

UNIVERSITY OF MISSOURI COLLEGE OF AGRICULTURE  
AGRICULTURAL EXPERIMENT STATION

ELMER R. KIEHL, *Director*

# Variations in Consumer Acceptability of Loins From Market Lots of Cattle

V. JAMES RHODES, H. D. NAUMANN, WILLIAM C. STRINGER, CHARLES L. CRAMER



(Publication authorized February 11, 1964)

COLUMBIA, MISSOURI

---

## SUMMARY

The consumer acceptability of each of 26 market lots of cattle was estimated from a carefully selected sample of four loins from each lot. The 104 sample carcasses included 41 Good, 57 Choice, and 6 Prime. The mean acceptability of these lots varied from 8.8 to 7.6 on a nine point scale in which nine was the best possible score. However, 15 of the 26 lots had acceptance means in the very narrow range of 8.0 to 8.2. The most frequent of the various criticisms was "not tender enough."

The relationships of carcass grade and of each of its components—maturity, marbling, and conformation—to consumer acceptance were very weak.

The linear correlation between acceptance and carcass retail yield was  $-.099$ . The "r" value indicates that there was a very slight tendency for higher yielding carcasses to be less acceptable to consumers than lower yielding carcasses.

## CONTENTS

Introduction .....	4
Design of Study .....	4
Results and Conclusions .....	7
Consumer Ratings by Lots .....	7
Variation Among Loins .....	10
Variation by Grades .....	10
Variation by Shear .....	12
Consumer Comments .....	14
Socio-Economic and Cooking Relationships .....	14
Household Performance .....	15

# Variations in Consumer Acceptability of Loins From Market Lots of Cattle

V. JAMES RHODES, H. D. NAUMANN, WILLIAM C. STRINGER, CHARLES L. CRAMER

## INTRODUCTION

In a recent study, 400 consumers ate and evaluated 5,600 loin steaks from 560 carcasses. This study provided a careful specification of the relationship of grade to consumer acceptance. In the summary of the study's implications, the authors said:

"Excessive finish is no longer regarded as an indicator of eating quality and is being discounted because it reduces trimmed retail yields. It is now apparent that marbling is not a sure indicator of eating quality. While there is certainly no implication that marbling should be avoided, it is possible that marbling could be overemphasized in certain breeding and feeding programs. What is even clearer is the need for an intensification of the breeding and meats research on heredity and other factors to determine which are basic to beef quality. It is now apparent that eating quality of beef largely depends on different and probably more complicated factors than those involved in the grading standards' definition of quality."<sup>1</sup>

The following report records the consumer acceptance found in a follow-up study, which was principally concerned with measuring variations in retail yield and the factors affecting it. However, because of the critical economic importance of determining the relationships of consumer acceptance and retail yield, the study was designed to accomplish this objective also.

## DESIGN OF STUDY

From commercial lots of cattle in the pens of the Kansas City Stockyards, 26 lots of cattle were selected over a several month period. A sample of eight cattle was chosen from each lot by skilled livestock men according to the following selection goals:

- (1) one pair of high retail yield,
- (2) one pair of low retail yield,
- (3) four cattle of average yield (with high, low and average referring to that particular lot).

<sup>1</sup>H. D. Naumann, *et al.*, *Consumer and Laboratory Panel Evaluation of Good and Choice Beef Loins*, Mo. Res. Bul. 777 July 1961, pp. 4-5.

A further selection was then made in the packer cooler from these eight carcasses to obtain the best, the poorest and two average yielding carcasses. Loins from these four carcasses were used for consumer acceptance testing.

This sampling method was obviously picked for its operational feasibility rather than its sampling efficiency. The sampling method should be kept in mind in evaluating the results.

With two exceptions the lots were steers; they ranged in number per lot from 16 to 124. Lots were selected to represent the supermarket type in both weight and quality. The 104 sample carcasses tested included 41 Good, 57 Choice and 6 Prime; 74 were in the typical supermarket range of Good Plus to Middle Choice.

