Covariation Assessment by Consumers

JAMES R. BETTMAN
DEBORAH ROEDDER JOHN
CAROL A. SCOTT*

Two experiments were conducted to investigate the effect of prior beliefs and information format on consumers' assessment of the relationship between price and quality for four frequently purchased grocery products. In these studies, consumers were shown sets of data, each of which presented ranks of 10 brands of a product category on price and quality. Contrary to prior research on illusory correlation, consumers' estimates of covariation were relatively accurate and unaffected by the availability of relevant prior beliefs about the nature of the relationship between price and quality for grocery products in general or by format manipulations that varied the ease or difficulty of processing the data. These findings are discussed in terms of the effect of detailed instructions, the availability of simple heuristics for processing rank-ordered data, differences between social and consumer perceptions, and the stages of consumer information processing most likely to be affected by prior beliefs.

Among the types of beliefs that consumers hold about the marketplace, those concerning covariation are particularly important because of their prevalence and potential impact on behavior (cf., Bettman, John, and Scott 1984; Duncan and Olshavsky 1982). Covariation beliefs, in general, refer to those beliefs regarding the degree of relationship or association between two events or concepts. Beliefs such as "foreign cars are made better than domestic ones," "big cars are safer than smaller cars," "a national brand name does not always guarantee freshness," and "you get what you pay for" are just a few examples of the types of associations consumers may form in daily interactions with the marketplace. Once in place, such beliefs can influence a wide range of consumer activity, as they may enable consumers to explain past events, control present outcomes, and predict future occurrences.

Questions regarding the formation, modification, and persistence of covariation beliefs have received surprisingly little attention from consumer researchers. In one of the few studies done on this topic, John, Scott, and Bettman (1986) found that consumers appeared to gather product information that was likely to be consistent with their prior beliefs. In particular, consumers who believed that higher prices were associated with higher quality sampled brands with higher average prices. However, there has not been any research to date explaining how consumers assess covariation given such data. The purpose of this article is to study several determinants of covariation assessment by consumers. Specifically, we report the results of two experiments designed to investigate the general proposition that consumers' covariation assessments are a function of prior beliefs *and* the ease of processing new or incoming data. This proposition was studied in the context of one particular type of covariation beliefs relevant to consumer choice: beliefs about price–quality relationships (see Monroe and Petroshius 1981 and Olson 1977 for reviews).

BACKGROUND

The most pervasive finding in the literature on covariation assessment conducted by social psychologists (see Alloy and Tabachnik 1984 and Crocker 1981 for reviews) is that individuals are often poor judges of covariation. The evidence suggests that individuals asked to assess the degree of covariation present in a given data set will tend to misinterpret the data in line with their previous beliefs. Note that in this line of research, participants are asked to assess the degree of relationship in a given set of data, not to combine their prior beliefs with the data to derive a posterior estimate. In reviewing

*James R. Bettman is Burlington Industries Professor of Business Administration, Fuqua School of Business, Duke University, Durham, NC 27706. Deborah Roedder John is Assistant Professor, Graduate School of Business, University of Wisconsin, Madison, WI 53706. Carol A. Scott is Associate Professor, Graduate School of Management, University of California, Los Angeles, CA 90024. This research project was supported by the Academic Senate of UCLA, the Fuqua School of Business at Duke University, the Graduate School at the University of Wisconsin, and the Division of Research at Harvard University. The authors are listed in alphabetical order. All editorial decisions concerning this article, including selection of reviewers, editing, and final acceptance, were carried out by Harold H. Kassarjian in his role as Co-Editor.

© JOURNAL OF CONSUMER RESEARCH • Vol. 13 • December 1986

much of this literature, Nisbett and Ross concluded that (1980, p. 10):

> Prior theories or expectations may be more important to the perceptions of covariation than are the observed data configurations.

