TECHNICAL BULLETIN NO. 6 JULY, 1939

OKLAHOMA
AGRICULTURAL AND MECHANICAL COLLEGE
AGRICULTURAL EXPERIMENT STATION

W. L. BLIZZARD, DIRECTOR LIPPERT S. ELLIS, VICE DIRECTOR

Physiological Factors Affecting Milk Flavor, with a Consideration of the Validity of Flavor Scores

BY
EARL WEAVER

STILLWATER, OKLAHOMA

SUMMARY

The Problem

Both milk producers and the processors of milk products are greatly concerned with the problem of securing milk which will have a flavor acceptable to consumers. To assure milk of an acceptable flavor, the causes of "off" flavors must be determined so that they may be prevented. Most of the studies of milk flavor have been directed at eliminating causes of off-flavors arising from strong-flavored feeds eaten by the cows, from odors absorbed into the milk after it is drawn from the cow, and from bacterial or chemical activity in the milk after it is drawn. This study is aimed at determining the effects upon milk flavor of the physiological condition of the cows themselves. It considers such factors as age of the cow, stage of lactation, and others. As a by-product of this study, it was found possible to make a statistical study of the probable validity of the method of scoring commonly used in judging milk.

Method of Analysis

Milk samples for flavor scoring were collected at the evening milking each Tuesday over a period of 140 weeks. Thirty-two Jersey cows in the station herd furnished a total of 1,641 samples. On the day the samples were collected the cows were so fed as to deprive them of all roughage for 5 hours previous to milking. This was intended as a precaution to reduce or eliminate feed flavors.

Samples for scoring were collected, without aeration, into 10-ounce milk bottles, were then capped and cooled in ice water to 50° with frequent shaking. When all cows were milked the samples were placed in the refrigerator at 40°, where they were held for 15 hours and were then heated to 95° for the flavor scoring. Other samples were collected for the determinations of the percentage of fat and solids-not-fat.

The flavor scoring was done by members of a judging panel consisting of seven staff members. An effort was made to enlist four men of this panel for the scoring of the samples each Wednesday morning. When one judge's score on a sample is considered as an observation, there were 6,373 such observations on the 1,641 samples. The samples were scored only for flavor, use being made of a scoring chart designed for the purpose.

Statistical analyses were made of various portions of the data to: (a) ascertain the suitability of certain procedures that were employed, (b) determine the influence on milk flavor of the physiological factors studied, and (c) test the validity of the scoring.

Conclusions

1. There is no significant difference in the flavor of morning and evening milk. Also, the flavor of milk secured at one milking each week is a reliable index of the flavor of a cow's milk for that week and pictures accurately the trends in flavor over prolonged periods.

2. The sense of smell is of greater importance than the sense of taste in evaluating milk flavor.

3. Even when the best known practices are adopted in handling cows, flavor defects occur in their milk with frequency. The most frequent defects in order of their frequency were: feed, cowy, rancid, stale, salt, flat, sweet, and bitter. These eight defects constituted well over 90 percent of all the criticisms that were made.

4. The frequency with which flavor defects were noted in this trial was greater than would ordinarily be observed because of three general reasons. The judges were extremely critical, the milk was not aerated, and the samples were heated to intensify the odors.

5. In addition to the three general reasons, there are other specific ones which likely explain the frequency of some of the defects. Though the alfalfa hay was fed more than 5 hours before and the silage 10 hours before milking, it is possible that these feeds caused some of the flavor. Some flavors identified as feed may have been due to bacterial action. The rancid defect was noted chiefly in the milk of cows of the Nancy family though another cow, Blanchia, showed it frequently in one of her lactations following an abortion. The defect was especially noticeable with cows that were "high testers." The salt taste was especially frequent in the milk of the one cow, Tilda Belle.

6. There are great variations among individual cows in respect to milk flavor. The flavor may tend definitely up or down as a lactation progresses and may vary between lactations of the same cow.

7. The age of the cow exerts only a slight effect on milk flavor, the flavor decreasing with age.

8. In general, flavor scores decline with advancing lactation and with advancing gestation.

9. There is a distinct tendency for flavor scores to decline as cows decrease in daily milk yield.

10. Cows that are characterized by tests which are below average for the breed produce milk superior in flavor to that produced by the high-testing cows. The percentage of solids-not-fat is likewise negatively correlated with milk flavor, though the correlation is lower than in the case of fat percentage.

11. In the month of April the cows produce milk of choicest flavor; the lowest scores were accorded in October. The general physical state of the cow is probably a factor here. The trends in flavor score from month to month are directly opposite to those in fat percentage.

12. Neither the atmosphere temperature nor the relative humidity exerts any effect on milk flavor.

13. Though the scoring of the samples was done by seven men, four of whom had no previous experience in such work, the scoring results are on the whole valid and dependable. However, when studied from every viewpoint, the two experienced judges proved to be more proficient than the others.

14. Individual judges may vary somewhat in their thresholds for different tastes and odors and some judges are consistently more critical or more liberal than others in their attitude toward certain defects.

CONTENTS

CONTENTS—(Continued)

PHYSIOLOGICAL FACTORS AFFECTING MILK FLAVOR, WITH A CONSIDERATION OF THE VALIDITY OF FLAVOR SCORES*

EARL WEAVER**
Dairy Department

INTRODUCTION

While the question of flavors is an important one in the production, processing and distribution of all foods, there is no phase of the food industries in which the flavor of the product plays a more significant role than in dairying. Technical and commercial dairy workers attach great importance to this property, as is attested by the fact the official score card for milk assigns to "flavor" 25 points out of the total 100 points for perfection in this product. The score cards for butter and cheese assign 45 points for flavor, while that for ice cream assigns 50 points.

Further interest attaches to the problem of flavors in the dairy industry insofar as the flavors of the different products vary widely. Fluid milk is characterized by a delicate, scarcely discernible flavor and any slight deviation from this ideal constitutes a defect. On the other hand, the other major dairy products carry distinct flavors ranging widely in intensity through the moderate flavors of butter, ice cream and the milder cheeses up to the extreme tastes and odors of certain varieties of rare cheese.

Even though the industry encounters this wide diversity in flavor demands, there yet prevails in every phase of the industry great concern in the problem of milk flavor. Whether the milk is to be employed directly for household use or for the manufacture of other products, its flavor must be acceptable.

HISTORY OF MILK FLAVOR STUDIES

In spite of the likelihood that the importance of flavors in milk has been recognized since the beginning of the present era of dairying some seventy-five years ago, it is of interest that only recently have research workers given more than passing attention to the topic. In the past, numerous investigations have been conducted to ascertain the effects of various factors on milk yield and on the percentages of the different components, but until recently little attention has been given the variations in flavor which may accompany changes in these various factors or changes in milk composition.

In the early work of Steinegger and Alleman (1905) it was established that the milk of diseased quarters is characterized by the salty taste, thus signifying an increase in chlorides. Gabathuler (1915) observed also that the milk of diseased quarters mainfested a great decrease in its content of citrates, but he made no attempt to relate these observations to the flavor of the milk other than to its saltiness. Later Kieferle, Schwaibald and

* From a thesis submitted by the author in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Minnesota, March, 1938.

** Since July 1, 1937, Professor of Dairying, Michigan State College.

Hackmann (1925) established that milk with an abnormally high chlorine-lactose ratio always shows a low content of critic acid. They proposed this relationship as an effective criterion of whether a sample of milk is normal, but the phase of the question in respect to the milk flavor was only incidental in their work. Koestler (1920) had previously called attention to the chloride-lactose number of milk.

Twenty-five years ago Eckles and Shaw (1913-a) commented on the abnormal flavor of the milk from certain cows when near the end of lactation, and Palmer (1922) established that this defect was due to an increased secretion of lipase in the milk of some cows as they approach the end of lactation. He also significantly noted that the advance in lactation is not the primary cause of the defect.

Prior to the past 10 years, but few investigations devoted to milk flavor had been conducted. In the past decade, however, greater interest has arisen in the topic. Koestler, Roadhouse and Loetscher (1928) reported that ovarian disturbanaces were involved in the secretion of milk which became rancid. Then Roadhouse and Koestler (1929) and Roadhouse and Henderson (1930) afforded great impetus to flavor studies in emphasizing the relationship of milk flavor to the chlorides and lactose. Particularly since 1935 a great many contributions have been made. Most workers have been interested in the oxidized flavor, though saltiness and rancidity have also received some attention.

GENERAL CONSIDERATIONS WITH RESPECT TO FLAVOR

DEFINITION OF THE TERM: "FLAVOR"

Flavor is that quality of a substance observed from the blending of the sensations of taste and smell. The term is so used throughout this work. Unfortunately, the dairy literature abounds with inaccurate usage of the word. The official score card for milk uses the expression "flavor and odor." This is a simple case of redundancy. Actually, the quality known as flavor also embodies "odor." A more fortunate choice of terms would be "taste and odor" or simply the one word "flavor." The score card for butter has the expression "flavor and aroma," while the score cards for cheese and ice cream use correctly the single term "flavor."

While usage of the term "flavor" frequently carries the implication that it is synonomous with "taste" and that "odor" is a separate quality, the derivation of the word reveals it is more nearly analogous to "odor" than to "taste." The word "flavor" came from the Old French "fleur," which referred to that quality of a substance that affected the sense of smell. Thus if one were disposed to restrict the term "flavor" to embrace only one of the sensations, he could find some license for restricting it to "odor"; but it is an error to restrict it to "taste" alone.

TASTE AND MILK FLAVOR

Th^ Four Kinds of Taste

It is generally accepted there are only four kinds of taste, or only four sensations that are perceptible through the stimulation of the taste buds. These four tastes are: sweet, sour, salt and bitter. The taste buds are unable to detect any other sensations. It is of interest that this limit in the number of tastes was recognized more than 50 years ago by Bernstein (1886).

An important consideration in respect to the taste a substance may exhibit is the fact that the active principle which yields the taste must be capable of solution in water. The taste buds are subject to no other stimulus than that of a substance dissolved in water.

The Normal Milk Taste and Its Source

To describe the ideal milk taste is difficult. Usually the sweet sensation is emphasized in describing this taste. However, it seems logical to postulate that the ideal milk taste may depend as much on the "salt" as on the "sweet" taste and that the slight, characteristic taste of milk is due to a blending of these two sensations.

In giving consideration to the salt taste, interest is directed to the chlorides as the significant component. However, it is conceivable that the citrates may also contribute to the normal taste of milk. Milk contains citrates in an amount equivalent to 0.2 percent of citric acid, an appreciable quantity.

There is a dearth of information as to the components of milk that contribute to the normal taste. Some useful data on the topic are those afforded by Roadhouse and Koestler (1929). They state, " . . . the ratio of the chlorides per cent to the loctose per cent in milk is one of the most important relations between the components of milk influencing its primary taste." This supports the idea that the ideal taste represents a blending of "salt" and "sweet."

Further evidence on the importance of "salt" in the ideal taste is furnished by Parker (1922), who insists that a substance is insipid if it contains less salt than does saliva. Obviously, unless milk possesses a slight degree of "salt" it will be criticised as flat.

The sour taste is not an element in the normal taste of milk because in milk at a pH of 6.6 the hydrogen ion concentration is not great enough to be detected by taste. Neither is any sensation of bitterness perceptible in normal milk.

The protein and fat make no contribution to the taste of milk. Neither of these components is sufficiently soluble in saliva to afford any stimulus to the taste buds. Furthermore, there is no reason to surmise that either component would impart any one of the four taste sensations. Roadhouse and Koestler (1929) found the non-dialysable compounds of milk were concerned but slightly, if at all, with the primary taste.

The Role of Taste in Constituting Milk Flavor

The sense of taste plays a relatively insignificant role in the scoring of milk flavor. Odor is far more important. In the trial reported here, 6,373 flavor observations were made, an observation being considered as one judge's score on a sample. In only 13.69 per cent of these observations was the flavor defect found to be one of the four tastes while in 78.43 per cent of the observations the defect that was noted was identified through the sense of smell. In 7.88 per cent of the observations no defect was noted.

ODOR AND MILK FLAVOR

The Kinds of Odor

While attempts at classifying tastes have been quite successful, similar efforts to classify odors have been fraught with difficulty. Crocker and Henderson in their earlier work with perfumes and essential oils (1927), and also more recently (1934), make a proposal that the following four groups of odors be accepted: fragant, acid, burnt, caprylic. These workers

establish that a given odor is comprised of a mixture of these four components, the particular odor being dependent on the intensity of each component. They establish further that the intensity of each component can be resolved into nine degrees ranging from zero to eight.

For a substance to manifest an odor, some component of it must be volatilized. There is also some evidence that volatility alone is not sufficient, but that the particles which gain entrance to the olfactory region must also be capable of solution in the mucus that abounds in that region and submerges the receptors.

The Normal Milk Odor and Its Sources

Immediately on withdrawal from the cow, milk has a pronounced odor. However, the principles which impart this odor are quite volatile, and they rapidly disappear on exposure, leaving a mild, delicate odor which characterizes superior milk. The odor of fresh milk is often attributed to carbon dioxide. It seems likely, however, that other components may be more largely responsible. There is reason to suppose certain volatile fatty acids are involved. They undoubtedly occur in abundance in the milk within the udder. On withdrawal of the milk they rapidly escape, particularly at the temperature of fresh milk.

The components that confer the odor of the normal milk after its aeration are also not definitely determined. However, it is commonly accepted that this odor is due to free fatty acids and to glycerides of the lower acids.

ACCEPTED CAUSES OF ABNORMAL MILK FLAVORS

Though many workers had previously investigated the causes of abnormal flavors in milk, Gamble and Kelly (1922) were the first to classify these causes. They attribute off-flavors to:

"1. The internal or physical condition of the individual cow.

"2. Those absorbed within the body of the cow from highly flavored feeds.

"3. Odors absorbed into the milk after production.

"4. Bacterial development within the milk on standing."

Recent work on the oxidized flavor has shown it can not be placed in any of these four categories; hence Chilson (1935) proposed a fifth group of off-flavors, namely, those caused by chemical action within the milk on standing and which may not be related to any bacterial activity.*

THE PROBLEM; OBJECTS OF THE TRIAL

The foregoing summary indicates that valuable data on the effect of various factors on yields and milk composition have been obtained by many authors. These data have made it possible to chart quite accurately the normal trends of yields and of the milk components with changes in these various factors. It therefore was felt desirable to secure data that might permit the establishment of normal trends for milk flavor with

* Since the trial reported in this bulletin was started, a great many reports of milk flavor studies have been made. These contributions receive mention in the appropriate sections of this bulletin. Among the contributions the oxidized flavor has had considerable attention. A few have dealt with rancidity and some with saltiness. Likewise some attention has been given to certain of the 10 factors considered in this study with respect to their possible influence upon these three specific defects.

changes in certain physiological factors that seemed to lend themselves to study. Accordingly the trial was inaugurated to ascertain the effects on milk flavor of the following:

I Individuality of the cow.
II Age of the cow.
III Stage of lactation.
IV Stage of gestation.
V Daily milk yield.
VI Percentage of fat.
VII Percentage of solids-not-fat.
VIII Season of year.
IX Atmospheric temperature.
X Relative humidity.