These lots varied significantly in dressing percent, ribeye area, fat thickness and retail yield. Other relevant measures were:

Variables	Mean	Standard Deviation
Live weight	1061.2 lbs.	127.88 lbs.
Dressing percent	63.0 <sup>a/</sup>	2.20
Carcass grade	5.0 <sup>b/</sup>	1.63
Conformation grade	4.03 <sup>b/</sup>	1.59
Percent kidney knob	2.8	0.90

<sup>a/</sup> Calculated from lot weights  
<sup>b/</sup> See grade code on page 11

In some cases, detailed feeding and breeding records were obtained from the cattle feeders. Details of the experimental procedure relating to the estimation and measurement of retail yield are recorded in a companion report.

Twelve steaks from each of these 104 loins were eaten and evaluated by consumers selected by probability sampling in Jefferson City, Mo. Methods of testing were generally similar to many previous tests. Steaks were cut three-fourths inch thick, individually wrapped and coded as to lot, loin, and steak position, and frozen.

The steaks were eaten and evaluated by husbands and wives in 144 households. The 12 steaks from each loin were eaten by 12 individuals (6 men and 6 women) in 12 households. The steaks from six entire lots (i.e. 24 loins) were eaten each week. Husbands and wives were informed that they were not eating steaks from the same animal and that, therefore, the steaks were not necessarily similar. Each person evaluated his steak on the schedule shown in Figure 1.

The particular rating scale had been used previously for eggs and pork at this Station but not for beef. The scale is so structured that the ratings of a well-liked product like beef steaks are greatly skewed toward the top rating of "9." While this structuring has a certain communications convenience as contrasted with the Quartermaster scale used previously in our work, the skewed distribution may possibly reduce somewhat the statistical reliability of the analysis of variance technique.

JEFFERSON CITY BEEF STEAK PANEL  
UNIVERSITY OF MISSOURI  
1962

The steak without a ring is for the wife. Please eat the whole steak and then fill out this schedule immediately.

DIRECTIONS FOR USING RATING SCALE: Please rate this steak on the basis of eating characteristics. Please ignore size and thickness in making your ratings.

1. Number 9 is the top rating. A steak given this rating should have everything that you expect from a steak.
2. Ratings 8-1 indicate decreasing levels of acceptance.
3. "0" indicates that the steak was completely unacceptable.

	Circle One Number	<u>If you rated the steak less than 9, please check reasons why:</u>
Entirely satisfactory, this is the kind of steak I like to eat.-----	<input type="text" value="9"/>	<input type="checkbox"/> Not tender enough
Steaks rated in this range are acceptable but they are not entirely satisfactory--the poorer the steak the lower the rating.	<input type="text" value="8"/>	<input type="checkbox"/> Not juicy enough
	<input type="text" value="7"/>	<input type="checkbox"/> Lacked flavor
	<input type="text" value="6"/>	<input type="checkbox"/> Poor flavor
	<input type="text" value="5"/>	<input type="checkbox"/> Too fat
	<input type="text" value="4"/>	<input type="checkbox"/> Too lean
	<input type="text" value="3"/>	<input type="checkbox"/> Accidentally cooked overdone
	<input type="text" value="2"/>	<input type="checkbox"/> Accidentally cooked too rare
Not acceptable.-----	<input type="text" value="0"/>	<input type="checkbox"/> Appearance
		<input type="checkbox"/> Other

4. How cooked? (Please check one)

- Moist heat (liquid added or lid on)  
 Dry Heat (no liquid and no lid)  
 Charcoal broiled

5. Doneness?  Well (no pink meat)  Rare (some pink meat)
6. Any flavoring added (catsup, steak sauce, barbecue sauce, etc.)?  
 Yes;  No. If yes, what? \_\_\_\_\_
7. Comments. (Both favorable and unfavorable comments are useful to us and are greatly appreciated.) \_\_\_\_\_

8. Name \_\_\_\_\_
9. Address \_\_\_\_\_
10. Date Eaten \_\_\_\_\_
11. HH No. \_\_\_\_\_

Fig. 1—Photograph of the schedule householders used in evaluating steaks.

