A Content Analysis of Bias in *Consumer Reports* on Automobiles

Klaus Krippendorff
*University of Pennsylvania, kkrippendorff@asc.upenn.edu*

Follow this and additional works at: [http://repository.upenn.edu/asc_papers](http://repository.upenn.edu/asc_papers)

Part of the *Communication Commons*

**Recommended Citation**


This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/asc_papers/220](http://repository.upenn.edu/asc_papers/220)

For more information, please contact repository@pobox.upenn.edu.
A Content Analysis of Bias in Consumer Reports on Automobiles

Disciplines
Communication | Social and Behavioral Sciences

Comments

This other is available at ScholarlyCommons: http://repository.upenn.edu/asc_papers/220
A CONTENT ANALYSIS

OF

BIAS IN CONSUMER REPORTS ON AUTOMOBILES

by
Klaus Krippendorff

The Annenberg School of Communications
University of Pennsylvania, Philadelphia

August 23, 1982
corrected September 17, 1982
A CONTENT ANALYSIS OF BIAS IN CONSUMER REPORTS ON AUTOMOBILES

Contents

Introduction 1
The Data Analyzed 3
The Measuring Instrument Used 6
Bias 15
Findings 24
References 35

Appendices
A: The Measuring Instrument (Coding Instructions)

B: Data Sheet

C: Results of Reliability Tests

D: Tables of Findings
Introduction

Consumer Reports is published monthly by Consumers Union of United States, Inc., a non-profit organization whose principal aim is to provide consumers with information and council on consumer goods and services. It accomplishes this aim through an extensive program of evaluating consumer goods and services, using laboratory tests, controlled-use observations, and expert judgement of purchased samples sold largely in the United States and by publishing its ratings in Consumer Reports. In it, "Consumers Union pledges that any opinions entering into its ratings of products shall be as free of bias as possible."

Inasmuch as consumers are often overwhelmed with conflicting and exaggerated claims of sales-serving product advertisement, Consumers Union's non-profit status and its unambiguous pledge to objectivity has made it, among several similar institutions, an important judge of the market place with considerable influence on consumer decision making. Naturally, manufacturers whose products earn low scores and drop in sales might feel judged unfairly and consumers who follow the recommendations and experience disappointment might feel deceived. Both might be led to ask themselves: what if Consumer Reports is not as free of bias as it claims?
This research inquires into whether or to what extent Consumer Reports exhibits systematic biases in presenting consumer related facts to its readers. By systematic bias we mean a distortion or deviation from the facts that cannot be explained by chance alone but by -- perhaps hidden but nevertheless statistically manifest -- prejudices, preferences or preconceptions. The inquiry is limited to the reporting on and the ratings of automobiles published between January 1981 and August 1982. The research is exemplary of the "spirit of consumerism" applied to communications rather than to products and services.
The Data Analyzed

Between January 1981 and August 1982, Consumer Reports published 15 reports of test results on a total of 59 automobiles. Each report presents findings on about 4 cars which are somewhat comparable in size, price, engine or design.

In total, our analysis included reports on 59 automobiles:

31 American
21 Japanese
7 European

American Automobile manufacturers contributed to our data:

1 American Motors
10 Chrysler
7 Ford
13 General Motors

Consumer Reports classified these 59 automobile as:

37 small
22 medium
0 large

and recommended 23 of the total in its April 1982 issue.

Each report presents essentially two kinds of information, the verbal account of the test experiences and the formal ratings, or text and ratings for short. The text of a report tells the reader what kind of car the report describes and interprets the test results in an evaluative language the consumer presumably understands. The formal ratings summarize numerous quantitative performance
measurements, which are partially presented in a table of "Facts and Figures," and include certain qualitative expert judgements. *Consumer Reports* presents these summaries in the form of five-point scales:

<table>
<thead>
<tr>
<th>Excellent</th>
<th>Very good</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>🍀</td>
<td>🌐</td>
<td>🌍</td>
<td>🌍</td>
<td>🌍</td>
</tr>
</tbody>
</table>

on the following 17 subject matter categories:

**Engine and transmission:** Fuel economy
                             Engine drivability
                             Shifting
                             Acceleration

**Handling and breaking:** Accident-avoidance ability
                            Handling precision
                            Braking

**Ride:**
            Ride
            Noise
            Driving position
            Front seating (for two)
            Rear seating (for two)
            Climate control

**Convenience:**
            Controls
            Displays
            Servicing ease

**Other:**
            Predicted repair incidence

The five point scales used in the reports seem easily interpretable although *Consumer Reports* does not reveal how it arrives at these ratings. It makes reference to a computer program, which appeals to a sense of objectivity, but it does not reveal the equation underlying the computation. Our measurement effort is devoted to the text of the report. The somewhat independently derived ratings serve
as the reference for our assessment of bias. *Consumer Reports'* practice of summarizing its technical findings in terms of rating scales started in January 1981 which marked the beginning of our sample.

Conforming to these categories, we define a unit of analysis as any portion of the text pertaining to, arguing about, describing or evaluating a subject matter category for which *Consumer Reports* publishes its formal ratings. Since Consumers Union assesses each automobile in terms of 17 separate ratings, there are at most 17 units per automobile or $59 \times 17 = 1003$ units in the whole sample. Because some of these subject matters are either not mentioned in the text or not rated, the number of units actually analyzed turned out to be 954 units. The choice of this unit of analysis is particularly justified in view of Consumer's Unions' claim that each rating represents the results of entirely different batteries of tests and from its practice to describe the results of these tests in somewhat separate sections of the report. There is all reason to suppose that these units can vary freely relative to each other.
The Measuring Instrument Used

The measuring problem on hand is easy to state but was far less easy to implement. In its April 1982, April 1981 and April 1980 issues, Consumer Reports details how Consumers Union tests and evaluates each car. In particular, it describes the kind of tests and examinations its 17 subject matter categories represent.

It was important for our study not to invent new terminology but to adopt Consumers Union's definitions of categories and scale points as literally as possible. However, since the material measured consists of verbal evaluations and arguments, not of cars, we had to assure that our measurement scales for the 17 categories (a) correspond in ordinality and in number of scale points to the available ratings, (b) can be coded reliably, and (c) resemble the evaluative logic in Consumer Reports as closely as possible.

The requirement (a) that our measurement scales and Consumers Union's ratings correspond in syntactical characteristics assures that differences between our scale points and the Consumers Union's ratings can be expressed meaningfully. In the absence of any knowledge about how Consumers Union computes these its ratings, we decided to make the safest assumption of an ordinal scale and took the same order and number of points for our measurement scales. With
this 1:1 syntactical correspondence we can express differences in terms of the number of ranks in which any two scale points differ. In our situation, these differences can range between -4 and +4. The extremes come about when one scale point is labeled "excellent" and the other "poor." Zero indicates a perfect match.

The requirement (b) that our process of measurement be reliable is the same as saying that the data be relatively free of irrelevant variation. Irrelevant variation in the form of measuring errors or careless judgements retard the interpretability of the data (see Krippendorff, 1980). Although we did not inquire into reader responses, irrelevant variation may also be indicative of the uncertainties readers experience who approach the text with similar categories in mind. Reliability is usually assessed by an independent replications of the same measuring process for a small portion of the data analysed and by determining the above-chance agreement achieved between the two results. Again, we adopted a safe strategy and used not two but three (and on occasions also four) trained coders and we tested the reliability not on a small portion but on nearly the whole set of data analyzed. This gave us the best assurance of the quality of our data.

In fact, we ran several consecutive reliability tests and modified our measuring instrument after each such test until we were sufficiently confident that the instrument was as reliable as we could make it. The first test involved 60 recording units using cars not in the sample and yielded so discouragingly low inter-coder agreement that we didn't even compute the reliability coefficient.
This result called for substantial revisions of our coding instructions, incorporating more and more of the logic of evaluation Consumer Reports seems to follow [see (c) below]. The second reliability test involved another 90 units and yielded a reliability of .707. For the third test we coded 198 units and found the reliability to be .745. The final reliability test used 935 units with a fourth version of the measuring instrument and yielded a reliability of .818. This figure is far above chance and exceeds the standards common in coding tasks of this kind. For an elaboration of reliability tests, coefficients and standards see Krippendorff (1980).

The requirement (c) that our measurement and coding resemble Consumers Reports' evaluative logic is intended to assure our instrument's validity over and above the required reliability. (An additional source of validity will be discussed below). It turned out that it was this logic which caused the initial difficulties in achieving the desirable level of reliability. To understand how our instrument incorporates this logic, we show how three initial difficulties were resolved.

We started out with the assumption, strongly suggested in Consumer Reports' April 1982 article on "How CU tests and rates cars," that, because each formal rating represents the results of a separate set of laboratory and driving tests, the 17 subject matters would also be described and evaluated separately. This turned out not to be so. We found that Consumer Reports often started with general observations and general evaluative statements in its
introduction, proceeded with the detailed descriptions of the experiences with each car and concluded the report with an over-all recommendation. Evaluative judgements included generalizations about the group of cars tested and generalizations about subject matters for which more than one rating was available. If this evaluative logic were applied consistently, we would not find anything wrong with it. But in the hands of Consumer Reports the reader is often lead into paradoxes. For example, we found several cases in which a car that is generally judged "very good" is subsequently described as having so many defects or negative points that this alone would discourage a consumer from acquiring that car. Thus the original generalization is contradicted. The way we resolved these problems in our instrument was to take the general evaluations as a starting point and then asked how these impact or modify the specific evaluations subsumed by it. For example, if a car was said to handle "very well" generally, the more detailed mention of a loss of "feeling for the road" was interpreted as reducing the scale value of the steering without affecting the other subject matters subsumed under "handling." This procedure is consistent with Consumers Union's reasoning of "subtracting points for features judged inadequate."