More recently, Alloy and Tabachnik (1984) have challenged this view. They argue that *either* prior beliefs *or* incoming data may dominate covariation judgments, depending on the relative strength and accessibility of each. This view, based on an extensive review of studies in the areas of human and animal information processing, rests primarily on the observation that people have been reasonably accurate estimators of covariation in situations wherein they have no strong beliefs (e.g., Beach and Scopp 1966; Erlick and Mills 1967). Accuracy appears to deteriorate, however, when prior beliefs conflicting with incoming data are present or made particularly relevant (e.g., Chapman and Chapman 1967; Jenkins and Ward 1965; Smedslund 1963; Ward and Jenkins 1965). Although the evidence is more limited, Alloy and Tabachnik (1984) also argue that the availability and diagnosticity of incoming data affect accuracy. Alloy and Tabachnik (1984) develop a 2 × 2 table by dichotomizing strength of prior beliefs and strength of incoming data. They argue that individuals' covariation assessments will be dominated more by prior beliefs if such beliefs are strong and available and incoming data are ambiguous or difficult to process. Conversely, if prior beliefs are not present or relevant and the data are clear and easily processed, covariation beliefs will depend more on the data. If both prior beliefs and data are weak, individuals are predicted to refrain from making covariation assessments or to do so with low confidence. Finally, if both prior beliefs and data are strong, assessments will be made with high confidence if beliefs and data agree; however, individuals will be faced with a cognitive dilemma if the priors and data disagree. As the brief review above indicates, in previous studies individuals often have resolved this dilemma by making assessments more in line with their prior beliefs.

This framework can be applied to the current studies. We might expect, for example, that consumers' prior beliefs about the relationship between price and quality would have differing degrees of influence on subsequent covariation judgments depending on the relevance and availability of these prior beliefs *and* the strength or ease of processing of incoming data depicting price-quality relationships. In particular, the research reported here provides a direct test of the relative influence of both prior beliefs and incoming data by manipulating the presence–absence of prior beliefs and the strength of incoming data within the context of the same study.

STUDY 1

The first experiment was designed to test the general hypothesis that assessment of covariation would be affected by prior beliefs about the relationship to a greater extent when these beliefs were made more available and the incoming data were difficult to process. When prior beliefs were absent, however, and/or the incoming data were easy to process, we hypothesized that covariation assessments should reflect the correlation inherent in the data and should not vary as a function of prior beliefs. Hence, the experimental design was developed so that the availability of prior beliefs, the ease of processing, and the specific levels of correlation in the incoming data were manipulated. In particular, the experimental design had three between-subjects and two within-subjects factors. The between-subjects factors were *prior beliefs* about price–quality relationships (positive or high versus neutral or low), the *timing* of the product category information (product category named either before or after each price–quality data set was presented), and the data-presentation *format* (random versus ordered). Each participant saw the same four sets of data twice, first *labeled* "*y*" and "*z*," and then labeled "*price*" and "*quality*," and the sets of data were characterized by two *levels* (moderately high and low) of actual correlation to make up the within-subjects factors. Two product categories were included as replicates for each correlation level to make up the four sets of ranks. For each set of data, subjects were asked to give their estimates of the degree of correlation between the two characteristics.

Thus, the availability of prior beliefs was manipulated in several ways. No relevant prior beliefs should be available for the first set of *y* and *z* estimates. For the second set, however, the *price* and *quality* labels should cue prior beliefs regarding price and quality covariation, while the timing of product-category information should make prior beliefs for the specific product categories more (product category named before data) or less (product category named after data) available when estimates were being made. Finally, the data were easier to process for half of the participants, who received data ordered according to one of the characteristics instead of in a randomly determined order. Thus, one main goal of the experiment was to manipulate the relative ease of accessing prior beliefs and the data.