Another innovation in our usual technique was to analyze the consumer scores in an "adjusted" form as well as an unadjusted one. The "adjusted" mean scores were computed from the middle 50 percent of the 12 scores of each loin, i.e. three high and three low scores were dropped. The rationale for this experiment runs as follows: If some consumers are either uncritical or supercritical on a particular test, then that loin receives a score not based entirely on its merits. If we exclude the extreme scores, we may obtain "better" measures of loin acceptability. As already suggested, most loins were rated quite favorably and consumer agreement was relatively high, so that the effects of the "adjustment" on loin means were relatively minor. However, the adjustment effects on the analysis of variance were considerable.

## RESULTS AND CONCLUSIONS

### Consumer Ratings by Lots

The majority of lots had very similar mean consumer ratings. The median of the unadjusted lot means was 8.1 with 15 of the 26 lot ratings in the range 8.0 to 8.2. These scores reflect a rather high degree of satisfaction. Almost exactly one-half of all ratings were the "9" rating, the highest possible on the scale and labeled "entirely satisfactory, this is the kind of steak I like to eat". The majority of the lots each received about 50% rating of "9."

Not all the lots of beef had the same level of acceptability, even though a majority did. Unadjusted means of lots varied from 8.8 for lot no. 12 to 7.6 for lots 6 and 21 while adjusted means varied from 8.9 to 7.8 (Table 1). Whether these lot differences are judged to be significantly different or not depends upon the choice of error term. Sources of variation were: between lots, between loins, and error. The F ratio of "between lots" to "error" was very significant for both the adjusted and unadjusted scores. However, a more conservative F test is the ratio of "between lots" to "between loins." This test was not quite significant for either the adjusted or unadjusted scores. On the basis of this and other evidence later presented, it seems reasonable to conclude a likely difference between the two lots at the one tail of the distribution and the two lots at the other tail, but no other differences between lots.

The animals in the two most acceptable lots were larger, more finished, "higher" grade, lower retail yield cattle than the two least acceptable lots (Table 2). A negative "r" of  $-.099$  was found when the acceptance means of all 104 loins were correlated with retail yields of the carcasses. This sort of negative but weak relationship is similar to and consistent with the relationship usually found between grade and acceptance within the normal range of supermarket beef.

Another method of comparing lots is in terms of the number of consumer criticisms received. Consumers were asked to check reasons why whenever they

TABLE 1 - VARIOUS CONSUMER RATINGS OF KANSAS CITY LOTS\*

Lot No.	Unadjusted Cons $\bar{x}$	Adjusted Cons $\bar{x}$	No. of 9's Per Loin Above (+) or Below (-) Mean No.	No. of Criticisms Per Loin Group Above or Below Mean No.	Mean Shear Values
12	8.8	8.9	14.2	-22.9	16.64
15	8.6	8.7	8.2	-17.9	18.42
16	8.4	8.8	8.2	-13.9	14.64
2	8.4	8.7	3.2	- 4.9	15.63
7	8.4	8.6	5.2	- 4.9	13.16
11	8.4	8.5	1.2	- 1.9	14.35
13	8.2	8.7	4.2	- 5.9	12.11
1	8.2	8.5	1.2	- 6.9	15.65
17	8.2	8.5	5.2	- 8.9	16.34
18	8.2	8.5	0.2	0.1	14.65
8	8.2	8.4	-1.8	- 2.9	17.14
19	8.1	8.5	1.2	- 1.9	16.20
5	8.1	8.4	0.2	- 0.9	15.96
23	8.1	8.4	-3.8	3.1	15.77
9	8.1	8.3	-1.8	- 0.9	16.60
26	8.1	8.3	-2.8	4.1	18.24
4	8.0	8.5	1.2	0.1	15.76
20	8.0	8.3	-4.8	8.1	16.86
10	8.0	8.2	-3.8	3.1	14.49
14	8.0	8.2	-3.8	6.1	16.01
22	8.0	8.1	-4.8	8.1	17.71
3	7.9	8.4	-5.8	9.1	15.64
24	7.9	8.3	-1.8	7.1	16.66
25	7.9	8.1	-5.8	10.1	15.58
6	7.6	7.9	-5.8	16.1	16.41
21	7.6	7.8	-10.8	19.1	17.88
Mean	8.1	8.4	23.8	36.9	15.94

\* Ranked as to unadjusted consumer means.

rated steak less than "9." Therefore, one loin may have more criticisms (reasons why) than another for either or both of two reasons: (1) more steaks rated less than "9" and therefore criticized, and (2) more criticisms per steak of those steaks which were criticized.