A second initial assumption, commonly adopted in content analysis, was that we could measure consumers union's evaluation by measuring the volume and direction of evaluative statements made concerning each subject matter category and by determining which kind of statements, favourable or unfavourable, and by how much either
kind dominates this volume. Indicators of this kind assume that a
writer who believes, for example, that a car's climate control system
is very good is most likely to make more favourable than unfavourable
statements about this system. But this too proved not to be so in
*Consumer Reports'* style of evaluation. The number of negative points
mentioned usually exceeded the positive ones by four-to-one. This
might suit *Consumer Reports'* role as a critic but it initially caused
difficulties for our analysis. One might speculate that this
imbalance also causes problems in the reader's understanding of the
evaluations. We found evaluation to be strongly influenced by (i)
the number of details that were not mentioned in a report but could
have been cited and were in fact mentioned in conjunction with other
cars and by (ii) the weight given to these details. In fact,
*Consumer Reports'* April 1982 issue explains the reason for (i): "if
the results of a test are unexceptional, we may not mention them in
our report." In other words, in *Consumer Reports'* evaluative logic,
a few minor complaints with nothing else worth mentioning may still
mean "basically o.k.," "good," which is the evaluation of the
unexceptional. To resolve this problem operationally we weighted the
volume of not mentioned and hence unexceptional details against those
written about giving due consideration to their significance. Our
coders had to develop a mental list of all praises and complaints
mentioned elsewhere about the subject matter, for example, about
climate control, and to assess how many of those and which kinds
*Consumer Reports* considered exceptional enough to be included in the
report.
A third assumption concerned the semantical designation of our scale values. Initially, we had hoped to employ in our coding the terms *Consumer Reports* uses to express its ratings, i.e. "excellent," "very good," "good," "fair," and "poor." However, *Consumer Reports* interchangeably utilizes numerous rhetorical devices to express its enthusiasm and disappointment. Sometimes it expresses its evaluation relative to the costs involved, sometimes it judges the availability of options not in fact tested and sometimes it expresses attitudes on several logical levels, for example, by noting as favourable that a manufacturer introduced certain changes, but then disapproving of the results. This mixture of evaluative forms could not be measured easily. A closer examination revealed that *Consumer Reports* uses three primary evaluative schemes which we could separate: appeals to standards, comparative modes of evaluations and absolute judgements.

**Absolute Judgements** are exemplified in *Consumer Reports*’ ratings between "excellent" and "poor". In our instrument, these formal terms were elaborated by similar absolute expressions we encountered in the text. When new expressions occurred we made a general decision as to which scale value it would correspond to and then proceeded to use them throughout the coding.

**Comparative Evaluations** make references to other cars, by comparing one to the average, to most, to few or to one specific other car. The cars described in our report often belonged to a group which *Consumer Reports* compared first as a whole group with other groups of cars and then by creating contrasts within this
group. We equated any notion of "average" or majority with a position between "good" and "fair" on the absolute scale indicating that there would be nothing exceptional to write about and ascertained by how much better or worse the subject matter was described. When the comparison concerned two cars only, we considered where the already evaluated car belonged and how large a difference was indicated.

Appeals to Standards were explicit in Consumer Reports' April 1982 issue which details what Consumers Union expects or requires. Some of these standards are stated as standards of adequacy which we mapped into the middle of our five-point scale because the car could deviate from it in both ways. Other standards are stated as standards of ideals which we considered equivalent to the "excellent" point of the absolute scale because they can only be approximated, not exceeded. Accordingly, and relying on Consumers Union's own description, "adequate acceleration" is considered equal to "good," since it can be exceeded by an even better acceleration, whereas effortless shifting without noise and lurch" etc. is considered equal to "excellent," since it is a non-exceedable ideal.

The following table shows how the evaluative schemes correspond:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Absolute judgements</th>
<th>Comparative evaluations</th>
<th>Appeals to standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>excellent</td>
<td>exceptionally good</td>
<td>(not exceedable) ideal</td>
</tr>
<tr>
<td>4</td>
<td>very good</td>
<td>clearly above average</td>
<td>more than adequate</td>
</tr>
<tr>
<td>3</td>
<td>good</td>
<td>more positive than negative</td>
<td>adequate</td>
</tr>
<tr>
<td>2</td>
<td>fair</td>
<td>(average &amp; unexceptional)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>poor</td>
<td>clearly below average</td>
<td>far below standards</td>
</tr>
</tbody>
</table>
By incorporating much of Consumer Reports' evaluative logic, our instrument, which is reproduced in Appendix A, entangles many of the difficulties encountered in the verbal evaluations. The instrument could not be perfect, however, because as we noted above, Consumer Reports does not apply uniform standards in its evaluations. It is full of evaluative inconsistencies and paradoxes which makes it indeed difficult, even for an ordinary reader, to get an unambiguous sense as to how good a car actually is. For example, we had originally tried to incorporate into our instrument Consumers Union's over-all judgement on a car as an additional subject matter variable, but never achieved acceptable levels of above-chance agreement among the three coders, making it totally useless as a datum for analysis.

Finally, to further check and improve on the validity of our data we asked all coders to go over their independently obtained data and, after discussing any differences that did arise during their separate efforts, to agree on a single value for each unit of analysis. To facilitate this, we used a computer program that identified those units on which independent judgements differed. Since reliability was .818, this was needed for about 19% of the units. Sometimes a difference was due to a mere clerical error, sometimes a coder missed a sentence in an odd section of the text, and sometimes there were genuine difficulties in interpreting conflicting evaluations. Going over these together removed most of the doubts and perhaps replicated a situation of several interested consumers discussing the merits and drawbacks of a car and reaching a consensus about what the report says. This was the final effort to
obtain data of optimum validity.

We should also mention that the coders never saw either the formal ratings or the list of "Facts and Figures" while analysing the text. Thus, their individual and joint measures of the verbal evaluations are entirely independent of any extraneous information.

Thus, all conceivable precautions were taken to obtain the most reliable and the most valid measuring instrument achievable.
Bias

The assessment of bias in reporting has a long history in communication research. Since the 1930s, numerous content analyses attest to this tradition. In these assessments, bias is largely based on the journalistic ideal of a "balance in reporting" on two sides of a controversy, on both candidates for a political office or on positive and negative aspects of a public issue. The only index of bias existing in the literature is Janis and Fadner's (1949) "Coefficient of Imbalance." It pitches the volume of favourable comments against the volume of unfavourable ones and expresses this difference as a proportion of the total volume (including neutral comments). The standard of balance built into the coefficient becomes inappropriate, however, when the two sides are essentially unequal.

Consumer products are essentially unequal and inequalities of this kind are the primary motivation for Consumers Union to undertake its testing program. Indeed, Consumer Reports should call things by their names and describe favourably those which test out well and unfavourable those which do not. Thus, the issue is not one of balance but one of unbiased reporting. Ideally, unbiased reporting should exhibit no differences between the verbal evaluations and the
formal ratings. However, to be fair, we must allow Consumer Reports to maintain its own journalistic style and to elaborate on positive and negative ratings in whichever evaluative language it chooses. Considering this, unbiased reporting would then be indicated if the same kind of distortions were applied uniformly to all subject matter categories and to all products subject to evaluation. It means that the evaluative content of the text is predictable from the formal ratings. Finally, we must also allow for some stylistic variation in the sense that verbal evaluations may deviate positively and negatively from how ratings are expected to be described, given the body of reports in hand. Together, these qualifications on what we mean by unbiased reporting would require that the modal distributions of differences between verbal evaluations and ratings are the same for all subject matter categories and products. Biased reporting would manifest itself when these differences are significantly larger for one category of consumer products than for another category, regardless of how large these differences actually are. A measure of bias that takes account of these qualifications boils down to measuring the degree of dissimilarity between two distributions of rank differences between verbal accounts and ratings for any one category of consumer products and for its complement in the data available. This is essentially a problem of comparing two samples of rank order differences.

The index used in this study assesses just this sampling bias. Our measure of bias is affected by two kinds of errors. One error is the sampling error which is a function of the numerical size of the
sample, is expected to be large when the sample size is small and becomes smaller as the sample size increases. The other error is the systematic sampling bias which arises when sampling is not entirely chance, when some members of a population have a higher probability of inclusion in the sample then others or, as we are concerned with here, when distortions found in the reporting on one group of cars are unlike those found in another.

Because it is not entirely possible to separate the two kinds of errors, when our index is relatively small, we must check its statistical significance. If the value of the index is not statistically significant then the bias is merely due to the sampling error (which is inevitable with small sample sizes and hence excusable). However, if the value of the index is statistically significant then the sampling bias exceeds the sampling error and must be due to the systematic sampling bias. One virtue of our index is that it lends itself to a non-parametric test of its statistical significance. The following defines this index.

Let the whole body of data have \( N \) units of analysis and let a sample drawn from it have \( n \) such units. Let the differences in ranks be denoted by \( i = (\text{textscore} - \text{rating}) = -4, \ldots, +4 \). With reference to these differences, our data have the following form:

<table>
<thead>
<tr>
<th>Differences</th>
<th>-4</th>
<th>\ldots</th>
<th>1</th>
<th>\ldots</th>
<th>+4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample from data</td>
<td>( n_{-4} )</td>
<td>\ldots</td>
<td>( n_1 )</td>
<td>\ldots</td>
<td>( n_4 )</td>
</tr>
<tr>
<td>Complement of sample</td>
<td>( N_{-4} - n_{-4} )</td>
<td>\ldots</td>
<td>( N_1 - n_1 )</td>
<td>\ldots</td>
<td>( N_4 - n_4 )</td>
</tr>
<tr>
<td>Body of data</td>
<td>( N_{-4} )</td>
<td>\ldots</td>
<td>( N_1 )</td>
<td>\ldots</td>
<td>( N_4 )</td>
</tr>
</tbody>
</table>
In these terms, the number \( U \) of differences in the sample that exceed the differences in its complement is:

\[
U = \sum_{i} n_i \left[ \frac{N_i - n_i}{2} + \sum_{j>1} (N_j - n_j) \right]
\]

\( U \)'s limits are:

\[
0 \leq U \leq n(N-n)
\]

Our coefficient of bias \( c \) in ordinal data is a standardization of \( U \):

\[
c = \frac{n(N-n)}{2} - U
\]

\[
c = 1 - \frac{1}{n(N-n)} \sum_{i} n_i \left[ N_i - n_i + \sum_{j>1} (N_j - n_j) \right]
\]

whose limits are naturally:

\[
-1 \leq c \leq +1
\]

Using Mann-Whitney's non-parametric test of the statistical significance of \( U \) (see Siegel, 1956):
\[ \mu_U = \frac{n(N-n)}{2}; \quad \sigma_U = \sqrt{\frac{n(N-n)}{12(N-N)^2} \left[ N^3 - 3N^2 - N \sum (N_i^3 - N_i) \right]} \]

we can express the z-value for the normal distribution as a function of our coefficient of bias:

\[ z = \frac{U - \mu_U}{\sigma_U} = c \sqrt{\frac{3(N-N)}{(N^3 - N) - \sum (N_i^3 - N_i)}} \]

With the help of a table of probabilities for the normal distribution, this z-value leads us to the probability of obtaining the coefficient of bias by chance and thus to a separation of sampling error and systematic sampling bias. This procedure is an approximation which becomes better when the number of differences increases.

