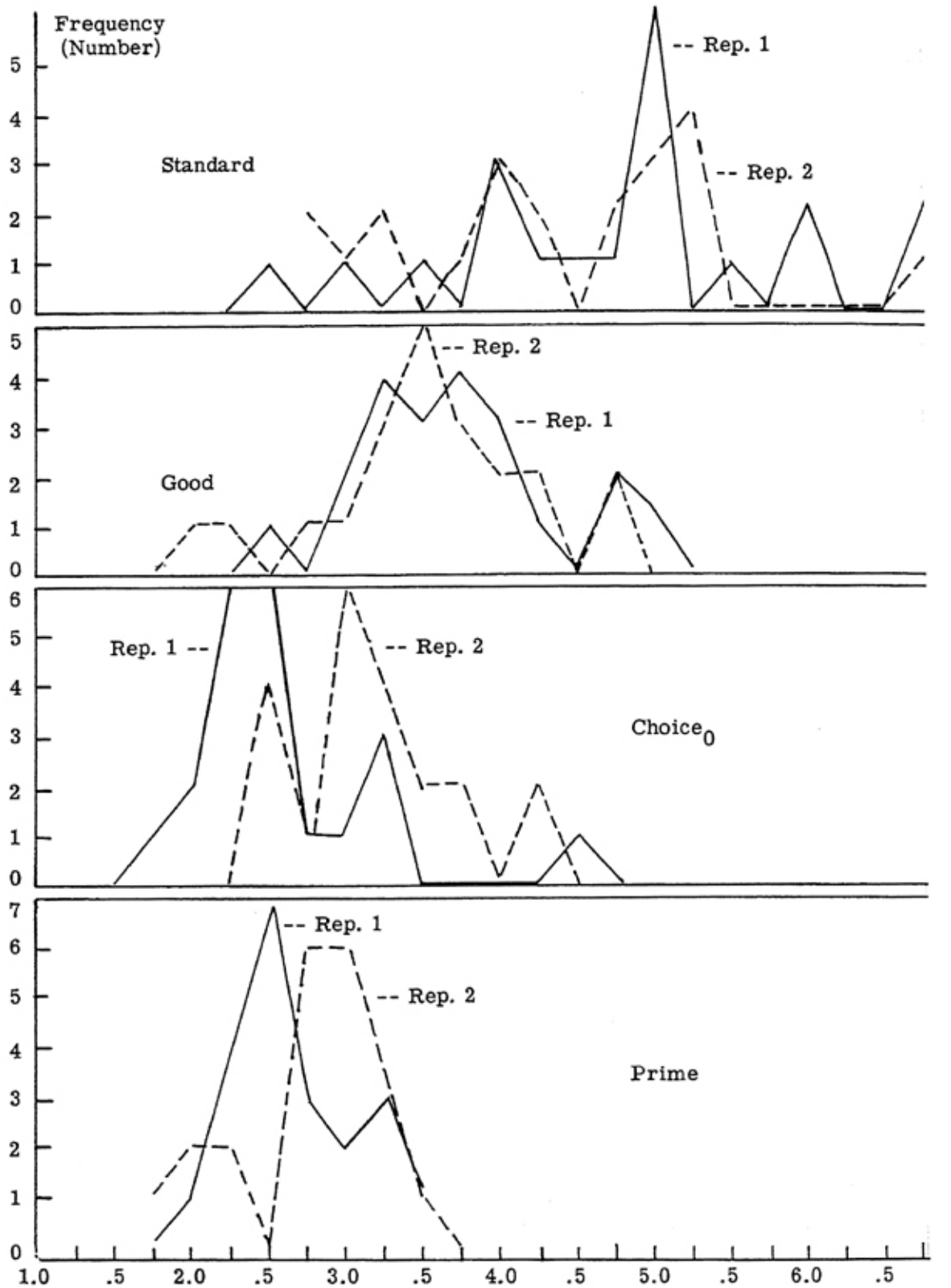
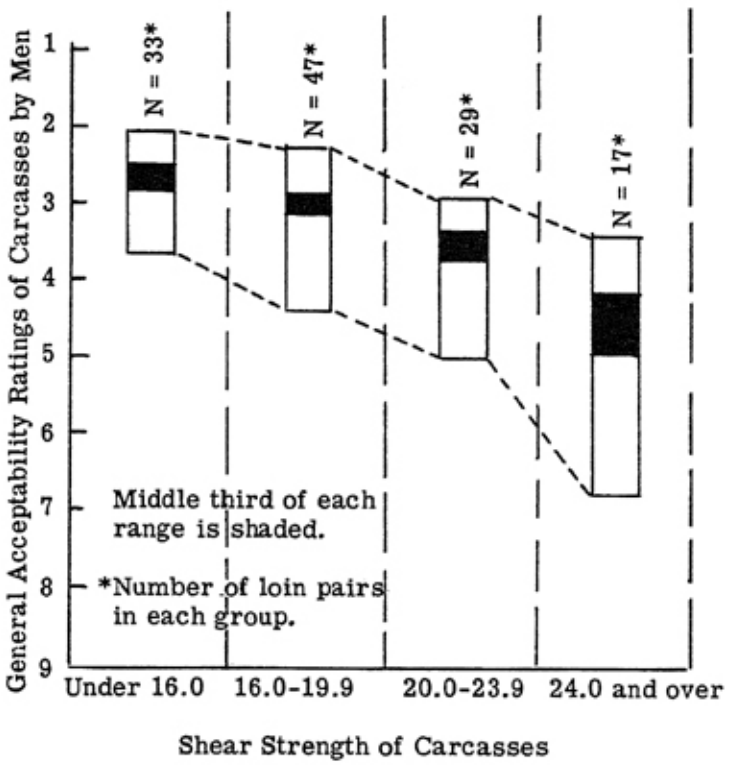


Fig. 5—Distribution of Carcasses (Loins) by Shear Group, St. Louis Panel. (Rating scale began with 1 as the best.)



Acceptability of Loins as Associated With Shear Values.

Shear-Acceptance: The smaller the carcass shear rating the smaller (better) in general was the carcass acceptability rating. The mean acceptability rating of the 33 carcasses shearing below 16 pounds was 2.73 with a range of 2.07 to 3.65, while the mean rating of the 17 carcasses shearing above 24 pounds was 4.61 with a range of 3.38 to 6.74 (Table 7 and Figure 5). The higher the shear group the more heterogeneous was loin acceptability. In general, prediction of acceptability by shear measurements alone would have been about as accurate as by grade alone. The grade method of prediction would have been least accurate at the leaner side of the grade scale, and the shear method would have been least accurate at the higher side of the shear scale.

Size of mean shear differences in paired carcasses was related to acceptability differences (Table 8). While the relationship was weak, it is of some usefulness. The association of changes in direction of the shear and rating differences in the $Choice_1$-$Choice_2$ comparisons is particularly interesting (Table 9). In 17 of the 21 pairs the Choice loin with the smaller shear had the better acceptance rating.

Shear-Grade: The lower mean shear values were associated with the two higher grades (Tables 7 and 10). In contrast to the shear values obtained in the sample of loins used for the laboratory panel[20] (see Appendix, Table 25) the mean shear values of the carcasses in this study were in grade-ranked order—that is, the low-

TABLE 7 -- MEAN RATINGS OF CARCASSES BY FOUR SHEAR GROUPS

Shear Pressure in Pounds

Under 16		16-19.9		20.0-23.9		24.0 and Over	
Shear	Rating	Shear	Rating	Shear	Rating	Shear	Rating
(Prime)		(Prime)		(Prime)		(Choice2)	
N- 8.2	2.17	P-16.5	3.00	U-21.4	2.89	M-25.1	3.39
T-10.8	2.07	W-17.7	2.85	(Choice0)		D-26.3	4.54
B-12.7	2.67	F-17.8	3.00	F-21.0	3.12	(Good)	
R-13.0	2.92	C-18.1	2.95	A-22.0	3.33	B-24.5	3.38
I- 13.2	2.79	L-18.7	2.79	(Choice1)		E-25.0	4.77
K-13.4	2.50	E-19.5	2.95	B-20.1	3.71	S-26.8	4.08
H-13.6	2.35	S-19.5	2.68	U-20.7	3.54	J-30.1	3.96
G-14.2	2.79	(Choice0)		T-21.2	3.74	U-30.1	4.22
J-14.2	3.13	C-16.5	2.45	D-21.9	3.17	(Commercial)	
V-14.4	2.56	L-16.7	3.00	E-22.1	3.17	L-25.5	5.61
M-15.0	3.17	H.17.0	2.76	P-22.3	3.39	K-25.7	5.52
A-15.5	2.08	I- 17.5	3.00	I- 23.2	3.50	N-26.7	4.96
D-15.7	2.83	W-17.5	2.74	(Choice2)		D-27.2	5.79
(Choice0)		R-17.6	3.89	S-20.0	2.89	A-28.3	3.71
M-11.3	2.43	K-18.0	2.74	I- 21.6	3.29	I- 28.4	4.12
J-11.8	2.91	N-18.0	3.08	U-21.7	3.35	C-28.5	4.00
D-12.4	2.62	S-18.4	2.86	L-23.4	4.18	P-28.6	4.64
B-13.9	3.65	P-18.6	2.96	N-23.8	3.48	S-29.6	4.89
G-13.9	2.46	E-18.8	3.22	(Good)		T-41.5	6.74
U-14.3	2.15	(Choice1)		P-20.2	3.33	Range	3.38 - 6.74
V-14.7	2.50	F-16.4	2.92	W-20.2	3.93	Mean	28.11 4.61
T-15.3	2.44	S-16.6	2.75	M-20.8	4.79	N = 17	
(Choice1)		R-16.8	2.81	N-21.4	3.61		
J-13.9	2.52	W-16.8	3.11	R-21.5	3.04		
M-15.4	2.57	L-17.0	2.27	F-21.6	3.54		
(Choice2)		G-17.5	2.86	A-22.6	3.58		
V-11.7	2.77	A-17.6	2.88	K-22.9	4.04		
K-12.3	2.54	K-17.6	3.38	(Commercial)			
B-12.7	2.83	N-17.6	3.00	U-20.8	5.33		
W-13.5	3.07	V-18.0	3.35	M-20.8	4.65		
(Good)		C-18.9	3.14	G-22.0	5.04		
G-12.5	3.17	H-19.3	2.91	W-22.3	4.96		
T-12.9	2.59	(Choice2)		F-22.6	3.08		
V-15.5	3.60	J-16.1	3.13	E-23.1	4.65		
L-15.8	3.38	T-16.1	2.52	Range	2.89 - 4.96		
(Commercial)		A-16.4	2.75	Mean	21.69 3.74		
B-15.0	2.65	P-16.7	2.64	N = 29			
V-15.0	3.11	C-17.0	3.00				
Range	2.07 - 3.65	R-17.3	2.59				
Mean	13.57 2.73	F-17.8	2.91				
N = 33		E-18.5	3.71				
		H-18.7	2.91				
		G-19.8	3.55				
		(Good)					
		H-17.5	3.58				
		I- 17.7	3.63				
		C-18.1	2.64				
		D-19.2	3.67				
		(Commercial)					
		J-17.5	4.39				
		R-17.5	3.48				
		H-19.5	4.14				
		Range	2.27 - 4.39				
		Mean	17.74 3.05				
		N = 47					

TABLE 8--CORRELATION COEFFICIENTS FOR SELECTED PHYSICAL AND HEDONIC SCALE RELATIONSHIPS[a]

		Correlation Coefficient	95% Confidence Limits[b]
Difference in Mean Shear of the Comparison to Difference in Mean Ratings of Carcasses			
Comparison:	N		
Standard - $Choice_0$	21	.75	.45 to .88
Good - Prime	21	.33	-.13 to .67
$Choice_1$ - $Choice_2$	21	.63	.26 to .83
All Grades	63	.67	.48 to .78
Mean Shear Value to Mean Ratings of Carcasses			
Grade:			
Standard	21	.64	.27 to .83
Good	21	.59	.19 to .81
$Choice_0$	21	.45	.01 to .74
$Choice_1$	21	.74	.44 to .88
$Choice_2$	21	.24	-.22 to .61
Prime	21	.53	.11 to .78

[a]Carcass ratings represent those by men only.

[b]Confidence limits should be interpreted cautiously because of the limitations upon the sampling techniques and the subjective nature of acceptance data.