The first seven of the ten factors enumerated above are directly related to the physiological activities of the cow. While the other three are of a seasonal nature it is not illogical to conjecture that such an influence on milk flavor as they might exert would not be a direct seasonal influence but an indirect one modifying the physiological activity of the mammary gland. Consequently all ten of the factors are considered to be of a physiological nature.

In the work reported here, attention was not restricted to any particular flavor. The observations embraced consideration of all flavor defects which might be anticipated.

It was found during the course of this trial that data were being obtained in such number as would make them valuable for some detailed analyses of the scoring as performed by the individual judges. Hence some attention is given here to the question of the validity of flavor scoring.

METHODS

The study reported here was inaugurated on August 1, 1933. Observations on the flavor of the milk from individual cows were continued until April 1, 1936. Attention to the data from this trial appears to have special justification in the fact that observations were made in a continuous series over a prolonged period of 140 weeks. The period embraced as many as three successive lactations of some cows. Furthermore, it embraced the succession of seasons from the summer of 1933 to the spring of 1936.

In planning this trial, major attention was focused only on that one category of causes of abnormal flavors which Gamble and Kelly suggested might be related to "the internal or physical condition of the individual cow."

An effort was made to eliminate the feed flavors such as fall in the second of the Gamble-Kelly categories. (See page 12, supra.) Results to be presented reveal that little success in this direction was achieved.

Only reasonable precautions as dictated by good dairy practice were taken to avoid odors that might be absorbed by the milk after its production. They were of no significance.

Through the expedient of cooling the milk immediately in ice water and holding it at 40° F. until ready for scoring it was hoped to avoid any interference from flavors due to bacterial action. It is not establshed that this precaution was entirely sufficient; bacterial action may have been involved in producing some flavors that were encountered.

The fifth possible cause of flavor defects; i. e. chemical changes in the milk, received no attention at the time the work was outlined. Care was exercised that the pails were properly tinned, but the chief object here was to avoid bacterial contamination. The role of copper in catalyzing the reaction that yields the oxidized flavor was not appreciated at that time. The milk was not protected against some copper contamination from the strainer that was used.

However, when the role of copper in producing this flavor was appreciated after the trial had been in progress for some time, it was deemed best not to make a radical change in procedure that would introduce a complication in presenting the data. In effect, if copper contamination had been a disturbance in the earlier observations it was conjectured that the continuance of such a disturbance as a possible constant effect throughout the trial was even preferable to any situation which would have invalidated comparisons between the earlier and later observations.

Several incidental observations were made to ascertain if the strainer that was used was involved in the frequent and intense flavor that the judges were identifying as "stale." It was suspected this flavor might be properly called "oxidized." It could not be demonstrated, however, that the strainer was thus involved. In repeated instances samples of milk from offending cows were collected by use of the strainer, while other samples were collected without straining. The unstrained samples showed the flavor defect as seriously as the others.

FREQUENCY OF SAMPLING

The question of the frequency with which it would be desirable to collect the samples for scoring provoked some deliberation in outlining this trial. It did not seem feasible to attempt the task of scoring the milk of each cow at each milking throughout the entire period of study. Furthermore, even though fat is the most variable milk component, previous studies had established that yields of fat on one day each month were quite reliable in computing annual yields. Accordingly, it was postulated that flavor scores of samples from both milkings on one day each week would picture accurately the flavor trend in the milk of a cow throughout a lactation. Tuesday was arbitrarily selected as the day of the week on which the samples would be collected.

Another aspect that demanded attention in outlining the trial was the number of samples that could be handled effectively each week. This determined the number of cows that could be used. With the labor, the assistance in scoring, and the facilities available, it appeared that approximately 20 samples a week was the maximum number that should be attempted. Considerable labor was necessary in feeding and handling the cows under the methods that were occasioned by the nature of the trial.

Also it was planned to use at least four staff members to score the samples collected each Tuesday, and it seemed inadvisable to contemplate the services of these men for more than an hour each week. Information was not at hand as to the number of samples such as this which a man could score accurately in an hour, but it was conjectured that more than 20 samples might be excessive. During the progress of the trial there were many days when the samples were of such a nature as permitted rapid work. On other days, the samples provoked such extreme care and deliberation that fatigue was quite evident; and on these days the limitation in the number of samples was a distinct advantage.

Reliability of Samples from One Milking a Day

As previously stated, when the trial was started, samples from both the morning and evening milkings on each Tuesday were collected and scored. After three weeks the data up to that period gave some intimation there was little, if any, difference in the flavor of a cow's morning milk and her evening milk. After the seventh week of the trial the data up to that point were analyzed. This indicated it was possible to rely with complete confidence on results secured from one sample a day instead of two. The sampling of both the morning and evening milk appeared superfluous. Thereafter samples were collected only on Tuesday evenings.

Group of Cows Based on Occurrence of Rancidity. The seven-week period used to ascertain if there was a difference between morning and evening milk embraced an entire month of August and continued into September. This is the season when many cows in the station herd are reaching the end of lactation. During these seven weeks, samples were collected from 11 cows at both milkings on Tuesdays.

Five of these cows showed no rancidity in their milk at any time during the seven weeks. On the other hand, six of the cows showed this defect one or more times during the period. It is necessary to consider these groups of cows separately, for there is a contrast between them. As revealed in Table I, the cows that never showed the rancidity received scores on their evening milk that were no different from their morning scores. With the cows whose milk was susceptible to the development of rancidity the difference in morning and evening scores was pronounced.

TABLE I. Average Flavor Scores of Morning and Evening Samples*

Group A. Five cows that never showed "rancid" in the 7-week period.

Group B. Six cows that did show "rancid" in the 7-week period.

Name of cow	Stage of lactation (weeks)	MORNING SAMPLES		EVENING SAMPLES		Excess of morning over evening
		Number	Score	Number	Score	
Group A						
Tilda Belle	19	7	20.5	7	20.3	0.2
S. F. Aileen	58	7	20.9	7	20.9	0.0
Little Jolly	20	7	21.0	7	20.7	0.2
Red Flag	46	5	20.6	5	19.8	0.8
Gypsy Lass	50	5	20.8	5	21.1	−0.3
All 5 cows	39	31	20.8	31	20.6	0.2
Group B						
F. F. L. Nancy	13	7	20.9	7	19.8	1.1
Fontaine	57	6	20.9	6	19.5	1.4
S. Aileen	25	7	20.1	7	18.2	1.9
R. F. L. Nancy	3	7	21.4	7	20.1	1.3
S. F. Nancy	53	2	21.9	2	17.6	4.3
S. F. Goo Goo	54	5	17.9	5	14.8	3.1
All 6 cows	34	34	20.5	34	18.3	2.2

* Scored on a basis of 25 points for perfect flavor, according to the American Dairy Science Association score card.

However, it is established to the author's satisfaction that the difference noted in the milk of the "rancid" group was not due to any inherent difference between morning and evening milk but to the fact the evening samples, under the procedure used, were not scored until the next morning after a 15-hour period of storage. The morning samples were scored after three hours; sufficient time did not elapse for rancidity to develop in these samples. It is generally accepted that 10 to 12 hours of storage are necessary for the defect to appear. Eckles and Shaw (1913-a) made this observation in their early work.

In the trial reported here, several incidental observations were made to ascertain if any of the milk was rancid immediately upon being drawn. Certain cows were consistently showing this defect after the necessary interval, and the judges were informed that the samples came from these suspected cows. The judges were also quite confident that the milk would be rancid 15 hours later, yet in no case was the defect ever detected in the milk when drawn. Such milk was often criticised as "cowy."

Analysis of Variance of Morning and Evening Samples. As shown in Table II, the data on each group of cows was subjected to an analysis of variance study to ascertain the significance of the difference between the morning and evening samples.

From Table I it may be noted that the number of samples was not the same for all the cows. Some of the cows dried off during the seven weeks; others freshened after a few weeks had passed. The unequal frequencies occasion somewhat awkward calculations in an analysis of variance and they preclude the possibility of "keying out" for the interaction of "cows × weeks." An alternative procedure here would have been to ignore the data

TABLE II. Analysis of Variance of Scores of Morning and Evening Samples

Group A. Five cows that never showed "rancid" in the 7-week period.

Group B. Six cows that did show "rancid" in the 7-week period.

Variation due to	Degrees of freedom	Sum of squares	Mean square	F value
For Group A				
Time of day	1	0.50	0.50	0.27
Cows	4	4.87	1.22	0.66
Time × cows	4	1.51	0.37	0.20
Error	52	96.24	1.85	
Total	61	103.12		
For Group B				
Time of day	1	22.25	22.25	5.27*
Cows	5	136.10	27.22	6.45**
Time × cows	5	67.08	13.42	3.18*
Error	56	236.32	4.22	
Total	57	461.75		

* The single asterisk adjoining an F value indicates it is significant to Snedecor's (1937) tabular 5 percent point.

** The double asterisk indicates high significance to the tabular 1 percent point.

These symbols are similarly used herein in all tables showing analyses of variance.

on those cows from which samples were not collected for the entire seven weeks; but, in view of the fact this woud have sacrificed some of the data, it was chosen not to use this alternative. The inability to establish this interaction of "cows × weeks" is not a serious handicap in interpreting the results. The trial reported here was of such a nature that the matter of frequencies could not be controlled. Consequently, the different phases of the trial that are to receive attention often present this problem of the unequal frequencies. Support for the procedure that was employed is afforded by Snedecor (1934) and Immer (1937).

From Table II it may be observed that for the cows that never showed the rancid flavor the F value for the "time of day" was only 0.27. This is far short of the point of significance; there is no significant difference in the scores of morning and evening milk from these cows.

With the other group of cows a different situation obtained. The computed F value of 5.27 exceeds the tabular 5 percent point and indicates a significant difference. In fewer than 5 cases out of 100 a variance as great as that obtained would be due to random sampling. However, the variance was undoubtedly due to the difference in the length of the storage period previous to scoring.

Another fact from Table II which merits attention is that pertaining to the variance due to "cows." With the "non-rancid" group the F value of only 0.66 indicates there was no significant difference in scores that could be attributed to the individual cows. With the other group the F value of 6.45 exceeds even the tabular 1 percent point. There is less than one chance in a hundred that a variance as great as that obtained would be due to random sampling; the variance among these cows was highly significant.

With both groups of cows the relatively low mean square for "time × cows" signifies that the cows all behaved quite regularly. Had it been relatively high it would have revealed that some cows at times showed a dfiferent relation between their evening and morning samples than they showed at other times.

Comparison of Tuesday Scores with the Average of Other Six Days

For various reasons it was thought desirable to collect the samples in this trial with no greater frequency than was necessary. Only that frequency of sampling was desired which would afford a reliable picture of the milk trends over the periods studied. As stated previously, the number of samples that could be effectively handled each week appeared to be limited. Out of this situation there arose two alternatives in procedure. On the one hand, more frequent samples could have been collected from fewer cows; this would have reduced the number of individual cows used in the trial. On the other hand, the less frequent sampling permitted the use of more cows.

Another consideration in respect to the frequency of sampling was the fact that some modification from herd practice was necessary in handling the cows to avoid as far as possible the disturbing effects of feed flavor. As will be explained, the cows were deprived of roughage for at least five hours previous to the milking at which samples were collected. While this procedure could scarcely be considered drastic when employed only one day each week, it was feared ill consequences might result if the cows were too frequently handled in an unusual manner.

The decision to resort to the collection of samples only on Tuesday was entirely an arbitrary one. It was hoped such a procedure would prove sufficient. Hence, as easily as possible in the study an incidental trial was conducted to ascertain if the procedure was adequate. Fortunately, the results of the trial established that such was the case.

Eight cows, including three of those used in the regular study, were used in testing this matter. The trial with these eight cows was continued for a period of 12 weeks, during which time the samples were collected at the evening milking each day.

Table III gives the results of this 12-week trial. Four of the cows showed lower scores for Tuesday than for the other six days of the week. Three of them show slightly higher scores on Tuesday, while one cow, Bernice, shows no difference. The average for all cows on Tuesday is 0.3 points lower than on the other six days.

In the analysis of variance given in Table IV, use was not made of the average scores given in Table III but of each Tuesday score against the average of the otner six days that week. In Table IV the F value of 3.79 for "Tuesday vs. other days" is not significant. The Tuesday score is not significantly different from the average scores on the other six days of that week.

The variance of the scores in the different weeks was highly significant, as shown by the F value of 4.09. There was enormous variance among the cows, with the F value of 255.99. This would be surmised from the data of Table III. R. Nancy, for instance, received scores of approximately 20.0, while the scores of F. F. L. Nancy were below 11.0.

The relatively high mean square for the interaction of "days × weeks" reveals the fact, which may be considered unfortunate here, that in some

TABLE III. Average score on Tuesday compared with the average of the other six days.

Name of cow	Average score on Tuesday	Average score on other 6 days	Excess of Tuesday over other 6 days
Creametta	18.6	19.2	−0.6
Fobes	17.9	18.1	−0.2
Dewdrop	19.7	19.5	0.2
Bernice	18.8	18.8	0.0
No. 302	19.7	19.5	0.2
Blanchia	17.3	17.7	−0.4
F. F. L. Nancy	10.7	11.2	−0.5
R. Nancy	20.1	20.0	0.1
All 8 cows	17.7	18.0	−0.3

TABLE IV. Analysis of variance of scores on Tuesday and average scores of other days.

Variation due to	Degrees of freedom	Sum of squares	Mean square	F value
Tuesday vs. other days	1	3.03	3.03	3.79
Weeks	11	35.94	3.26	4.09**
Cows	7	1431.73	204.53	255.99**
Days × weeks	11	24.55	2.23	2.79**
Days × cows	7	4.77	0.68	0.85
Weeks × cows	77	101.62	1.32	1.65*
Error	77	61.52	0.80	
Total	191	1663.16		

weeks the Tuesday samples were scored higher and in other weeks lower than the samples for the other days of the particular week. Likewise the high interaction of "weeks×cows" shows a lack of constancy. Some cows scored high in certain weeks; other cows scored high in other weeks. The low mean square for "days×cows" indicates that even though the cows varied greatly themselves, each cow was quite constant in showing the same relationship between her Tuesday scores and her score for other days that week.

THE COWS USED

Thirty-two purebred Jersey cows in the station herd were used in this study. The number of samples from the cows varied with the individuals. Certain cows were dry for longer periods than others. In two instances, with Blanchia and R. Nancy, the cows were under veterinary treatment for two to four weeks and their milk was rendered temporarily unsuitable for the trial because of the treatment employed. A third cow, S. F. Goo Goo, aborted and was kept in isolation for six weeks during which time observations on her milk were omitted.

At no time throughout the trial was the milk of any cow discarded because of mastitis. The cows of the herd were notably free of any external manifestations of mastitis infection; however, they were not all subjected to mastitis tests. Within a few weeks after the trial started the great frequency of saltiness in Tilda Bell's milk aroused interest. She was found to be consistently positive to the brom-thymol blue test but her milk was regularly sampled and scored.

FEEDING THE COWS

While an effort was made to feed the cows in such a manner as to eliminate feed flavors in the milk, some disappointment was experienced in the apparent inability to accomplish this end. In spite of the precautions taken to avoid feed flavors, the results show that in 36.2 percent of all observations the criticism of "feed" was indicated. It is of interest to note that recently Roadhouse and Henderson (1937) have shown the possibility that this problem of excessive feed flavors may be met by giving cows constant access to their hay.