Our coefficient of bias is new, but for data with only two ranks, its numerator resembles several well known coefficients such as Yule's Q, Kendall's \( \tau \), and Pearson's correlation coefficient:

\[ c = -\frac{ad-bc}{(a+b)(c+d)} \]

This fact may illustrate what the coefficient measures. When all entries are in the a-d-diagonale of a 2-by-2 table, the coefficient is -1. When all entries are in the b-c-diagonale, it is +1. And when the proportions a/c and b/d equal the proportion n/(N-n), the
coefficient is zero. Thus, the magnitude of the coefficient indicates the extent of the bias (dissimilarity between the sample and the remainder of the data) and its sign indicates the direction. Positive values indicate that the differences between the verbal accounts and the ratings in a report are more favourable than expected, given the style of evaluation, negative values indicate that these differences are less favourable than expected. Note that absolute values of these differences do not enter the coefficient, only their rank ordering, thus the coefficient is not affected by the distortions characteristic of that publication's writing style.

Although the coefficient is designed to indicate the bias relative to the data as a whole, we may also use it to express bias in absolute terms, thereby ignoring what the foregoing considered a virtue, the reference to a publication's writing style. This is accomplished by adding to the data a "sample" of size \((N_m - n_m) = 1\) as marker of the balanced or middle position \(m\) (or of any other position chosen as a standard). With this modification, the coefficient \(c\) becomes one that pitches differences that exceed the chosen value \(m\) against those exceeded by that value:

\[
c_m = \frac{\sum_{i>m} N_i - \sum_{i<m} N_i}{N}
\]

This form corresponds to Janis and Fadner's (1949) above mentioned Coefficient of Imbalance. Unlike Janis and Fadner's coefficient, it provides for choices among reference points and unlike our relative
coefficient of bias, it additionally ignores the statistical expectations of the rank differences to which it is applied here.

To demonstrate the properties of these coefficients, we use our data with one hypothetical example and four actual findings:

<table>
<thead>
<tr>
<th>Differences</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>119</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complement</td>
<td>8</td>
<td>61</td>
<td>30</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All cars</td>
<td>0</td>
<td>8</td>
<td>61</td>
<td>30</td>
<td>58</td>
<td>119</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ c = 1.000 \]
\[ p = .0000 \]

In this hypothetical example, all cases in which the evaluations in the text are one or more ranks more favourable than the ratings appear in the sample and all other cases turn up in its complement.

The situation is extreme and the coefficient registers plus one.

<table>
<thead>
<tr>
<th>Differences</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chrysler</td>
<td>22</td>
<td>38</td>
<td>75</td>
<td>23</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All other cars</td>
<td>8</td>
<td>39</td>
<td>267</td>
<td>383</td>
<td>96</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All cars</td>
<td>0</td>
<td>8</td>
<td>61</td>
<td>30</td>
<td>58</td>
<td>119</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ c = .002 \]
\[ p = .4860 \]

Here the differences for Chrysler cars and those for the remaining set of our data have nearly the same modal distribution. Bias is practically absent and the coefficient near zero, which is further indicated by the near absence of statistical significance.
<table>
<thead>
<tr>
<th>Differences</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMC + Ford + GM</td>
<td>8</td>
<td>26</td>
<td>120</td>
<td>151</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td>338</td>
</tr>
<tr>
<td>All other cars</td>
<td>35</td>
<td>185</td>
<td>307</td>
<td>86</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>616</td>
</tr>
<tr>
<td>All cars</td>
<td>0</td>
<td>8</td>
<td>61</td>
<td>305</td>
<td>458</td>
<td>119</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ c = -0.126 \]
\[ p = 0.0002 \]

Here the differences for three American automobile manufacturers are seen to be more on the negative side than the differences for the remainder of the data. Cars by these manufacturers are not only described in terms less favourable than called for by their ratings, but also more unfavourable than expected, given Consumer Reports' style. Systematic sampling bias is indicated by the relatively large value of the coefficient and by the small probability of obtaining or exceeding this value by chance.

<table>
<thead>
<tr>
<th>Differences</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ford</td>
<td>13</td>
<td>38</td>
<td>50</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>111</td>
</tr>
<tr>
<td>GM</td>
<td>8</td>
<td>12</td>
<td>77</td>
<td>91</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td>211</td>
</tr>
<tr>
<td>Ford + GM</td>
<td>0</td>
<td>8</td>
<td>25</td>
<td>115</td>
<td>141</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ c = -0.016 \]
\[ p = 0.4009 \]

Here the sample of cars from Ford is compared with the sample of cars from General Motors. The two distributions are remarkably similar which is reflected in the small and statistically insignificant bias against Ford (and by implication in favor of General Motors).

Apparently, the two groups of cars are not treated differently.
<table>
<thead>
<tr>
<th>Differences</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>American cars</td>
<td>8</td>
<td>48</td>
<td>158</td>
<td>226</td>
<td>56</td>
<td>1</td>
<td></td>
<td></td>
<td>497</td>
</tr>
<tr>
<td>Marker at &quot;0&quot;</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Sum</td>
<td>0</td>
<td>8</td>
<td>48</td>
<td>158</td>
<td>227</td>
<td>56</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ c = -0.316 \]

Here the absolute bias with reference to the zero-difference between text score and rating is expressed for all American cars.

Unfavourable distortions dominate the favourable distortions by nearly four-to-one and the absolute form of the coefficient, here of balance, is negative and large. (Statistical significance is not computed in this case).

With these preliminaries and "previews" we now consider the findings.
Findings

The first question we asked was whether Consumer Reports' descriptions of the test results is balanced in the sense that favourable and unfavourable deviations of verbal evaluations from formal ratings cancel each other. The answer is negative. The distributions of these deviation is skewed. Negative differences exceed positive ones by a ratio of three-to-one. The coefficient of bias, using the zero-difference as a standard, reflects this observation:

Absolute bias for all 59 cars: \( c_m = -.264 \)

However, as we have argued above, this finding may be somewhat inconsequential insofar as the manifest preferences "to paint things worse than they are" may merely reflect Consumer Reports' generally critical style of writing. Consumer Reports is entitled to such a style provided it is applied consistently to all products tested and reported therein. By expressing biases relative to this critical style, we will now ask how fairly various categories of automobiles are treated in this publication and briefly return to some absolute measures in the conclusion.
One statistically significant result, though not the most important finding, concerns Consumer Reports' bias towards cars it recommends to the readers of its April 1982 issue. The biases towards recommended and to not recommended cars are:

23 recommended cars: $c = +.119$, $p = .0004$

36 not recommended cars: $c = -.119$, $p = .0004$

We will present such findings by means of bar graphs whose width indicate the number of cars involved and whose height, above and below the zero-line, indicates favourable and unfavourable biases respectively:

![Bias Towards Recommended/Not Recommended Cars](image)

Figure 1
We start deliberately with this somewhat simple result because it might explain what our coefficient measures and thus facilitate the interpretation of subsequent findings. To begin with, we reemphasize that our coefficient measures neither how positive or negative (recommended) cars are rated nor how favourable or unfavourable the verbal evaluations of these cars turned out in the reports. Clearly, good cars are expected to have positive ratings and favourable write-ups. Our coefficient measures the difference between these two separate data on a scale of -1 to +1. Considering Consumer Reports' generally critical reporting style and that its verbal accounts might vary in both directions from what would be a consistent application of this style, a coefficient of zero would indicate that the differences (between verbal accounts and ratings) found in reports on one group of (recommended) cars are just as expected or have the same modal distribution as the differences in its complement (of not recommended cars) and in fact in the data as a whole. A negative coefficient would indicate that in the reports on these cars' verbal evaluations are unexpectedly more unfavourable than expected, whereas a positive value indicates the opposite, that the reports for this group systematically upgrade the test results and make the cars involved in this group appear better than what one would predict from their formal ratings. The latter is the case for automobiles Consumer Reports recommends to its readers.

A word on the statistical significance may also be in place. All statistical results are subject to uncertainties. The p-value given here indicates the probability with which a random sample drawn
from our data would exhibit a coefficient equal to or more extreme
than the one found. For recommended cars, the value of \( p = .0004 \)
indicates that a bias of \( c = +.119 \) would be equalled or exceeded in
only 4 out of 10,000 randomly drawn samples from our data. Inasmuch
as we compare a sample with its complement, whose respective
coefficients of bias differ only in their sign, a two-tailed test of
significance is appropriate. The statistical significance for the
magnitude of the coefficient is therefore one minus twice the \( p \)-value
given in Appendix D and in Figures 1 through 5. For recommended
versus not recommended cars, the statistical significance is
therefore .9992. Such certainty is rarely achieved in the social
sciences and this bias can therefore hardly be a chance event.

We also tested the homogeneity of our data by contrasting every
second car and every second unit and found the biases in both cases
near zero and statistically insignificant (see Appendix D).

With these comments in mind, we conclude beyond reasonable doubt
that \textit{Consumer Reports}' verbal accounts are strongly biased in support
of its own recommendations, thus showing evidence that unfounded
preferences govern the verbal evaluations contained in the reports.
It is not farfetched to conjecture that \textit{Consumer Reports}'
recommendations may be the subject of self-fulfilling preconceptions.

A more important finding concerns biases towards the
manufacturers' country of origin: France, Germany, Japan, Sweden and
the United States. Since the number of European cars in our data is
relatively small, Figure 2 lumps them into one group:
Biases Towards Manufacturers' Country of Origin

Figure 2

There is a small bias in favour of Japanese cars: Honda, Mazda, Nissan/Datsun, Toyota and Subaru. However, from its lack of statistical significance (.8626), one must conclude, given its critical style, that Consumer Reports does not present Japanese cars unfairly. This slightly favourably bias towards Japanese cars is largely due to Consumer Reports' statistically significant (.9984) and unusually favourable (c=.180) treatment of Toyota.

Reports on European cars, Peugeot, Renault, Saab, Volkswagen and Volvo are more clearly biased in their favour although we will qualify this statement below. We conclude (with a statistical significance of .9964) that, relative to its generally critical style, Consumer Reports presents European cars significantly better.
than expected. Here it is the statistically significant (.9924) bias (c=.207) in favour of Volkswagen that dominates this group.