While a main effect of level of actual correlation was expected (higher estimates should be given for higher actual levels), the more theoretically interesting hypotheses are those that involve the relationship between prior beliefs and data. Specifically, we expected a prior belief × label interaction, with estimates more in line with prior beliefs (e.g., higher estimates for those with beliefs that price and quality are related) only when the data were labeled "*price*" and "*quality.*" Further, when the data were labeled "*price*" and "*quality,*" we expected that prior beliefs would affect covariation estimates to a greater extent when the product category was named prior to seeing the data than when the category was named later, because prior beliefs would be more available in the former condition (cf., Crocker and Taylor

1978). This implies a prior belief × category timing interaction within the *price–quality* conditions. Finally, both of these effects may be qualified by the interaction of a third term, the data presentation format. That is, prior beliefs may affect covariation estimates only when prior beliefs are available (price–quality label, product category named before data) and the data are difficult to process (i.e., are presented in random order). When the data are easier to process, there may be little, if any, effect of prior beliefs. This implies a prior belief × category timing × format interaction within the *price–quality* condition.

Method

One hundred and seventy consumers were recruited to participate in this study. Individuals were solicited by telephone by a marketing research firm, with only those who reported doing most or all of the grocery shopping for their family or household asked to participate. Participants were scheduled at a central research facility in groups of approximately 20 and were each paid $10 for 45 minutes of their time. Because all experimental manipulations were included in the questionnaire booklet, individuals (and not groups) were randomly assigned to all experimental conditions except those for prior beliefs. Prior beliefs in this study were measured rather than manipulated, as discussed further below.

Procedure. Participants were seated at tables and given a questionnaire booklet reflecting the experimental condition to which they had been assigned. Detailed instructions on the task were included in the first part of the booklet to explain ranking and introduce the idea of assessing relationships between sets of rank-order data. The reason for using rank-order data was our belief that quality information is usually available only on an ordinal or ranking basis. Several examples of relationships between two sets were presented. For each example, subjects were shown two sets of ranks and the rating on a −10 to +10 scale, labeled at each end (−10 = perfectly negative relationship; +10 = perfectly positive relationship) and in the middle (0 = no relationship). The degree of relationship between ranks shown on the scale was based on Spearman's rho, but the formula for this computation was not included in the instructions. Participants were told that they would be judging relationships similar to these and were instructed to be as accurate as possible in their responses. No mention was made of price–quality relations in these instructions. These detailed instructions and examples and the emphasis on accuracy were provided to eliminate two possible artifactual reasons for lack of accuracy in judging covariation: motivation, and lack of understanding regarding the concept of relationships (Alloy and Abramson 1979; Shaklee and Tucker 1980).

TABLE 1

DATA SETS FOR STUDY 1

Data set characteristics	Set 1	Set 2	Set 3	Set 4
Ranks on characteristic *z* (*y* and *z* condition) or on price (*price–quality* condition)	1	3	3	1
	3	2	10	4
	6	9	8	9
	7	6	2	2
	2	4	5	5
	4	10	4	7
	8	5	7	3
	9	7	9	8
	10	8	6	6
	5	1	1	10
Spearman's ρ	.64	.10	−.16	.58
Kendall's τ	.56	.11	−.11	.47
Product category for *price–quality* condition	Beef hot dogs	Creamy peanut butter	Vegetable oil	Italian salad dressing

NOTE: Assume the ranks on Characteristic *y* (*y* and *z* condition) or on quality (*price–quality* condition) are from 1 to 10 consecutively.

Next, participants were given four sets of data, each set consisting of 10 items ranked on two characteristics labeled "*y*" and "*z*." For each set, subjects examined the two columns of rank-order data, turned to the next page, and estimated the degree of relationship for the preceding data on the −10 to +10 scale just described. No mention was made of price–quality relationships at this point to ensure that these ratings would be made without reference to subjects' prior price–quality beliefs. Subjects were asked to estimate only the relationship shown in the data and were not asked for a combined estimate of any prior opinion and the data. The order in which the four sets of data appeared was counterbalanced to control for possible order effects. These four sets included two sets with actual rank-order correlations near zero and two sets with rank-order correlations near 0.6 (see Table 1 for actual sets used). For half of the participants, the data were ordered from 1 to 10 on characteristic *y* (the ordered format). The other half saw the pairs of ranks listed in a random order (the random format).