It was the plan in this trial to feed the cows their silage on Tuesday morning immediately after the milking was completed, about 6:00 a. m. Darso silage was used. It was usually eaten within an hour, but if not eaten in that time it was removed from the mangers. The silage thus was fed about 10 hours before the experimental milking at 4:00 p. m., and it was hoped that therefore this silage would not affect the milk flavor. Some other results which the author has reported (1935) showed a six-hour interval between the feeding of darso silage and the milking was sufficient to avoid the flavor. However, Gamble and Kelly (1922) found that corn silage fed immediately after one milking might affect the flavor at the next milking, but it is believed that corn silage is more likely to affect milk flavor than is the darso silage used in this trial.

After the silage was eaten or removed from the manger, the cows were usually turned out into a lot. The lot had a small quantity of bermuda grass. In severe cold or wet weather, which was infrequent, the cows were kept in the barn.

At 9:00 o'clock on Tuesday morning alfalfa hay was fed. This was all consumed by 11:00 o'clock. At that time, the cows were brought into a dry lot or, if the weather was quite cold or quite hot, into the barn. In no case were the cows fed roughage after the morning feeding of hay. Thus they were deprived of all roughage for at least a five-hour interval previous to the experimental milking at 4:00 p. m.

It is possible that this interval was not long enough and that the frequently observed feed flavor may have been due to the hay. The work by the author (1935) which has just been cited indicated an interval of five hours is not definite assurance against slight flavors from alfalfa hay.

However, in handling a herd it is not feasible to deprive cows of their roughage for too long a period. Furthermore, it is desirable in a trial of this nature to avoid any pronounced abnormalities in the feeding regime which might affect the milk flavor through possible modification in the yield or composition of the milk.

In any event, though flavors from the silage or alfalfa hay may have been detected, there is some assurance that these flavors exerted a constant effect throughout the trial and interfered little, if any, with the effort to measure the effects of the particular factors being studied.

It is emphasized here that the cows were fed the same on each Tuesday, regardless of the season. The silo was kept open the year 'round. As a matter of fact the cows also received silage on the other days than Tuesday throughout the year. On the other days than Tuesday they did have access to pasture during the pasture season. However, such pasture was of little consequence except in early spring and following fall rains in October.

Concentrates were fed during the milking period. The concentrate mixture consisted of corn, bran, oats and cottonseed meal. When feed flavors were encountered so frequently, some suspicion was directed against the concentrates. However, repeated observations wherein these feeds were used singly and in mixture established they were not responsible for the flavors.

Roadhouse, Regan and Mead (1926) had shown that no effect on milk flavor was exerted by a grain mixture of rolled barley, bran and linseed meal. Also McCandlish and Leitch (1932) had reported that concentrates have no effect on milk flavor unless they are of an unusual nature such as fish meals.

THE COLLECTION OF SAMPLES

The cows were milked by hand. As soon as each cow was milked the milk was carried to the milk room and weighed. It was then poured twice from one pail to another for satisfactory mixing and was then sampled with a pint tin-cup. The sample was then poured through a strainer into a 10-ounce milk bottle. As soon as the sample was collected the bottle was capped. It was immediately cooled in ice water to 50° F., the bottles being shaken every two or three minutes to facilitate cooling.

It is to be noted that this procedure did not permit of aerating the milk. Such aeration would have removed or decreased some of the off-flavors and would have resulted in higher scores. The author (1935) has shown aeration will remove about half of the off-flavors imparted to milk by alfalfa hay. Gamble and Kelly (1922) likewise found aeration would materially reduce the flavors from corn silage; and Babcock (1923-a, 1923-b, 1924, 1925) found that similar efficiency in aeration reduced the intensity of flavors from the series of feeds he studied.

However, in the flavor trial reported here the object was to determine the flavor of the milk as produced by the individual cows. It was necessary to avoid aeration to accomplish this object.

When the milking was completed on Tuesday evening, at about 5:00 o'clock, the cooled samples were placed in the refrigerator at 40° F. and were held there until 8:00 o'clock the next morning. Thus they were held for 15 hours before they were scored.

The samples for the fat and solids determinations were collected in half pint bottles. They were not cooled in ice water but were placed in the refrigerator along with the other samples. The determinations by use of the Majonnier tester were made on each Wednesday afternoon, 20 hours after milking.

SCORING THE MILK

The Judges

An effort was made to have four of the Dairy Department staff members score the milk, which was always done on Wednesday morning at 8:00 o'clock. The scoring was not a prescribed duty for any four particular men but was done by any of the seven staff members who were considered as qualified. These seven men constituted a panel of judges, and the samples were scored by those members of this panel who were available and who were not indisposed by colds, throat ailments, or any other condition that might have impaired the senses of smell and taste. Incidentally, the men exercised especial precaution in choosing their diets for the Wednesday morning meal; each evolved certain choices that usually enabled him to proceed with the scoring with considerable confidence in the score he accorded.

Frequent reference will be made to the seven judges in respect to the scores they accorded, to the frequency with which they detected the various defects, and to the agreement among them. Table V gives some information about these judges. It is seen from this table that judges 1 and 5 had had considerable experience in scoring milk previous to the start of this trial. With the exception of judge 6, who had done some work of this nature, all the others were inexperienced.

The ages of the judges are given because of a quite common belief that younger persons possess more ability than older persons in detecting tastes and smells. Miss King (1937) recently reported a study on judging bread wherein she found those persons under 30 years of age were more accurate in the judgments they rendered than persons over 30. The superior ability of the younger judges lay especially in their lower thresholds for the sour taste. Dukes (1935) reports the taste buds are more numerous in young individuals than in old.

However, Trout and Sharp (1936) state that the proficiency of specialists in the art of tasting is not due to increased sensitiveness but to the acquired "knowledge of what signs to look for and how to interpret these signs." It would logically follow that young judges do not necessarily enjoy an advantage. The results in this trial show the age of the judge was not a factor in the accuracy of the scoring.

It is often alleged that the smoking habit impairs one's ability to detect flavors. Miss King, however, found this was not a factor in judging bread. The results here do not indicate that the judges who smoked were less proficient than the others.

Decision to Use All Seven Judges. In outlining this project the necessity arose for a decision as to whether all seven of the available staff members should be used or whether the two experienced judges should be assigned the responsibility for all the scoring. The former alternative was selected. It was hoped the inexperienced judges would rapidly acquire such proficiency as would enable them to score the samples satisfactorily. Whether this hope was realized is answered as fully as possible in subsequent sections here. Suffice to say the results indicate that the inexperienced judges did perform the scoring with considerable accuracy and dependability; nevertheless, it will be apparent the two experienced judges performed in a somewhat superior manner.

TABLE V. The judges.

Judge's number	Smoking habit	Approximate age	Previous experience
1	None	34	Considerable
2	Pipe	48	None
3	Pipe and cigars	41	None
4	None	35	None
5	Cigarettes and pipe	40	Considerable
6	Cigarettes	25	Some
7	Cigarettes	42	None

Trout and Sharp (1936) found that young college men with little or no experience rapidly acquired the ability to employ taste accurately. Platt (1931) in considering the scoring of foods in general emphasizes that "experts are not necessarily the best judges." Of course, considerable latitude exists here in the concept as to what constitutes an "expert" or an "experienced" judge. Miss King (1937), in the work previously cited, determined the taste abilities of 64 persons, both men and women. She employed pure solutions of sodium-chloride, sucrose, lactic acid, and caffeine. From the 64 persons, 14 with superior ability were selected. However, the final results in judging bread did not reveal the selected 14 were more capable than the original 64.

In making the decision to use all the seven available staff members for the scoring in this trial it can not be denied there was some disposition to impute a certain degree of importance to sheer numbers of observations. Admittedly this attitude is subject to some criticism. A great number of unreliable observations can not compensate for fewer dependable observations. However, the hope prevailed that all the judges would rapidly acquire proficiency so their observations would be highly acceptable.

Trebler (1935) is one worker who definitely challenges the attitude wherein too much import is attached to the virtue in numbers. In reference to milk scoring he insists that the results from one good taster are more valuable than opinions of several people who are not especially interested.

A further consideration in the decision to use the several judges in this trial lay in the belief that the greater number of persons whose scores were obtained would more nearly represent the viewpoint of the usual consumer than is generally true in the judging of dairy products. It was felt that the importance of this viewpoint has been minimized too often, both in the commercial scoring of dairy products and in some of the experimental work that has been reported. Platt (1931, 1937) repeatedly emphasizes this aspect of the question in respect to foods in general.

Insofar as this trial contemplated the observations for a period of 140 weeks, it was surmised that if all the scoring were delegated to only two men it would constitute a burden that seemed unnecessary. In the event of the sickness of one judge or his enforced absence, the entire responsibility would have devolved upon the other. There appeared the possibility also that both judges might have been incapacitated at the same time, and this would have interrupted the sequence in the observations. The decision to use the greater number of judges embraced the desire to avoid such possible interruptions.

Identification of the Samples

For the scoring, the samples were identified only by numbers painted on the bottles. The number for a given cow was varied from week to week. While this changing of numbering was designed to safeguard the identity of the cow whose milk was being scored, it is believed the effort in this direction was largely superfluous. The judges rapidly acquired the ability to identify many of the cows from the flavor of their milk.

The Scoring Technic

On each Wednesday morning shortly before 8:00 o'clock the samples were removed from the refrigerator, taken to the laboratory and heated to 95° F. in water. The water was then adjusted to hold the milk at this temperature until the scoring was completed.

When the judges were assembled, a sample was removed from the water bath and mixed by inverting the bottle. The cap was then removed. Each judge, after observing the odor, poured a portion of the milk into a 50 m.l. beaker. He did his scoring independent of the others. The usual practice was to take 8 to 12 m.l. of milk into the mouth, though in some cases, as when the odor revealed rancidity, the quantity taken was quite small. In a few cases of distinct rancidity the judges refused to taste the samples; their scores in such cases were based entirely on the odor. Some judges consistently used larger portions of milk than the others.

An effort was made to avoid distraction in the laboratory during the scoring. The judges were disposed to employ quite complete introspection in evaluating the flavor scores. Trout and Sharp (1936) have shown that distraction impairs one's ability to detect and identify flavors. A further precaution was taken to avoid odors in the laboratory during the scoring.

Each judge was encouraged to take as much time as he desired in scoring a sample. At the start of the trial the time to score the samples was recorded, though this plan was abandoned after three weeks. The judge was at liberty to repeat the tasting and smelling of a sample as frequently as he desired in order to reassure himself. After he reached his decision on a sample, he recorded his score on a special form provided him. When all scores on a sample were recorded it was the usual practice for each judge to announce his score, though no judge was permitted to change his score after any had been announced. The average of the scores by all the judges on a sample was accepted as the score on that sample.

It was realized that this procedure could result in marked variation in the individual observations on a sample. Such variation did exist. No such success was enjoyed in this trial as that reported by Roadhouse and Koestler (1929) wherein they state, "In few instances was the variation greater than two-tenths of a point in score." As a matter of fact, in the trial reported here the use by the judges of fractional points was discouraged. It was not believed the scoring was sufficiently precise to justify their use; further there appeared to be no virtue in employing any subterfuge in an effort to intimate that the scoring was possessed of extreme precision.

It is entirely likely that such variation as existed in the individual scores on a given sample could have been greatly reduced had the judges striven meticuously for a "composite" score. However, such a composite score was not desired; the opinion of the individual judge was the objective. It may or may not have been unfortunate that these opinions frequently varied.

The Flavor Scoring Chart

It is to be admitted that scoring of milk for flavor can not give wholly exact and uniform results but is subject to limitations imposed by the fact it depends on individual opinion. It was hoped, however, in conducting this trial that greatest reliability in the scoring might be attained through the use of a flavor-scoring chart as described by Fouts and Weaver (1935). The chart was suspended on the wall of the laboratory and was before the judges during the scoring. It was designed to accomplish some unification of the terminology employed, and to afford guidance in evaluating the seriousness of a given defect when encountered. Though the chart was especially helpful to the inexperienced judges, the experienced men also enjoyed some benefits in its use.

This flavor-scoring chart was a modification of the "general guide for scoring flavor and odor" as proposed by Babcock and Leete (1929) and embraced certain suggestions by Lucas (1929) regarding the numerical scores for various defects. Platt (1931) had proposed a similar chart for scoring foods in general, and later Thomsen (1936) submitted an analogous one for scoring the flavor of butter.

The chart, which is given in Table VI, suggests that a sample may be placed in any one of the five quality groups as proposed by Babcock and Leete. The milk is grouped as **excellent** when it scores 23 or above, **good** when the score is 21 or 22, **fair** when 18 to 20, **poor** when 12 to 17, and **bad** when 11 or below.

In the first column of the chart, 18 flavor defects are suggested. It is recognized this list was only an arbitrary one arising out of experience in judging mik; many workers would suggest other terms, but this list proved quite adequate. There was insistence at all times that unless a judge felt satisfied one of the suggested terms properly described the defect which he

TABLE VI. The chart used to aid in scoring milk flavor.*

Quality group	Excellent	Good	Fair	Poor	Bad
Score	23 or above	21 or 22	18 to 20	12 to 17	11 or under
Defect					
Acid				s. acid	acid or v.
Bitter				s. or bitter	v. bitter
Cooked		s. cooked	cooked	v. cooked	
Cowy		s. cowy	cowy	v. cowy	
Disinfectant				s. dis.	dis. or v.
Feed		s. feed	feed	v. feed	
Flat		s. flat	flat	v. flat	
Metallic			s. metal	metal or v.	
Musty				s. musty	musty or v.
Nutty			s. nutty	nutty or v.	
Oxidized			s. oxid.	oxid. or v.	
Rancid				s. rancid	rancid or v.
Salt			s. salt	salt	v. salt
Sharp		s. sharp	sharp	v. sharp	
Stale			s. stale	stale or v.	
Sweet		s. or sweet	v. sweet		
Watered			s. water.	water. or v.	
Weedy				s. weedy	weedy or v.

* s. indicates slight; v. indicates very.

noted, he was to employ one of two plans. He was encouraged either to coin any descriptive term of his own which he felt was satisfactory, or, in the event he could not coin such a term, he was to indicate the defect as "unidentified." Among the coined terms, "oily" was used most frequently. In addition, there were certain inelegant terms such as are often employed in scoring dairy products.

It is of interest, however, that coined terms were used only 26 times, or in only 0.41 percent of the total observations. In 90 of the observations, or 1.41 percent, the judges were unable to describe the defect to their satisfaction and indicated "unidentified." However, the total number that fell into these two categories was only 116, or 1.82 percent. The list of defects on the scoring chart was evidently quite adequate.

RESULTS IN RESPECT TO SCORES AND DEFECTS

DISTRIBUTION OF THE NUMERICAL SCORES

The judges were at liberty to accord numerical scores ranging from zero to 25.0. However, the scores that were given were distinctly skew in their distribution. Of the total 6,373 observations, 19.63 percent fell in the class for the score of 21.0. This was the class with the highest frequency. Only 14.51 percent of the scores were 22.0, 7.27 percent were 23.0, 0.60 percent were 24.0 and 0.02 percent were 25.0.

Proceeding downward from the high class at the score of 21.0 the decline in frequency in successive classes was quite gradual. Six-tenths of one percent of all the observations fell in the class for the score of zero.