TABLE 9--DIFFERENCE IN MEAN RATINGS OF MEN AND DIFFERENCE IN SHEAR VALUE OF LOIN PAIRS BY COMPARISONS

	Standard-$Choice_0$		Good-Prime		$Choice_1$-$Choice_2$	
Loin Pair	Difference in Mean Rating[a]	Difference in Shear Value[b]	Difference in Mean Rating[a]	Difference in Shear Value[b]	Difference in Mean Rating[a]	Difference in Shear Value[b]
A	0.38	6.3	1.50	7.1	0.13	1.2
B	-1.00	1.1	0.71	11.8	0.88	8.0
C	1.55	12.0	-0.31	0.0	0.14	1.9
D	3.17	14.8	0.84	3.5	-1.37	-4.4
E	1.43	4.3	1.82	5.5	-0.54	3.6
F	-0.40	1.6	0.54	3.8	0.01	-1.4
G	2.58	8.1	0.38	- 1.7	-0.69	-2.3
H	1.38	2.5	1.00	3.9	0.00	0.6
I	1.12	10.9	0.84	4.5	0.21	1.6
J	1.48	5.7	0.83	15.9	-0.61	-2.2
K	2.78	7.7	1.54	9.5	0.84	5.3
L	2.61	8.8	0.59	- 2.9	-1.91	-6.4
M	2.22	9.5	1.62	5.8	-0.82	-9.7
N	1.88	8.7	1.44	13.2	-0.48	-6.2
P	1.68	10.0	0.33	3.7	0.75	5.6
R	-0.41	- 0.1	0.12	8.5	0.22	-0.5
S	2.03	11.2	1.40	7.3	-0.14	-3.4
T	4.30	26.2	0.52	2.1	1.22	5.1
U	3.18	6.3	1.33	9.3	0.19	-1.0
V	0.61	0.3	1.04	1.1	0.58	6.3
W	2.22	4.8	1.08	2.5	0.04	3.3
Grade	1.70	7.7	0.94	5.4	-0.02	0.2

[a]A negative difference in mean rating denotes that the first grade in the comparison had a lower (better) mean rating than the second grade.

[b]A negative difference in shear indicates that the first grade in the comparison had the smaller shear.

TABLE 10--MEAN SHEAR VALUES OF CARCASSES BY GRADES

Carcass	Mean Shear[a]	Standard Deviation
Commercial	24.1	5.93
Good	20.8	4.91
$Choice_0$	16.4	2.98
$Choice_1$	18.6	2.78
$Choice_2$	18.4	4.09
Prime	15.4	3.03

[a]Mean shear values represent the average of 9 core readings; 3 dorsal, 3 medial and 3 ventral positions on each steak.

est mean shears were for Prime and the highest for the Standard. This may have been in part due to actual variation in the cattle. The cattle in the laboratory phase were obtained in the spring season while in this phase they were obtained in the early fall season.

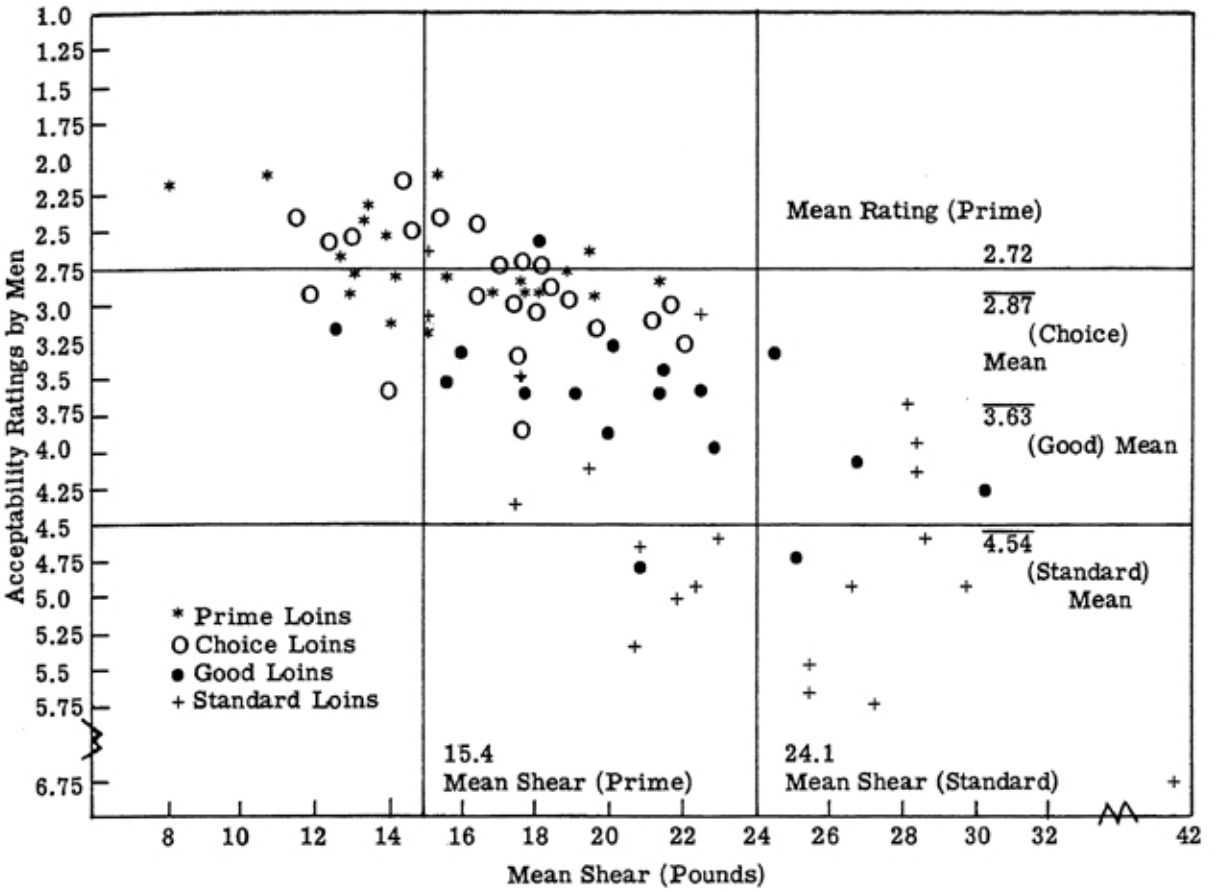
Reference has been made to the relatively great heterogeneity of the Standard and Good grades. Carcass ratings from the Standard and Good grades were classified according to two groups of mean shear values—i.e., less than 22.0 pounds and 22.0 pounds and over. This division of shear values at 22.0 pounds produced a disproportionate distribution of the 42 carcasses: 14 Good and 7 Standard carcasses had shears less than 22 pounds, while 7 Good and 14 Standard carcasses had shears 22 pounds or greater.

An analysis of variance indicates that both grade and shear variation were significant at the .05 level. It is apparent that both shear and subjective grade criteria were independently related to carcass rating evaluations. This suggests that using both physical shear and subjective grade criteria in the leaner grade carcasses would significantly increase homogeneity of ratings of the "grade". It is suspected that the problems of homogeneity of the leaner grades may have a number of additional dimensions that need not be considered in connection with the fatter grades.

Shear-Grade-Acceptability: While shear and grade were somewhat related on the average, there was sufficient independence that use of the two factors together increased ability to predict consumer ratings. This was especially true for the leaner carcasses. Figure 6 shows the values of shears and mean ratings for all carcasses in the Standard-$Choice_0$ and Good-Prime comparisons. The smaller mean ratings represent superior evaluations of rating characteristics of carcasses and the smaller shear values represent a physical measure of carcasses having greater tenderness. On the diagram are lines showing the mean ratings and also the mean shear values of Standard and Prime carcasses.

Two important generalizations may be drawn from Figure 6. First, the leaner grades are more heterogeneous as measured by both shear and ratings. Prime and $Choice_0$ grade carcasses did not exceed 23 pounds shear or a 3.89 rating. Second, an important segment of the carcasses in the leaner grades had low shears and superior acceptability ratings. For example, 14 of the 21 Good carcasses rated better than 3.89 which was the rating of the poorest $Choice_0$ car-

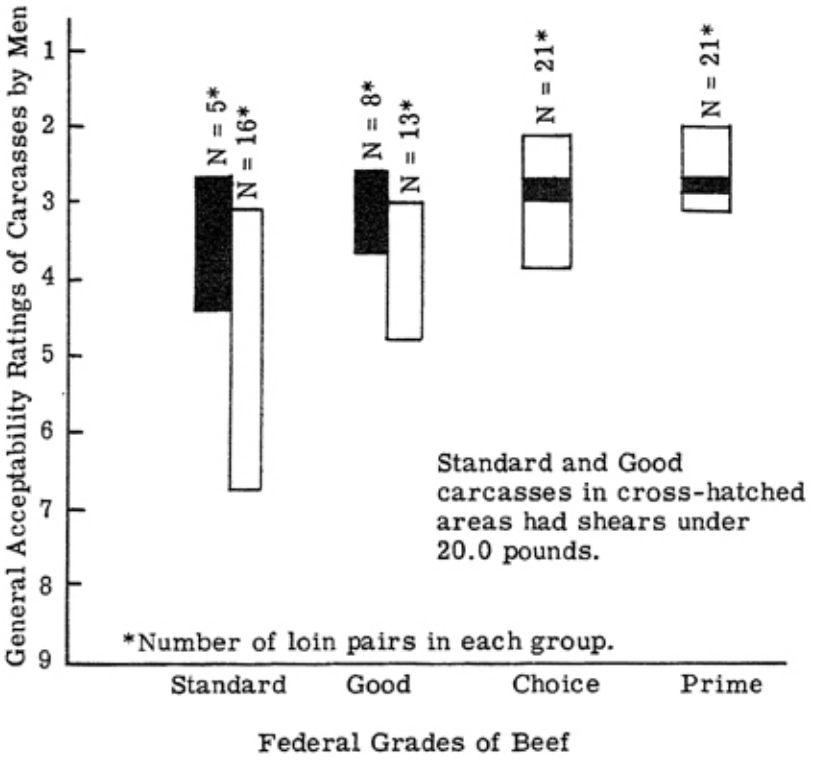
Fig. 6—Relationship of Mean Ratings of Loins by Men to Shear and Grade.



cass. It might be added that 2 of the Choice$_2$ carcasses rated poorer than 3.89 (Table 4). Moreover, 5 of the 21 Standard carcass mean ratings were better than 3.89. It is possible that the boundary line for very acceptable carcasses in this study should be at 3.25 or some other number slightly smaller than 3.89. However, it is obvious that almost all of the carcasses of the Choice and Prime grades, a small portion of the Standard grade, and one-half to two-thirds of the Good grade were considered highly acceptable by most consumers. The large market price differentials associated with the division of *these highly acceptable carcasses* into four different grades were not economically justified. This indicates the need for a method of separating the highly acceptable lean carcasses from the less acceptable lean carcasses. While virtually all the fatter carcasses were highly acceptable, this question should be raised: to what extent was the high acceptance caused by feeding and management and to what extent was it caused by the genetic make-up of the animals?

Inference regarding the carcass population is severely limited by the small size of the sample. The meager evidence from the two shear tests indicates that there is probably considerable shear variation within the population which is independent of degree of finish or grade.

Fig. 7—Relationship of Carcass (Loin) Ratings to Grade and Shear, St. Louis Panel.



There is an errors-in-data problem as well as the sampling problem. It has been shown that such errors not only reduce the correlation between two or more variables but also may give a biased estimate of the regression relationship.[26] More precise measurements of acceptability and of objective tenderness are greatly needed. Interpretation of correlation data is also hampered by the small numbers in the samples and the non-normality of the population.

Prediction of acceptability of leaner loins would have been better through use of both shear and grade rather than either alone. As indicated in Figure 7, those Standard and Good loins with shear measurements of less than 20 pounds were all nearly as acceptable as the Choice loins. The results of this type of two-variable prediction are certainly less than ideal but they represent a real improvement.

Recombination of Grades: It was shown that the Choice and Prime grades were more homogeneous than Standard or Good whether measured by shear or ratings of eating characteristics. The possibility of combining Choice and Prime (the two fatter grades) was explored in a grade classification—that is, the ratings of the 42 Choice$_0$ and Prime carcasses were combined and considered as one "grade." In this analysis the carcass ratings of Prime-Choice combined, Good, and Standard provided for a three-"grade" classification. The analysis shows that variation within a grade by this classification was significantly less than varia-

TABLE 11--SOURCE OF VARIATION IN RATINGS OF CARCASSES CLASSIFIED ACCORDING TO VARIOUS METHODS

Source of Variation Classification	Degrees of Freedom	Mean Square	Observed F Ratio[a]
Three "Grades" Classification:			
Grade	2	43.25	83.01**
Replicate	1	.09	
Interaction	2	2.03	
Within	162	.502	
Four "Grades" Classification:			
Grade	3	29.15	55.95**
Replicate	1	.09	
Interaction	3	1.727	
Within	160	.498	
Three "Grades" Classification:			
Shear	2	34.70	55.34**
Replicate	1	.10	
Interaction	2	.630	
Within	162	.627	

**Significant at .01 level.

[a]Error term = Interaction + Within Sum of Squares divided by combined Degrees of Freedom.

tion among grades (Table 11). A similar analysis of ratings in relation to the four grades indicated that grade was a significant classification of the rating evaluations.