In contrast, there is a considerable bias against US cars. The high significance (.9990) leaves no doubt that Consumer Reports, given their generally imbalanced style, presents American cars consistently worse than predictable from their formal ratings.

We also explored whether the apparently unfair reporting on American cars would be upheld when US produced US brands are separated from captive imports (foreign made US brands) on the one side and contrasted with imports and US produced foreign cars on the other side. Figure 3 answers this question affirmatively:

![Diagram showing biases towards country of production and brand](image-url)
Since the group of US produced cars consists of Volkswagens only, the aforementioned bias in their favour cannot be generalized beyond this one German car manufacturer. There is some evidence for bias in favour of imported cars. Even though this bias is not large, it can still be considered statistically significant (.9682). But, even after removing Chrysler's captive import from the group of American cars, Consumer Reports' bias against US produced US brands of cars is still evident. We conclude that, compared with their ratings and considering what the journalistic style leads us to expect, Consumer Reports significantly (.9972) downgrades US designed cars.

Finally, Figure 4 depicts the biases towards individual US automobile manufacturers:
The coefficient of bias for Chrysler automobiles is practically zero. There is no evidence of any bias in favour or against this manufacturer. Reports on these cars are as critical as on all cars in our data. We conclude that Consumer Reports characterizes Chrysler automobiles as fairly as can be.

The bias against American Motors is not insubstantial but at the same time not significant (.3480). We are therefore unable to judge whether more data would confirm the negative bias or show the findings to be statistically rare. Thus, we are lead to conclude that Consumer Reports' bias against automobiles by American Motors is inconclusive.

The data are clearer for General Motors and Ford. Biases against the automobiles of these two manufacturers are similar in magnitude and statistically significant (.9820 for GM and .9510 for Ford). Our unambiguous conclusion is that Consumer Reports unfairly characterizes automobiles from General Motors and Ford, Ford being treated somewhat more unfairly than General Motors, but the difference between the two is not statistically significant (.1982) (see fourth example in the chapter on Bias).

Probably the clearest summary indication of where Consumer Reports directs its most favourable and its most unfavourable bias is found in Figure 5. It presents the group of 30 cars produced by American Motors, Ford and General Motors on the one side and the group of 10 cars by Saab, Toyota and Volkswagen on the other:
Groupings with Most Favourable and with Most Unfavourable Biases

Figure 5

Whereas Saab, Toyota and Volkswagen are found to be the most favorably treated in comparison by Consumer Reports' evaluative writing style, American Motors, Ford and General Motors experience the most unfavourable distortions in the reporting of test results by Consumer Reports. Both of these results are statistically significant at a level beyond reasonable doubt (.9999 for Saab, Toyota and VW, .9996 for AMC, Ford and GM).

While the bias in favour of Volkswagen seems to be the largest for a single manufacturer and resembles that in favour of Saab, Toyota and Volkswagen combined, in absolute terms, this does not
appear unfair at all. Looking at the data for Volkswagen (see Appendix D), there are 29 cases in which verbal evaluations and formal ratings perfectly coincide, 10 in which the text is more favourable than the ratings and 11 in which it is the other way around. This nearly ideal symmetrical distribution around the zero-difference is reflected in a very small coefficient of balance (the coefficient of bias with the zero-difference as standard) of \( c_0 = -.020 \). With reference to the ideal of approximating "the truth" or a balance around "the facts" which ignores what we considered seriously in the foregoing analysis, Consumer Reports' generally critical style, we find the following absolute biases:

\[
\begin{align*}
\text{Volkswagen alone} & \quad c_0 = -.020 \\
\text{Saab, Toyota, and VW} & \quad c_0 = -.061 \\
\text{All cars in the data} & \quad c_0 = -.264 \\
\text{AMC, Ford and GM} & \quad c_0 = -.358
\end{align*}
\]

Thus, considering "fairness" in absolute terms, it does not appear that Consumer Reports presents Volkswagen including Saab and Toyota more favorably than justified by their ratings. Evidently, the reporting on these cars is least distorted, most balanced around the zero-difference between test scores and ratings, and hence most fair. Consumer Reports' generally critical style of reporting, which we took as the baseline of our bias measure, is unfair to most other automobiles tested, by presenting the results in terms worse than their ratings would justify. But in the reporting on cars by
American Motors, Ford and General Motors combined, unfavorable distortions are far more severe than could be explained by this style. Thus we conclude that, also in absolute terms, Consumer Reports clearly treats automobiles by Ford and General Motors most unfairly.

As a final comment, we assumed throughout that Consumers Union's formal rating would be independent of the verbal part of the reports and that only the verbal evaluations could be subject of prejudiced distortions whereas the ratings could not. However, since we are not informed exactly how the ratings come about, for example, whether the writers that exhibit some of the preferences we found in the foregoing also test-drive and thus influence the ratings, the ratings could be prejudicially distorted as well. If this were to be the case our results may not indicate the full strength of the bias actually existing.
References


APPENDIX A

The Measuring Instrument
(Coding Instructions)
For each automobile evaluated in Consumer Reports there is a separate data sheet. The coder completes boxes 1 through 8 and the 17 boxes of the column labeled "text." The four boxes to the left of that column are for convenience only and the remainder of the boxes are completed in a separate step.

Following are the codes for the first 8 boxes:
<table>
<thead>
<tr>
<th>Coder Identification Number, Column 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Jane Jorgenson</td>
</tr>
<tr>
<td>2  Tom Potterfield</td>
</tr>
<tr>
<td>3  Deb Giffen</td>
</tr>
<tr>
<td>4  Cathy Kirkland</td>
</tr>
<tr>
<td>5  Klaus Krippendorff</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>June 82</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>July 82</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Aug, 82</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Car Name, Columns 2-3</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Jan. 81</td>
</tr>
<tr>
<td>01 Dodge Aries / Plymouth Reliant, 2.2-liter sedan, automatic</td>
</tr>
<tr>
<td>02 , , , , , 2.6-liter wagon, automatic</td>
</tr>
<tr>
<td>03 , , , 2.2-liter coupe, manual</td>
</tr>
<tr>
<td>04 Buick Skylark / Chevrolet Citation, 2.5-liter Four sedan, automatic</td>
</tr>
<tr>
<td>05 , , , 2.8-liter V6 hatchback, autom.</td>
</tr>
<tr>
<td>Febr. 81</td>
</tr>
<tr>
<td>06 Buick Century V6</td>
</tr>
<tr>
<td>07 Oldsmobile Cutlass V8, diesel</td>
</tr>
<tr>
<td>08 Ford Granada 6</td>
</tr>
<tr>
<td>09 Chrysler LeBaron 6</td>
</tr>
<tr>
<td>March 81</td>
</tr>
<tr>
<td>10 Toyota Corolla Tercel 4-dr</td>
</tr>
<tr>
<td>11 Mazda GLC 4-dr hatchback</td>
</tr>
<tr>
<td>12 Plymouth Horizon 4-dr hatchback</td>
</tr>
<tr>
<td>13 Ford Escort 4-dr wagon</td>
</tr>
<tr>
<td>14 Ford Escort 2-dr hatchback</td>
</tr>
<tr>
<td>May 81</td>
</tr>
<tr>
<td>15 Volkswagen Jetta</td>
</tr>
<tr>
<td>16 Renault 181</td>
</tr>
<tr>
<td>June 81</td>
</tr>
<tr>
<td>17 Honda Prelude</td>
</tr>
<tr>
<td>18 Toyota Corolla SR5</td>
</tr>
<tr>
<td>19 Datsun 200SX</td>
</tr>
<tr>
<td>20 Mazda 626</td>
</tr>
<tr>
<td>July 81</td>
</tr>
<tr>
<td>21 Peugeot 505S</td>
</tr>
<tr>
<td>22 Saab 900S</td>
</tr>
<tr>
<td>23 Volvo 240S</td>
</tr>
<tr>
<td>24 Toyota Cressida</td>
</tr>
<tr>
<td>25 Datsun 810 Maxima</td>
</tr>
<tr>
<td>Aug. 81</td>
</tr>
<tr>
<td>26 Toyota Starlet</td>
</tr>
<tr>
<td>27 Plymouth Champ</td>
</tr>
<tr>
<td>28 Honda Civic 1300</td>
</tr>
<tr>
<td>29 Subaru</td>
</tr>
<tr>
<td>30 Datsun 210 MPV</td>
</tr>
<tr>
<td>Sept. 81</td>
</tr>
<tr>
<td>31 Volkswagen Rabbit, diesel</td>
</tr>
<tr>
<td>32 Chevrolet Chevette, diesel</td>
</tr>
<tr>
<td>Jan. 82</td>
</tr>
<tr>
<td>33 Chevrolet Cavalier CL, 4-door</td>
</tr>
<tr>
<td>34 Chevrolet Cavalier, Station wagon</td>
</tr>
<tr>
<td>35 Chevrolet Cavalier ME, 2-door</td>
</tr>
<tr>
<td>36 Pontiac J2000 SE, Hatchback</td>
</tr>
<tr>
<td>Febr. 82</td>
</tr>
<tr>
<td>37 Dodge Aries SE</td>
</tr>
<tr>
<td>38 Pontiac Phoenix</td>
</tr>
<tr>
<td>39 Ford Fairmont Futura</td>
</tr>
<tr>
<td>40 AMC Concords DL</td>
</tr>
<tr>
<td>March 82</td>
</tr>
<tr>
<td>41 Volkswagen Rabbit L</td>
</tr>
<tr>
<td>42 Mazda GLC Custom L</td>
</tr>
<tr>
<td>43 Mazda GLC Custom</td>
</tr>
<tr>
<td>44 Plymouth Horizon Miser</td>
</tr>
<tr>
<td>45 Mercury Lynx L</td>
</tr>
<tr>
<td>May 82</td>
</tr>
<tr>
<td>46 Nissan Stanza</td>
</tr>
<tr>
<td>47 Honda Accord</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>21</td>
</tr>
<tr>
<td>22</td>
</tr>
<tr>
<td>23</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>26</td>
</tr>
<tr>
<td>31</td>
</tr>
<tr>
<td>41</td>
</tr>
<tr>
<td>42</td>
</tr>
<tr>
<td>51</td>
</tr>
<tr>
<td>52</td>
</tr>
</tbody>
</table>
Production/Design or Brand, Column 6