With the distinctly a-normal distribution of the scores on these samples, some concern is felt that the statistical treatment of the data may not be valid in all cases. However, no method to meet this situation seemed to be available and it is not believed that the situation imposes a material handicap in the interpretation of the data.

FREQUENCY OF THE FLAVOR DEFECTS

One of the quite startling results of this trial was the high frequency with which the flavor defects were noted. In only 502 observations, or 7.88 percent, was no defect noted. While Roadhouse and Koestler (1929) had reported their study wherein only 9.8 percent of the cows produced milk that was "excellent," it was anticipated that the controlled feeding in this trial would afford milk considerably above average.

As was stated previously, in 1.41 percent of the observations the defect was unidentified and in 0.41 percent the judges noted some defect other than the 18 that were suggested. In most of the observations, i. e. 68.49 percent, only one flavor defect was noted, while in 21.81 percent there were more than one. In most of the latter cases two defects were detected such as "feed and cowy" or "bitter and rancid." In a few cases three defects were indicated, and in rare instances there were four. These results are shown in Table VII.

Out of the consideration of Table VII the inquiry arises as to which of the specific defects were noted most commonly in this trial. Table VIII gives results on this aspect of the problem. Weaver, Fouts and McGilliard (1935) have reported some of the preliminary results.

Data were obtained to show the frequency with which each of the defects was noted alone in a sample and also the frequency with which each was noted with any other defect. It is not essential, however, to present here separate tabulations of these data. Table VIII shows the frequency

of each defect both when it was used as the sole criticism and when used in combination. The list of defects includes the 18 that were suggested on the chart, together with "oily."

TABLE VII. General summary of the scoring.

	Number of observations	Percentage of all observations
Observations in which:		
No criticism was noted	502	7.88
The defect was unidentified	90	1.41
Defect was not one of the 18 enumerated	26	0.41
One defect was noted alone	4,365	68.49
Two or more defects were noted	1,390	21.81
Total	6,373	100.00

TABLE VIII. Frequency with which defects were noted both alone and in combination.

Rank in frequency	Defect	Number of times noted	Percentage of all observations
1	Feed	2,305	36.17
2	Cowy	1,347	21.92
3	Rancid	910	14.28
4	Stale	872	13.68
5	Salt	568	8.91
6	Flat	457	7.17
7	Sweet	164	2.57
8	Bitter	135	2.12
9	Sharp	42	0.66
10	Weedy	25	0.39
11	Cooked	24	0.38
12	Oxidized	24	0.38
13	Metallic	21	0.33
14	Nutty	18	0.28
15	Oily	14	0.22
16	Watered	11	0.17
17	Acid	6	0.09
18	Musty	4	0.06
19	Disinfectant	3	0.05

The criticism of "feed" was used in 36.17 percent of the observations. The first four defects embrace a majority of the observations; only the first eight were used in more than one percent. The remaining 11 defects were not significant.

Attention is called to the fact the figures in Table VIII are of a non-additive nature and totals at the bottom of the columns would be ambiguous. For instance, "feed" and "cowy" were often noted together. In such a case this one observation is included opposite each defect. In cases where three defects were noted, the one observation is included in three places. Hence, the total would exceed the actual number of observations.

Possible General Reasons for the High Frequency

There are three possible general reasons why so many defective samples were noted in this trial. In the first place the judges were keenly alert to detect any flavor defect that might exist. Nelson and Trout (1934) urge a caution in judging dairy products to the effect the samples be given the benefit of the doubt, but it appears that the judges in this trial violated this caution. Seldom was the sample given the benefit. Likely enough the judges observed defects that would have passed unnoticed by even the most fastidious consumer of milk.

A second reason that defects were noted with such frequency is that the milk was not aerated. The object was to ascertain the flavors of the milk as produced by the cow.

The third reason is that the samples were heated to 95° F. for scoring. This was done in conformity with the usual practice in scoring milk. Incidentally, it is proposed that this is an unfortunate practice; room temperature would be more appropriate.

Comparison of Warm and Cool Samples. An incidental trial was conducted to ascertain the difference in flavor scores when samples were scored both warm and cool. Twelve of the cows were used on three different dates. Duplicate samples were collected. One of the duplicates was heated to 95° F. before scoring; the other was scored at 70° F. In practically all cases the warm samples received the lower score. The average score of the 36 warm samples was 18.8, while that for the cool samples was 20.0. The analysis of variance revealed this difference was highly significant.

Frequency of Specific Defects and Possible Reasons

Feed. As shown previously, the feed flavor was the one detected in these samples with greatest frequency. Of all the observations, 36.2 percent revealed this defect either alone or in combination. Roadhouse and Henderson (1935) have reported that in some years the milk samples at the California State Fair have included as many as 23.1 percent which were criticised for a feed flavor. While this high figure they report did tend to allay the concern over the great number of "feed" observations in this trial, the explanation of the causes was yet perplexing. It is possible the alfalfa hay fed seven hours before milking or the darso silage fed ten hours before may have been involved.

But it would seem desirable to call attention to another possible cause for the numerous samples which exhibited the feed flavor. It is conceivable that bacterial action may have produced some of the flavors that were not identified as such but were considered by the judges as "feed." The samples were cooled immediately in ice water and were held at 40° F. until they were scored. Some reassurance was felt that this procedure would prevent any flavor development due to micro-organisms, but it is not definitely established that the precaution was effective.

To check if the milk were absorbing flavors from the barn atmosphere, an incidental trial was conducted. Fresh silage was thrown from the silo; cows were placed in this silage at the foot of the chute and were milked in this atmosphere charged with silage odors. At no time did the milk so produced show any more flavor than that from the same cows milked in the yard. At times the air of the barn was surcharged with odors, but the milk at these times showed no more pronounced feed flavors than when the barn was completely open and free of smell. These observations dispelled the possibility "barny" flavors were a factor. Furthermore Gamble and Kelly (1922) had shown that milk is far less susceptible to barn odors than is often alleged.

Cowy. The cowy flavor often appeared as a forerunner of rancidity. It was not unusual for the judges to detect cowiness in a sample and thereby predict the cow involved would show "rancid" the next week. Because of the frequent rancid samples the numerous notations of "cowy" are not unexpected.

Rancid. While the number of observations of "rancid" were lower than those of "feed" or "cowy," this particular defect was the most serious one encountered, because if milk is rancid the score is materially reduced. In this trial, "slight rancid" connoted **poor** milk which was scored from 12 to 17. The higher intensity led to scores of 11 or below. "Rancid" was noted in 14.3 percent of the observations.

At the start of the trial it was realized that rancid and bitter milk was not uncommon. Roadhouse and Koestler (1929) had found that 11.0 percent of the cows used in their study produced rancid milk. Incidentally Roadhouse and Henderson (1935) have later shown that in some herds up to 23.5 percent of the milk samples were criticised for this defect, so the frequency of rancididty in this trial was not especially out of line. Nevertheless it had been anticipated that the likelihood of rancid milk from the College herd would be lower than was found to be the case.

A conspicuous feature of this trial was the fact that the cows which were the most serious offenders in the matter of rancid milk were with few exceptions members of the so-called Nancy family. Some pertinent data about the cows in this family are given in Table IX.

Among all the cows in this trial, the greatest frequency of rancidity occurred with F. F. L. Nancy. R. F. L. Nancy ranked third in this respect; S. F. Nancy was fourth, and R. Nancy eleventh. These four cows were responsible for 34.9 percent of all the observations of "rancid" in the trial, yet they produced only 20.1 percent of the samples. It is due almost entirely to the frequent occurrence of rancidity that the milk of these cows was scored unusually low. Among the entire group of 32 cows, R. F. L. Nancy ranked 28th in respect to the average score of her milk; F. F. L. Nancy ranked 31st; S. F. Nancy 29th; and R. Nancy 15th.

TABLE IX. "Rancid" observations and fat percentages of members of the Nancy family.

Cow	Relationship	Percentage of observations of "rancid"	Average score	Percentage of fat
Nancy Family				
R. F. L. Nancy	Dam of F. F. L. Nancy	26.7	18.7	7.1
F. F. L. Nancy	Daughter of R. F. L. Nancy	34.5	17.2	6.0
S. F. Nancy	Her dam was 100 percent the same blood as F. F. L. Nancy	18.3	18.6	5.4
R. Nancy	Daughter of S. F. Nancy	8.1	20.1	5.7
All 32 cows		14.3	19.6	5.4

There is an additional feature of this trial which may provoke some question in respect to the frequency with which the rancid defect was noted. Several references in the literature lead to a conclusion rancidity will not develop at a temperature as low as 40° F. Davies (1931), for instance, says lipase activity is suppressed at 0° to 3° C. (32° to 37° F.). In this trial samples held at 40° F. for 15 hours were far more intensely rancid than duplicates held at 70°. Sharp and de Tomassi's (1932) results are in agreement with those of this trial and Palmer (1922) has stated that the lipolytic hydrolysis occurred "even at fairly low temperatures."

The fat percentage of the cow's milk seems to be related to the frequency of rancidity. As shown in Table IX, the "rancid" cows were "high testers." One would scarcely postulate that the tendency to produce milk which will develop rancidity is an inherited character. Nevertheless, there is a distinct effect of inheritance in determining the percentage of fat in milk, and it is established here that there is a correlation between the percentage of "rancid" observations on a cow's milk and her percentage of fat. The correlation coefficient is 0.403 ± 0.076, which, being more than five times its standard error, is considered significant. The regression coefficient is 0.489.

For Figure 1 the percentage of each cow's observations which showed "rancid" was plotted against her fat percentage, giving the regression curve that is shown. The equation for this curve is of the form $Y = \bar{y} + b(x - \bar{x})$. Only 29 cows are included in this phase of the study. Three cows had fewer than 10 samples each and they are omitted.

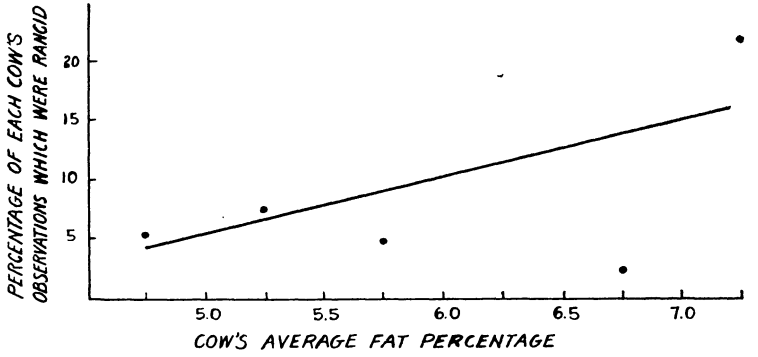

Fig. 1. Regression curve for each cow's percentage of rancid observations on her fat percentage.

The number of cows in each class being quite small, the observed percentages of rancidity are irregular. The curve fails to fit these observed values as closely as would be desired for a clear demonstration of relationship, but the upward trend in rancidity with the higher fat percentages is nevertheless unmistakable.

Stale. The flavor designated as stale ranked fourth in frequency. Unfortunately there is a possibility the judges were confused in some instances in differentiating between the stale and oxidized flavors. However, the belief that such confusion did not exist is supported by one important fact: The stale flavor was far more conspicuous in summer than in winter, and this is contrary to the seasonal trend in the oxidized flavor. During

the period of this trial, milk from the herd was pasteurized, bottled and distributed to consumers. In winter this pasteurized milk often exhibited the typical oxidized flavor, but the stale flavor of the raw samples subsided. It appears, therefore, that the judges distinguished between the two flavors.

Salt. In 8.91 percent of the observations the salt taste was noted, which figure appears in line with results of other workers. In this trial, the frequency of the criticisms of saltiness is due largely to Tilda Belle. This cow was responsible for 52.8 percent of all such criticisms, though the observations on her milk comprised only 7.1 percent of all the obserations. Forty-eight percent of this cow's samples were salty.

Flat. The flat flavor was noted in 7.17 percent of the observations. It exhibited no particularly irregular behavior worthy of comment.

Sweet. The sweet taste was detected in 2.57 percent of the observations, but, like the previous defect, it was of little consequence.

Bitter. With the mention of the bitter defect there have been included all those defects which were noted in more than 1.0 percent of the observations. Bitter was noted in 2.12 percent. It is of interest this defect was seldom noted except in combination with rancid. Anderson (1934) says that cows seldom produce bitter milk and then generally in advanced lactation. It is likely that such milk is evidence of quite severe udder disturbances to the point where distinct secretory abnormalities arise. As a matter of fact, Kieferle (1933) has reported that mastitis produces extreme changes in the nitrogen-bearing components of milk. Such changes would conceivably give rise to the bitter flavor.

RESULTS IN RESPECT TO THE SPECIFIC FACTORS

INFLUENCE OF INDIVIDUALITY ON MILK FLAVOR

Eckles and Shaw (1913-b) were among the first workers to indicate that cows vary in respect to the flavor of their milk. Their observations related to the bitter and rancid flavor. Palmer (1922) dwelt at greater length on this question in connection with his work establishing the role of lipase in producing this flavor in the milk of some cows. The same year Gamble and Kelly (1922) took occasion to mention this variation among cows in respect to abnormal flavors in general. Later Roadhouse, Regan and Meade (1926) made similar observations; and, still more recently, Olson, Totman and Wallis (1936) remarked that the individuality of the cow was more significant in determining milk flavor than was the use of tankage in the ration. In the last few years several workers have mentioned this factor in connection with the oxidized flavor in milk. Among these workers are Guthrie and Bruckner (1933), Brown, Thuston and Dustman (1937), and Dahle and Palmer (1937).

In the trial reported here, nearly all analyses of the different phases of the data have shown conspicuously that cows do vary in the flavor of their milk. The cow, Bluefeather, received the highest average score of all cows. The average score on all her samples was 21.2. At the other extreme was the cow, Trixie, with an average score of only 17.1.

Some results showing the particular defects in the milk of individual cows are given in Table X. It appeared inadvisable to present here a table extensive enough to include all cows, so only a third of the cows are incuded. All the 32 cows were listed in order of their average scores and every third cow was selected for inclusion in Table X. The exclusion of the other cows does not militate against the value of the results.

TABLE X. Percentage of all observations on each cow that were criticised as indicated.*

Cow	Average score	FLAVOR DEFECT				
		None	Feed	Rancid	Stale	Salt
Bluefeather	21.2	27.3	21.8	1.8	8.2	0.9
S. S. Jolly	20.7	1.1	54.7	5.3	12.6	1.1
S. S. Goo Goo	20.5	5.3	65.8	9.2	11.8	11.8
S. S. Girl	20.4	8.2	37.6	3.8	18.5	3.3
Marianna	20.0	7.6	38.1	15.5	7.6	5.7
Okaileen	19.7	4.7	44.9	15.9	9.3	0.0
Blanchia	19.5	14.3	26.6	22.0	6.9	1.8
Red Flag	18.9	2.4	27.5	28.2	11.5	2.4
R. F. L. Nancy	18.7	9.8	16.5	27.5	9.0	10.2
F. F. L. Nancy	17.2	5.4	19.2	44.5	8.0	4.0
All 32 cows	19.6	7.9	36.2	14.3	13.7	8.9

*See text, page 30.