After completing all four sets of *y* and *z* estimates, participants were asked to indicate their beliefs about price–quality relationships for several types of familiar grocery products on the same type of −10 to +10 scale used previously. For this task, subjects were not given sets of rank-order data, but rather were asked to give their ratings based on their current beliefs about price and quality. These ratings, obtained for eight specific grocery product categories and for grocery products in general, were needed to split the sample into different prior belief groups in subsequent analyses.

Following these ratings, subjects completed several filler tasks intended to clear memory for the four data sets rated earlier. These included a set of 25 self-description, agree–disagree statements and several sets of questions about product opinions and usage. The product questions included sets of questions about price and quality differences in the various product categories used. Together with the prior beliefs measures and the price–quality labels, such questions were intended to make prior price–quality beliefs salient for the participants.

Next, subjects were asked to rate price–quality relationships for four sets of data representing four different product classes. These data sets were the same as those previously presented, but this time subjects were told that the rankings represented price and quality rankings for brands in their area. Brands were identified by the letters A through J. Quality ranks for these brands were provided in the left column, and price ranks were listed in the right column. Although these price–quality rankings were taken from actual *Consumer Reports* data, the source was not identified to eliminate possible bias due to subjects' beliefs about *Consumer Reports.* Respondents rated these four data sets in an order different from the one used for the first rating task, but saw the data in the same format. That is, a respondent saw both the y and z and the *price–quality* data in either a random or an ordered format. In addition, the specific product category name (beef hot dogs, creamy peanut butter, etc.) was placed at the top of the data set page for half of the participants (product category identified *before* data presentation). For the other half, the specific product category was not named until subjects turned to the page containing the scale (product category identified *after* data presentation). Participants completed these tasks and filled out the questionnaire at their own pace. Most respondents completed the questionnaire in about 35–45 minutes. When finished, subjects were debriefed, thanked, and paid for their participation.

Product Selection. Four product categories were selected, two with levels of price–quality correlations near zero and two with levels near 0.6. The levels of correlation were computed from data in *Consumer Reports.* The brands rated by *Consumer Reports* were checked for availability in local supermarkets to ensure consistency between the *Consumer Reports* data and brand information potentially available to local shoppers. The four product categories selected were: beef hot dogs (Spearman $\rho = 0.64$), creamy peanut butter ($\rho = 0.10$), vegetable oil ($\rho = -0.16$), and Italian salad dressing ($\rho = 0.58$). Both beef hot dogs and Italian salad dressing had positive correlations significantly different from zero, whereas the correlations for creamy peanut butter and vegetable oil were not significantly different from zero.

Dependent Measures. Two types of dependent measures were included in this study: covariation estimates and task ratings. Covariation estimates were collected by having subjects mark a line 100 millimeters in length on a -10 to $+10$ scale. For more precision, these markings were measured to the nearest millimeter and then converted back to the -10 to $+10$ range.

A second set of measures examined subjects' perceptions of the difficulty and complexity of the task. Subjects rated the task on six semantic differential scales: complex–simple, pleasant–unpleasant, easy–hard, confusing–clear, hard to follow–easy to follow, and difficult to complete–not difficult to complete. These measures were included to check subjects' reactions to the format conditions.