The number of criticisms per lot averaged 36.9. Since about 24 steaks per lot were criticized the average number of criticisms per criticized steak was 1.5. The number of criticisms per lot varied significantly at the 0.1 percent level according to a chi square analysis. The range per lot was from only 14 criticisms of lot 12 to 56 criticisms of lot 21 (Table 1). The variation in number of criticisms was nicely associated with mean acceptance scores. It seems possible that a count of number of criticisms may be as sensitive a measure of acceptance as the regular scoring method.

There were 335 criticisms of "not tender enough" for a lot average of almost 13. Criticisms per lot varied from a low of 6 each for lot numbers 12, 15 and 16 to a high of 25 for lot 21. These frequencies may be appraised in terms of the maximum possible criticisms per lot of 48 (on any one characteristic such

TABLE 2 - DESCRIPTION OF FOUR LOTS OF CATTLE

Lot No.	Adjusted Consumer Rating	No. of Criticisms on			Live Weight Range	Carcass Grades
		Tenderness	Juiciness	Flavor		
Two Most Acceptable Lots						
12	8.9	6	1	3	940- 1115	2 Ch 1 Ch- 1 Gd+
16	8.8	6	4	8	1140- 1305	1 Pr 1 Ch 1 Ch- 1 Gd+
Two least Acceptable Lots						
6	7.9	21	13	10	735- 1075	2 Gd- 1 Gd 1 Gd+
21	7.8	25	12	13	1070- 1160	1 Gd 3 Gd+
Ribeye	Fat Over Rib Thickness	Retail Yields		Breed	Days on Feed	
(sq. in. per cwt.)	(in. per cwt.)	%				
Two Most Acceptable Lots						
1.66	.08	66.2, 71.0 71.1, 72.6		Hereford	175	
1.63	.09	62.2, 63.5 68.5, 72.3		Angus	NA.	
Two Least Acceptable Lots						
1.99	.04	73.5, 75.9 77.3, 79.6		Brahma X	180	
1.69	.07	68.6, 70.3 71.3, 74.1		Hereford	NA.	

as the lack of tenderness). Chi square analysis showed a significant difference in the number of criticisms per lot.

There were 250 criticisms of "lacked flavor" or "poor flavor" for an average of 9.6 per lot. The top acceptance lot, number 12, received the low of 4 such criticisms while lot 20 received a high of 16 criticisms. However, the differences between lots were not significant.

There were 201 criticisms of "not juicy enough" for an average of 7.7 per lot. The top acceptance lot, number 12, received the low of one such criticism while lot 6 received the high of 13 criticisms. However, the differences between lots were not significant.

There was a tendency for some lots to receive proportionately more specific criticisms on one characteristic than another. For example, lot number 8 received 5 above the average number of criticisms of tenderness, but 2.7 below the average number of criticisms of flavor and juiciness. However, a chi square analysis found no significant difference between lots in these proportions.

### Variation Among Loins

Two loins received perfect acceptance scores with means of "9." They were loin 3 in lot 12 and loin 2 in lot 15. At the other end of the range was loin 2 in lot 6 with a mean score of 6.7. It is interesting to note that the poorest loin in high acceptance lot number 12 had a mean of 8.4. Analysis of variance showed a very significant difference between loins on the basis of adjusted scores but not on the basis of unadjusted scores. This discrepancy is not unexpected since the adjustment reduced the error substantially.