Furthermore, it was inadvisable to include figures on all the flavor defects. Figures for "feed" are included because this was the most frequent defect. Those for "cowy" are omitted because this defect did not vary greatly among the individual cows, while those for "rancid," "stale" and "salt" were significant in determining the scores of the milk.

In Table X the figures for each cow may be compared with those for other cows and with those in the bottom row for all cows. Bluefeather's milk was scored high because in 27.3 percent of the observations on her milk no defect was noted. The average of all cows shows only 7.9 percent of the observations which indicated no defect. Bluefeather was extremely low in her frequency of rancidity. The milk of S. S. Jolly scored high, but this was not due to a high proportion of her samples in which no defect was noted. She showed the feed flavor with great frequency; however, this was not a serious defect and does not reduce the scores materially.

The only consistent trend in frequency of any of the defects is "rancid." As the cows decline in score the percentage of "rancid" rises rapidly.

Variation Between Lactations of the Same Cow

Although considerable attention has been given to the fact that individual cows vary greatly in the flavor of their milk, another aspect of the question is important. The flavor of milk of a given cow may vary greatly from one lactation to another. Table XI illustrates this fact. The three cows shown in the table are not selected as typical of all the cows; they are used to emphasize the point. In most instances the milk of the cows did not vary greatly between lactations.

Table XI shows the distinct variation in the different lactations. With Tilda Belle there was a decline in each lactation. As mentioned previously, this cow showed the highest frequency of saltiness. In her first lactation 27.0 percent of the samples were so criticised. In her second lactation the figure was 56.5 percent, and in her third it was 61.9 percent. It is entirely possible that the advancing age of the cow aggravated an udder disorder.

F. F. L. Nancy declined from her first to her second lactation, but recovered in her third. Throughout the trial the rancidity of this cow's milk was conspicuous, but was worst in her second lactation when 64.1 per-

cent of her samples were so criticised. The explanation of this cow's behavior is not apparent. She exhibited no physiological disorder and conceived on first service early in her second lactation. Koestler, Roadhouse and Loertscher (1928), however, emphasize that the rancid milk may arise from some disorder that cannot be detected on cursory examination.

The cause for Blanchia's behavior is entirely apparent. She aborted between her lactations. Koestler, Roadhouse and Loertscher say lipase may increase in the milk under certain conditions because of modified hormone activity. They also quote Grutter and Wigger to the effect that rancidity accompanies nymphomania and that such a condition is also accompanied by higher butterfat tests.

TABLE XI. Scores of the milk from different lactations of the same cows.

Cow	Lactation	Cow's age at start Yr.:Mo.	Number of samples	Average score
Tilda Belle	First	12:2	33	20.3
	Second	13:0	49	19.0
	Third	14:0	33	18.4
	All		115	19.2
F. F. L. Nancy	First	5:11	30	18.3
	Second	6:9	41	15.4
	Third	7:9	37	18.3
	All		108	17.2
Blanchia	First	11:8	41	21.8
	Second	12:7	82	18.4
	All		123	19.5

During Blanchia's second lactation she developed cystic ovaries, and, although she never appeared as a nymphomaniac, she was bred repeatedly and failed to conceive until 16 months after the abortion. During this second lactation, 35.0 percent of her samples were rancid; in her first lactation only 2.5 percent received this criticism.

INFLUENCE OF AGE OF COW

From the data just given relative to the change in the flavor score of milk of different lactations, some indication was obtained that the scores might decline with advancing age. This was particularly shown in the case of Tilda Belle. Blanchia, too, produced lower-scoring milk during her second lactation in the trial.

The question arises if the tendency for the score of the milk to decline with advancing age is at all characteristic of cows in general. The data indicate there is such a slight tendency.

Attention is called to the fact that no experimental work has been reported in which this factor of age has had direct consideration. However, as has been mentioned, a diseased udder leads to the production of salty milk, and certain physiological disturbances contribute to the development of rancidity. Also, many workers have adduced evidence that ration deficiencies such as a lack of certain vitamins or of mineral components may induce undesirable flavor changes. It is readily conceivable that all these circumstances may be aggravated in older cows and thus result in more defective milk.

The usual practice in dairy records is to base a cow's age on her age at the time a lactation starts. Such was not the exact procedure here. The age at which a lactation was started was not considered. Until a cow attained her third birth anniversary, for instance, she was considered a two-year-old. Thereafter she was a three-year-old until her next birth anniversary. Thus in a single lactation some samples were placed in two different age classes. In a few cases a cow with an unusually long lactation had samples from that lactation placed in three successive age classes.

In presenting the data on the influence of age it seemed advisable to depart somewhat from the type of statistical treatment employed in earlier sections here and resort to the computation of statistics from which a regression curve might be plotted.

It was found that the correlation coefficient between flavor score and age of the cow was −0.081 with a standard error of 0.025. This correlation is so low that it possesses no practical importance. Yet the correlation coefficient is more than three times its standard error and is accepted as significant. The regression coefficient was only −0.072.

Figure 2 gives the regression curve for the flavor score on the age of the cow together with the observed values in the different age classes. There were 1,641 samples included in this phase of the study, yet the observed values do not coincide satisfactorily with the curve.

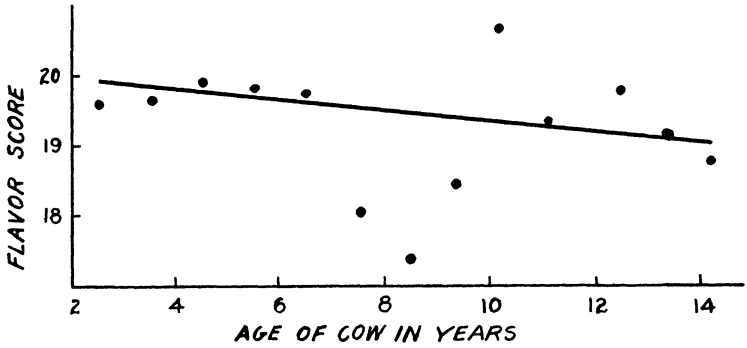

Fig. 2. Regression curve for flavor score on age of cow.

INFLUENCE OF STAGE OF LACTATION

Groups of Cows Based on Frequency of Rancidity

The results in respect to the influence of stage of lactation on milk flavor show an extreme variation among cows in their response to this factor. There recurs here the necessity that the cows be divided into groups based on the frequency with which rancidity was detected in their milk. The cows of the group whose milk showed no rancidity at any time maintained essentially the same level of flavor score throughout the lactation; those in whose milk the rancidity appeared showed a regular decline in scores as lactation progressed. Furthermore, this decline was greatest with those cows whose milk was characterized by the greatest frequency of rancid criticisms.

Four groups were established. The non-rancid group embraced six cows that never showed rancidity at any time during the trial. The slight-

rancid group included 16 cows in whose milk "rancid" was noted in fewer than 10 percent of the observations. The moderate-rancid group had 7 cows that ranged from 10 to 20 percent in frequency of rancidity, and the high-rancid group had 3 cows which exceeded 20 percent.

Statistics for the Groups

With the realization that the four groups of cows behaved differently in respect to the stage of lactation, it appeared advisable to maintain the grouping for the statistical studies that were made. Table XII gives some of the statistics that were computed, though the figures for the slight- and modern-rancid groups are omitted because they make no contribution to the interpretation of the results. The behavior of these two imtermediate groups was nearly identical to that of the entire 32 cows.

In handling the data relative to the influence of the stage of lactation, it was arbitrarily decided to use no samples from cows that were beyond the 64th week, or 16th month of lactation. Too few samples were obtained beyond this stage to justify their use. Three cows during the trial continued in milk beyond this stage, but their samples during this extra period were omitted from this study. The elimination of these samples reduced the number to 1,611. This also reduced the number of observations down to 6,273.

Explanation must also be made of the term "lactation-month." The samples were collected each week and the data are available by "lactation-weeks." However, they are too extensive to be included here. Thev were combined into "lactation-months" wherein are included four successive weeks. This usage of the term, lactation month, is preserved throughout.

TABLE XII. Statistics computed for the groups to show the relation of stage of lactation and criticisms of "rancid."

	Non-rancid group	High-rancid group	Entire 32 cows
Number of cows	6	3	32
Number of samples	91	264	1611
Average score	21	18.3	19.6
Average stage of lactation-month	3.86	6.48	5.94
Correlation coefficient	−0.004	−0.313	−0.236
Standard error±	0.075	0.038	0.016
Reg. coef. of scores on lactation	−0.002	−0.257	−0.172

From Table XII it may be noted that there are contrasts in the average stage of lactation of the groups. The non-rancid group embraced cows that furnished samples only during relatively early lactation. This is a weakness in these data. The results do not indicate whether the cows of this group would have exhibited the rancid defect had they been in more advanced lactation. All this group were young cows. They had freshened with first calves only in the later months of the trial and had not progressed into advanced lactation when the trial was terminated.

It is important to remark that these six cows were all daughters of one bull. His daughters were typically "low testers." Previous evidence has shown that such cows are less likely to exhibit rancidity, so it is possible these cows would have continued free from the defect. One other piece of evidence to be presented indicates also that these cows behaved

in a manner that supports the conjecture they would not have exhibited the rancidity even if observations had been continued into their later lactation.

The other group of cows was in more advanced lactation, the average stage being 6.48 months for the three cows of the high-rancid group.

The correlation coefficients for the scores and stage of lactation are of interest. The coefficient is of no significance for the non-rancid group. While that for the high-rancid group is low, it is nevertheless significant, being -0.313, or about 10 times its standard error.

Figure 3 gives the regression curves for the entire group of cows and for the two extreme groups. This figure reveals the influence of stage of lactation on milk flavor. There is no influence in the case of cows whose milk is not disposed to the development of rancidity. It is pronounced, however, with other cows.

Observed Scores in Successive Months

It was hardly feasible in Figure 3 to plot the observed scores in successive months for the different groups of cows; hence, these scores are given in Table XIII. The scores as given in the table are somewhat irregular even for a given group of cows. This was especially true for the months when the samples were less frequent. There is obviously no definite trend indicated in the scores for the non-rancid group. For the high-rancid group the trend is definitely downward to the low point of 14.8 in the eleventh month. Thereafter the scores are irregular but the numbers of samples are lower and little effect is exerted on the position of the regression curve. However, the fact is not to be ignored completely that the highest average score on the milk of these cows occurred in the 21st month of lactation. This matter will be mentioned subsequently.

Frequency of Flavor Defects

The frequency with which the different flavor defects were noted in successive stages of the lactation period is a matter of importance.

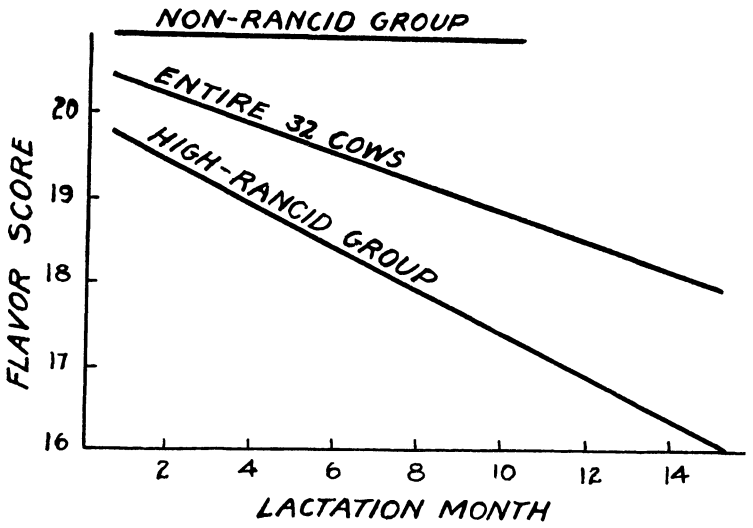

Fig. 3. Regression curves for flavor score on lactation month.

TABLE XIII. Observed scores in successive months during one lactation, by non-rancid and high-rancid groups.

	NON-RANCID GROUP		HIGH-RANCID GROUP		ENTIRE 32 COWS	
Lactation month	Number of samples	Average score	Number of samples	Average score	Number of samples	Average score
1	9	20.3	16	20.4	132	20.3
2	16	20.8	20	20.1	168	20.2
3	16	21.0	20	20.5	154	20.4
4	14	21.2	23	19.9	145	20.4
5	10	20.9	23	19.6	132	19.9
6	8	21.8	24	17.4	137	19.6
7	6	21.0	24	17.8	148	19.9
8	4	21.5	24	16.6	136	19.7
9	3	21.5	24	16.5	120	18.9
10	2	19.5	21	16.5	93	18.2
11	3	19.5	13	14.8	60	17.8
12	0		5	18.9	39	19.0
13	0		8	19.4	39	19.2
14	0		7	18.8	39	18.4
15	0		8	18.8	42	18.7
16	0		4	21.3	27	18.2
All	91	21.0	264	18.3	1611	19.6

The greatest number of samples in which no defect was noted occurred in early lactation. The number of these excellent samples declined as lactation progressed. The feed flavor is most frequent in early lactation and gradually subsides to about the twelfth month. With the cows that continued beyond this stage there was a tendency for the feed flavors to increase in frequency. These trends in frequency of feed flavors were directly the reverse of those for rancidity. In the period embracing the first to fourth months of lactation, 7.2 percent of all observations revealed rancidity; from the fifth to eighth months the figure was 12.4 percent; and from the ninth to twelfth months it was up to 27.9 percent. But in the thirteenth to sixteenth months it declined to 21.0 percent. This reduced rancidity is reflected in the higher scores of the high-rancid cows as shown in Table XIII.

The earliest investigations involving the rancid flavor all showed its greater frequency in later stages of lactation. It is worthy of emphasis, however, that the defect may appear in the first few days after freshening. Blanchia, in her second lactation, affords an example of this. During the first four months, 71.6 percent of the observations on her milk revealed rancidity; in the thirteenth to sixteenth months this figure was only 18.4 percent.

The frequency of the stale flavor was not affected by the stage of lactation. Neither was the oxidized flavor found to vary. This is in agreement with most workers on the oxidized flavor, though Gondos (1934) from somewhat fragmentary evidence concluded it was associated with the stage of lactation.

There was a distinct tendency for saltiness to increase as lactation advanced. In the first four months of lactation, 5.4 percent of the observations revealed "salt"; this increased to 17.8 percent in the thirteenth

to sixteenth months. Trunz (1901) was the first worker to show this effect of lactation on the saltiness of milk. Sharp and Struble (1935) observed that the chloride content increased perceptibly in the last 10 percent of the lactation period.

INFLUENCE OF STAGE OF GESTATION

Ii connection with the possible influence upon milk flavor of such a factor as the stage of lactation, interest also arises as to the role which may be played by changes in the stage of gestation. It is readily appreciated that the influence of one factor is not easily separated from that of the other. Throughout the major part of a normal lactation period of a cow the stage of gestation must exist as a concurrent factor.

Eckles and Shaw (1913-a) in their early work showed that pregnancy was a significant factor in producing the changes in milk composition that occurred with advance in lactation. Farrow cows failed to manifest the usual changes. Palmer and Eckles (1917-a, 1917-b) showed that the stage of gestation exerts its effects by hastening the termination of lactation. Sharp and Struble (1935), however, concluded from their data that "gestation exerts a negating tendency on the increase in chlorides due to progressing lactation."