Mean ratings of the 126 carcasses in the study were classified according to three groups of shear values. The three shear groups were: less than 16.5 pounds, 16.5 to 20.7 pounds, and 20.8 pounds and over. Again the classification of carcass ratings by shear values was significant.

A significant observed F test would suggest that the mean ratings came from different populations. In the analysis above several classifications of ratings either by shear or by grade proved to be significant; hence these classifications were useful in sorting the total population into groups. Had group classifications not been significant, sorting into either these grades or these shear groups would have been of little use in predicting ratings of eating characteristics.

Since the four-grade, three-"grade," and three-shear groups all proved to be significant at the 1 percent level, the question arises concerning which method is superior. A criterion of choice among several ways of classifying data has been suggested by Rosander.[27] He states that a superior statistical criterion is that classification which maximizes the variation between classes and minimizes the variation within classes. The larger ratio of the between variation to within variation expresses the better classification.

In the three alternative methods the classification of the three "grades"—in which Prime and Choice grades were combined—was superior to either the four-grade or three-shear classifications (Table 12). There was sufficient overlap in the ratings of the Prime-Choice grades to render the additional grade classification from three to four less efficient overall in classification. Generally, it would

TABLE 12--COMPARISON OF MEAN SQUARES OF THREE METHODS OF CLASSIFICATION OF CARCASS RATINGS

Method of Classification	Mean Square Between Classes	Mean Square Within Classes	Ratio Between/Within Mean Square
Three "Grades"	43.45	.502	86.6
Four "Grades"	29.15	.498	58.5
Three "Shears"	34.70	.627	55.3

be expected that additional groupings, each being fairly homogeneous, would improve classification efficiency. However, panel members in the aggregate apparently found little difference between Choice and Prime. This raises a number of questions concerning the division of the fatter segment of beef into two grades. Is market differentiation in terms of two grades justified? Probably only a few connoisseurs would be able to detect differences between them. Is the function of the Prime grade primarily one of obtaining product differentiation and premium pricing for part of the fatter end of beef?

Subjective measures—i.e., grading of carcass—compared favorably with objective measures in obtaining "grade" homogeneity of ratings. However, it has been shown that either method departs considerably from the ideal in assuring homogeneity of eating characteristics. It is believed that improvement in grade homogeneity will result from further development of techniques of objective measurement of tenderness and other sensory factors associated with eating satisfaction. It is doubtful that subjective measures can be refined to obtain the necessary degree of precision to assure homogeneity.

Other Predictors of Consumer Acceptance.

Socio-Economic Characteristics of Panel Members: The possibility was anticipated that age, education, and income levels of panel members might influence both the level of over-all ratings and differences among grades. Information on age of the housewife was obtained and related to ratings of steaks by grades in the two non-adjoining grade comparisons. Younger housewives—those under 25 years of age—tended to give poorer general over-all ratings while older women in general rated steaks better. This difference reflects either a more favorable attitude toward steaks by older women or a difference in their interpretation of the rating scale as compared with that of younger women. The nature of the interaction of age and grade ratings is indicated in Figure 8.

Housewives with fewer years of formal education tended to rate all grades and particularly the lower grades of steaks a little better than did those with more formal education (Figure 9).

There appeared to be an inverse relationship of over-all mean ratings to income levels of the household—that is, the ratings of all grades combined were substantially better for the lower-income groups while the higher-income groups (over $4,000) tended to rate all beef grades at a lower hedonic-scale rating. The ratings of the Standard grade among the three income groups was substantially

Fig. 8—Mean Ratings (Second Replicate) of Grades of Steaks by Age of Housewife.

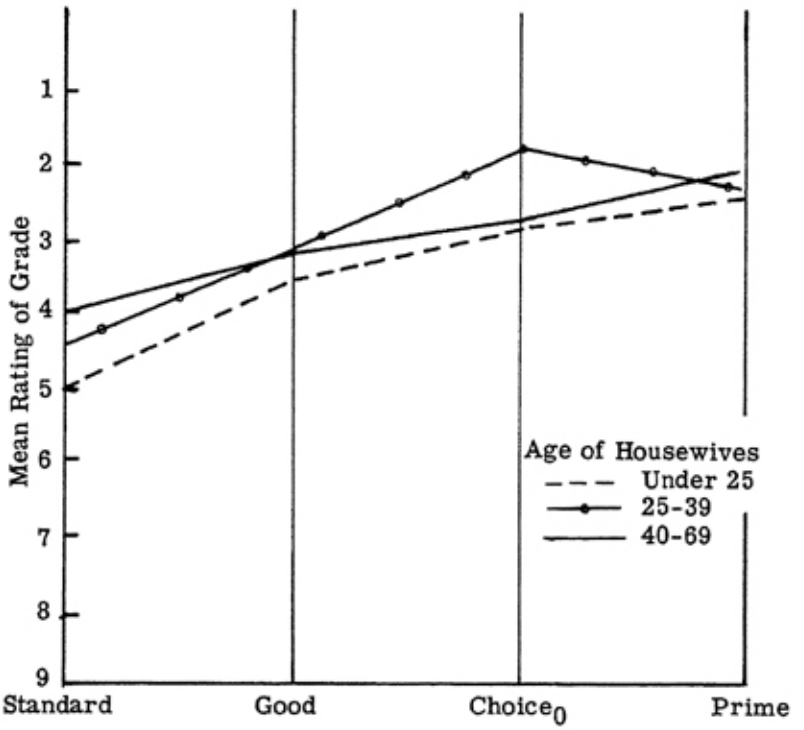


Fig. 9—Mean Rating (Second Replicate) of Grades of Steaks by Educational Level of Housewives.

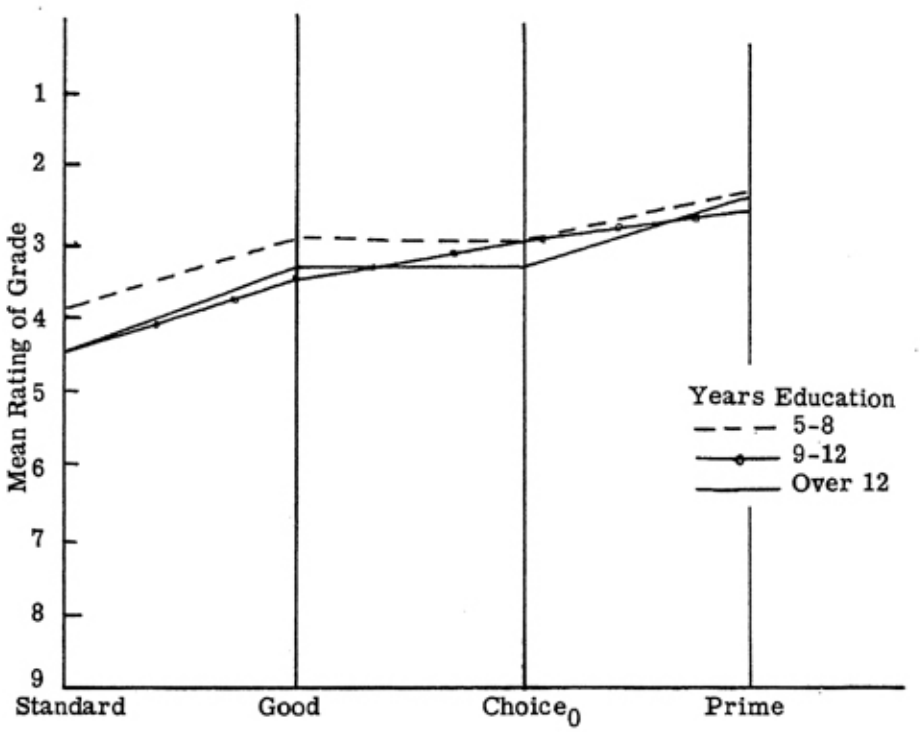


Fig. 10—Men's Mean Ratings of Grades of Steak by Household Income.

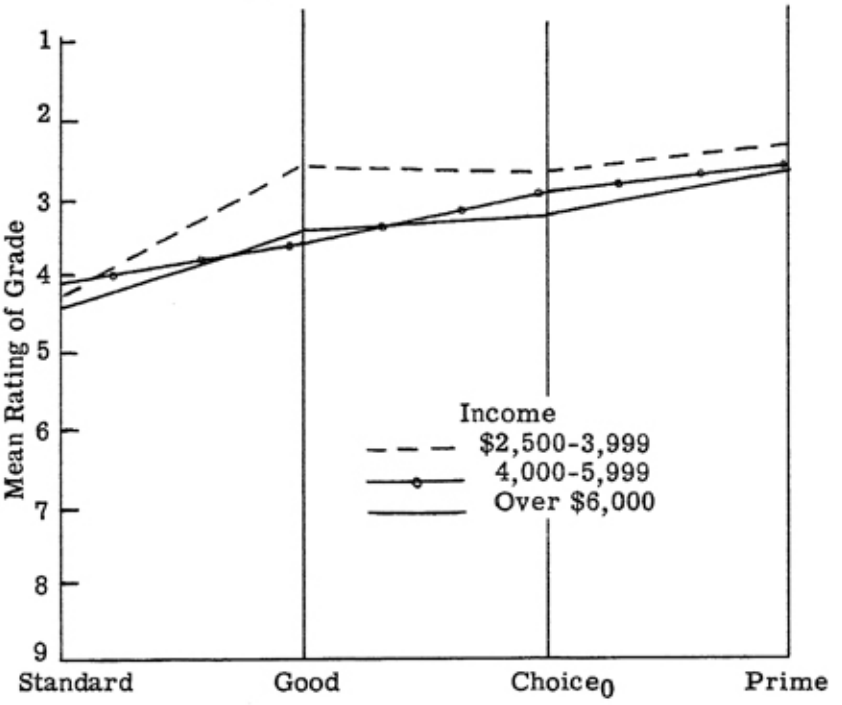


TABLE 13--MEAN RATINGS OF GRADES OF STEAKS ACCORDING TO INCOME OF THE HOUSEHOLD[a]

Grade	Annual Income[b] $2,500 3,999	$4,000 5,999	Over $6,000
Men:			
Standard	4.31	4.21	4.45
Choice$_0$	2.79	3.06	3.42
Good	2.73	3.64	3.58
Prime	2.44	2.78	2.85
Average mean rating[c]	3.00	3.42	3.57
Women:			
Standard	4.13	4.36	4.51
Choice$_0$	2.63	3.04	3.34
Good	2.85	3.60	3.34
Prime	3.13	2.70	2.65
Average mean rating[c]	3.01	3.42	3.44

[a]Second replicate ratings.
[b]Annual income of household after taxes. Households with incomes below $2500 were excluded from the panel.
[c]Weighted average.

the same for men while women in the lowest income group tended to give higher ratings to that grade than did the higher-income groups (Table 13 and Figure 10).

Numerous studies have shown that income and educational levels are positively associated. This relationship is apparent in the similarity of the ratings of the higher income and higher educated. Conversely those with lower incomes and fewer years of formal education were similar in absolute over-all ratings.

An analysis in which income was included as a possible source of variation among individual rating scores showed that income was a very significant influence. The higher the income of male respondents, the poorer the rating given *all* grades. However, the small size of the income-grade interaction indicates that grade acceptability was not related to income. Neither sex of respondents nor replication contributed significantly to the variation in the individual rating scores of steaks. Grade-replication was the only significant interaction.

The reader should be cautioned that these income results are biased by product variation unless such variation was equally distributed across income groups. While the random design should have given a nearly equal distribution of heterogeneity, there is no way of knowing its degree of success. The absence of product homogeneity within grades tremendously complicates the analysis. Moreover, the difference between the mean acceptance ratings of men and women might have been larger or smaller if they had made their ratings completely independent of each other.

Cooking Methods and Degree of Doneness: It is recognized that cooking methods can materially affect the flavor, tenderness, juiciness, and general acceptability of steaks. The panel members were instructed to cook the steaks by their customary method and length of time. No recommendations were made to them concerning the methods of cooking in relation to degree of finish. It would have been impractical to indicate differential cooking methods to panel members and it would have been difficult to determine if such instructions had been carried out. Furthermore, such instructions would have involved identifying the higher degree of finish and, consequently, would have meant a large modification of the design and approach of the study. Broiling and pan-fry methods of cooking the steaks were by far the most popular and represented about 93 percent of the total cooking methods used. Braising was used by 5 percent of the households. The method of cooking did not change appreciably during the test—that is, the proportion of households broiling, frying, or braising the steaks did not change materially from replicate to replicate or among the three comparisons (Table 14).