Results

Two groups with different prior beliefs regarding the degree of relationship between price and quality were formed for each product category using the responses to the category-specific -10 to $+10$ scale and for grocery products in general using the response to the general product -10 to $+10$ scale. Respondents below the median value for the total sample on each scale were assigned to the "low" prior belief group, and respondents above the median were assigned to the "high" prior belief group.[1]

Task Ratings. Task ratings were analyzed to determine how the format manipulation affected subjects' ability to process the rank-order data. A 2 (low or high priors) $\times$ 2 (ordered or random format) $\times$ 2 (timing of category information) between-subjects analysis of variance was performed on the six task-rating questions. The random format was seen as more complex than the ordered format ($\bar{X} = 2.9$ vs. 3.4, $F(1,158) = 5.54$, $p < 0.02$), harder than the ordered format ($\bar{X} = 4.9$ vs. 4.2, $F(1,157) = 6.13$, $p < 0.02$), and harder to follow than the ordered format ($\bar{X} = 3.7$ vs. 4.2, $F(1,160) = 3.72$, $p < 0.06$). In addition, subjects felt the timing-after condition was harder to follow than the timing-before condition ($\bar{X} = 3.7$ vs. 4.3, $F(1,160) = 4.98$, $p < 0.03$). Using the formula $r = [(F)\ (df\,\text{effect})/((F)\ (df\,\text{effect}) + df\,\text{error})]^{1/2}$ as a measure of effect size, the values for r are 0.18, 0.19, 0.15, and 0.17 for complexity, hard, hard to follow, and timing hard to follow, respectively.

Estimates of Covariation. Two sets of analyses were performed to examine predictions about relative effects of the presence or accessibility of prior beliefs and the strength or ease of processing of incoming data. In the first set of analyses, the effects of prior beliefs were examined by comparing covariation estimates when the data were labeled "*price*" and "*quality*" vs. "y" and "z." In the second set of analyses, the effects of prior

[1]The median values, on a $-10/+10$ scale, were: 5.4 for beef hot dogs (52.3 percent of the sample), 4.6 for creamy peanut butter (51.8 percent of the sample), 3.6 for vegetable oil (50.6 percent of the sample), 1.6 for Italian salad dressing (51.8 percent of the sample), and 4.6 for grocery products in general (50.9 percent of the sample).

TABLE 2

MEAN VALUES FOR COVARIATION RATINGS FOR STUDY 1

	Low prior beliefs				High prior beliefs			
	Low correlation		High correlation		Low correlation		High correlation	
Format	*y–z*	Price–quality	*y–z*	Price–quality	*y–z*	Price–quality	*y–z*	Price–quality
Ordered								
Mean values	−.19	.36	2.96	3.45	.11	.55	4.08	4.12
N	88	88	88	88	80	80	80	80
Random								
Mean values	−.26	−.15	3.17	3.61	−1.19	.43	3.47	3.99
N	84	86	84	86	84	84	84	84

NOTE: The mean values are for ratings on a −10 to +10 scale, with −10 equal to perfectly negative relationship and +10 equal to perfectly positive relationship.

beliefs were examined by comparing covariation estimates for the *price–quality* data when the product category was named before *price–quality* data were presented vs. after *price–quality* data were presented.

The first analysis was a three within-factor (*price–quality* or *y–z label,* moderately positive or low actual *level* of correlation, and *replicates* or products nested within levels), two between-factor (positive or neutral general *prior beliefs,* ordered or random *format*) mixed analysis of variance, with the dependent measure being the subjects' ratings on the −10 to +10 scale. Prior beliefs about grocery products in general were used because participants might not fall in the same prior belief category (high vs. low) for all four of the sets of data. Thus, use of product-specific prior beliefs would require four separate analyses of variance. In addition, as Alloy and Tabachnik (1984) note, it is not clear whether general or specific prior beliefs are most relevant. Analyses were also carried out with product-specific prior beliefs, but the results were similar; thus, only the general prior belief results are presented here. Timing of product category information was not included since it was not relevant for the y and z condition. Estimates given for price–quality data sets are collapsed over the timing variable for this analysis. Mean scale values for the cells in this design are shown in Table 2.