1. US production of a US designed car
2. US production of a foreign designed car
3. Foreign production of a US designed car
4. Captive imports (foreign production of a foreign car marketed as a US brand)
5. Imports (foreign production of a foreign designed car marketed as a foreign car)
Car Size, Column 7

Here consult the classification of cars in Consumer Reports' April 1982 issue and code:

1 small cars
2 medium cars
3 large cars
Here consult Consumer Reports' April 1982 issue and code:

0    not recommended (car name in black)
1    recommended (car name in red)

### At a glance: CU's preferred cars
(Listed in alphabetical order)

<table>
<thead>
<tr>
<th>SMALL CARS</th>
<th>MEDIUM CARS</th>
<th>LARGE CARS</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUDI 4000</td>
<td>TOYOTA COROLLA</td>
<td>PEUGEOT 505</td>
</tr>
<tr>
<td>CHEVROLET CHEVETTE</td>
<td>TOYOTA COROLLA TERCEL</td>
<td>SAAB 900</td>
</tr>
<tr>
<td>DATSUN 200SX</td>
<td>TOYOTA CORONA</td>
<td>TOYOTA CRESSIDA</td>
</tr>
<tr>
<td>HONDA ACCORD</td>
<td>TOYOTA STARLET</td>
<td>VOLVO</td>
</tr>
<tr>
<td>HONDA CIVIC</td>
<td>VOLKSWAGEN JETTA</td>
<td></td>
</tr>
<tr>
<td>HONDA PRELUDE</td>
<td>VOLKSWAGEN RABBIT DIESEL</td>
<td>LARGE CARS</td>
</tr>
<tr>
<td>MAZDA 626</td>
<td>MEDIUM CARS</td>
<td>BUICK ELECTRA</td>
</tr>
<tr>
<td>MAZDA GLC</td>
<td>AUDI 5000</td>
<td>BUICK LE SABRE</td>
</tr>
<tr>
<td>NISSAN STANZA</td>
<td>DATSUN 810 MAXIMA</td>
<td>CHEVROLET IMPALA/CAPRICE</td>
</tr>
<tr>
<td>TOYOTA CELICA</td>
<td>MERCEDES-BENZ 240D</td>
<td>OLDSMOBILE DELTA 88</td>
</tr>
<tr>
<td></td>
<td>MERCEDES-BENZ 300D</td>
<td>OLDSMOBILE 98</td>
</tr>
</tbody>
</table>
Now we consider what is said about various aspects of each car and describe it in terms of 17 subject matter categories. The definitions of these subject matter categories are taken from Consumer Reports' April 1982 article on "How CU tests and rates cars" (enclosed as part of the coding instructions). The coder must thoroughly familiarize himself with this document and consult it and the following elaborations as often as needed and particularly whenever uncertainties arise. All verbal evaluations and arguments that are relevant to a subject matter category and contained in a report on the car in question is to be considered. These verbal descriptions may be found in the introduction, in the final recommendations and particularly in the main body of text dealing with each car separately.

We elaborate on the scales of measurements used, give special rules for summarizing the verbal content, define how each subject matter should be handled and finally include the above mentioned article for reference.
All descriptive accounts are assumed to have some evaluative information. By Consumer Reports' own characterization, if it were not outstanding in any way it wouldn't be worth mentioning it. So, a strictly neutral statement is not to be expected.

We define a single evaluative scale in three ways, corresponding to the logic of the description underlying the particular verbal report. Evaluations seem to be made (a) relative to a standard, (b) in terms of absolute judgements and (c) in a comparative mode.

(a) Relative to a standard. If this evaluation scheme is employed, we compare the standard spelled out in Consumers Reports April 1982 issue with the way the subject matter is characterised in the report and record whether and how the car deviates from the standard. Categories are:

(5) far exceeds standard, commendable, exemplary, ideal (non-exceedable)
(4) somewhat above standard, better than expected, more-than-adequate
(3) matches standard, adequate
(2) somewhat below standard, worse than expected, less than adequate
(1) far below standard, objectionable

(b) In terms of absolute judgements. Here we attempt to map the attributes used to describe a subject matter into a kind of polar scale of favourable-unfavourable terms or expressed preferences. In determining which scale value is appropriate we consider how much better or worse the writer can describe the car.

(5) excellent
(4) very good
(3) good (2.5) is considered the neutral point of unexceptional = not mentioned facts
(2) fair
(1) poor

(c) In a comparative mode. This scheme is applied when the writer compares two cars, gives his rank order of the car in the context of all others or indicates where in the distribution of (all or the majority of) other cars the one is located. Here we think of all cars as grouped into five categories and make an effort to record in which category the writer sees the car being evaluated with reference to other cars.

(5) the highest scores, in the best group, exceptionally good
(4) clearly above average, largely positive points
(3) positive points exceed negative ones
(2) negative points exceed positive ones average & unexceptional
(1) clearly below average, largely negative points, exceptionally bad

Note that if a report states that one car is better than the other, this does not mean that they have to be given different scale values. They might still end up in the same group of cars when compared with all the other cars.
Rules and Procedure

(1) In all three ways of defining the scale, 2.5 is a kind of neutral point which should not occur in a single statement because this would carry no evaluative information and should be rare when there is a group of statements, perfectly balanced.

(2) When the verbal descriptions seem to be a clear tie between two neighboring categories, we take the scale value of the more extreme category. E.g. a statement "... is judged to be between poor and fair" becomes (1).

(3) When nothing is said concerning a subject matter to be judged, we enter a blank or a (0). However, this zero should not be confused with a scale value or interpreted as the opposite of (5).

(4) There is no need to assume that the three ways of defining a scale are mutually exclusive. We examine the statements made concerning the subject matter to be recorded and decide which ways of defining the scale points are appropriate. If more than one are applicable we apply them all and take their separate results into account. Actually, there rarely are conflicting results. Usually one of the definitions is more appropriate with the other reinforcing or modifying the former's value. The three ways are intended to accommodate the writer's idiosyncratic style of evaluative reporting.

(5) Usually, a report on a set of cars starts with an introduction and concludes with recommendations. Both contain general statements about the car as a whole and outline the advantageous and disadvantageous features. These are important in setting the stage of the more detailed criticism and praise. Usually the more detailed accounts for each car separately is merely modifying the over-all judgements offered in the introduction and summarized in the recommendations. To come as close as possible to the writer's style of evaluative reporting, if he makes generalizations, we first describe evaluations in these categories:

- engine and transmission
- handling and braking
- comfort
- convenience

and then go into the details. (E.g. an "extremely comfortable car" should not end up having only below average scale values in ride, noise, driving position, passenger's seats and climate control. We expect that evaluators and testers are picky. In finding how an over-all scale value is modified we consider how many things are often complained about and relative to that, how many or few the writer finds worth mentioning in describing the car in question.

(6) When general statements are absent, we judge both the severity of the criticism and praise as well as the number of things talked about. When Consumer Reports doesn't like the horn switch on the handle of the signal and writes about it at length, this doesn't make all controls bad, as there are many others which are assumed to be (2.5) when not mentioned. Balancing positive points against negative ones must consider the weights given to each argument and consider those not mentioned.

(7) The availability of options as such is neutral except when the manufacturer recommends them highly and when Consumer Reports tested a car that had them and evaluated them in the report. Note that the reports are always about a particular car Consumers Union chose to purchase and test.
Special reference points in the definition of SCALES for subject matter evaluations largely obtained from Consumer Reports April 1982. This source should be consulted in case uncertainties arise.

**Fuel Economy**

CR judges this generally in a comparative mode. Accordingly small cars can reach a higher fuel economy (5) than medium size cars (3) or large cars (1). CR usually notes differences within these groups which will change the scale values in the direction indicated in the report.

**Engine Drivability**

Ideally, (5), a car's engine has "excellent drivability." It starts quickly, without stalling. It doesn't stumble or hesitate even while it's cold.

Any deviation from this ideal might reduce the scale value depending on the significance of the fault found.

(3) is reserved for a generally adequate engine, one that performs well, maybe after some minor starting difficulties.

(1) is reserved for engines that hesitate to acceleration, lose power easily or splutter in deceleration etc.

**Shifting**

CR defines two standards as its ideal (5). A manual transmission should shift easily with little effort and precisely and the clutch should engage smoothly rather than abruptly. An automatic transmission should shift without making a lurch.

Any faults found reduce the scale value below the ideal with a mere "adequate transmission" given a (3).

**Acceleration**

Acceleration is generally judged by comparison (there is no upper limit, e.g. "hot-rod acceleration" which CR does not demand) but there is also a minimum acceptable standard, given a (3) which CR defines as "to merge safely with expressway traffic and to pass slow vehicles without undue drama."

Favourable attributes like "peppy" add to the value of (3), unfavourable attributes subtract from it. Cars with the best acceleration get a (5).

**Accident-avoidance ability**

Consult CR for the tests of handling under emergency conditions described in this subject matter category.

(5) refers to cars that negotiate their course even at relatively high speeds with no noticeable instabilities.

(3) refers to cars that are essentially judged safe under emergency driving conditions. There may be some deviations from the ideal but this is not judged significant.

(2) refers to cars where safety is said to be a problem.

Note that "plowing off course instead of recovering from a skid" (3) is considered to be less of a failure than if "a car's rear end swings out and threatens to spin" (2).
Handling Precision refers to steering and control of the car under normal driving conditions.

(5) means not exceedable in responsiveness without being over-sensitive, provides excellent feel for the road while easy to handle. The car should track perfectly and maintains its course even in crosswinds.

(3) means handling is "competent", car's handling is still safe despite noted imperfections.

Braking

(5) would indicate that a car is outstanding in the qualities listed in (4).

(4) refers to a car that stops quickly, its brakes are not touchy and do not fade when warm. They do not squeek. Brakes in this "very good" category do not show any weaknesses.

(3) refers to a car whose brakes are described as merely adequate, perhaps with some less desirable defects noted but not unsafe or seriously defective.

Ride

(5) means riding comfort is described as excellent and under nearly all driving conditions.

(4) means riding is very comfortable with some problems judged as minor (e.g. bumpiness on rough roads).

(3) means an essentially good and steady ride. Harsh responses to road conditions, jolting, etc. reduces the scale values.

Noise

Since the ideal of a total absence of noise is hardly achievable, we rely heavily on absolute judgements and comparisons made in the report. (3) is the acceptable or slightly above average level of noise.