TABLE 14--PERCENTAGE OF HOUSEHOLDS REPORTING METHOD USED IN COOKING BEEF STEAKS

Method	Standard-Choice (Percent)	Good-Prime (Percent)	$Choice_1$-$Choice_2$ (Percent)	Mean- (Percent)
Broil	48	45	46	46
Pan-Fry	46	47	47	47
Braise	5	6	4	5
Other	1	2	3	2

Nearly three-fourths of the panel households cooked the steaks well done. This agrees with the 75 percent of the St. Louis sample in 1954 who reported a preference for the well or medium-well degrees of doneness.[14] It should be recognized that these are subjective evaluations on the part of the respondents. For

example, what would be rare to a particular family might be considered well-done by another family. The schedule defined well-done as the absence of any internal pink. There were differences with respect to doneness according to certain socio-economic characteristics. The degree of doneness was related to income. Rareness was more popular among the higher income groups and among those with higher education. Degree of doneness was not associated with age of housewife.

A larger proportion of those braising steaks than those using other methods cooked the steaks to well-done. Those that preferred rare beef appeared to use the broiling method most frequently (Table 15). Broiling was much more popular among the younger housewives while the frying method was more frequently used among the older housewives. A larger proportion of the higher income families and those housewives with more years of formal education broiled the steaks, while the frying methods were more popular with those with lower incomes and less education. No relationship of cooking methods to whether or not the housewife was employed was found.

TABLE 15--PERCENTAGE OF HOUSEHOLDS REPORTING DEGREE OF DONENESS ACCORDING TO METHOD OF COOKING

Method of Cooking	Degree of Doneness		
	Well	Rare	Both
Broil	68	28	4
Pan-Fry	77	21	2
Braise	83	15	2
Other	--	--	-
Mean	72	24	4

The ¾-inch steaks were possibly considered too thin for broiling rare by some people who prefer that method and degree of doneness. Thus, providing a standard steak thickness may have confounded slightly the degree of doneness, method of cooking, and degree of acceptability.

Analysis of the source of variation, using individual rating scores of the panel members rather than mean ratings, revealed that differences in the rating scores of persons who broiled and those who did not were significant at the 5 percent level. (Table 16) Those who broiled gave slightly less favorable ratings on the average to all steaks than did those who cooked by other methods. Since a larger proportion of people with higher incomes cook steaks by broiling, the less favorable ratings of those who broil and those with higher incomes probably reflect a single factor in attitudes of the same general group. The interaction of the acceptability of grades and method-of-cooking was small and insignificant. Differences in rating scores by sex of panel member and by replicates were not significant. As expected, rating scores by grade were an important source of variation. The grade-replication interaction was large and significant; it reflected a significant shift in individual score ratings from the first to the second replicate, particularly in the Standard-Choice$_0$ comparison.

TABLE 16--SOURCE OF VARIATION OF INDIVIDUAL RATINGS IN THE STANDARD-CHOICE$_0$ AND GOOD-PRIME COMPARISONS

Source of Variation	Degrees of Freedom	Mean Square	Observed F
Method of Cooking	1	4.63	5.58*
Sex	1	1.20	1.45
Replication	1	.46	.55
Grade	3	129.72	156.48**
Interaction			
(Grade-Replication)	(3)	(9.14)	(11.03)**
(Other Interaction)	(22)	(.19)	(.23)
Within	640	.829	

*Significant at .05 level.
**Significant at .01 level.

Laboratory Discrimination Tests: Can laboratory evidence of differences in products be used to predict differences in consumer acceptability of those products?

The general design provided for laboratory taste tests of each of 14 pairs of loins used in the larger consumer panel. Six steaks from each pair of loins (A through N) were reserved and stored for the laboratory panel. These steaks were placed in bags, coded for identification, and stored at 0° F until they were withdrawn for panel testing.

The laboratory panel was organized in January and tasting of samples was continued through February, 1956. The panel was composed of eight student members who were employed as tasters. The panel was divided into two sections, four tasting in the morning and four in the afternoon with each section composed of two male and two female tasters. Panel procedure was quite similar to that of previous panels.[20] General procedure and some methodological comments are in the Appendix.

The trio comparison method was used in these discrimination tests. The comparisons were identical with those in the consumer panel in St. Louis—that is, Choice$_0$-Standard, Prime-Good, Choice$_1$-Choice $_2$, and matched samples of steaks from the same loins compared in the St. Louis panel.

The six steaks from each pair of loins of each grade of the comparisons provided for four trio tests on each pair of loins. As there were 14 pairs of loins, 56 trio tests were planned for each grade comparison, such as the Choice$_0$-Standard. The number of trio tests in the three grade comparisons totalled 168. Each trio test involved four individual taste comparisons by individual panel members and hence there was the possibility for a total of 672 individual taste comparisons. The actual total number of individual taste comparisons was 635. The difference is accounted for, in part, by absences of panel members and also by the unsuitability of steak samples from one pair of loins for tasting.

A total of 387 discriminations out of the 635 individual taste comparisons were successful. The over-all average percentage of correct discriminations was 60.94. Of all discriminations, an estimated 41.4 percent could not be explained

by chance. The proportion of successful discriminations was highest for the $Choice_0$-Standard and lowest for the $Choice_1$-$Choice_2$ comparisons. The net percentage of successes (percentage exceeding that explained by chance) was greatest for the $Choice_0$-Standard comparison and was significantly greater than the percentage of successful discriminations in the Prime-Good and $Choice_1$-$Choice_2$ comparisons. (Table 17) The panel was able to discriminate between the steaks of two non-adjacent grades in the $Choice_0$-Standard comparison in about one-half of the cases. In the identical grade comparison ($Choice_1$-$Choice_2$) the panel found differences in about 37 percent of the taste trios.

TABLE 17--PERCENTAGE OF SUCCESSFUL DISCRIMINATION ACCORDING TO GRADE COMPARISONS

Comparison	Total Number of Trials	Actual Percentage Success	Net Percentage Success[a]
$Choice_0$ - Standard	218	65.6	48.9
Prime - Good	202	59.4	39.6
$Choice_1$ - $Choice_2$	215	57.7	36.1

[a]The percentage of successful discrimination which is not explained by chance. Using Formula C = 3/2 (O-E), where C = percent correct above chance, O = observed percent correct, and E = expected percent correct by chance.[28]

The level of discrimination varied among loin pairs. This variation was due to the extent of differences in sensory characteristics among the samples that made it either possible or impossible for panel members to find differences. Apparently, loins paired in the comparison were, in some instances, quite similar in sensory characteristics, even though the samples of the loins were from non-adjacent grades. The number of significant discriminations in the $Choice_0$-Standard comparisons was nine out of a possible 14 (Table 18). Ideally, in an effective grading system, discrimination between all the 14 pairs of loins would have been expected. Similarly, all instead of six significant discriminations in the Prime-Good comparison would have been made. In the identical or within-grade comparison ($Choice_1$-$Choice_2$) no discriminations would be expected if the grading system was ideal. But panel members discriminated between six of the 14 pairs of loins. In both the non-adjacent grade and identical grade comparisons the number of successful discriminations by this particular panel departed considerably from the chance level of successful discriminations.

There does appear to be a relationship between degree of laboratory discrimination and differences in consumer acceptability ratings. The following points stand out in Table 21:

(1) 21 pairs of carcasses had mean acceptability differences of less than 1.0. Only eight of these were discriminated between significantly in the laboratory;

(2) 14 pairs had acceptability differences of 1.0 to 2.0, inclusive. Eight of these were discriminated between significantly;

(3) 5 pairs had acceptability differences exceeding 2.0 and all 5 were dis-

TABLE 18--SUCCESSES BY LOINS ACCORDING TO GRADE COMPARISONS

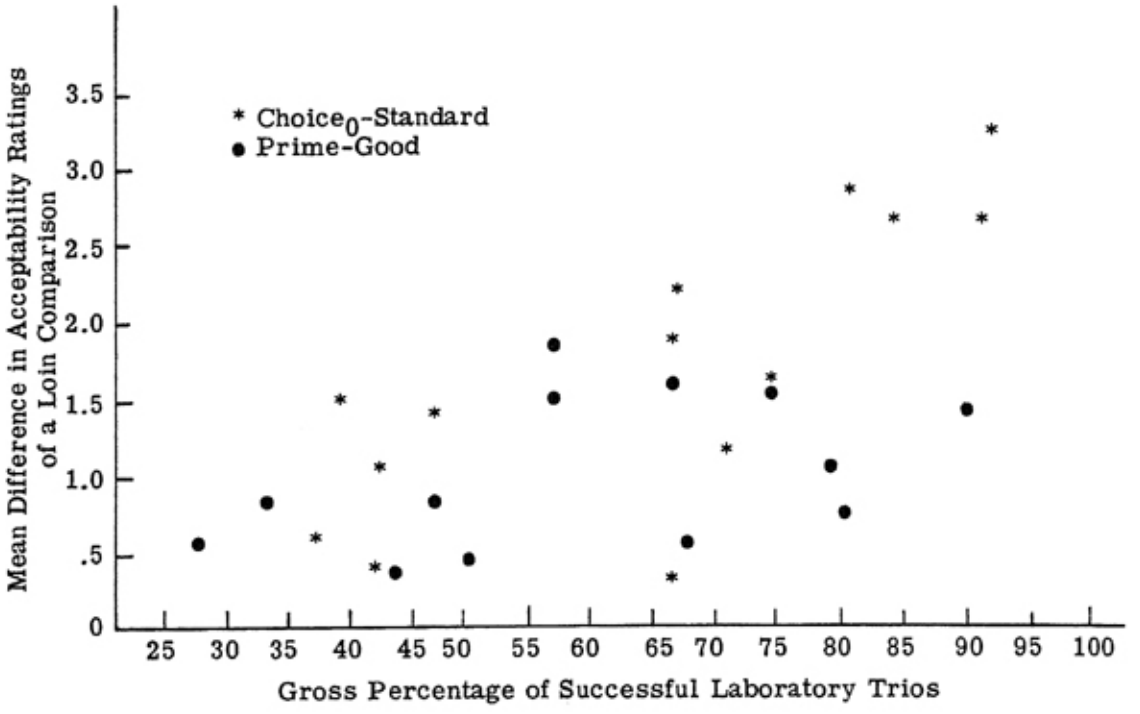
Loin Pair	Choice0-Standard No. of Attempts	No. of Successes	Difference in Ratings[a]	Prime-Good No. of Attempts	No. of Successes	Difference in Ratings[a]	Choice1-Choice2 No. of Attempts	No. of Successes	Difference in Ratings[a]
A	16	11**	.4	16	9	1.5**	15	13***	.1
B	16	7	-1.0*	16	13***	.7*	16	7	.9
C	16	12***	1.6*	16	8	- .3	15	9*	.1
D	16	15***	3.2**	16	5	.8*	16	10*	1.4**
E	16	7	1.4**	16	9	1.8**	16	11**	.5
F	16	6	- .4	16	11**	.5	16	6	.0
G	16	15***	2.6**	16	7	.4	15	8	.7
H	15	7	1.4*	15	12***	1.0*	15	6	.0
I	15	11**	1.1*	15	7	.8*	16	9	.2
J	15	6	1.5*	b/			15	7	.6
K	16	13***	2.8**	15	11**	1.5*	15	12***	.8
L	15	13***	2.6**	15	4	.6*	15	8	1.9*
M	15	10**	2.2**	15	10**	1.6**	15	10**	.8*
N	15	10**	1.9**	15	14***	1.4**	15	8	.5
Totals	218	143		202	120		215	124	

[a]A minus sign indicates that the rating of the leaner grade was better than the rating of the fatter grade.

[b]Sample unsuited for tasting.

*.05 Significance level.
**.01 Significance level.
***.001 Significance level.

Fig. 11—Relationship of Acceptability Differences to Laboratory Discrimination of 27 Loin Pairs.



criminated between significantly.

What kind of a prediction may be made about differences in consumer acceptance from laboratory discrimination results? These experiments suggest that only very cautious inferences can be made (Figure 11). Ten pairs in this laboratory experiment were discriminated between at the .001 level, 11 at the .01 or .05 level, and 20 were not significant. The following relationships were found:

(1) .001 significance of trio discrimination; acceptability differences: range of 0.1 to 3.2, median 1.5, mean 1.68.

(2) .01 or .05 significance of trio discrimination; acceptability differences: range of 0.1 to 2.2, median 1.1; mean 1.09.

(3) Non-significant trio discrimination; acceptability differences: range of 0.0 to 1.9, median .75, mean .84.