Contrary to expectations, only three effects were significant. First, there was a strong main effect of level of actual correlation ($F(1,1165) = 367.4$, $p < 0.001$, effect size = 0.49) such that higher levels resulted in higher estimates ($\bar{X} = 3.50$) than did lower actual levels ($\bar{X} = -0.05$). Thus, participants did discriminate between levels. Second, there was an effect of replicates ($F(2,1165) = 16.1$, $p < 0.001$, effect size = 0.16), which indicates that estimates were not the same for product categories within actual correlation levels ($\bar{X} = 0.46$ and -0.55 for peanut butter and vegetable oil at the low levels, and $\bar{X} = 3.02$ and 4.17 for beef hot dogs and Italian salad dressing at the positive levels). Note that these estimates are correctly ordered for the peanut butter and vegetable oil sets, but that the order of the estimates is opposite to that of the actual correlations for the beef hot dogs and Italian salad dressing sets. Finally, there was a main effect of label. Estimates of covariation were higher when the data were labeled "*price*" and "*quality*" ($\bar{X} = 2.04$) than when they were labeled "y" and "z" ($\bar{X} = 1.51$) ($F(1,1165) = 7.67$, $p < 0.006$, effect size = 0.08). This effect was not qualified by prior beliefs, as the prior belief × label interaction was not significant ($F(1,1165) < 1$, n.s.). Further, there was no significant main effect of format ($F(1,165) < 1$, n.s.) and no significant interactions involving format.

The second set of analyses examined the effects of prior beliefs and data format when the relative availability of prior beliefs was manipulated by naming the specific product category either before (more available) or after (less available) the rank-order data were presented. These analyses were conducted only within the price–quality data label conditions, and again used responses to the −10/+10 scale as the dependent measure. The actual design was a two within-subjects (low or high actual *level* of covariation, *replicates* within levels), three between-subjects (positive or neutral general *prior beliefs,* ordered or random *format,* and *timing* before or after data presentation) analysis of variance. The mean values for these conditions are shown in Table 3.

The results of this set of analyses were similar to those of the first set. As before, there was a large main effect of level of actual correlation ($F(1,497) = 181.5$, $p < 0.001$, effect size = 0.52), with low-level sets estimated at 0.03 and high-level sets at 3.78. There was also an effect of replicates similar to that found in the previous analysis ($F(2,497) = 8.35$, $p < 0.001$, effect size = 0.18). In addition, however, there was a level × timing interaction ($F(1,497) = 7.50$, $p < 0.007$, effect size = 0.12). When the product category was named before subjects saw the data, the ratings were higher for the high levels and lower for the low levels ($\bar{X} = 4.28$ and 0.01) than when the product category was not named until after the data were presented ($\bar{X} = 3.28$ and 0.49). No other

TABLE 3

MEAN VALUES FOR COVARIATION RATINGS FOR TIMING ANALYSES

	Low prior beliefs				High prior beliefs			
	Low correlation		High correlation		Low correlation		High correlation	
Format	Before	After	Before	After	Before	After	Before	After
Ordered								
Mean values	.28	.46	4.20	2.62	.30	.80	4.90	3.35
N	46	42	46	42	40	40	40	40
Random								
Mean values	−.24	−.06	3.89	3.33	.05	.77	4.21	3.79
N	44	42	44	42	40	44	40	44

NOTE: The mean values are for ratings on a −10 to +10 scale, with −10 equal to perfectly negative relationship and +10 equal to perfectly positive relationship.

effects reached significance. In particular, there was no evidence of a prior belief × timing interaction ($F(1,161) < 1$, n.s.) or for any interactions involving format.

Discussion

In general, the results show that subjects can discriminate quite clearly between levels of rank-order correlation. Further, subjects' estimates are reasonably accurate. The actual values of subjects' estimates, Spearman's rho, and Kendall's tau respectively for the four sets of data, converted to a −1 to +1 scale for ease of comparison, are 0.30, 0.64, 0.56 for beef hot dogs, 0.05, 0.10, 0.11 for creamy peanut butter, −0.05, −0.16, −0.11 for vegetable oil, and 0.42, 0.58, and 0.47 for Italian salad dressing. Contrary to our hypotheses, there were virtually no effects of prior beliefs or format on the dependent variables. These results are quite surprising in view of the fact that prior covariation assessment studies have typically found strong effects of prior beliefs on covariation estimates. However, there are several features of this experiment that may account for these results.