Occasional noises (e.g. when starting a diesel engine) weigh less than permanent noises (e.g. engine and transmission noise during highway driving).

Driving position

refers to the driver's ability to maneuver comfortably, his leg room, head room, seat, etc. When the driver's seat is not especially mentioned but the front seat (for two) is, we take the description of the front seat into consideration when coding the driving position.

The ideal (5) is ample room, extremely comfortable seat giving good support, etc. It is easy to get in and out. Seatbelts are easy to handle and do not hinder the driver's mobility.

Front Seats (for 2)
Rear Seats (for 2)

(5) sitting comfort, leg and head room is mentioned in superlatives, upholstery is velour, seats are adjustable for people of various sizes, seats give excellent support, seatbelts are easy to handle.

(3) is reserved for seats that are judged basically adequate with still tolerable drawbacks.
Climate Controls

The ideal (5) is a climate-control system that is powerful, draft-free, and responsive to its controls. The fresh-air ventilation outlets should send plenty of air evenly throughout the car. The heater should cope with 0°F weather without creating hot spots. If the car has an air-conditioner, it should keep you cool when the temperature is in the 90's. If the air-conditioner is integrated with the heater, so much the better; on cold days, you can then quickly and comfortably defog the windows.

Note that the availability of an air conditioner is not a plus. Its performance decides the value given. Any conditions that are described to be less satisfactory reduces the scale value, including when the climate control system doesn't respond quickly and predictably to changes in control settings.

Controls

refer to the steering wheel, and all levers, pedals and switches (including for the radio) the driver has to operate before and after driving the car.

(3) is reserved for essentially good controls with few if any outstanding features noted.

Positive or negative features are weighted by the magnitude of the consequences of failures as well by the frequency of use.

Displays

refers to the speedometer, indicators for turn signal, gas, battery, etc. essentially all information-bearing configurations the driver has to observe.

(5) is reserved for displays that are as clear as can possibly be. There are no distracting reflections, excellent illumination at night, no obstacles to reading them, unambiguous in meaning.

Any deviation from the ideal reduces the scale value.

(3) is reserved for essentially adequate well-designed displays.

Servicing Ease

The ideal (5), is that all parts including the engine that need checking, servicing or some maintenance work the car owner might engage in himself are easily accessible, that a manual provided for such care is clear, and the operations required are simple and call for few uncomplicated tools. The jack is easy to use. The spare tire can be ready without undue effort.

The lack of adequate instructions (absence of or unclear manual), obstructions and any negation of the features deemed ideal reduce the scale value.

Mechanical Reliability

refers to the amount of trouble the owner of a car is likely to experience. In the absence of data, CR sometimes refers to past experiences with a similar model which are taken here instead of the test results. All judgments are taken to be comparative:

(5) means the car has an excellent (i.e. extremely low) frequency of repair record, rarely needs repair work during the first few years and comes with few sampling defects.

(3) has an average frequency of repair record.

Note: the number of sampling defects weigh less than the record of repairs. Generally, we look for how much trouble the car makes for the driver.
Finally, completely separate from the independent coding task outlined above, and after joint sessions during which consensus on a single value for each unit has been reached among all coders involved, we enter the formal ratings into the data sheet using the following convention:

<table>
<thead>
<tr>
<th>Excellent</th>
<th>Very Good</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>⭐️⭐️⭐️⭐️⭐️</td>
<td>⭐️⭐️⭐️⭐️⭐️</td>
<td>⭐️⭐️⭐️⭐️</td>
<td>⭐️⭐️⭐️</td>
<td>⭐️⭐️</td>
</tr>
</tbody>
</table>

(5) (4) (3) (2) (1) (0)=rating unavailable

We then compute and enter the difference: text score less rating in the rightmost column of the data sheet. These differences may range between -4 and +4.

As a special rule, if either the value for the text and/or the value of the rating is zero (meaning unavailable or no information), enter the number "9" instead of the difference.
How CU tests and rates cars

CU's auto-test engineers are driving their lives away. At least, sometimes it seems that way. They test more than 35 cars a year for our reports, driving each car for at least 300 hours.

While other drivers may spend time behind the wheel listening to the radio or taking in the scenery, CU's drivers are constantly making mental notes about the car. They consider aspects of performance, safety, comfort, and convenience that please or disappoint them. The tested cars are on trial at all times—whether they’re being driven down the road to pick up lunch or swerving around traffic cones at the test track.

The chosen ones

The cars we choose to test are generally small and medium-sized models built to accommodate at least four adults. This year, we’re looking at the new models coming out of Detroit: the Chrysler K-car, reported on in January, and the Ford Escort, covered in March. Of course, we also test imported cars, such as the Toyota Corolla Tercel and the Mazda GLC (also in our March report). Few large cars are sold these days. But from time to time we examine what’s available.

No matter what the size of the car, we try to buy a variety of body styles: sedans, hatchbacks, station wagons.

Some sporty cars also go through our tests, as do some luxury cars. Occasionally, we look at vans or small trucks.

We also test diesel-powered cars, of course. And we always try to do new models. When General Motors' J-cars are introduced later this year, we hope to be the first on our block to buy one.

We test the cars in groups. To rate one car fairly against another, we try to test each car in a group with a comparable engine, transmission, and suspension. We aren’t fussy about options. We’ll buy a car that has some frills we wouldn’t ordinarily recommend and note how the extras affect the car’s character. Sometimes we seek out a bare-bones, base-price car, to see what that’s like.

We never test a demonstration model or a preproduction car put together for “press introductions.” We test only cars consumers can buy. And we buy our own from dealerships. If we’re lucky, the car we want is in stock. If not, we order.

Getting to know them

As the cars arrive at CU’s auto-test division, the unusual ones often find themselves the center of attention for a while. But with all cars, the attention goes beyond the superficial.

The auto-test staff carefully inspects every system in the car—more than 50 items in all—to be sure each works properly. The staffers try the lights, radio, heater and fan, door locks and windows, rear-view mirrors, seat adjustments, and safety belts. They check fluid levels, from engine oil to windshield-washer fluid. And they note cosmetic flaws—poorly finished interiors and improperly attached trim.

Anything that isn’t as it should be is considered a sample defect. Serious problems that could affect the car’s performance in our tests require a trip back to the dealership for repairs.

Measuring comes next. That supplies much of the information for our monthly Facts and Figures chart. We measure people room and cargo room. We weigh the car with a full tank of fuel and determine front and rear weight distribution. We verify the high-gear ratio, which is the relationship between how hard the engine works and how fast the wheels turn. That affects fuel mileage, and we want to be sure we’ve received the gear we ordered.

Finally, we make sure the car is tuned to factory specifications.

On the road

Then the driving begins. Our engineers put at least 2000 miles on each car before doing any formal testing. The first 500 miles or so includes as much variable-speed, stop-and-go driving as possible. That helps to give the engine, transmission, and brakes a proper break-in.

The cars receive no special tender, loving care. They rarely see the inside of a heated garage. That way we can find out how they start when cold. We don’t let them idle until they’re warmed up. We want to know how the cold engine performs. (Besides, idling wastes fuel.)

The engineers use the cars as their personal cars. They switch cars every day so that each engineer spends about the same amount of time—and develops the same level of familiarity—with each car.

The auto testers take the cars to the supermarket and on family outings. They find out what it’s like to load groceries into the trunk, what happens when a lanky son tries to get in or out of the rear seat, and how the air-conditioner works when it’s 90°F and the car is stuck in beach traffic. Those are the kinds of real-world experiences ordinary drivers go through every day.

When there’s enough mileage on the car, it’s readied for formal testing.

In those tests, the engineers evaluate each car in 15 areas, covering more than 80 characteristics. Here’s how we arrive at our judgments:

Economy and performance

Fuel economy is one of the most important characteristics of a car. With a special meter hooked up to the car, we test fuel mileage in three ways: on a 195-mile trip, in simulated city driving, and at a constant expressway speed.

The 195-mile trip takes us over highways and back roads. Groups of cars travel in a convoy. City driving is done over a one-mile course. Each car goes through it at the same speed and with the same stops and starts. The highway driving is done at a constant 55 mph over a fairly flat, 2-mile stretch. We cover the
course in each direction to minimize the effects of wind and grade variations.

To calculate cruising range, our engineers multiply the fuel mileage on the trip by the amount of fuel the car holds. They subtract 30 miles to allow enough fuel to get to a filling station.

To give an idea of what it will cost to run each car for a year, we note in Facts and Figures the number of gallons of fuel required to go 15,000 miles, typically about a year’s worth of driving. (We calculate the figure from equal parts of our three mileage tests.) To get a dollar figure, multiply that number by whatever fuel costs in your area.

To measure acceleration, we run the cars on the straightaway of a test track. A fifth wheel (shown in the photo) accurately measures the car’s speed. We record times for 0 to 30 mph, 0 to 60 mph, and, for passing, 45 to 65 mph.

We aren’t looking for jackrabbit starts. We do look for acceleration that’s strong enough to permit the car to merge easily, pass safely, and get out of the way of a crazed 18-wheeler.

No special tests are needed to evaluate a car’s transmission. The everyday driving tells our engineers what they need to know. In a car with an automatic transmission, shifting should be imperceptible. A manual should shift crisply, and the clutch should engage smoothly.

A car shouldn’t need constant steering corrections to maintain a straight-ahead course. That ability is called tracking, and a car that does it well scores points.

A car shouldn’t be severely deflected from its course by bumps in the road or by crossing. It should respond to the driver’s steering instructions—whether those are normal maneuvers or emergency measures—without overreacting.

Our 195-mile trip course has an assortment of bumps and sharp turns, and enough wide-open road for our engineers to evaluate a car’s normal behavior. At the test track, the engineers see how the car responds in an emergency. Several different tests are run at the track.

In one test, we use traffic cones to denote lanes. Our test drivers take each car through the course at gradually increasing speeds. At a predetermined point, they swerve left (without braking or accelerating) to avoid hitting an imagined object. Then they swerve back into the lane. The faster the car can negotiate the course without knocking down cones, and the more controllable it remains, the better it scores.

Braking tests are done at the track, too. We accelerate until the car reaches 60 mph, then measure how short a stop it can make without having its wheels lock, causing the car to skid. Cars that swerve under hard braking don’t score very well in this test. Touchy brakes that throw passengers around earn demerits, too.