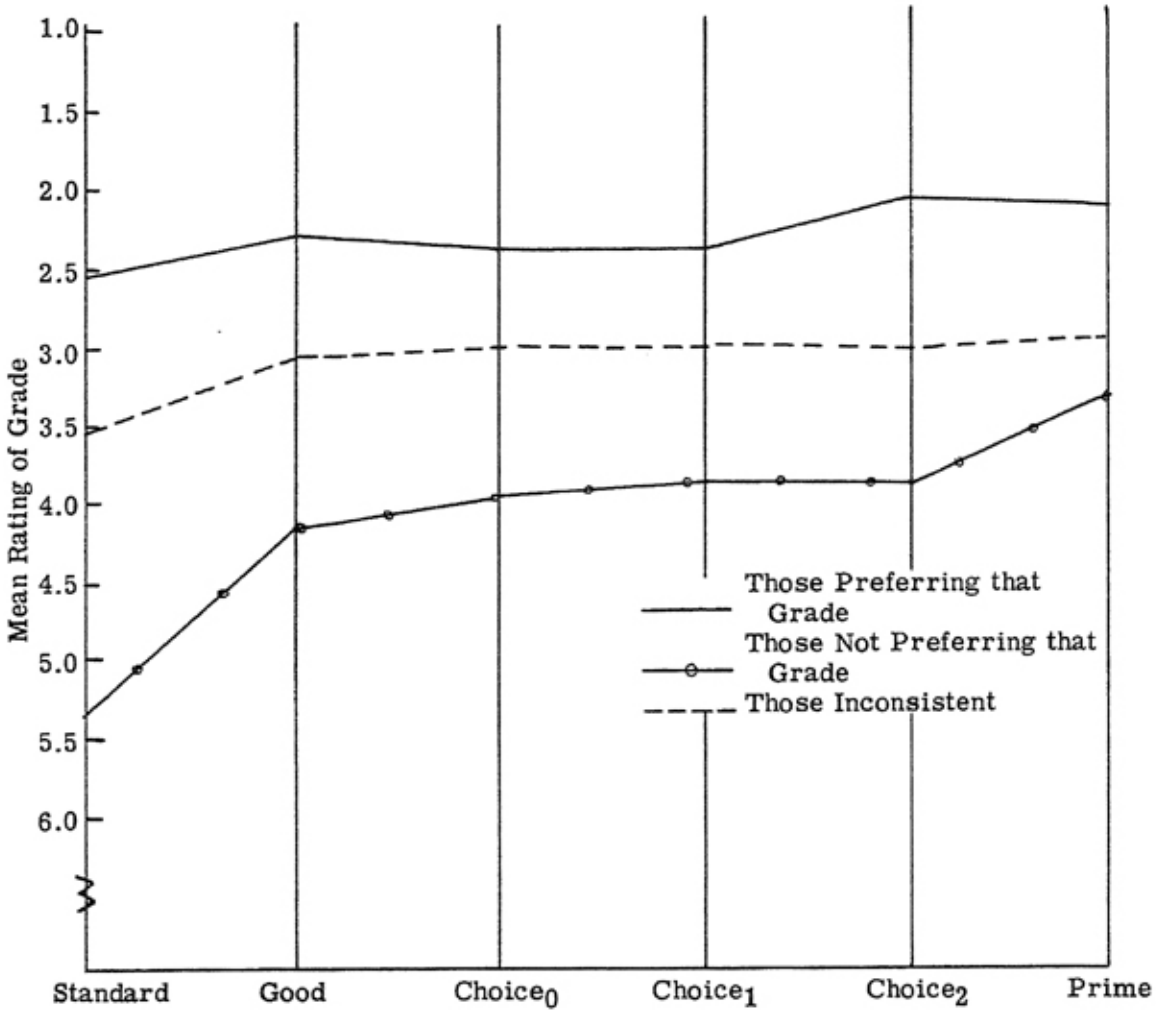
It is possible that some of the small acceptability differences associated with significant laboratory discrimination occurred because of conflicting preferences of consumers. The acceptability means would thus be close together because of differing likes and not because of consumer inability to detect a difference. However, there is evidence that this conflict in preferences rarely occurred. There was some consistent disagreement in preference for pair A of the $Choice_0$-Standard comparison and pair A of the $Choice_1$-$Choice_2$ comparison. Both pairs were discriminated between significantly but had very small mean differences in acceptability (Table 19).

TABLE 19--NUMBER OF MEN PREFERRING, NOT PREFERRING AND INCONSISTENT COMPARED TO AN ARRAY OF CARCASS MEAN RATINGS OF A GRADE IN THE COMPARISON

Standard		Number			Difference in Pair Mean Ratings	Good		Number			Difference in Pair Mean Ratings	Choice[1]		Number			Difference in Pair Mean Ratings
Carcass	Mean Rating	Preferring	Not Preferring	Inconsistent		Carcass	Mean Rating	Preferring	Not Preferring	Inconsistent		Carcass	Mean Rating	Preferring	Not Preferring	Inconsistent	
B	2.65	5*	0	6	-1.0***	T	2.59	3	3	7	.5	L	2.27	10**	1	0	1.9
F	3.08	4	1	6	- .4	C	2.64	2	0	6	- .3	J	2.52	4	0	6	.6
V	3.11	0	5	9	.6	R	3.04	3	4	5	.1	M	2.57	6*	0	6	.8
R	3.48	3	2	7	- .4	G	3.17	1	4	7	.4	S	2.75	3	0	7	.1
A	3.71	2	4	6	.4	P	3.33	1	5*	6	.3	R	2.81	1	4	7	.2
C	4.00	0	7**	3	1.6	H	3.35	0	6*	5	1.0	G	2.86	3	2	7	.7
I	4.12	0	5*	6	1.1	B	3.38	1	5*	6	.7	A	2.88	3	3	6	.1
H	4.14	1	6*	3	1.4	L	3.38	1	4	5	.6	H	2.91	2	2	7	.0
J	4.39	1	5*	4	1.5	F	3.54	0	4	7	.5	F	2.92	3	0	8	.0
P	4.64	0	8**	5	1.7	A	3.58	0	7**	4	1.5	N	3.00	4	2	5	.5
E	4.65	1	7*	4	1.4	V	3.60	0	7*	6	1.0	W	3.11	2	4	6	.0
M	4.65	0	7**	3	2.2	N	3.61	1	6*	5	1.4	C	3.14	2	1	6	.1
S	4.89	0	11**	3	2.0	I	3.63	1	7*	4	.8	D	3.17	6*	0	5	1.4
W	4.96	0	8**	4	2.2	D	3.67	0	5	7	.8	E	3.17	4	1	5	.5
N	4.96	0	6*	6	1.9	W	3.93	1	8**	5	1.1	V	3.35	1	5	7	.6
G	5.04	0	10**	1	2.6	J	3.96	1	4	6	.8	K	3.38	1	5*	5	.8
U	5.33	0	10**	2	3.2	K	4.04	1	6*	5	1.5	P	3.39	2	5	6	.8
K	5.52	1	8**	3	2.8	S	4.08	0	7**	5	1.4	I	3.50	0	3	7	.2
L	5.61	0	7**	4	2.6	U	4.22	0	8**	4	1.3	U	3.54	2	1	10	.2
D	5.79	0	9**	2	3.2	E	4.77	0	8**	2	1.8	B	3.71	1	3	7	.9
T	6.74	0	12**	1	4.3	M	4.79	0	6*	6	1.6	T	3.74	0	10**	4	1.2

*.05 level.
**.01 level.
***Minus sign indicates the rating of the leaner loin was superior.

Fig. 12—Mean Acceptability Ratings of Grades According to Nature of Preferences of Male Respondents.

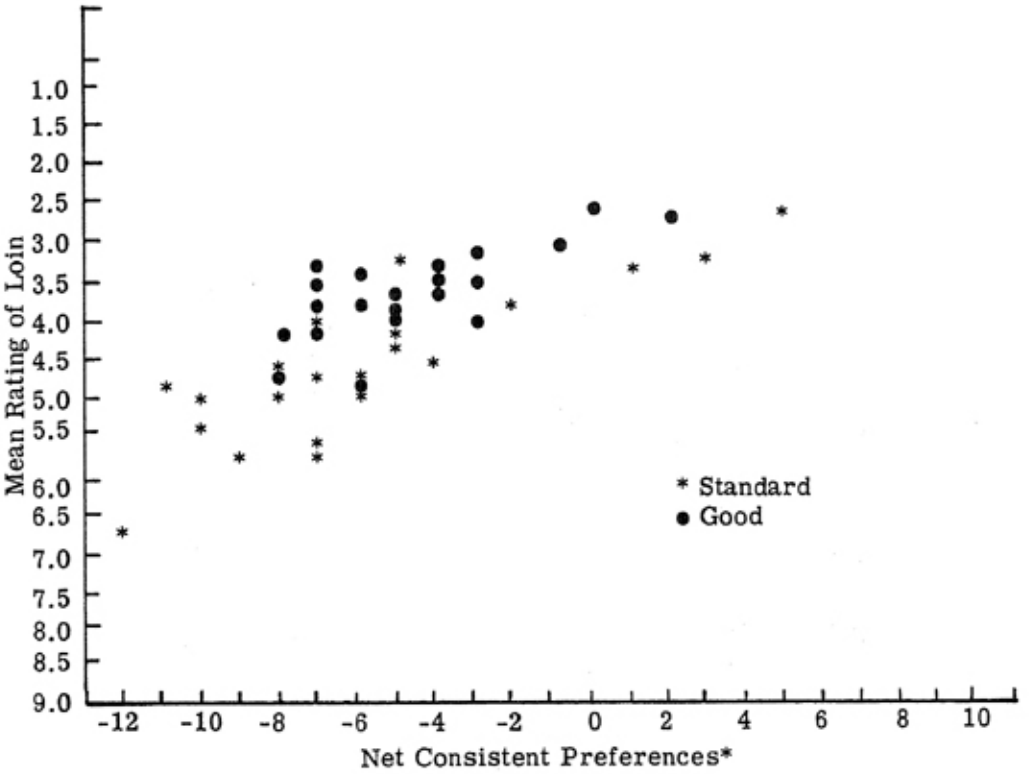


It had been hoped that discrimination tests could be used to screen out products for consumer testing which would have distinguishable differences.[29] While of some aid in predicting acceptability differences, they were not as accurate as desired.

Preferences: After rating each pair of steaks, consumers gave a preference for one over the other. The complete analysis of these preferences and their relation to discrimination results are presented in a forthcoming bulletin. Of interest here is the relationship between preferences and acceptability ratings. Some relationship is virtually necessary because of the logical relation of acceptability and preferences—that is, when a preference is expressed for A over B, then the rating of A must be as good or better than B, or else there is a contradiction in meaning. This contradiction did occur on about 1 percent of the schedules, causing their omission from the analysis.

The size of the difference in mean acceptability scores given by those preferring a product and by those not preferring that product is of considerable inter-

Fig. 13—Relationship of the Acceptability Rating of a Loin to Its Net Consistent Preferences.



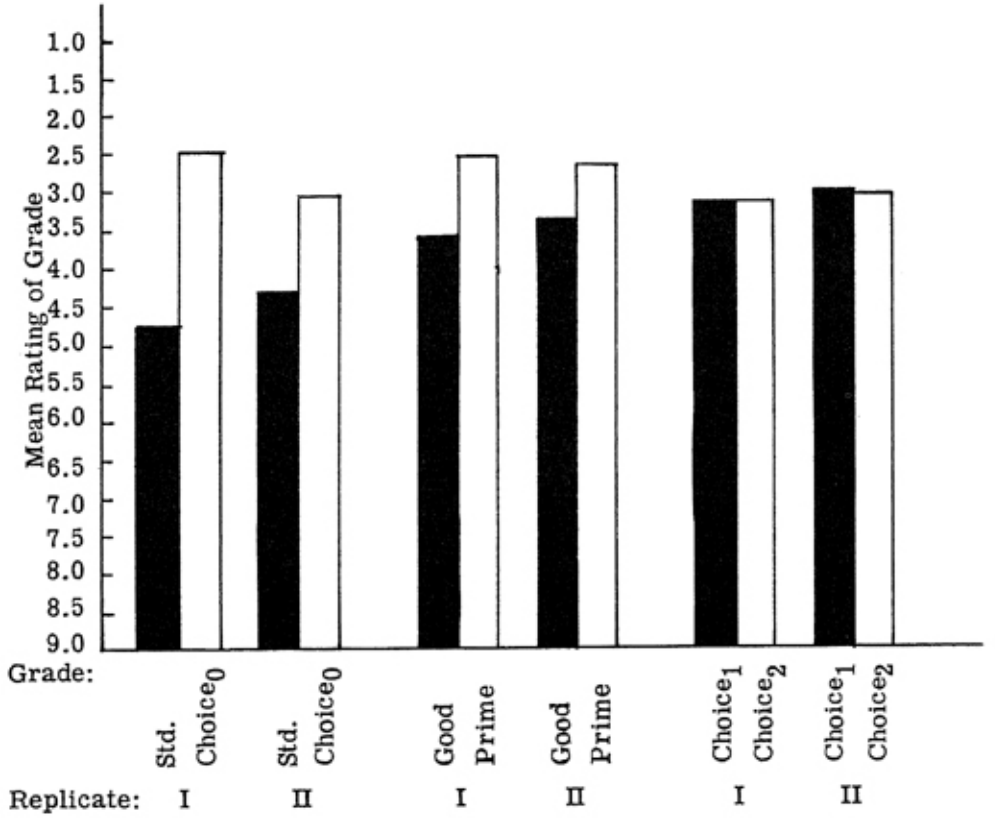
*Net Consistent Preferences equal the total Consistent Preferences for a Loin Minus the Total Consistent Preferences for the Other Loin in the Pair.

est. The column of "those preferring" contains the mean ratings of those people who preferred that grade both times. As indicated in Figure 12, this mean difference varied from 1.2 for Prime to 2.8 for Standard. Preferences were expressed at two different times for each comparison. Many consumers who said they preferred, say, $Choice_0$ the first time they rated it, switched their preference to Standard the second time. The mean ratings of those who switched preferences are in the column labeled "those inconsistent." There was a positive relationship between net consistent preferences and mean ratings (Figure 13).