Loins varied widely in the number of specific criticisms received. The number of criticisms varied on *tenderness per loin* from 0 to 9, *juiciness* from 0 to 7, *lack of flavor* from 0 to 7, *poor flavor* from 0 to 2, *too fat* from 0 to 3, *too lean* from 0 to 3, accidentally cooked *overdone* from 0 to 2, and on accidentally cooked *too rare* from 0 to 1 (only 6 such criticisms in total). The number of the 104 loins receiving *no criticism* on a specific point varied as follows:

Not Tender Enough	7 loins
Not Juicy Enough	19 loins
Lack Flavor	16 loins
Poor Flavor	54 loins
Too Fat	62 loins
Too Lean	60 loins

Another way to look at the situation is in terms of the number of loins having four or more specific criticisms (12 maximum possible). These numbers varied from 40 loins *not tender enough* to 11 loins *not juicy enough* and 10 loins *lack flavor*, to zero loins in the other categories.

Note that *not tender enough* clearly ranked by any measure as the most frequent criticism while *too fat* and *too lean* were fairly infrequent criticisms.

### Variation by Grades

The carcasses were graded in the plant by a federal grader. The acceptance results by grade were as follows (Table 3).

While the weak relationship between consumer acceptance and sections of the Good and Choice grade has been well established previously, the relationship in this study is unusually weak.

TABLE 3 - ACCEPTANCE BY GRADE

Grade	Grade Code	Number	Mean Score	Range
Gd.-	(8)	8	8.1	6.7-8.7
Av. Gd.	(7)	8	7.9	7.4-8.4
Gd.+	(6)	25	7.9	7.1-8.8
Ch.-	(5)	25	8.2	7.4-8.8
Av. Ch.	(4)	24	7.9	7.6-9.0
Ch+	(3)	8	8.0	7.6-8.3
Pr. & Pr.-	(1&2)	6	8.5	8.2-9.0

The degree of maturity of each carcass was indicated by a federal grader. The acceptance results were:

Degrees Maturity	(Code)	n	Mean Score	Range	Limits of Top Quartile
A A+	(2 & 3)	47	8.11	6.7-9.0	8.4-9.0
B- B	(4 & 5)	57	7.98	7.2-9.0	8.5-9.0

We have no previous results with which to compare. The use of maturity as a palatability measure in the federal grades is a relatively recent innovation. While the innovation probably has some validity when comparing baby beef with old, lean steers, it probably has little usefulness within the general range of block beef. In this test of 104 animals varying considerably wider than the usual weight and grade range of supermarket beef, only 4 of the 12 possible degrees of maturity were found, and the acceptability relationship to maturity appeared negligible.

The degree of marbling of each carcass was indicated by a federal grader. The acceptance results were:

Degrees Marbling	(Code)	n	Mean Score	Range	Limits of Top Quartile
Abundant and Moderately Abundant	(3&4)	9	8.43	8.0-9.0	8.6-9.0
Moderate and Modest	(6&7)	50	8.18	7.2-9.0	8.5-9.0
Small Amount and Slight Amount	(8&9)	43	8.00	6.7-8.8	8.4-8.8

These results are consistent with previous results of our work. The influence of a large amount of marbling was to raise considerably the lower limit of the

range and to improve the mean somewhat. However, there was almost complete overlapping of the top quartiles of all three marbling classifications. In this case, marbling appears to have been a sufficient condition for high acceptance but not a necessary condition.

The degree of conformation of each carcass was indicated by a federal grader. The acceptance results were:

Degree Conformation	(Code)	n	Mean Score	Range	Limits of Top Quartile
Pr. and Pr-	(1&2)	16	8.31	7.5-9.0	8.6-9.0
Ch.+ and Ch	(3&4)	48	8.06	7.1-9.0	8.5-9.0
Ch.- and G+	(5&6)	35	8.16	7.2-8.8	8.5-8.8

Conformation, in itself, is not generally claimed to be related to eating quality. The new grading proposals of the USDA, which include conformation as a "quality" factor explicitly disclaim any implication "that variations in conformation are either directly or indirectly related to palatability." (*National Provisioner*, September 21, 1963). In this group of 104 cattle, conformation is not quite as related to consumer acceptance as is marbling but is more related than maturity. The relationship may well be an indirect one—for all three: conformation, marbling, and maturity.