When brakes become hot from repeated use, they often fade, or lose efficiency. Then the driver has to step harder on the brake pedal to stop the car. To test for fade, we accelerate to 60 mph and then brake firmly, but not electromatically. We do this 10 times in a row. Each time, an instrument measures the pedal effort required. If we note serious fade or other problems, we downgrade braking.

People pleasers

The overall comfort of a car depends on many factors: ride; noise; seat comfort and roominess; heating, air-conditioning, and ventilation. We check all those.

Driving over a 20-mile course that has a variety of road surfaces tells us a lot about the comfort of the ride. We first make the run with only two people in the car. We do it again with the maximum load specified by the automaker. The ride can be soft or firm and still be good. But it should never be harsh or jolting.

We use a high-quality tape recorder and microphone in the car to record engine, transmission, wind, and road sounds. The tapes are analyzed by computer to determine relative noisiness.

The testers who evaluate seating range in height from 5 feet to 6 feet 1 inch and in weight from 90 to 250 pounds. They judge the comfort of the seats and seatbacks and the ease of getting in and out, front and rear. They also evaluate the comfort of the driving position and the accessibility of the pedals and controls.

In a car’s interior climate, we look for uniform temperature distribution, freedom from drafts, and responsive controls. Defrosters should clear windows quickly and keep them clear for safety.

For convenience’s sake, gauges, warning lights, and other sources of vital information should be easily seen and legible day and night. Controls should be logically laid out, accessible, and easy to operate. If they aren’t, we report their shortcomings. We favor features such as intermittent settings on wipers, high beams that can be flashed when the headlights are on or off, and, in many hatchbacks and station wagons, rear-window wipers and washers.

An otherwise fine car can be marred by stingly trunk space. We stow luggage and put tall grocery bags into the car’s trunk to see how much it holds.

We try changing a tire with the car’s own equipment. And, for do-it-yourselfers who go beyond tire changing, we point out inconveniences we find under the hood. In the published report, we also note the automaker’s recommended routine servicing schedule.

Safety and reliability

Safety is one of our chief concerns. Readers say reliability is one of theirs.

Tests of all sizes try on a car’s safety belts. We watch for belts that are positioned so they won’t inflict injury while trying to prevent it. Belts should also be convenient to use and comfortable to wear. And they should be able to secure a child’s safety seat or infant carrier.

Head restraints should protect tall people as well as short people. If the restraints can do that without causing blind spots for the driver, so much the better. (Most cars have blind spots, often caused by roof pillars. We call them as we see them.)

The fuel tank should be well protected from damage in a crash. Since the left side of a car is more often hit in sideswipe accidents, the best location for the fuel filler is high on the right side of the car.

Federal regulations require that a car’s bumpers adequately protect body, grillwork, and safety components such as lights in 5-mph front and rear impacts, and in 3-mph corner impacts. And the bumpers must do it without suffering significant damage themselves. Our bumper basher, a massive steel battering ram, simulates such impacts. We measure and record the damage done. We also note how much the car rebounds. Too vigorous a response could cause injuries to people in the car.

The best safety information would come from actual crash-testing. The National Highway Traffic Safety Administration has made a commendable beginning on this addition to available consumer information. But, as we explain in the report on page 188, the testing is not yet at the point where it can contribute to our overall evaluation of cars.

For a line on the potential reliability of cars, we depend on you. This year, our Annual Questionnaire generated information from our readers on 360,000 cars. The experiences they have had with car repairs give us a good clue as to what to expect from cars old and new. The information is part of each car’s report and is compiled each April as our Frequency-of-Repair Records (see page 225).

Tallying up

We conclude our reports by itemizing sample defects—those on the car when we received it and those that show up along the way. Such defects are peculiar to the sample we purchased, not the model. While the number of defects may be an indication of an automaker’s quality control, they don’t count against the car in our overall judgment. We have yet to buy a perfect new car.

All the judgments about a car are assigned a numerical value, given an appropriate weight, and fed into a computer. The computer tells us the car’s overall score. Specific characteristics are judged on a scale from excellent to poor. That information is included in each month’s Ratings chart.
CU's test program for the 35 or so models we test each year always begins the same way—we buy the cars from dealers. Like the average car buyer, we haggle over price, and we return the car to the dealer for service and repairs. That way, we get an ordinary assembly-line sample to test—and we gain insight into the problems that the average person has when dealing with a car dealer.

Because of the importance of fuel economy, we concentrate on small and medium-sized models. Sedans, hatchbacks, station wagons—we test all body styles. Occasionally, we test large models, sporty specialty cars, even pickup trucks and vans.

Generally, we test a group of several similar models for a monthly report. Sometimes, we buy more than one sample of the same model—perhaps a "loaded" version and a stripped one—so we can determine which options are useful and how they affect performance.

Once we take delivery, our engineers and technicians swarm over the car to see that its systems are working according to the manufacturer's specifications. They also take the measurements that appear in the Facts & Figures chart with each annual test report.

Some assembly-line errors that they find are merely amusing. Several years ago, we took delivery of a car that had Plymouth trim on one side and Dodge trim on the other. Some errors, however, could distort our test results if undiscovered. For example, the wrong rear-axle ratio would affect our acceleration and fuel-economy tests.

The break-in period

No matter how carefully a car is manufactured, it takes a while for its many components to "set," to wear in a little so the car performs at its best. The break-in process applies not only to the engine, but to the transmission and axle gears, the brakes, and even the tires. For that reason, we drive each car 2000 to 4000 miles before beginning our formal testing.

For those miles, our engineers and technicians drive each car as if it were their own. They use the cars for supermarket shopping trips, weekend visits to the in-laws, and shuttle service to the dentist and piano teacher. They switch cars frequently so they can compare the models within a test group.

That sort of day-to-day driving tells us what it's like to live with each car. Is the car maneuverable in heavy downtown traffic? Is it easy to park? Is the trunk deep enough to hold grocery bags upright? Will all 260 pounds of Uncle Fred fit into the back seat?

The formal tests

After the break-in period, we check each car again and make adjustments so it conforms to the manufacturer's specifications. Then our formal tests begin.

We run some tests—for ride comfort and fuel economy, for example—on public roads. We also spend several days at a test track, doing the tests that are too dangerous for public roads. An added benefit of testing at the track is that conditions are the same for each car.

If the results of a test are unexceptional, we may not mention them in our report. But every car is tested the same way. Here are the main areas we check for our report each month:

Engine and transmission: Ideally, a car's engine has excellent "drivability." It starts quickly, without stalling. It doesn't stumble or hesitate, even while it's cold.

A manual transmission should shift easily and precisely, and the clutch should engage smoothly rather than abruptly. An automatic transmission should shift without making a lurch.

We don't demand hot-rod acceleration from the cars we test, but we do expect them to be peppy enough to merge safely with expressway traffic and to pass slow vehicles without undue drama. We record acceleration times from 0 to 30 mph, 0 to 60 mph, and, for passing, from 45 to 65 mph. And we see what speed a car can reach from a standing start on a quarter-mile straightaway. A fifth wheel clamped to the rear of the car accurately measures the car's speed for these tests.

Fuel mileage counts heavily in a car's overall test score. With a precise flow meter installed in the fuel line to measure fuel consumption, we drive the cars in a convoy on a 195-mile trip that includes stop-and-go driving, rural roads, and expressways. We also perform a city-traffic mileage test, which simulates downtown traffic. And we measure fuel mileage at a constant 55 mph over a fairly flat four-mile stretch; we run the test in both directions to minimize the effects of wind and slight variations in grade. The results of those tests are combined in our estimate of fuel used over 15,000 miles, a typical year's worth of driving.

Handling and braking: Most motorists rarely get to explore the high-speed handling ability of their car. They're more concerned with how the car handles in normal driving. There, a car should be nimble and responsive to its steering without being twitchy and overly sensitive. The steering shouldn't be too heavy, nor should it be so light that the driver can't tell from the "feel" of the wheel when the front wheels lose traction.

A car should "track" well. It should maintain its course, even in a crosswind, without the need for constant corrections with the steering wheel. A bumpy corner shouldn't make the car's front or rear end skitter to the side.
Although the average motorist may not often be concerned with a car's high-speed handling, CU's engineers are. Even the most conservative drivers may need to make an abrupt avoidance maneuver if they're cut off by another driver. We run several different emergency-handling tests at the track.

One course, outlined by traffic cones, tests a car's response to a simulation of the classic hazard of a child chasing a ball into traffic. The maneuver comprises a sharp swerve into the left lane and an equally sharp swerve back to the right lane. To eliminate as many variables as possible, our drivers go through the course without using the accelerator or the brakes. The drivers gradually increase the speed at which they enter the course until the car can no longer go through without knocking down cones.

The speed at which the car can negotiate the course is an important test criterion. So is the car's mode of failure. Some cars blow off course instead of recovering from the second swerve. We consider that the safest type of failure. Not as safe is the car whose rear end swings out, threatening a spin.

Another test that reveals a lot about a car's emergency-handling ability involves driving at various speeds through the test track's assortment of turns. Some cars "run wide" during turns—that is, they don't respond enough to the steering wheel and tend to slide toward the outside of the corner. Other cars respond too strongly to the steering wheel and turn in too sharply.

One of our braking tests, run at the track, involves accelerating to 60 mph and seeing how quickly the car can stop without having any of its tires skid; cars that do some skidding lose points.

Brakes become hot when they're used repeatedly or continuously, especially at high speed or down a long incline. Heat causes brakes to "fade," or lose their efficiency; that means you have to press harder and harder on the pedal to get the same amount of braking. In extreme cases, the car may not stop at all, no matter how hard you apply the brakes. To test for fade, we repeat this sequence 10 times: We accelerate to 60 mph, travel one-third of a mile, and brake moderately. One instrument measures braking effort; another, the rate of slowing. A sharp increase in the pedal effort required means a bad mark for the brakes.

Brakes can also be too sensitive. In some cars, a tap of the brake pedal results in a disproportionately hard stop. Touchy brakes are downgraded.

**Comfort:** A 20-mile circuit of roads with a variety of surfaces tells us a lot about ride comfort. We run through the course twice—once with two people in the car, the second time with the maximum load of people and luggage specified by the automaker.

On another ride, a high-quality tape recorder in the car picks up noise from engine, drivetrain, wind, road, and tires on both smooth and coarse pavement. Later, a computer helps us determine the relative noiseiness of the ride.