Further Acceptability Results.

Shifts in Acceptability Ratings Between Replicates: It appears quite probable that consumer acceptability ratings will vary on the same product from one test to the next.[30] This variation may be caused by changes in consumer attitudes—perhaps as a result of eating experiences—or by changes in consumer interpretation of the rating scale, or by changes in the environment of consumption, or by random factors. Generally, the greater the reproducibility of ratings over time, the greater the confidence of researchers in the technique. However, if it is as-

Fig. 14—Mean Acceptability Ratings of Grades by Replicate, Men and Women Combined.



sumed that consumer attitudes may change with consumption experience, then a lack of reproducibility does not necessarily raise questions concerning the accuracy of the measuring technique.

This assumption that consumer attitudes can change with experience raises some difficult questions in methodology concerning the meaning of preference and acceptance data. Resources permitted the use of only two replicates in this experiment, so only meager evidence is available concerning these questions. Further work on the problem is underway at this Station.

Mean acceptability differences narrowed considerably from the first to the second replicate for both the Standard-Choice and the Prime-Good comparisons (Table 20 and Figure 14). This diminution of differences was associated with better ratings for the leaner grades and poorer ratings for the fatter grades on the second replicate.

In the instance of the identical grade comparison, $Choice_1$-$Choice_2$, the mean ratings adjusted to a slightly higher level for both grades from the first to the second replicate. Thus the convergence phenomenon between the ratings of the two grades did not appear here and the differences in mean ratings between the two identical grades were not significantly different. This striking similarity in

TABLE 20--MEAN RATINGS OF BEEF STEAKS BY REPLICATES ACCORDING TO GRADE COMPARISONS

	Mean Ratings								
	Men			Women			Combined		
Comparisons	Replicate 1	Replicate 2	Difference	Replicate 1	Replicate 2	Difference	Replicate 1	Replicate 2	Difference
Standard-Choice									
Standard	4.74	4.30	-.34	4.83	4.33	-.50	4.78	4.31	-.47
$Choice_0$	2.54	3.15	+.61	2.44	3.08	+.64	2.49	3.11	+.62
Difference	2.20	1.15		2.39	1.25		2.29	1.20	
Good-Prime									
Good	3.72	3.54	-.18	3.61	3.40	-.21	3.66	3.47	-.19
Prime	2.63	2.76	+.13	2.49	2.66	+.17	2.56	2.71	+.15
Difference	1.09	.78		1.12	.74		1.10	.76	
Choice-Choice									
$Choice_1$	3.24	2.94	-.30	3.06	3.04	-.02	3.15	2.99	-.16
$Choice_2$	3.19	3.05	-.14	3.12	3.04	-.08	3.15	3.04	-.11
Difference	.05	.11		.06	.00		.00	.05	

the ratings would be expected since it is presumed that the two groups of randomly sorted Choice steaks in this comparison were generally similar in eating characteristics.

The shifts in mean ratings varied considerably in direction and in degree. Table 21 shows the shift in the mean ratings by men of the 21 sets of loins for

TABLE 21--DIFFERENCES OF THE MEAN RATINGS BY MEN OF LOINS BY GRADES FROM REPLICATE 1 TO REPLICATE 2

Loin Pair	Grade						Total Difference
	Standard	$Choice_0$	Good	Prime	$Choice_1$	$Choice_2$	
A	- .75	+ .83	- .17	0	- .09	- .16	- .34
B	+ .38	-1.21	- .09	+ .50	- .58	- .17	-1.17
C	0	+ .37	- .62	+ .27	- .27	+1.34	+1.09
D	-1.75	+ .91	- .50	+ .50	-1.50	- .25	-2.59
E	+1.19	+1.85	- .09	- .45	+ .33	- .08	+2.75
F	0	- .09	+ .75	0	-1.00	-1.00	-1.34
G	+ .25	+ .42	+1.17	+1.08	+1.18	+ .36	+4.46
H	- .68	+ .69	- .49	- .15	- .38	+ .35	- .66
I	- .25	+ .84	- .09	- .25	- .16	- .08	+ .01
J	- .92	- .88	+ .78	- .60	+ .22	- .25	-1.65
K	-1.00	+ .33	+ .25	- .34	- .25	+ .42	- .59
L	- .64	+ .52	+ .25	+ .42	0	+ .54	+1.09
M	- .55	+ .91	- .25	+ .67	- .56	- .40	- .18
N	- .08	+1.17	- .47	- .16	- .26	- .22	- .02
P	- .86	+ .93	+ .20	- .44	+ .07	- .29	- .39
R	- .48	+ .66	- .56	+ .65	+ .66	- .10	+ .83
S	- .36	+1.00	0	+ .29	+ .64	- .64	+ .93
T	- .21	+1.30	- .99	- .44	-1.28	- .71	-2.33
U	- .20	+ .46	- .90	- .51	-1.84	- .23	-3.22
V	- .64	+ .14	- .52	+ .69	-1.06	- .89	-2.28
W	- .22	+ .98	- .45	+ .73	+ .07	- .57	+ .54
Mean Algebraic Difference	- .37	+ .58	- .13	+ .12	- .29	- .14	
Mean Absolute Difference	.54	.78	.46	.44	.59	.43	

*Negative difference = higher (better) hedonic rating; a positive difference = lower (poorer) hedonic rating.

all grades from the first to the second replicate. Despite the preponderant appearance of a negative shift—i.e., to higher ratings for the second replicate—the average shift per loin pair for the 126 loins amounts to 0.04. This amount was not significant. However, there was a preponderant negative shift in the case of Standard and a positive one for $Choice_0$. In general, there were both negative and positive shifts within any group of households and set of loins (designated by a letter). However, the shifts for all grades in the G set were to lower mean ratings.

The shift toward the lower mean ratings of $Choice_0$ in Table 22 was large enough to indicate a significant difference in the mean ratings of $Choice_0$ between replicates. The shift in the mean ratings of the other grades between replicates was not great enough to be significant. The loins in the second replicate were the paired right loins. The shift in ratings of $Choice_0$ should not be construed as a difference in eating characteristics between the right and left loins of the same carcass. Rather, it indicates a shift in ratings of $Choice_0$.

TABLE 22--SOURCE OF VARIATION IN MEAN RATINGS BY MEN OF REPLICATES OF LOINS BY GRADES

Source of Variation	Degrees of Freedom	Observed F Ratio by Grade					
		Standard	$Choice_0$	Good	Prime	$Choice_1$	$Choice_2$
Replicates of Loins:							
Between Replicates	1	1.31	10.94**	.44	.82	3.0	.63
Within Replicate Loin Means	40						

**Significant at the .01 level.

The first comparison made by the household panel was the first replicate of the Standard-$Choice_0$ comparison. The second replicate of this comparison was tasted in the second week of the study. It may be that the large differences in the mean rating in the first replicate resulted from inexperience with handling the rating technique and represents, in part, an adjustment in the handling of this technique on the part of the panel members.

It should be recognized that there are physical differences in the meat within the same grade. But it is not likely that the physical differences between left and right loins of the same carcass for the 21 pairs used in the test would produce the significant difference in the means.

In general, it appears that consumers did re-evaluate the leaner grades in more favorable terms on the second replicate. Whether or not this difference would have narrowed still further with additional replications is an important but, unfortunately, unanswered question.

$Choice_1$-$Choice_2$ ratings in Table 21 also indicate that mean ratings by as small a group as 12 consumers are somewhat unstable. The median absolute change in the ratings of $Choice_1$ and $Choice_2$ loins between replicates was 0.35. As indicated in Table 20 these shifts in $Choice_1$ and $Choice_2$ ratings averaged

out in the whole sample so that replicate means differed by only 0.16 and 0.11, respectively.

Variations in Acceptability Ratings Among Households: Steaks from paired sets of loins in each comparison were tasted by a particular group of households. For example, the steaks from the loin pairs lettered A of all three comparisons were tasted by households A_1 through A_{12}. These 12 households, however, were located in 12 different sampling clusters. Similarly, the paired loins B through N were each tasted by 12 households, each located in 12 different sampling clusters. Steaks from paired loins P through W were tasted by 14 households located in 14 different sampling clusters. (See Table 1) Differences in mean ratings of steaks of particular loins could arise from variations in household characteristics or situations. Inasmuch as steaks from the same lettered pairs of loins were distributed to the same households throughout the test, differences of mean ratings of particular loins could possibly arise either from variations among these household groups or from variations in loins. To isolate the possible influence of panel group variation, an analysis of variance was made of the combined mean group ratings for all comparisons.

It was reasoned that if any "group" of 12 or 14 households happened to register acceptability ratings different from other groups, this group would have a mean rating of *all* grades different from other groups. There was some variation in these group means with a range of 2.98 to 3.77 (Table 4), but it was not statistically significant. However, there is no way of proving that the 12 meat samples of each group were of exactly the same *average* acceptability. There was much variation in the ranks of a particular group's mean ratings of the various grades, which is further evidence that most of the variation in carcass ratings should be attributed to the products and not to the people doing the rating (Table 4).

Table 23 and Figure 15 indicate the variances of carcass ratings by men. While there were large intra-grade variations in variances, there was clearly more heterogeneity in the Standard grade than in the fatter grades. The simple coefficients of correlation between loin variances and loin means for the first replicate for all grades was 0.648. This relationship indicates that the poorer the average acceptability of a carcass, the more the dispersion of ratings. This may indicate that consumers really agree less about the poorer carcasses than the better ones. It may be that much of the dispersion of ratings of carcasses is caused by different interpretations of the rating scale and particularly of the "dislike" end. Consumers are probably more accustomed to thinking in terms of degrees of liking than degrees of disliking. However, the variances of mediocre carcasses may be expected to be larger than either extremely acceptable or unacceptable carcasses because there is more "space" for variation in the middle of the scale.

Variations in Ratings Among Steaks Positions: Steaks cut from a strip-loin have a slightly different appearance and shape proceeding from the anterior to

Fig. 15—Distribution of Men's Rating Variances on Single Loins by Grades.

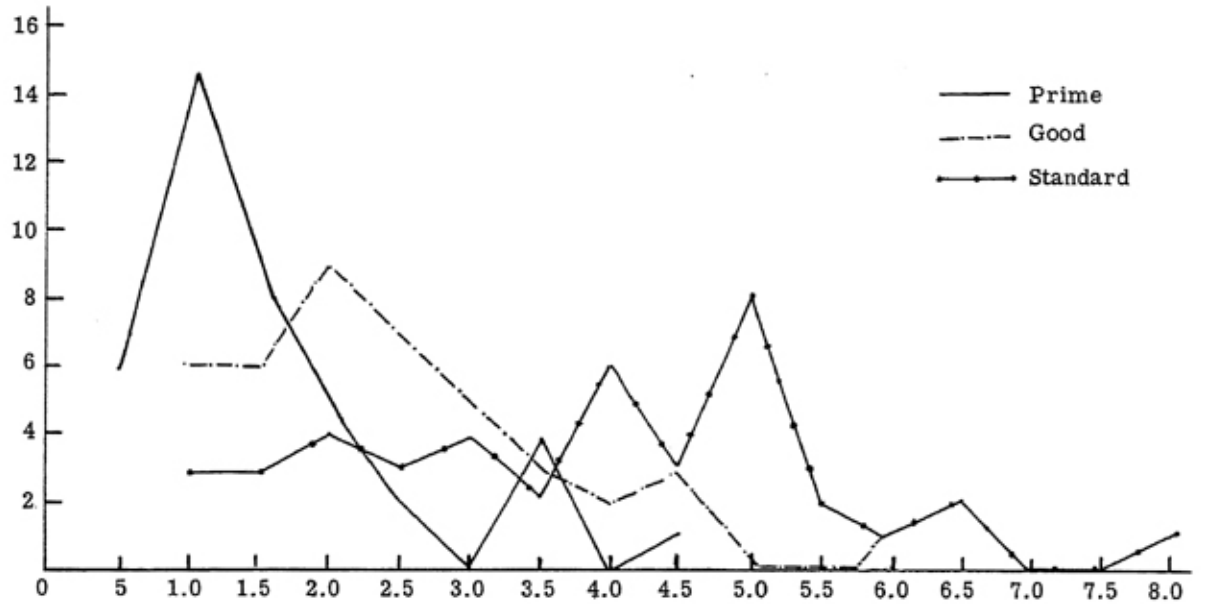


TABLE 23--VARIANCES OF CARCASS RATINGS BY MEN

Carcass	Grade					
	Standard	$Choice_0$	Good	Prime	$Choice_1$	$Choice_2$
A	2.54	1.72	1.91	.66	1.40	1.10
B	2.84	3.79	1.57	2.06	2.79	3.22
C	2.18	1.61	1.87	.77	1.46	2.86
D	5.75	2.07	1.56	1.56	2.39	4.00
E	3.18	3.91	3.18	1.41	1.81	2.63
F	1.58	2.11	3.25	2.33	2.24	2.24
G	4.37	1.58	2.67	2.50	2.77	3.98
H	5.46	1.42	3.10	1.10	2.63	2.63
I	3.03	1.75	1.74	2.00	3.58	2.29
J	4.50	3.04	3.00	2.55	1.38	2.03
K	7.39	3.84	5.04	2.33	3.32	2.42
L	6.41	2.52	2.48	1.50	.47	4.33
M	4.83	1.03	3.33	.97	1.20	2.15
N	5.21	2.49	3.11	1.19	1.35	2.34
P	3.87	3.32	2.30	1.33	2.38	1.73
R	1.95	3.28	1.48	1.59	3.14	1.06
S	3.67	2.69	2.71	1.18	1.76	3.10
T	2.78	2.69	2.24	.81	2.86	2.03
U	4.81	1.16	3.65	4.02	4.17	3.30
V	1.73	1.82	1.20	1.45	3.53	2.10
W	3.29	1.82	2.14	1.46	1.45	1.57
Unweighted Mean of Variances	3.87	2.36	2.55	1.66	2.29	2.53