In the trial reported here, it was found that the milk flavor declined somewhat more rapidly with advancing gestation than with advancing lactation. All the statistics that were computed establish such to be the case. However, the behavior in respect to gestation is so similar to that for lactation that it is scarcely advisable to repeat here all the tabulations in such detail as were just presented in respect to the changes in stage of lactation.

The correlation coefficient between the flavor score and gestation month for the entire group of cows was -0.269 ± 0.029 as compared to the companion figure of -0.236 ± 0.016 for the stage of lactation as given in Table XII. The regression coefficient of scores on stage of gestation was -0.373, which is much higher than the -0.172 for the stage of lactation.

The partial correlation coefficients also show that the stage of gestation exerts a slightly more marked influence on milk flavor than does the stage of lactation. The partial correlation of the flavor scores and stage of lactation, when gestation is held constant, was -0.152 ± 0.031. That for the scores and stage of gestation when lactation is held constant was -0.193 ± 0.030.

As was true in respect to the influence of the stage of lactation, the high rancid group of cows showed the distinct effect of stage of gestation, while the non-rancid cows were not influenced by this factor.

INFLUENCE OF DAILY MILK YIELD

In consequence of the results just given wherein it is shown that the flavor scores decline with advancing lactation and gestation, the question arises: Will these scores also decline with decreasing milk yields? Such was found to be the case. Sharp and Struble (1935) determined that the content of chlorides rose markedly when the daily yield of Jersey cows dropped below 15 pounds.

The results of this trial established the correlation coefficient of 0.327 ± 0.022 between the flavor scores and daily milk yield. The regression coefficient was 0.128.

The regression curve is shown in Figure 4. It reveals the decline in scores as the daily yields decreased. The observed values plotted on the figure lead to the idea that they fit the curve quite poorly. However, the frequencies in the two higher and two lower yield classes were small. Of

the total 1,641 observations, only four occurred in the class above 40 pounds of milk and only seven below five pounds. Ninety-two percent of all the frequencies are in the five classes from 10 to 35 pounds. The observed values in these classes which predominate in number determine the position of the curve and they fit quite satisfactorily.

A suspicion also may arise from cursory examination of Figure 4 that the data are of a curvilinear nature. A test of this, however, established that such was not the case. Here again the low frequencies in the extreme classes fail to modify the data.

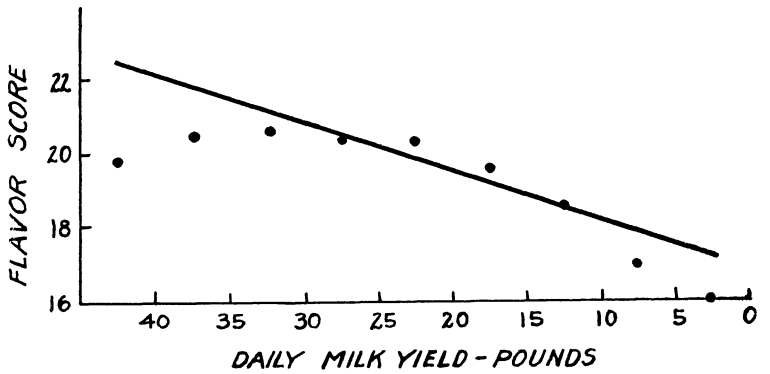

Fig. 4. Regression curve of flavor score on daily milk yield.

INFLUENCE OF PERCENTAGE OF FAT

Every indication from this trial leads to the conclusion that with increases in the percentage of fat in the milk the flavor scores decline. Though this is contrary to the general belief on the question, such experimental data as have been reported by other workers support the conclusion.

In the extensive work by Overman, Sanmann and Wright (1929) it was found that with an increase in the fat of milk the ash content increased and the sugar decreased. These workers made no observations on the flavor of the milk, but there is every reason to surmise that such changes in composition would cause a decline in flavor. Kelly (1934) has stated, "The butterfat content is not the predominating factor affecting the palability of milk." He had found that judges were often unable to distinguish by taste between high-test and low-test milk. Tracy and Ruehe (1931) found that an increased fat percentage in milk rendered it more susceptible to the development of the oxidized flavor. Their results were confirmed by Roland and Trebler (1937), though Garrett, Bender and Tucker (1937) did not find such an influence.

In this trial the correlation coefficient between the scores and fat tests was -0.236 ± 0.024. As in other cases, this correlation is not high but it is significant being nearly 10 times its standard error.

The regression curve is shown in Figure 5. It establishes quite definitely that the flavor scores decline as the fat tests increase. Here again the observed values in the extreme classes are not of importance because of the small numbers in these classes. Out of the total 1,520 samples in

this phase of the study, only eight tested below 3.0 percent of fat. Three tested from 9.00 to 10.0 percent, and only two above 10.0 percent. Obviously with these Jersey cows most of the samples tested from 4.0 to 7.0 percent. Hence, the observed values in these classes largely determine the position of the regression curve.

In connection with this phase of the study the question arose whether there was any difference in the behavior of characteristic "low testers" and "high testers." The average fat percentage of all samples from each cow was computed. It was found that eight cows tested below 5.00 percent while five tested above 6.00 percent. The former group averaged 4.8 percent on all their samples, and the average flavor score of all these samples was 20.3. The average fat percentage of all samples from the "high testers" was 6.3 percent, but the average flavor score was only 18.2. There seems to be no question that a "low testing" cow produces milk of flavor superior to that of a "high tester."

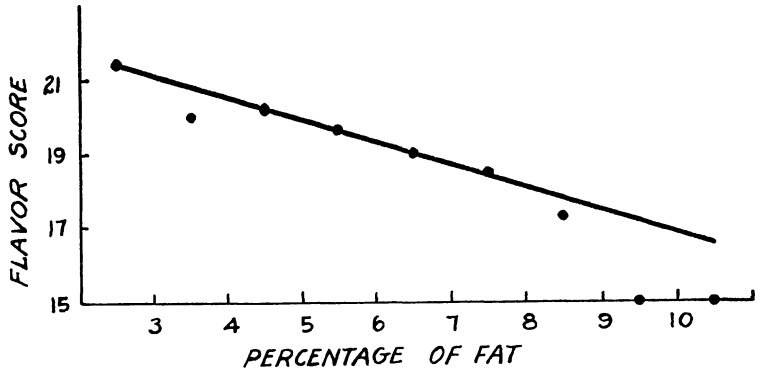

Fig. 5. Regression curve of flavor score on fat percentage.

Figure 1, which was presented earlier, established the fact that the cows in whose milk the rancidity was noted with greatest frequency were those with the higher fat tests. This explains the lower flavor scores on the milk of such cows.

INFLUENCE OF SOLIDS-NOT-FAT

It is generally recognized that the percentages of fat and of solids-not-fat in cow's milk are correlated. In this trial the correlation coefficient was 0.253 ± 0.024. From the facts just established to the effect that the fat percentage modifies the flavor score it is logical to conjecture that the percentage of non-fatty solids would exert a similar influence. Results here show that such is the case. However, the flavor is not so closely correlated with the percentage of solids-not-fat as with the fat percentage.

The correlation coefficient between the flavor score and the percentage of solids-not-fat was -0.167 ± 0.025. It is to be recalled that this correlation is lower than that for the fat percentage. The regression coefficient for scores on solids-not-fat was -0.381.

There seems to be a plausible explanation of the fact that the high fat test exerts a more depressing effect on flavor than does the high percentage of solids-not-fat. Antognistic forces may be in operation. Overman, Sanmann and Wright (1929) have shown that lactose increases with the percentage of solids-not-fat. This would tend to improve the flavor and to combat the depressing influence of the high fat percentage.

Figure 6 gives the regression curve for the flavor score on percentage of solids-not-fat. The curve fits the observed values quite well, it being necessary to recognize again that frequencies are low in those classes wherein the solids-not-fat exceeds 10.0 percent.

INFLUENCE OF SEASON OF YEAR

The results of this trial show that the season of year exerts a marked effect on milk flavor. In Palmer's (1922) earliest work on the rancid flavor he noted that that flavor was most likely to occur in the fall of the year while pastures were yet suffering from the usual summer drouth. Hileman and Courtney (1935) have also noted the seasonal effects upon the lipase in milk. Their work was done in New York where they noted the minimum content of lipase in early summer, during June and July, and the maximum in early winter, or December. Anderson, Hardenberg and Wilson (1936) ascribe to the carotene in the ration a chief role in determining whether milk will exhibit the rancid flavor.

The various workers who have studied the oxidized flavor have shown the greater frequency of the defect in winter. Most workers credit Kende (1931) as the first to make this observation. Some contradictory evidence, however, has been advanced since that date in the attempts to explain the cause underlying the more frequent manifestation of the defect in winter. Kende (1931) in his first publication noted that the feeding regime of winter was a factor in the case, and Majer (1931) had studied the role of certain feeds. Richter (1931) expressed a belief the "tainted" milk was an accompaniment of reduced content of CaO in relation to P_2O_5 in the milk and that the ratio could be improved by feeding chalk and calcium chloride. Even in Kende's later publication (1932) when he attached considerable significance to "oleinase" he did not abandon the idea of the seasonal feed influence.

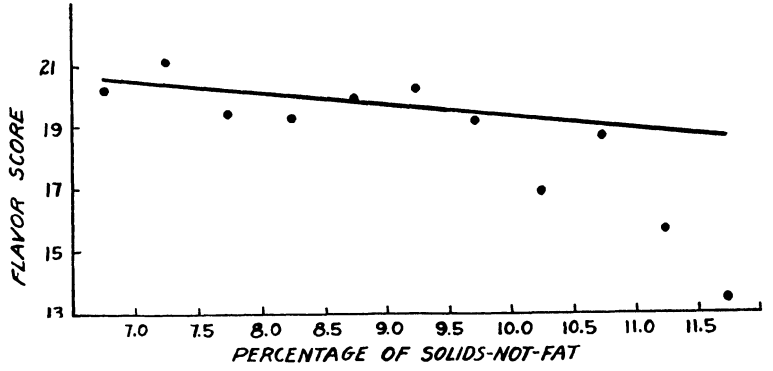

Fig. 6. Regression curve of flavor score on percentage of solids-not-fat.

Kertesz (1932) also admitted the role of feeds with their probable seasonal deficiencies in giving rise to the flavor, but he emphasized especially the micro-biological aspects of the problem. Csiszar (1932) reported his belief that feed was not a factor but that the manifestation of the defect was related instead to metabolic changes in the cow. He was confirmed in this by Chilson (1935), who concluded there was a purely seasonal effect entirely apart from the feeding.

Guthrie and Brueckner (1933) were convinced from their study that feed was not a sole factor in the development of the flavor, and they acknowledged the seasonal influence. Thurston (1935) supported their findings. Prewitt and Parfitt (1935) were unable to establish that the choice of concentrated feeds in the ration was involved in any changing susceptibility of the milk. They thus confirmed the results of Frazier (1928) that the flavor developed no more rapidly when the cows were "fed heavy rations of cottonseed and linseed oil cake." Then Anderson, Hardenberg and Wilson (1936) proposed that the seasonal effect was due to variation in the carotene content of the feeds, while about the same time Brown, Thurston and Dustman (1937) and Sharp, Trout and Guthrie (1936) had reached the conclusion that the lack of vitamin C in the ration contributed to the development of the flavor.

Garret, Bender and Tucker (1937) found a high correlation between the flavor score of milk of individual cows and the vitamin C content. Dahle and Palmer (1937) were unable to establish such an intimate relationship between the vitamin C content of milk and the development of the oxidized flavor, though they observed that the addition of the vitamin to milk prevented the flavor.

Whitnah, Martin and Beck (1937) observed no relation between the vitamin C content and the oxidized flavor in milk of individual cows of a given breed. They reported that feeding dehydrated young oats reduced the flavor and that the use of a carotene concentrate eliminated it. Webb and Hileman (1937) indicated that "seven or eight" factors might be involved in the development of the oxidized flavor, and Rogers (1937) claimed that all these possible factors exert their effect through their influence on the oxidation-reduction potential of the milk. Most recently Anderson, Dowd and Stuewer (1937) conclude that the flavor is associated with the higher titratable acidity which they report is characteristic of winter milk.

Leach (1934) has said a salty flavor in milk may accompany a bodily depletion of calcium and phosphorus, a condition that is subject to a seasonal influence. Turner (1936) has called attention to the fact that summer drouth conditions may exert pronounced effect on milk flavor through the enormous increase in chlorides that may occur.

The marked effects of the season of year as obtained in this trial are shown in Figure 7 along with the observed values in the different calendar months. The distinct curvilinear relationship is apparent. The curve in Figure 7 is of the form $Y=\bar{y}+b\ (x-\bar{x})\ +\ C\ (x^2-\bar{x^2})$.

It is to be noted that the highest average score occurred in the month of April and the lowest in October.

Frequency of Defects in Different Seasons

In view of the fact that the flavor showed a distinct seasonal variation, some interest arose in the question whether changes in the frequency of different defects were responsible. Accordingly, the defects were tabulated by months. The results of this tabulation are condensed into seasons for presentation in Table XIV. This table shows the percentage of the total observations each season in which the particular defect was noted. There is a predominance in spring of the observations in which no defect was noted. This explains in part the high average scores at this season. The feed flavors are slightly more conspicuous in spring and winter, though the tendency is not distinct. The cowy flavor is distributed quite evenly throughout the year.

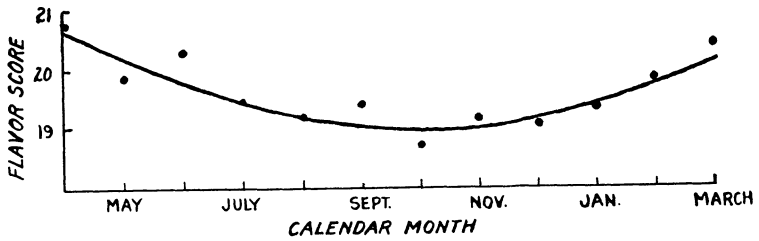

Fig. 7. Regression curve of flavor score on calendar month.

A notable fact from Table XIV is the infrequency of "rancid" in the spring and its predominance in the fall. In the latter season, this defect was noted in 22.2 percent of the observations that were made. It has been mentioned previously that the stale criticisms predominated in summer. They were least frequent in the winter.

The saltiness did not show any pronounced seasonal trend. It has been indicated that severe drouth conditions or a deficiency of calcium and phosphorus may contribute to the frequency of salty milk. This trial embraced the severe drouth of 1934, yet no material change is evident in the occurrence of salt in the milk.

The bitter flavor showed the same trend as rancid, for when observed it was usually in combination with rancid. The oxidized flavor was noted only 24 times in this entire trial. Consequently, the percentages showing its frequency are of doubtful significance.

Fat Percentage and Season of Year in Relation to Flavor

It is generally recognized that the percentage of fat in milk is subject to a seasonal influence, being higher in winter and lower in summer. The preceding material having shown that milk flavor also varies with the season gives rise to a conjecture that there may be an interrelationship among these factors. Figure 8 indicates such is the case.

TABLE XIV. Percentage of the total observations each season that were criticised as indicated.