Some cars' driving position and seating accommodations are comfortable only for average-sized people; others favor short or tall riders. And some seats are designed with enough adjustments to make almost any-sized person happy. The testers who evaluate interior comfort and convenience range in height from 5 feet to 6 feet 1 inch and in weight from 90 to 210 pounds.

A car's climate-control system should be powerful, draft-free, and responsive to its controls. The fresh-air ventilation outlets should send plenty of air evenly throughout the car. The heater should cope with 0°F weather without creating hot spots. If the car has an air-conditioner, it should cool you when the temperature is in the 90's. If the air-conditioner is integrated with the heater, so much the better; on cold, dank days, you can then quickly and comfortably defog the windows. We take off points if a climate-control system doesn't respond quickly and predictably to changes in its control settings.

**Convenience:** For safety as well as convenience, controls should be easy for the driver to understand, reach, and use. As an example, we consider the steering wheel a better place for the horn button than the end of the turn-signal stalk. Gauges should be clear and easy to read, unblocked by the steering wheel and free of distracting reflections.

While the size of a car's trunk is indisputably important, so is its shape. A boxy trunk with a flat floor has more useful luggage space than, say, a long, shallow trunk with a multilevel floor. A large opening and a low sill make loading and unloading easier, and a trunk lid that opens high enough so you can bash your head on it is an added blessing.

In a hatchback or a station wagon, the rear seat should be easy to fold down. Here, too, the opening should be large and low, and the seat should fold flat.

In some cars, you can remove the spare tire without first unloading most of the luggage in the trunk. We give extra points for such convenience. We also check whether the jack that comes with the car is easy to use. We prefer the old-fashioned full-service spare, which is found on fewer and fewer cars these days, to the spares designed to limp only as far as the nearest service station.

For motorists who perform at least some of the regular maintenance on their car, we also point out servicing inconveniences that we find under the hood and elsewhere.

**Other considerations:** In this category, we detail important safety evaluations, predictions about mechanical reliability, and any findings that don't fit neatly into our other categories.

We check the front and rear safety belts to make sure they're convenient to put on and comfortable to wear—and to judge their potential for protecting the wearer in a crash. We try the belts on a child safety seat, to see whether they can secure it. And we see whether the head restraints are high enough to protect even tall riders from whiplash. We note any obstructions, such as wide roof pillars and high windshields, that might interfere with the driver's view.

Until this year, we checked our cars' bumper protection by simulating a 5-mph crash with a hydraulic bumper basher. But the bumpers on current models are so effective that we've discontinued the test. We credit the Government's bumper standard for the improved performance (the standard requires cars to survive 5-mph front and rear crashes without significant damage). However, if the standard is weakened, as is possible under the present Administration's proposals, we may have to resume our bumper tests.

Mechanical reliability is probably the most important criterion to many car buyers. For information on that, we depend on you, the reader. This year, our Annual Questionnaire elicited reports on the repair experiences of 351,000 cars. The data collected from current and previous questionnaires allow us to predict how the current version of a car will hold up in the years to come. If we have reports on enough samples of a particular model, the information is included in the auto report and is compiled each April as our Frequency-of-Repair charts (see page 198).

We conclude each auto report by totaling the number of sample defects—those we discover when we take delivery of the car and those that show up during our ownership. Although another sample of the same model may not suffer from the same defects, our list does give the prospective new-car buyer an idea of what could go wrong. We have yet to buy a car that doesn't have at least a few defects. Some cars arrive with dozens.

After our testing is done, we score the cars in each of 76 categories, weight each score according to the importance of the test, and then feed the numbers into a computer. From that, we get the cars' scores and an overall Ratings order for the group tested. To make it easier for you to shop, we also list base prices and prices of major options for the models tested and some of their competitors.
APPENDIX B

Data Sheet
<table>
<thead>
<tr>
<th>Coder ID</th>
<th>Car name(s)</th>
<th>Manufacturer's label</th>
<th>Production</th>
<th>Size by CU's classification</th>
<th>CU's 1982 recommendation</th>
<th>blank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Engine &amp; transmission</th>
<th>fuel economy</th>
<th>engine drivability</th>
<th>shifting</th>
<th>acceleration</th>
<th>accident avoidability</th>
<th>handling precision</th>
<th>breaking</th>
<th>ride</th>
<th>noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>engine &amp; transmission</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comfort</th>
<th>driving position</th>
<th>front seat (for 2)</th>
<th>rear seat (for 2)</th>
<th>climate control</th>
<th>controls</th>
<th>displays</th>
<th>servicing ease</th>
<th>repair incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Convenience</th>
<th>displays</th>
<th>servicing ease</th>
<th>repair incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C

Results of Reliability Tests
For the agreement statistics used see Krippendorff (1980). The tests involved three independently coding analysts.

The instrument was developed in conjunction with three preliminary reliability tests. The fourth represents the final over-all data reliability:

- 1st test using 60 units of analysis: a was not computed
- 2nd test using 90 units: a = .7067
- 3rd test using 198 units: a = .7447
- Final test using 935 units: a = .8175

The last test was completed before publication of the August 1982 issue of Consumers Reports which contained 4 more cars and units of analysis. The over-all data reliability of .8175 thus represents 93% of the data actually analysed.

To test whether the data contained major discontinuities, we computed the agreements reached on all subject matter categories separately involving 55 units (= automobiles in the reliability sample) each:

- Fuel economy: a = .7733
- Engine drivability: a = .8489
- Shifting: a = .7638
- Acceleration: a = .8131
- Accident-avoidance ability: a = .7616
- Handling precision: a = .7334
- Braking: a = .8468
- Ride: a = .6005
- Noise: a = .6944
- Driving position: a = .7508
- Front seating: a = .7264
- Rear seating: a = .7094
- Climate control: a = .7522
- Controls: a = .6792
- Displays: a = .7447
- Servicing ease: a = .7382
- Predicted repair incidence: a = .9491

Since there were no strong discontinuities among these variables we included the measures on all 17 subject matter evaluations in our data and accepted the results as sufficiently reliable. During the first three tests three additional variables had been included. They did not achieve acceptable levels of reliability.

Finally we further improved the quality of the data by asking the three coders, after all cars had been coded, to discuss all disagreements that emerged during the separate process and agree on a single value, reconsidering all details a coder may have omitted and correcting any deviance from the collective judgement. It is therefore safe to conclude that the data we analysed are reliable at least at the a=.8175 level and most probably exceed this figure.
APPENDIX D

Tables of Findings
<table>
<thead>
<tr>
<th>n</th>
<th>Major Groups of Cars</th>
<th>Frequencies of Differences</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-3</td>
<td>-2</td>
</tr>
<tr>
<td>59</td>
<td>All cars, the data:</td>
<td>8</td>
<td>61</td>
</tr>
<tr>
<td>37</td>
<td>Size - small</td>
<td>8</td>
<td>39</td>
</tr>
<tr>
<td>22</td>
<td>- medium</td>
<td>22</td>
<td>99</td>
</tr>
<tr>
<td>23</td>
<td>Recommended cars</td>
<td>5</td>
<td>124</td>
</tr>
<tr>
<td>30</td>
<td>Production - US made US</td>
<td>8</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>- US made foreign = VW</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>- Foreign made US</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>25</td>
<td>- Imports</td>
<td>11</td>
<td>138</td>
</tr>
<tr>
<td>31</td>
<td>Maker - American cars</td>
<td>8</td>
<td>48</td>
</tr>
<tr>
<td>21</td>
<td>AMC+Ford+GM</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>American Motors</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>Ford</td>
<td>13</td>
<td>38</td>
</tr>
<tr>
<td>13</td>
<td>General motors</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>Chrysler</td>
<td>22</td>
<td>38</td>
</tr>
<tr>
<td>21</td>
<td>- Japanese cars</td>
<td>8</td>
<td>117</td>
</tr>
<tr>
<td>5</td>
<td>Honda</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>Mazda</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>Nissan-Datsun</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>1</td>
<td>Subaru</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>Toyota</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>7</td>
<td>- European cars</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>- French cars</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>Peugeot</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>Renault</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>- German Volkswagen</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>- Swedish cars</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>1</td>
<td>Saab</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>Volvo</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>Saab + Toyota + VW</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td>29</td>
<td>Every second car</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Every second unit</td>
<td>4</td>
<td>34</td>
</tr>
</tbody>
</table>

\[ c_0 = \]
The following lists measures of bias separated by subject matter categories:

<table>
<thead>
<tr>
<th>Measures</th>
<th>Frequencies of Differences</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(-3) (-2) (-1) (0) (1) (2)</td>
<td>(c) (p) (c) (0)</td>
</tr>
<tr>
<td>Fuel economy</td>
<td>4</td>
<td>28 27</td>
</tr>
<tr>
<td>Engine drivability</td>
<td>4</td>
<td>9 31 13 2</td>
</tr>
<tr>
<td>Shifting</td>
<td>7</td>
<td>15 28 8</td>
</tr>
<tr>
<td>Acceleration</td>
<td>3</td>
<td>4 12 20 1</td>
</tr>
<tr>
<td>Accident-avoidance ability</td>
<td>1</td>
<td>20 34 3 1</td>
</tr>
<tr>
<td>Handling precision</td>
<td>18</td>
<td>34 7</td>
</tr>
<tr>
<td>Braking</td>
<td>1</td>
<td>22 35 1</td>
</tr>
<tr>
<td>Ride</td>
<td>1</td>
<td>5 38 15</td>
</tr>
<tr>
<td>Noise</td>
<td>2</td>
<td>15 40 2</td>
</tr>
<tr>
<td>Driving position</td>
<td>5</td>
<td>34 19 1</td>
</tr>
<tr>
<td>Front seating (for two)</td>
<td>1</td>
<td>24 33 1</td>
</tr>
<tr>
<td>Rear seating (for two)</td>
<td></td>
<td>38 20 1</td>
</tr>
<tr>
<td>Climate control</td>
<td>8</td>
<td>32 15 4</td>
</tr>
<tr>
<td>Controls</td>
<td>7</td>
<td>19 30 2</td>
</tr>
<tr>
<td>Displays</td>
<td>4</td>
<td>13 31 10</td>
</tr>
<tr>
<td>Servicing ease</td>
<td>3</td>
<td>31 19 6</td>
</tr>
<tr>
<td>Predicted repair incidence</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>All cars</td>
<td>8</td>
<td>61 305 458 119 3</td>
</tr>
</tbody>
</table>