*Note:* the distributions of carcasses as to conformation, marbling, and maturity were:

Conformation	Marbling	Maturity
9 Pr.	2 Abundant	7 A
7 Pr.-	7 Moderately Abundant	40 A+
22 Ch+	1 Slightly Abundant	48 B-
26 Ch	23 Moderate	9 B
24 Ch-	27 Modest	
11 G+	31 Small	
3 G	12 Slight	
1 G-	1 Traces	
1 Std+		
104	104	104

### Variation by Shear

In Table 4, the loins are classified by shear force and grade. Those loins shearing under 14 pounds did have a higher consumer mean than the other groups, even though the distribution of grades was biased upwards only slightly in that group. Moreover, the poorest acceptance loin in the low shear group was only 7.9, suggesting that low shear within the grade range of this experiment may be a sufficient condition for satisfactory consumer acceptance, although it is clearly not a necessary condition. Note also that those Good grade loins with

TABLE 4 - CONSUMER MEAN RATINGS\* OF LOIN BY SHEAR

Less than 14	14.0 - 15.9	16.0 - 16.9	18 and Over
(Pr. & Pr.-) 8.1 8.4	(Pr. & Pr.-) 8.2 8.5	(Pr. & Pr.-) ---	(Pr. & Pr.-) 8.3 9.0
(Choice+) 8.1	(Choice+) 7.5 7.8 8.1 8.3	(Choice+) 7.5 8.2	(Choice+) 8.0
(Choice) 7.9 8.0 8.2 8.4 8.5 8.9 8.8	(Choice) 7.5 7.7 7.9 8.0 8.2 8.8 9.0	(Choice) 7.6 7.7 7.8 8.0 8.0 8.5 8.8	(Choice) 8.0 8.1 8.5
(Choice-) 8.1 8.4 8.5 8.5 8.5 8.7	(Choice-) 7.7 7.8 7.8 8.0 8.3 8.5 8.5 8.8	(Choice-) 7.5 7.8 8.0 8.1 8.2 8.2 8.4 8.7	(Choice-) 7.4 8.4 8.5
(Good+) 8.0 8.4	(Good+) 7.1 7.1 7.4 7.6 7.7 7.8 7.8 7.9 8.0 8.0 8.1 8.4 8.6 8.7	(Good+) 7.7 8.1 8.4	(Good+) 7.0 7.5 7.7 7.7 8.0 8.2
(Good) ---	(Good) 8.2	(Good) 7.5 7.6 8.1 8.4	(Good) 7.5 7.9 8.0
(Good-) 8.1 8.5 8.6	(Good-) 8.5	(Good-) 6.6 7.7 8.0 8.5	(Good-) ---
Mean 8.39	8.09	8.04	8.01
Range 7.9-8.9	7.1-9.0	6.6-8.8	7.0-9.0

\*Unadjusted scores

the highest shears (18 pounds or more) tended to have acceptance means below 8.0.

### Consumer Comments

Many consumers wrote comments on their schedules. Whenever these comments involved a criticism that had not been checked on the schedule, the criticism was checked in the editing process. Analysis of comments is difficult because of their unsystematic nature, and analysis was less needed here because of the system of obtaining criticisms on the schedule. However, some of the comments are included below. While they represent a range of opinions they are in no sense a careful sample of the comments.

*Good flavor and appearance. Not tender enough.*

*Very good. When preparing it appeared bloody, but maybe that's the way it should be.*

*Somewhat tough. Good flavor.*

*The fat seemed tough; and there was some gristle or very tough part to the fat. Not tender enough.*

*This steak in greater thickness would be an excellent steak.*

*The steak was not objectionably tough, but the tenderness didn't measure up to the excellence of flavor, appearance, and juiciness.*

*Looked stringy but was OK in taste.*

*The meat was beautiful. I would pick it up in a store for its appearance, however, I would have been unhappy about it. It was too "dry" and lacked the flavor I enjoy.*

*This is the kind I enjoy eating, but the appearance before cooking made it such that I would not pick it out at the store.*

*Had excellent flavor, but needed more fat to keep it moist while cooking.*

*Had strong taste which was not consistently strong in some bites, tasteless in others.*

*Did not have texture of quality meat; flavor a little vague.*

*Not tough in ordinary sense but it seemed to me like I had to chew each tissue.*

*Baby beef flavor.*

### Socio-Economic and Cooking Relationships

The number of criticisms did not vary significantly among groups classified by income or education or age of housewife.