First, one might argue that the sample used for this experiment represented very positive and moderately positive prior belief levels rather than positive and low or neutral levels. The median scores for all product categories were relatively high (1.6 to 5.4 on a −10 to +10 scale). Although there were people in each belief category who reported a belief corresponding to no relationship or lower (from 15 percent to 31 percent of the sample across various categories), most people felt that there was some degree of positive relationship. Small cell sizes precluded splits at much lower levels, but additional analyses were conducted to examine the interpretation problem caused by this overall positive tendency. A four-way split on prior beliefs and prior beliefs as a continuous (ANCOVA) variable were used in analyses that produced results similar to those reported for the two-group split. Thus, while different findings might be obtained if more subjects with lower prior beliefs were included, there is no evidence for this in the data.

Second, it may be that subjects were accurate in this experiment because the task in all conditions was simply too easy. That is, there were only two levels of correlation and the difference between these levels may have been obvious, particularly since a limited number of rank pairs was presented for each product and all pairs were presented simultaneously. Our second experiment included seven levels of correlation with much smaller intervals between levels.

Finally, one might argue that subjects were able to use simple heuristics to assess covariation for the type of data used here (rank-ordered data) that are less applicable or useful for interval-scaled continuous data. While particularly well suited to marketing contexts, where products are often described in relative terms such as "better quality" or "more expensive," rank-ordered data may have facilitated the use of simple heuristics that happened to produce accurate results with the particular data sets used in the first study. For example, individuals may have simply added up the absolute differences between the ranks on price and quality for the brands ranked highest and lowest on quality. Or, subjects might have added up the absolute differences between the ranks on price and quality for all 10 brands listed. Subjects could then use such sums to estimate the degree of relationship, with large sums of differences leading to lower estimates. By reducing the difficulty of estimating relationships in this manner, heuristics such as these may also reduce the difficulty of dealing with certain information formats and may have contributed to the lack of findings for the format manipulation in the first study. Although subjects in the random format generally rated the task as more difficult than did those in the ordered format, the differences in ratings were relatively small, as indicated by the effect sizes. Subjects' reactions were presumably a function of both the format and the heuristics used.

TABLE 4

STIMULUS SETS OF RANK-ORDER DATA FOR STUDY 2

Data set characteristics	Set number													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Rank on price[a]	10	7	8	4	10	7	9	5	4	1	7	3	4	1
	9	10	10	9	7	5	2	3	6	4	2	1	1	2
	8	9	2	10	5	9	4	9	2	6	1	9	2	3
	5	8	9	6	2	8	3	6	3	9	5	2	3	6
	4	6	5	7	1	6	8	7	5	10	4	6	5	7
	7	5	4	8	9	3	5	8	8	2	3	7	6	4
	6	4	6	3	6	4	10	1	7	5	8	5	7	5
	2	2	7	1	8	1	7	4	10	3	10	4	9	9
	3	1	3	5	4	2	1	2	9	7	6	8	10	8
	1	3	1	2	3	10	6	10	1	8	9	10	8	10
Spearman's ρ	−.89	−.89	−.59	−.60	−.30	−.30	−.01	−.01	.30	.30	.60	.59	.89	.89
Kendall's τ	−.78	−.78	−.42	−.42	−.29	−.29	.02	.02	.29	.29	.42	.42	.78	.78
Discrepancy at extremes[b]	18	13	16	11	16	6	12	4	12	2	7	2	5	0
Discrepancy condition	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
Total discrepancy[c]	46	50	40	46	38	40	30	34	22	30	20	20	10	10

[a] Assume the ranks on quality are from 1 to 10 consecutively.
[b] Discrepancy at the extremes was computed by adding the absolute differences between the ranks on quality and price for those cases ranked 1 and 10 on quality.
[c] Total discrepancy is the sum of the absolute differences on quality and price for all ten cases.