	Spring	Summer	Fall	Winter	All
None	12.7	4.9	6.1	7.3	7.9
Feed	42.4	29.1	27.8	43.1	36.2
Cowy	21.1	21.1	26.6	19.0	21.9
Rancid	6.3	9.4	22.2	17.0	14.3
Stale	11.1	27.6	11.4	8.8	13.7
Salt	6.3	9.4	9.7	10.0	8.9
Bitter	0.2	1.3	4.9	1.8	2.1
Oxidized	0.6	0.2	0.3	0.4	0.4

In Figure 8 the curve for the fat percentage was fitter according to the equation $Y=\bar{y}+b\ (x-\bar{x})+c\ (x^2-x^2)$. The curve for the flavor score is a cubic parabola of the form $Y=18.33+1.22x+0.22^2+0.01x^2$. It was fitted by W. D. Baten of the department of mathematics, Michigan State College.

It is noted from this figure that flavor scores rise in winter and early spring as the tests decline. In the fall, when the tests are about midway up in their seasonal ascent, the flavor scores decline to the lowest point.

Apparently the condition in spring which gives rise to the declining tests also results in the production of milk of a superior flavor. It is not inconceivable that cows may exhibit some inherent capacity in spring to attain a high nutritive state and this improved state induces both the reduced fat percentage and the superior flavor of the milk. It would be further conjectured that an opposite situation prevails in the fall.

INFLUENCE OF ATMOSPHERIC TEMPERATURE

The results of this trial indicate that atmospheric temperature is not a factor that affects the flavor of milk. Of course, the knowledge that fat tests decline in hot weather and that the scores rise with declining tests might lead to a conjecture that the scores would be higher during hot weather. It appears, however, certain antognostic influences are in operation. Rancidity becomes more conspicuous in August; and the stale flavor was especially notable in July and August. Both defects are so serious as to obscure any slight improvement in flavor that might accompany the characteristically low tests in these hot seasons.

In this trial the mean temperature each Tuesday was plotted against the scores of the samples collected that day. It was found that no correlation exists between these variables.

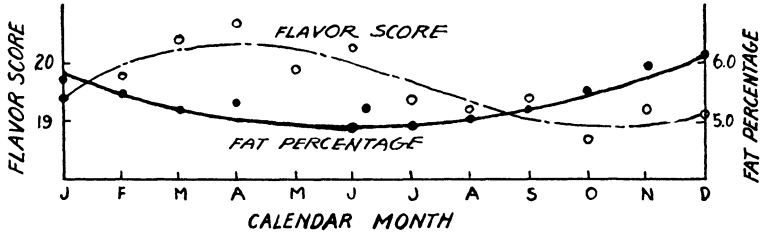

Fig. 8. Regression curves of flavor score and fat percentage on calendar month.

INFLUENCE OF RELATIVE HUMIDITY

In contemplating the possible factors that might have an influence on milk flavor, some attention was directed to the relative humidity. The results establish quite definitely that this is not a factor. No correlation was found between the mean of the humidity readings each Tuesday and the flavor scores of samples collected that day.

RESULTS IN RESPECT TO THE VALIDITY OF THE SCORES

In conducting the trial reported here, some apprehension was felt from the first that the scoring of the samples might not be performed with sufficient accuracy to insure dependability in the results. It was recognized the scoring was a fundamental feature of the trial. Platt (1931) emphasizes the necessity for caution in such a trial. He states, "In experimental work on foods, accurate scoring is often the only means of measuring the results obtained, so that all the conclusions depend upon the reliability of the scoring."

In general, the procedure for the scoring in this trial conformed closely to that commonly accepted by dairymen. However, it had not been definitely established that such procedures are entirely trustworthy; there remained the possibility that the confidence some persons evince in results of milk scoring was not entirely justified, or conversely, that the skepticism some persons express was well founded.

A quite detailed analysis of the scoring results from this trial was made in the effort to ascertain if the "experimental technic" employed was dependable. In general the analysis reveals that such was the case.

COMPARISON OF AVERAGE SCORES BY INDIVIDUAL JUDGES

A difficulty which arose in the effort to ascertain the validity of the scoring was the selection of a suitable criterion with which comparisons could be made. One possibility in this direction is to compare the average scores by each judge. Roadhouse and Henderson (1937) used this method of comparison in a study they recently reported, as did also Olson, Totman and Wallis (1936).

Table XV gives the average scores by the different judges, together with the scores by other judges on the same samples. No judge scored all the samples that were collected. For instance, Judge 1 scored 961 samples.

TABLE XV. Each judge's average score and the average by the other judges on the same samples.

Judge	Samples scored by each judge	Judge's average score on these samples	Average score by the other judges on the same samples	Difference
1	961	19.6	19.3	+0.3
2	962	19.1	19.4	−0.3
3	957	19.8	19.4	+0.4
4	1104	19.7	19.5	+0.2
5	970	19.3	19.3	0.0
6	166	19.6	20.2	−0.6
7	1253	19.3	19.2	+0.1
All	6373	19.5	19.4	+0.1

Working with him on these samples were one or more of the other six judges, depending on which men were available when the samples were ready on each Wednesday morning. Thus the average score by any of the judges is not directly comparable with the score by any other judge as shown in the same column. The judges did not all score the same samples.

However, a comparison between the score accorded by a judge and the average score by other judges on the same samples has some validity. The average score by judge 1 on his 961 samples was 19.6. The other judges who scored these same samples accorded them the average score of 19.3. Thus judge 1 accorded scores that were higher than those by other judges on the same samples. Judge 2 was consistently more critical in his scoring than the other judges. The same situation existed with judge 6.

On the other hand judge 3 was consistently more liberal than the others, his average score being 0.4 higher. Judge 5 was in complete agreement with the other judges while judges 7 and 4 ranked next in this comparison. It will be recalled from Table V that judges 1 and 5 were experienced; judge 6 had had some experience.

ANALYSIS OF VARIANCE OF THE SCORES

Enough samples were scored in this trial and enoungh individual observations were made to permit an analysis of variance on each of several groups of data. Such analyses lend far more confidence in the results than does the comparison as presented in Table XV. In fact they show that, except in one instance, the differences in scores of the individual judges were not statistically significant.

As has been stated, an effort was made to have at least four judges score the samples that were ready each Wednesday morning; but in many cases this was not possible. At the start of the work six judges were used in the effort to unify the scoring. As shown in Table XVI, there were 156 samples scored by the two combinations of 6 judges. On several Wednesdays 5-judge combinations were used. Though the 4-judge combinations were desired, they scored fewer samples than 3 judges; and in a few instances, with 106 samples, only 2 judges were available.

Table XVI affords some logical groups of data on which separate analyses of variance can be made. It would appear logical, for instance, to investigate the scores by each judge in the 6-judge combination which scored 80 samples. The scores by these judges and those of certain other combinations are given in Table XVII. It scarcely seemed advisable to present the scores of the judges in all the combinations shown in Table XVI, so it was arbitrarily decided to restrict the analyses to those combinations of judges which scored more than 100 samples each. They alone are included in Table XVII.

Attention to Table XVII reveals considerable difference among the judges in the 6-judge combination that worked on the 80 samples. It so happened these 80 samples were all scored during the first six weeks of the trial. They comprised the first milk samples ever scored by judges 2, 3 and 4. No surprise is felt in the fact that the scores varied.

The first few weeks of the trial embraced an entire month of August and the first part of September. The rancid defect was frequently noted during this period, for the cows were in advanced lactation. From the first, judge 2 exhibited an extremely low threshold for "rancid"; he was particularly severe in his scores on these early samples in comparison with the other judges. The frequency with which he criticized the samples as "rancid" and the severity of his scores on such samples resulted in his low average score.

On the other hand, judge 4 had established an ideal milk flavor that was extremely liberal toward "feed." Until he attained a more critical attitude toward feed flavors, his scores were high.

Thus the difference in the scores by judges 2 and 4 is not surprising. The scores of these inexperienced judges during their period of "apprenticeship" could have been omitted from the report of this trial, but such omission seemed unnecessary. It is not believed the inclusion of these scores detracts from the data presented.

TABLE XVI. Number of samples scored by different combinations of judges.

Judges who comprised the combination	Samples scored by the combination
6-judge combination	
1-2-3-4-5-6	80
1-2-3-4-5-7	76
All	156
5-judge combination	
1-2-3-4-5	115
1-2-3-4-6	10
1-2-3-4-7	56
1-2-3-5-6	24
1-2-3-5-7	51
1-2-4-5-7	21
1-3-4-5-7	5
2-3-4-5-7	32
All	314
4-judge combination	
1-2-3-4	26
1-2-3-5	5
1-2-3-7	24
1-2-4-5	8
1-2-4-7	38
1-2-5-6	15
1-2-5-7	34
1-3-4-5	11
1-3-4-6	6
1-3-4-7	42
1-3-5-7	13
1-4-5-7	58
2-3-4-5	4
2-4-5-6	23
2-4-5-7	127
3-4-5-7	26
All	460
3-judge combination	
1-2-3	4
1-2-7	24
1-3-7	33
1-4-5	7
1-4-7	73
1-5-7	102
2-3-4	9
2-4-7	22
2-5-7	76
3-4-7	203
3-5-7	30
4-5-7	22
All	605
2-judge combination	
2-3	29
2-7	29
3-4	4
3-6	3
3-7	36
5-6	5
All	106

TABLE XVII. Average score by each judge in each of the combinations being analyzed.

The combination	Judge	His average score	The combination	Judge	His average score
6 judges on 80 samples	1	20.6	4 judges on 127 samples	2	18.6
	2	19.9		4	19.0
	3	20.7		5	18.9
	4	21.1		7	18.9
	5	20.6			
	6	20.4		All	18.9
	All	20.5	3 judges on 102 samples	1	19.7
6 judges on 76 samples	1	18.7		5	19.7
	2	18.2		7	19.7
	3	18.9		All	19.7
	4	19.3	3 judges on 203 samples	3	20.5
	5	18.6		4	20.4
	7	18.6		7	20.4
	All	18.7		All	20.4
5 judges on 115 samples	1	20.5			
	2	20.5			
	3	20.7			
	4	20.3			
	5	20.1			
	All	20.4			

The first section of Table XVII presents another interesting fact. Judges 1 and 5 were the two experienced judges. On these 80 samples scored by this 6-judge combination, these two judges accorded identical average scores, namely 19.6. Such consistency is noteworthy.

In the other sections of Table XVII the differences in the scores by the individual judges are not so large as those in the first portion, though throughout there is the tendency for judge 2 to be most critical and judges 3 and 4 to be most liberal. In general, it would be said that the scoring appears to have been quite valid.

An advantage which arises out of the plan to group the data here according to the combination of judges which scored the same samples lies in the fact that such a grouping facilitates an analysis of variance study. Such an analysis can reveal if differences among the judges were significant and if the scoring were valid.

An analysis was made of the scores by each judge in each of the six combinations. The results of these six analyses are presented in Table XVIII. The particular combination under consideration in the table can be identified by remembering that the "degrees of freedom" as given is one less than the number of judges or the number of samples.

The first section of Table XVIII confirms the suspicion previously expressed that there was significant variance in the 6-judge combination that scored the 80 samples. The F value of 5.22 indicates there was less than 1 chance in 100 that variance as great as that obtained was due to

random sampling; it was highly significant. However, in no other analysis was the variance among the judges significant. They varied but little in the scores they accorded.

TABLE XVIII. Analyses of variance of scores by judges in each combination.

Variation due to:	Degrees of freedom	Sum of squares	Mean square	F value
Judges	5	52.57	10.51	5.22**
Samples	79	1654.25	20.93	10.36**
Error	395	797.26	2.02	
Total	479	2503.08		
Judges	5	46.75	9.35	1.13
Samples	75	741.27	9.88	1.20
Error	375	3088.92	8.24	
Total	455	3876.94		
Judges	4	23.25	5.81	1.59
Samples	114	2750.23	24.12	6.61**
Error	456	1662.35	3.65	
Total	574	4435.83		
Judges	3	9.88	3.29	2.04
Samples	126	5946.83	47.20	29.32**
Error	378	610.37	1.61	
Total	507	6567.08		
Judges	2	0.37	0.19	0.20
Samples	101	1901.64	18.83	20.03**
Error	202	188.96	0.94	
Total	305	2090.97		
Judges	2	1.02	0.51	0.37
Samples	202	2193.37	10.86	7.93**
Error	404	551.65	1.37	
Total	608	2746.04		

There was great variance due to "samples." This of course devolved on the variation among cows. In every instance except in the second section of Table XVIII relating to the 6-judge combination that scored the 76 samples, the scores on samples were significantly different. A reason for this exception is not apparent.

AGREEMENT BETWEEN THE TWO EXPERIENCED JUDGES

In the discussion of Table XVII, mention was made of the instance in which two experienced judges 1 and 5 accorded the exact average score of 20.6 on 80 samples. It was desired to ascertain if their scores showed a similar agreement on all the samples which both of them scored. On scoring the data it was found that there were 625 samples on which both these men worked.

The result of the study was quite astounding. The average score by judge 1 on these 625 samples was 19.49; that by judge 5 was 19.47. Of course an analysis of variance study on these scores was totally superfluous; the variance was too minute. Apparently these experienced judges exhibited an extremely high degree of proficiency.

SOME INDIVIDUAL DIFFERENCES IN THE JUDGES

Insofar as this trial afforded a great many data on the behavior of different judges in the matter of scoring milk, and since such data are not generally available in the literature, there appears some justification in presenting further certain aspects of the topic here. The data can be resolved into two phases. The first relates to the "quality groups" in which each judge placed the various samples; the second relates to the frequency with which he detected the different defects.

In Assigning Samples to the Different Quality Groups

The flavor scoring chart used in this trial suggested five "quality groups" based on the numerical score and into which the samples were assigned. The judges were disposed to give considerable attention to the grouping of the samples. Table XIX indicates the samples which each judge placed in each of these groups. However, instead of giving actual frequencies in the table, the figures represent the percentage of the judge's total observations that he assigned to each group.

As has been mentioned, the scoring chart was designed to give the judges some guidance as to the score that should be assigned when any of the flavor defects were detected in any degrees of intensity. In spite of this guidance, however, some of the judges showed consistent tendencies for a more liberal or a more critical attitude toward the samples and especially toward certain defects.

Table XIX reveals some of these tendencies. In giving attention to this table it is well to contrast the percentage for each judge not only with the others but particularly with the percentage for all judges as given in the bottom row. Judge 2 was lowest in the samples he considered excellent and good. He was highest in the number of fair and poor samples. It is recalled his scores were consistently low.

Judge 3 was extremely liberal in his scoring. In 13.1 percent of his observations he considered the samples excellent, while the figure for all judges indicates only 7.9 percent were so grouped. On the other hand this judge was relatively low in the number of "fair" and "poor" observations; his average score on all samples was the highest of all judges.

Judge 4 likewise exhibited some liberality in according the higher scores. On the other hand, he was unusually high in the number of samples he considered "bad"; thus his average score was not out of line with the other judges. He, along with judge 6, succeeded better than the others in avoiding the unduly critical attitude toward the superior samples and undue liberality toward the inferior ones. He was not at all averse to an extreme range in his scores.

Judge 7 succeeded 6 on the panel of judges a few months after the trial was started. As indicated in Table V, judge 7 was one of the inexperienced judges; yet he was extremely consistent. If figures such as given in Table XIX can be accepted as a criterion, it would be conjectured that judge 7 ranked along with the experienced judges, 1 and 5, in the consistency which characterized their scoring. They were likely the most proficient of the judges.