Total criticisms were significantly related to cooking methods used. The ratios of criticisms to steaks were 62.7% for moist heat methods, 65.5% for charcoal broiled, and 90.7% for other dry heat methods. The percentages of steaks cooked various ways were moist heat 29%, charcoal 25%, other dry heat 46%. On another experiment conducted simultaneously of a parallel panel of 90 families,

a similar relationship of criticisms to cooking methods was found: 71.8% for moist heat, 60.5% for charcoal broiled, and 81.4% for other dry heat methods.

The proportions of specific criticisms did not vary significantly by cooking methods.

The total number of ratings of "9" also varied significantly by cooking method in a manner consistent with the variation in total criticisms. The ratios of "9"-rated steaks to total steaks were: moist heat 58.4%, charcoal broiled 55.4%, other dry heat 42.1%.

Some 62% of the steaks were cooked well-done (no pink) and 38% rare. The total number of ratings of "9" did not vary significantly by degree of doneness. However, there were significantly more total criticisms of the well done steaks (81.7%) than of the rare steaks (66.7%). The proportions of specific criticisms also varied significantly by degree of doneness. The well done steaks had proportionately more criticisms of "not juicy enough" and proportionately less criticisms of "lack flavor" than the rare steaks.

Extra flavoring (other than salt and pepper) was added to 32% of the steaks. However, the addition of flavoring did not affect the proportions of specific criticisms.

### Household Performance

A total of 581 steaks of the 1151 steaks evaluated by all households were rated "9." (Lots 25 and 26 involving 96 more steaks are not included in this analysis since they were in an asymmetrical part of the panel design. Included are 8 steaks each for all 144 households).

The possible number of "9" ratings per household ranged from 0 to 8. The actual number ranged from 0 to 8 with a mode and median of 4 and a mean of 4.5 ratings of "9" per household. The complete distribution was:

Per Household	Number of Households
0	5
1	9
2	23
3	19
4	34
5	12
6	26
7	11
8	5
Total	144

The five households at each tail of the distribution must be regarded as extremely critical and extremely uncritical for giving no "9" and all "9," respectively. An examination of the mean ratings of all the loins rated by each of these 10 households indicates that six of the 10 were usually a bit out-of-line with the ratings of other households while the other four were in line.

As another attempt to measure consistency of performance, all the individual scores on each loin were examined. People in 31 households were found who had given scores completely out of line (generally, three or more units below the loin mean and often two or more units below the other scores). However, of these 31 households only nine had such occasions for both members once or for one member twice; only three had such occasions for a total of three times; the remaining 19 people were "out-of-line" only once. The very small number of such occurrences is further evidence that these households were performing fairly conscientiously. Perhaps, it is better stated to say that intensive examination revealed almost no households that could be suspected of performing carelessly or in a completely biased way. However, there was no way of detecting minor carelessness or inattention to the evaluation task.

The distribution of households making criticisms merits a slightly different type of consideration. First, criticisms were requested and relevant only in the approximately one-half of the cases which were rated below 9. Second, the number of criticisms might sometimes reflect individual values as logically as differences in the meat. For example, a household with a preference for extreme tenderness might criticize almost every loin on that basis.

The number of possible criticisms per household on a specific characteristic varied from zero to 8. The distributions for the three more important characteristics were:

No. Criticisms	Not Tender Enough No. Households	Not Juicy Enough No. Households	Lacked Flavor No. Households
0	24	53	50
1	36	36	42
2	33	35	27
3	21	8	15
4	17	8	9
5	10	3	1
6	2	1	0
7	1	0	0
8	0	0	0

The 24 households criticizing none of their 8 steaks on tenderness and those 13 families criticizing more than one-half of their 8 steaks may indicate differences in family attitudes, although some of the variation is undoubtedly due to the particular steaks that they happened to receive.