the posterior part of the loin. To avoid the possibility of position affecting the eating characteristics or of appearance influencing the rating of the eating characteristics, steaks of each comparison were from the same position or location on the loin. For example, steak number 1, in terms of position, was always compared with the number 1 steak in the other loin of the comparison. Steaks were numbered from 1 to 14 from the anterior end of the loin. Analysis of the mean ratings of steaks by position on the loin showed no significant difference in the mean ratings.

Evaluation of St. Louis Consumer Panel

The Paired Steaks Technique: The techique of obtaining preferences and acceptability ratings by delivering pairs of steaks to households does work. During the enlistment of the panel, the refusal rate was rather high in the poorer income areas. Once enlisted, however, only 13 households dropped out before or immediately after the first delivery. These were immediately replaced and there were only 2 more drop-outs, which were caused by illness. A later experiment has indicated that drop-outs can be held to 2 or 3 percent for a delivery period of as long as 10 weeks. It is difficult to obtain cooperation and make deliveries to families in which all adults work during the day. More needs to be known about the extent of any bias which results from undersampling such families.

Delivery of only a pair of steaks at a time and quick pickup of the completed schedules did maintain considerable control over the panel. In general, the results indicated honest and careful cooperation by most families, although a few instances of careless schedule completion were noted.

The paired steak technique was obviously a test situation. Two marked steaks were consumed by the two adults—often with the children looking on a bit enviously. Moreover, a direct comparison was forced between the two steaks. While such a comparison was necessary to obtain preferences, it was not necessary, and was perhaps detrimental, in obtaining unbiased acceptability ratings. Bayton has argued that single stimulus testing in a situation as casual and unobtrusive as possible is a better type of testing.[31]

The mean ratings of the Choice loins which were compared together were better than the mean ratings of the $Choice_0$ loins which were compared with Standard loins. This difference indicates that the ratings of $Choice_0$ may have been slightly biased downwards. There is no proof of this, however, since there were no independent ratings of these Choice loins for comparison. Some researchers have found evidence that ratings of products are affected by the products with which they are compared.[32,33]

Another difficulty of the paired comparison technique is identification of the products. A comparison of the mean ratings of samples by men and women according to their identification by ring and no-ring indicates a significant bias toward higher mean ratings for the ring samples. The design, however, was such that the steaks of each of the various grades were identified by ring and no-

ring throughout the study an equal number of times. In the first replication of the Standard-Choice$_0$ comparison, the Standard steak samples were not identified with a ring while the Choice$_0$ steak samples carried the ring. In the second replicate of this comparison the Standard steaks were identified with a ring while the Choice$_0$ were not. A similar procedure was followed in the Good-Prime and Choice$_1$-Choice$_2$ comparisons. The apparent bias was largely associated with a shift in the mean ratings from the first to the second replicate in the cases of those comparisons involving the two non-adjoining grades (Standard-Choice$_0$ and Good-Prime). In the identical grade comparison, Choice$_1$-Choice$_2$, the higher mean ratings by women were for the ring sample in both replicates, and for the no-ring sample by men in both replicates. However, when the mean ratings of the ring samples and no-ring samples were combined for both replicates of the Choice$_1$-Choice$_2$ comparisons, the mean ratings were almost identical and not significantly different.

Position Bias: A difficulty which faces any consumer acceptability study is the confounding effect of thickness and size of cut. Use of a median size and thickness will satisfy many people. However, the median cut will be the wrong size for some people. This source of dissatisfaction will probably affect cooking methods and degree of doneness and will almost certainly affect general level of satisfaction.

The Sampling of Carcasses: Some problems of sampling inherent in the make-up of carcasses need to be recognized. One problem concerns the selection of the wholesale cut of the carcass from which the samples are taken. There are differences in eating characteristics among chucks, rounds and loins within the carcass. This difference is partly reflected in the market prices among those wholesale cuts within a grade. The greatest difference in market prices between grades is in the loin cut. Chuck prices in the market are not greatly differentiated between grades. In fact in most markets, Prime and Choice chuck are sold at the same prices; in some instances the fatter grade is sold for less.

Another aspect in sampling is the matter of the number of steak samples that can be obtained from the cut. Here the short loin again has advantages over other wholesale cuts. It yields from 14 to 18 fairly uniform steak samples. In the market, however, these steaks from the short loin are differentiated by the names of Porterhouse, T-bone and Club steaks. To make the samples uniform in appearance, the tip of the steak was trimmed and the tender-loin muscle was removed. Thus, the steaks had a fairly uniform appearance from the anterior to the posterior of the short loin. Some members of the consumer panel commented that they had not seen steaks like these before. This may have been an advantage as it removed possibility of confounding cut preferences for a Porterhouse or Club steak with grade preferences.

To increase the sample number of steaks from each carcass, both left and right loins were used. The number of sample steaks available from a carcass varies from 28 to 36 steaks, the practical limit from two loins. This limitation on

samples of a particular carcass imposes itself in the design of a beef preference study.

Selection of carcasses representative of a grade also introduces a number of special problems. Each grade has arbitrary boundaries along a continuum of characteristics which are subjectively determined by the graders. Within each grade it is possible to select, subjectively, carcasses falling within parts of the grade. However, the number falling within, for example, each decile within the grade would describe a positively skewed distribution within the higher grades and a negatively skewed distribution within the lower grades. In addition the boundaries between grades is a slightly nebulous area. In this study, specifications to graders were that the carcasses be selected from the middle one-third of the grade. It is possible that this procedure generally resulted in the selection of the median type of carcasses representing the grade rather than the modal type.

There is some evidence from this study and the previous laboratory experiment 20 that there may be differences in carcasses within grades over time and among general geographical sources of cattle. It appears that these differences cannot be detected in the subjective grading process. Apparently, genetic, feeding and other factors influence eating characteristics. Purchase of carcasses from packers serving different cattle-producing areas would maximize differences associated with the different types of breeding and feeding programs prevailing in the areas and would thereby approach a more representative national sample. The major feeding areas market cattle in different seasons of the year and in varying proportions by grades. Therefore, selection of a sample of carcasses at a particular point in time will reflect the particular cross-section of characteristics at that time.

Carcasses for these studies were purchased at two different seasons of the year: in April, 1955, for the large laboratory panel, and in September, 1955, for the household consumer panel and the subsequent laboratory panel in February, 1956. The difficulty of obtaining carcasses "typical" of the grade according to the specifications of this study forced procurement of carcasses from widely separated markets.

Beef carcasses were purchased from packers located in Kansas City; Omaha, Nebraska; Ottumwa, Iowa; and St. Louis. However, no systematic sampling procedure of regions nor of the universe of breeding and feeding practices as they might influence variation of eating characteristics was possible. Little is known concerning the nature of this variation; consequently, reliance was placed on the uniformity of interpretation of the federal grade standards by the graders operating in the various markets. The cost would be tremendous for a sample of carcasses according to territories and season of the year. The modest objective of these studies to discover the nature of the variation between and within grades probably was not seriously impaired by not introducing these refinements in sampling.

Market Significance of Variations in Acceptability Ratings: While it seems obvious that the popularity of beef is related directly to its acceptability, there is virtually no evidence as to the precise nature of that relationship. There are no proven answers to such questions as these: What is the level of eating acceptability below which as much as 5 percent of a typical supermarket's customers would have complaints about their loin steaks? How wide can be the range of eating acceptability in a grade without causing many consumers to consider that grade non-homogeneous?

A carefully designed experiment measuring both sales and acceptability ratings for many people over a long period might answer these questions. There is little indirect evidence from this study. Choice grade is sold exclusively by many retail stores and is generally considered fairly homogeneous and satisfactory by the trade. Mean acceptability ratings of 63 carcasses in the middle third of the Choice grade ranged from 2.15 to 4.54 or a difference of 2.39 points (Table 4). Even if this difference were halved because of a feeling that the middle third of the Choice grade was not sufficiently homogeneous, a mean difference of 1.2 or less between two carcasses may not indicate any difference in consumer acceptability.

How great should be the difference in the means of two large groups of carcasses (for example, two grades) to attach market importance to the difference? Certainly a difference which is just statistically significant is a minimum difference. It may need to be much larger to have market significance. Undue importance, unfortunately, has been placed on differences of minimum statistical significance by some researchers. The fact that the mean rating of 21 Choice loins by 266 women varied 0.64 of a point from the first replicate to the second indicates that intergrade differences of at least that magnitude probably have no market significance (Table 20). The evidence on the heterogeneity of eating acceptability of carcasses in the leaner grades suggests considerable limitation of the usefulness of even comparing the means of grades as presently defined.

If grades were to be redefined to obtain more intra-grade homogeneity in terms of acceptability ratings, various alternatives would be possible. One possibility would be:

(1) Prime, best three-fourths of Choice, best one-half of Good, and best one-sixth of Standard;
(2) remainder of Choice and Good, and next best one-half of Standard;
(3) poorest one-third of Standard.

A simpler, two-grade, alternative system would be:

(1) All Prime and Choice, best two-thirds of Good, best one-third of Standard;
(2) poorest one-third of Good, poorest two-thirds of Standard.

While shear measures would be of considerable aid in dividing the above grades into a "best" portion and a poorer portion, better grading methods are needed.

Samples are small and data are presently too inadequate to argue strongly for any redefinition of grades. However, this lack should not prevent recognition of the tremendous economic significance of the possible redefinition suggested above. It should be readily apparent which parts of any revised grade could be produced most cheaply. The production of fewer well-fed animals would be greatly encouraged. This revision of grades would affect not only type of production but also areas of production. Moreover, it would emphasize breeding and processing to produce acceptably leaner carcasses.

APPENDIX

Sampling Procedure

The first stage in the procedure was the determination of the cluster areas. Data from the 1950 Census of Population were arranged by tracts to show (1) proportion of white and negro population, (2) number of households, and (3) median household income.[34] Tracts having a population of more than 50 percent negro and/or having a median annual income of less than $2,000 were eliminated. This removed 22 tracts with 53,929 households in St. Louis and 4 tracts and 2,912 households in St. Louis County.

The remaining Census tracts of the city of St. Louis were combined into contiguous units of approximately 3,000 households each. Adjustments were made for rapid population growth since 1950 in the suburban county area. Information obtained from the Market Research Department of the *St. Louis Globe Democrat* and data on building permits were used to make an adjustment. From 1950 to 1954, 32,878 residential building permits had been issued by municipalities and 11,912 had been issued by the county in areas outside of the city proper. These were allocated to tract areas, based on street addresses and a comparison of detailed county street maps for July, 1950, and July, 1955, showing the development of new areas.[35]

The tracts in St. Louis city were combined to make up 67 clusters with an average of 3,052 household units. St. Louis County tracts, adjusted for population growth, were combined into 51½ clusters with an average of 2,924 household units. The 118½ city and county clusters averaged 2,992 households. The original cluster units were then divided into equal units of approximately 1500 households each. Numbers were assigned to the 237 cluster units and 38 of these were selected randomly as the sample cluster units.