TABLE XIX. Percentage of each judge's total observations that he assigned to each quality group.

Judge	QUALITY GROUP					
	Excellent 23 or above	Good 21 or 22	Fair 18 to 20	Poor 12 to 17	Bad 11 or below	All
1	5.7	38.1	38.6	15.2	2.4	100
2	5.1	29.4	44.2	18.4	2.9	100
3	13.1	38.5	31.1	14.7	2.6	100
4	10.3	35.3	38.2	12.7	3.4	100
5	7.3	30.4	43.0	16.8	2.5	100
6	13.3	39.8	32.5	10.8	3.6	100
7	5.3	32.6	43.9	15.3	3.0	100
All	7.9	34.1	39.8	15.3	2.8	100

In Detecting the Flavor Defects

Table VI has indicated the 18 flavor defects to which the judges gave especial attention in scoring the samples. From the first it was apparent that the judges varied in their thresholds for the different tastes and odors. Likewise they varied somewhat in their tolerance for them, and this is reflected in the severity of their scores.

In Table XX is indicated the percentage of the total observations of certain defects that was noted by each judge. It is not feasible to present the results for each of the 18 defects, so only five are included in the table.

TABLE XX. Percentage of the total observations of certain defects that were noted by each judge.

Judge	Judge's percentage of all observations	FLAVOR DEFECT				
		Feed	Rancid	Stale	Salt	Flat
1	15.1	15.1	14.0	16.6	11.6	24.7
2	15.1	11.4	20.4	27.2	13.9	21.2
3	15.0	13.8	10.3	9.9	10.9	13.3
4	17.3	20.2	17.9	12.4	20.6	20.8
5	15.2	11.5	15.1	8.3	21.7	3.9
6	2.6	3.0	1.0	0.9	1.4	9.1
7	19.7	25.0	21.3	24.8	19.9	12.0
All	100	100	100	100	100	100

Judge 1, for instance, made 15.1 percent of the observations; likewise he was responsible for 15.1 percent of the "feed" criticisms. In fact, his proportion of responsibility for each of the defects was quite regular except in the case of "flat." He was responsible for 24.7 percent of all the cases in which this defect was noted.

Similarily, the records of the other judges as shown in Table XX reveal that they exhibited individual differences in identifying and evaluating the defects.

In spite of these differences, however, there seems to be no occasion to challenge the validity of the scoring results. The figures of Table XX

merely explain some of the data presented earlier. For instance, judge 2 was consistently low in the scores he accorded. His proclivity to note the "rancid" and "stale" defects is the cause. Judge 3 was consistently liberal. He noted none of the defects with such frequency as the other judges. The fact that all the judges exhibited consistency in their scoring affords added confidence in the results.

BIBLIOGRAPHY

Anderson, E. O., Dowd, L. R. and Stuewer, C. A.
1937 Relation of acidity of milk to oxidized flavor.
Food Res. 2, 143-150.

Anderson, J. A.
1934 Bitter milk and cream.
Proc. Inter. Assoc. Milk Dealers.
27th Ann. Convention Lab. sect., 17-27.

Anderson, J. A., Hardenberg, J. K. and Wilson, L. T.
1936 Concerning the cause of rancid and oxidized flavor of bovine origin.
J. Dairy Sci. 19, 483-184.

Babcock, C. J.
1923-a Effect of feeding green alfalfa and green corn on flavor and odor of milk.
U. S. Dep. Agr. Dep. Bull. 1190.

Babcock, C. J.
1923-b Effect of feeding turnips on flavor and odor of milk.
U. S. Dep. Agr. Dep. Bull. 1208.

Dukes, H. H.
1935 The physiology of domestic animals.
Third edition, Ithaca, New York, Comstock Publishing Company, 167 illus.

Eckles, C. H. and Shaw, R. H.
1913-b The influence of breed and individuality on the composition and properties of milk.
U. S. Dep. Agr., Bur. Ani. Ind., Bull. 155.

Eckles, C. H. and Shaw, R. H.
1913-b The influence of breed and individuality on the composition and properties of milk.
U. S. Dept. Agr., Bur. Ani. Ind., Bull. 156.

Fouts, E. L. and Weaver, Earl.
1935 A chart to aid in scoring milk flavor.
J. Dairy Sci., 18, 51-54.

Frazier, William C.
1928 Defect in milk due to light.
J. Dairy Sci., 11, 375-379.

Gabathuler, A.
1915 Die Bedeutung des Milchzucher fuer die hygienische.
Bourteilung der Milch.
Zeit. Fleisch-Milchhyg., 25, 97-140.

Gamble, J. A. and Kelly, Ernest.
1922 Effect of silage on the flavor and odor of milk.
U. S. Dep. Agr. Dep. Bull. 1097.

Garrett, O. F., Bender, C. B. and Tucker, H. H.
1937 Relation of grass silage to the color, vitamin C and flavor in milk from individual cows.
J. Dairy Sci., **20,** 426-427.

Gondos, M. A.
1934 Alterations oxidatives du lait et des products laitiers (gout huileux-suffieux).
Lait. **14,** 25-30.

Guthrie, E. S. and Brueckner, H. J.
1933 The cow as a source of oxidized flavors of milk.
N. Y. (Cornell) Agr. Exp. Sta. Bull. 606.

Hileman, J. L. and Courtney, Eleanor.
1935 Season variations in the lipase content of milk.
J. Dairy Sci., **18,** 247-255.

Immer, F. R.
1937 Information to the author.

Babcock, C. J.
1924 Effect of feeding cabbage and potatoes on flavor and odor of milk.
U. S. Dep. Agr. Dep. Bull. 1297.

Babcock, C. J.
1925 Effect of feeding green rye and green cowpeas on flavor and odor of milk.
U. S. Dep. Agr. Dep. Bull. 1342.

Babcock, C. J. and Leete, C. S.
1929 How to conduct milk and cream contests.
U. S. Dep. Agr. Circ. 384.

Bernstein, J.
1886 The five senses of man.
New York D. Applegate and Company 91 illus.

Brown, W. C., Thurston, L. M. and Dustman, R. B.
1937 IV Studies of the relation of the feed of the cow to oxidized flavor.
J. Dairy Sci., **20,** 133-145.

Chilson, W. H.
1935 A study of the oxidized flavor of market milk.
Milk Plant Monthly, **24,** number 11, 24-28; number 12, 30-34.

Crocker, E. C. and Henderson, L. F.
1927 Analysis and classification of odors.
Perfumer, **22,** 325-330.

Crocker, E. C. and Henderson, L. F.
1934 Whiff numbers.
Tech. Rev. **36,** 5-15.

Csiszar, J.
1932 Die Milch mit fettspaltenden Eigenschaften als Ursache der "oelig-ranzigen" Milch. Milchwirtschaft Forsch., **14**, 288-297.

Dahle, C. D. and Palmer, L. S.
1937 The oxidized flavor in the milk from the individual cow. Penn. Agr. Exp. Sta. Bull. 347.

Davies, W. L.
1931 The inactivation of lipase in dairy products by traces of heavy metals. J. Dairy Res. **3**, 254-260.

Kelly, E.
1934 Effect of butterfat content on milk palatability. U. S. Dep. Agr., Bur. Dairy Ind., Ann. Rep., 15.

Kende, S.
1931 Reasons for and combating of "oily" milk defects. "Oleinase" a new enzyme in milk. Proc. ninth Intern. Dairy Cong., Sec. 3, Paper 137. (summary), supplement. 55.

Kende, S.
1932 Untersuchungen ueber "oelig-talgige," "schmirgelige" Veraenderungen der Milch. Milchwirtschaft. Forsch. **13**, 111-143.

Kertesz, T.
1932 Die Verhuetung des "Oeligwerdens" der Milch und Verbesserung der Olmilch durch bakteriologische Behandlung. Molkerei-Zeit. (Hildesheim). **46**, 687-688.

Kieferle, F., Schwaibald, J. and Hackmann, Ch.
1925 Der Gehalt der Kuhmilch an Citronensaeure und dessen Bezeihungen zur Chlorzucherzahl als Kriterium fuer normale b. z. w. anormale Milch. Zeit. Physiol. Chem., **145**, 18-38.

Kieferle, F.
1933 Gestoerte Sekretion. 3. Verhalten der stickstoffhaltigen Bestandtiele der Milch, insbesondere des Reststickstoffs bei gestorter Sekretion. Milchwirtschaft Forsch., **14**, 567-581.

King, Florence B.
1937 Judging flavors in food. Food Res., **2**, 207-219.

Koestler, G.
1920 Zum Nachweis der durch Sekretionsstoerung veraenderten Milch. Mitt. Lebensm. **11**, 154-169.

Koestler, G. G., Roadhouse, C. L. and Loertscher, W.
1928 Zur Kenntnis der Sekretion Lipolytisch aktiver sogenannter "ranziger" Milch. Lantwirt. Jahrb. Schweiz, **42**, 937-966.

Leach, Carl A.
1934 Saltiness in milk.
Dairy Goat Journal, number 7, 14.

Lucas, P. S.
1929 Many factors cause abnormal milk flavors.
Mich. Quar. Bull., **12**, 18-20.

McCandlish, Andrew C. and Leitch, R. H.
1932 Feeding the dairy cow and foods and flavors of milk.
West. Scot. Agr. Col. Bull., 126.

Majer, G.
1931 Ursachen des Schmirgelgeschmakes in der Milch und dessen Abhangigkeit von der Fuetterung.
Moekerei-Zeit. (Hildesheim) **45; 1882-1884.**

Nelson, John A. and Trout, G. Malcolm.
1934 Judging Dairy Products.
Milwaukee, The Olsen Pub. Company, 22 illus.

Olson, T. M., Totman, C. C. and Wallis, D. C.
1936 The effect of tankage on the flavor of milk.
J. Dairy Sci., **19**, 313-315.

Overman, O. R., Sanmann, F. P. and Wright, K. E.
1929 Studies on the composition of milk.
Ill. Agr. Exp. Sta. Bull. 325.

Palmer, Leroy S.
1922 Bitter milk of advanced lactation.
J. Dairy Sci., **5**, 201-211.

Palmer, Leroy S. and Eckles, C. H.
1917-a The influence of the stage of gestation on the composition and properties of milk.
J. Dairy Sci., **1**, 185-198.

Palmer, Leroy S. and Eckles, C. H.
1917-b Gestation versus lactation as factor causing abnormal composition of cow's milk.
Am. J. Diseases Children, **13**, 413-422.

Parker, G. H.
1922 Smell, taste and allied senses in the vertebrates.
Philadelphia, J. B. Lippincott Company, 37 illus.

Platt, Washington.
1931 Rational methods of scoring food products.
Food Ind., **3**, 108-117.

Platt, Washington.
1937 Some fundamental assumptions pertaining to the judgment of food flavors.
Food Res., **2**, 237-249.

Prewitt, E. and Parfitt, E. H.
1935 Effects of feeds on oxidized flavors in pasteurized milk.
J. Dairy Sci., **18**, 468.

Richter, K.
1931 Untersuchungen ueber die Entwicklung der Bakterienflora im Milch and ihre abhangigkeit von der Pufferung. Habililationsarbeit. Univ. Kiel.

Roadhouse, C. L., Regan, W. M. and Mead, S. W.
1926 Research in dairy industry. Calif. Agr. Exp. Sta. Rep. 62-64.

Roadhouse, C. L. and Koestler, G.
1929 Contribution to the knowledge of the taste of milk. J. Dairy Sci., **12**, 421-437.

Roadhouse, C. L. and Henderson, J. L.
1930 The influence of molasses and sodium chloride on the lactose and chloride content and taste of milk. J. Dairy Sci., **13**, 117-123.

Roadhouse, C. L. and Henderson, J. L.
1935 Flavors of milk and their control. Calif. Agr. Exp. Sta. Bull. 595.

Roadhouse, C. L. and Henderson, J. L.
1937 Regulating the feeding of certain roughages to minimize their influence on the flavor of milk. J. Dairy Sci., **20**, 679-683.

Rogers, L. A.
1937 Chemistry and bacteriology of milk. U. S. Dep. Agr., Bur. Dairy Ind. Rep. 24.

Roland, Chas. T. and Trebler, H. A.
1937 The effect of fat content on oxidized flavor in milk and cream J. Dairy Sci., **20**, 345-350.

Sharp, Paul and de Tomassi, J. A.
1932 Increase in non-lactic acidity in raw cream and its control. Int. Assoc. Milk Dealers, Lab. sect. 25th. Ann. Proc. 3-20.

Sharp, Paul F. and Struble, Earle B.
1935 Period of lactation and the direct titratable chloride value in milk. J. Dairy Sci., **18**, 527-538.

Sharp, Paul F., Trout, G. Malcolm and Guthrie, E. S.
1936 Vitamin C, copper and the oxidized flavor of milk. Proc. N. Y. Assoc. Dairy and Milk Inspectors Tenth Ann. Convention, 153-164.

Snedecor, George W.
1934 Calculation and interpretation of analysis of variance and covarience. Ames, Ia. Collegiate Press, Inc.

Snedecor, George W.
1937 Statistical Methods. Ames, Ia. Collegiate Press, Inc.

Steineggar, E. R. and Alleman, O.
1905 Beitrag zur Kenntnis der Beschaffenheit salzigbitterer Milch. Landwirt. Jahrb. Schweiz **19**, 527-529.

Thomsen, L. C.
1936 How to judge butter.
Nat. Butter and Cheese J. **27**, 8-13.

Thurston, L. M.
1935 Oxidized flavor in milk.
Proc. Intern. Assoc. Milk Dealers 28th Ann. Convention, Lab. sect., 121-141.

Tracy, P. H. and Ruehe, H. A.
1931 The relation of certain plant processes to flavor development in market milk.
J. Dairy Sci., **14**, 250-267.

Trebler, H. A.
1935 The effect of metals on flavor of dairy products.
Proc. Intern. Assoc. Milk Dealers, 28th Ann. Convention, Plant sect. 107-111.

Trout, G. Malcolm and Sharp, Paul F.
1936 The reliability of flavor judgments with special reference to the oxodized flavor of milk.
N. Y. (Cornell) Agr. Exp. Sta. Memoir **204**.

Trunz, A.
1903 Ueber die mineralischen Bestandteile der Kuhmilch und ihre Schwankungen im Verlauf einer Laktationsperiode.
Zeit. Physiol. Chem. **40**, 263-310.

Turner, C. W.
1936 Factors affecting the composition of milk.
Mo. Agr. Exp. Sta. Bull. 365.

Weaver, Earl, Fouts, E. L. and McGilliard, P. C.
1935 Frequency of the flavor defects of milk.
J. Dairy Sci., **18**, 467-468.

Weaver, Earl, Kuhlman, A. H. and Fouts, E. L.
1935 The effect of alfalfa hay on milk flavor.
J. Dairy Sci., **18**, 55-61.

Webb, R. E. and Hileman, J. L.
1937 The relation of the oxidition-reduction potential of milk to the oxidized flavor.
J. Dairy Sci., **20**, 47-57.

Whitnah, C. H., Martin, W. H. and Beck, G. H.
1937 Oxidized milk flavor as related to carotene, lecithin and vitamin C.
J. Dairy Sci., **20**, 433-434.