The seven household units in each cluster were serially selected from a random point to reduce travel within each cluster below that required by a random selection of the seven in each cluster.

Five experienced women interviewers were employed to handle both the recruitment and the taste panel phases of the study. After a two-day training period, these interviewers began recruitment for the panel under the supervision of a staff member.

A detailed street map was prepared for each of the sample clusters. A street intersection was chosen randomly within each of the sample clusters (from which point interviews to recruit the household units began). Beginning at this point and circling the blocks clockwise, the interviewer interviewed the first household and every seventh household thereafter to determine its eligibility and whether or not members would cooperate. For the clusters located in the irregular county areas, specific directions were given on the maps to indicate rotation among streets from which to recruit the household panel. One call back was ordinarily made if household members were not at home. Interviews of every seventh household continued until seven panel households had been obtained in each cluster sample. Of the households contacted, about 29 percent were ineligible and about 28 percent refused to be interviewed or refused to cooperate. A total of 266 household panel units, each with two adult tasters, were thus obtained.

Laboratory Panel Procedure

After the loins were aged 7 to 9 days the tenderloin muscle and ventral, vertebral processes were removed. Steaks numbered 13, 14, and 15 were cut ¾ inch thick and stored at 0° F from October until tasted in February.

After steaks had been thawed in a 38° F cooler for 24 hours, they were cooked to medium-well-done on a thermostatically controlled electric grill. In practice, this was interpreted to mean termination of cooking as soon as all internal pink color was gone.

Steaks for the shear analysis were cooked to well done.[2] One-inch cores were taken from the medial, central and lateral positions of the longissimus dorsi muscle of the loin steaks. Shear determinations were made from each core sample while the samples were at serving temperatures.

Two panels of four members each were utilized in making the taste comparisons. Both panels consisted of male and female students with no previous tasting experience. The two panels met daily but at different times. However, they tasted parallel samples so their results could be compared and combined. The panel members tasted in separate darkened booths.

The triangle or trio test was used in these laboratory discrimination tests. Briefly it might be described in this manner. Three samples of food are tasted by each panel member or judge. Two of these samples are alike while the third presumably is different. The panel members are requested to identify the sample which is different. The degree to which the proportion of successful determinations of the odd samples in a series of repeated comparisons exceeds chance becomes a measure of discriminable sensory differences between the compared items. The necessity for rigid control of variables concerning the product or procedure is obvious. Lack of controls can and does influence the proportion of successes, as panel members are likely to use any clue to select the odd sample in the trio.

Three loin steaks were used for each comparison. Two adjoining steaks were

cut from one loin and one steak (presumably different) of the same relative position from another loin. Each panel member received bites from the same relative position on the compared steaks throughout the experiment. These locations were selected by random numbers. Therefore, each person received three samples, each from a different steak. The use of two samples from the same steak possibly could have aided discrimination by providing extraneous clues such as similarity in degree of doneness and/or in thickness of the samples.

The order of tasting the various comparisons was randomized subject to the following restrictions:

(1) The order for each panel was the same.

(2) The three grade comparisons were distributed uniformly throughout the period so that fatigue (physiological and psychological), learning, or other time factors would not affect comparability of results.

Comparison of Results of the Two Laboratory Panels

Comparison of the results of the April-May, 1955, panel (reported in Bulletin 612) with the February, 1956, panel reveals some striking differences in the percentage of successful discriminations. The former panel's net over-all discrimination percentage was 25.8 percent which for the latter was 41.4 percent. The former panel was most successful in the Good-Prime comparison (49.6 percent) while the latter was most successful in the Standard-Choice comparison (48.9 percent). Both panels found differences in the identical grade comparison (Table 24).

TABLE 24--COMPARISON OF SUCCESSFUL DISCRIMINATION OF APRIL-MAY 1955 AND FEBRUARY 1956 LABORATORY PANELS

	April-May 1955 Panel[a]		February 1956 Panel	
Comparison	Actual Percentage Success	Net Percentage Success	Actual Percentage Success	Net Percentage Success
Commercial[b]-Choice$_0$	52.0	28.5	65.6	48.9
Good-Prime	66.1	49.6	59.4	39.6
Choice$_1$-Choice$_2$	46.6	20.4	57.7	37.1

[a]Data of the inexperienced panel group. This group had the better level of performance.

[b]Standard in 1956 panel.

These results largely confirm those of the earlier laboratory panel concerning the heterogeneity within grades. However, the level of performance with respect to comparisons varied widely. This variation is further evidence of the heterogeneity of the grades. Apparently the sampling variation of beef carcasses representing the various grades was such that it produced inconsistent performance between the two panels. Seasonal factors in the production and marketing of cattle probably were the underlying causes. Therefore, a sample of the populations of beef carcasses would have to account for seasonal variation.

A clue to physical heterogeneity in the two samples of beef carcasses used

in the two panels is found in the comparison of the shear values. The former panel had the highest proportion of successful discriminations in the Good-Prime comparison where the average difference between the mean shears of these grades were greatest (5.03). They did poorly in the Commercial-Choice comparison where the average difference in mean shears was 0.5 and where the mean shear of Commercial was the smaller of the two grades (Table 25).

TABLE 25--COMPARISON OF SHEAR VALUES IN SAMPLE OF CARCASSES USED IN TWO LABORATORY PANELS

	Shear Values			
	April-May, 1955[a] Sample		February, 1956[a] Sample	
Grade	Mean	Standard Deviation	Mean	Standard Deviation
Commercial[b]	16.88	5.93	24.1	5.93
Good	17.47	4.66	20.8	4.91
Choice	17.38	4.62	17.8[c]	3.33
Prime	12.44	3.47	15.4	3.03

[a]Carcasses for the April-May, 1955, panel were purchased in late March and April, 1955; for the February, 1956, panel they were purchased in September, 1955.

[b]Standard in 1956 panel.

[c]Mean of $Choice_0$, $Choice_1$, and $Choice_2$ carcasses.

The February, 1956, panel had the highest proportion of successful discriminations in the Commercial-Choice comparison in which the average difference in mean shears of the grades compared was largest (6.3). Conversely, where the average difference of mean shears was less in the case of the Good-Prime comparison the level of performance in discrimination was less. Apparently, a large share of the differences in the performances of the two panels arose from sampling variation of the beef carcasses.

The correlation between the net percentage of trio discriminations and the mean shear difference of the carcasses compared was 0.80 ($N = 27$) for between-grade comparisons of the 1956 panel. However, this same coefficient was only 0.35 ($N = 40$) for the 1955 panel and was 0.57 ($N = 67$) for the combined panels.

Since it was thought that shear differences might possibly be somewhat relative, the shear difference was computed as a percentage of the larger shear. The coefficient of correlation of net percentage trios and percentage difference in shear for between-grade comparisons in the 1955 panel was 0.33 ($N = 40$). Thus, conversion of the shear difference to a percentage difference did not improve the correlation with trio discrimination. The coefficient of correlation of net percentage trios and shear difference for within-grade comparisons in the 1955 panel was only 0.07 ($N = 80$). While shear differences of pairs within grades were quite small—a mean difference of 3.24 compared to 6.5 between grades, 1955 panel—it is difficult to account for such a weak within-grade relationship. How-

ever, the relationship of shear within the Good and Standard grades to acceptability ratings was shown above to be sufficient to improve prediction of acceptability.

It is also possible that part of the difference in the level of performance between the panels is due to differences in judges or to some unknown and uncontrolled variation in panel techniques.

REFERENCES

1. E. G. Nourse, "The Proper Sphere of Governmental Regulation in Connection with the Marketing of Farm Products," *American Economic Review,* March, 1923, XIII, Page 200.
2. L. S. Tenny, "Standardization of Farm Products," *The Annals of the American Academy of Political and Social Science,* May, 1928, CXXXVII, Pages 205, 212.
3. Kelsey Gardner and Lawrence Adams, *Consumer Habits and Preferences in the Purchases and Consumption of Meat,* USDA Dept. Bulletin No. 1443, November, 1926.
4. R. C. Ashby, *et al., Retailer and Consumer Reaction to Graded and Branded Beef,* Illinois Bulletin 497, August, 1941.
5. Margaret I. Liston, *A Study of Quality Demands in Household Buying of Food,* Vt. Bulletin 415, 1937.
6. T. O. Meyer and M. E. Ensminger, "Consumer Preference and Knowledge of Quality in Retail Beef Cuts," Washington State Stations Circular No. 168.
7. George W. Campbell, *Consumer Acceptance of Beef,* Arizona Report No. 145, December, 1956.
8. Raymond Seltzer, *Consumer Preferences for Beef in Phoenix, Arizona,* Arizona Station Bulletin No. 267, October, 1955.
9. *Beef Consumer Use and Preferences,* Colorado Station Bulletin No. 495-S, April, 1956.
10. Ira Stevens, "How Do You Like Your Beef?" *Wyoming Roundup,* October, 1955, pp. 4-5.
11. Robert E. Branson, *The Consumer Market for Beef,* Texas Agricultural Experiment Station Bulletin 856, April, 1957.
12. Fred G. Lasley, Elmer R. Kiehl, D. E. Brady, *Consumer Preference for Beef in Relation to Finish,* Mo. Res. Bulletin 580, March, 1955.
13. V. James Rhodes and Elmer R. Kiehl, "Are Consumers Getting the Beef They Want?" *Meat and Food Merchandising,* March, 1955.
14. V. James Rhodes, Elmer R. Kiehl, and D. E. Brady, *Visual Preferences for Grades of Retail Beef Cuts,* Mo. Res. Bulletin 583, June, 1955.
15. Sylvia Cover, O. D. Butter, and T. C. Cartwright, "The Relationship of Fatness in Yearling Steers to Juiciness and Tenderness of Broiled and Braised Steaks," *Journal of Animal Science,* 15:464-72, 1956.
16. O. G. Hankins and N. R. Ellis, *Proc. Am. Animal Prod.,* 1939, Page 314.
17. J. J. Wanderstock and J. I. Miller, "Quality and Palatability of Beef as Affected by Method of Feeding and Carcass Grade," *Food Research,* 13:291-303.
18. S. A. Husaini, F. E. Deatherage, and L. E. Kunkle, "Studies on Meat: II. Observations on Relation of Biochemical Factors to Changes in Tenderness," *Food Tech.,* 4:366-69.

19. V. James Rhodes, "The Measurement of Consumer Preferences," *Journal of Farm Economics*, November, 1955.
20. V. James Rhodes, Elmer R. Kiehl, Neil B. Wilson, D. E. Brady, and E. B. Birmingham, *Consumer Preferences and Beef Grades*. Mo. Research Bulletin 612, 1956.
21. "U. S. Grades for Beef," U.S.D.A. Leaflet Number 310.
22. "Grading Has Proved Its Worth But Packers Aid Can Make It Better," *National Provisioner*, March 9, 1957, Pages 66-67.
23. Sleeter Bull, *Meat for the Table* (New York: McGraw-Hill, 1951), Ch. 7.
24. L. J. Bratzler, "Determining the Tenderness of Meat by Use of the Warner-Bratzler Method," *Recip. Meat Conf. Proc.*, 2:117-128.
25. H. Hurwicz and R. G. Tischer, "Variation in Determinations of Shear Force by Means of the 'Bratzler-Warner Shear,' " *Food Tech.*, 8:391-93.
26. Guy H. Orcutt, "Measurement of Price Elasticities in International Trade," *The Review of Economics and Statistics*, May, 1950, Pages 117-132.
27. A. C. Rosander, *Elementary Principles of Statistics*, (New York: D. V. Nostrand Company, Inc., 1951), Pages 532-534.
28. N. T. Gridgeman, "Taste Comparisons: Two Samples or Three?" *Food Technology*, 9:148-150, 1955.
29. V. James Rhodes and Elmer R. Kiehl, "On Consumer Grades for Foods," *Journal of Farm Economics*, February, 1956.
30. David Peryam and Francis Pilgrim, "Hedonic Scale Method of Measuring Food Preferences," *Food Tech.*, 11:9-14, Symposium, September, 1955.
31. James Bayton, "Statistical vs. Psychological Bias in Consumer Taste Preference Research," *Middle Atlantic Conference Transactions*, American Society for Quality Control, Washington, D. C., 1956, Pages 23-27.
32. E. E. Lockhart and Jean M. Gaines, "Acceptance of Fresh, Frozen, Canned Orange Juice," *Food Research*, 14:429-33, 1949.
33. H. L. Hanson, *et. al.*, "Sensory Test Methods, II. Effect of Previous Tests on Consumer Response to Foods," *Food Technology*, 9:56-59, 1955.
34. United States Census of Population 1950, *Census Tract Statistics St. Louis, Missouri and Adjacent Area, 1950. Population Census Report*, Vol. III, Chapter 47.
35. Henry E. Gross, Street Maps St. Louis and St. Louis County. 406 Market, St. Louis, 1950 and 1955.