Informal evidence for use of such heuristics emerged in pilot studies, but more direct evidence was sought in Study 2. Diagnostic information on the types of heuristics being used was pursued in the second study by varying additional properties of the data. The purpose of Study 2, therefore, was to provide further insights into possible explanations for subjects' accurate performance in Study 1 by (1) examining a task requiring subjects to discriminate more levels of correlation, and (2) offering some initial ideas about the heuristics subjects use for estimating covariation in rank-order data.

STUDY 2

In Study 2, subjects' prior opinions about the relationship between price and quality for grocery products in general were measured on a -10 to $+10$ scale like that used in Study 1. They then were given 14 different sets of rank-order data, which they were told represented different types of grocery products. These data were presented on a computer monitor in either ordered or random format, and the sets varied over seven different levels of actual rank correlation (-0.9, -0.6, -0.3, 0, 0.3, 0.6, and 0.9).

Diagnostic information on heuristic use was obtained by manipulating one property of the data. Jennings, Amabile, and Ross (1980) argue that people use the extreme values in a set of data as a simple heuristic for assessing relationships. Use of this heuristic was tested in the present study by varying the "discrepancy" at extreme ranks as shown in the stimulus data sets in Table 4. For example, one can see that set 5 has a greater discrepancy value (16) than set 6 (6), where these discrepancy values were obtained by taking the sum of the absolute differences between ranks for quality and price for those cases ranked 1 and 10 on quality. Based on Jennings et al. (1980), one might predict that higher levels of discrepancy at the extremes would lead to lower covariation estimates (i.e., less correlation), since higher differences would be the focus of attention. This might be true especially for the ordered format, where the extreme values are more apparent and easier to spot.

On the other hand, participants may make use of all the data in a stimulus set by focusing on the total sum of the absolute differences between ranks. As can be seen by inspecting the pairs of stimulus sets with equal rank correlations in Table 4, higher discrepancy at the extreme ranks generally results in lower total discrepancy and vice versa. For example, sets 9 and 10 have discrepancies at the extremes of 12 and 2 and total discrepancies of 22 and 30, respectively. Therefore, if subjects follow a heuristic based on the total sum of the absolute differences, higher discrepancy at the extremes (and hence lower total discrepancy) could yield higher covariation estimates. The discrepancy manipulation, then, has the potential to be diagnostic about two of the heuristics subjects may use.

Thus, the factors involved in Study 2 were two within-subjects factors (seven *levels* of actual correlation and low or high *discrepancy* within each level) and two between-subjects factors (ordered or random *format* and low or high *prior beliefs* regarding price and quality relationships). Based on the rationale for Study 1 and the arguments just made, main effects for level and discrepancy were expected, as well as an interaction of prior beliefs with format (i.e., prior beliefs would affect covariation estimates when the data were presented in a random, difficult-to-process format). In addition, a discrepancy × format interaction might be expected, with extreme ranks used more frequently when the data are ordered.

These hypotheses refer to the estimates of correlation for the 14 data sets. One can also measure accuracy for each subject due to the fact that there are now 14 data points per person. Within-subject Pearson correlations were computed between the subject's estimates and the actual values of the correlation for the 14 sets (using both Spearman's rho and Kendall's tau as the actual values). These correlations were then used to assess accuracy as a function of prior beliefs and format. While the random format should decrease accuracy in general, it is not clear what hypothesis should be made about the effects of prior beliefs. Those subjects with high prior beliefs may tend to give higher estimates than those with lower prior beliefs, but which group is likely to be more accurate is difficult to predict and may depend upon the level of the true correlation.

Method

Subjects were 32 graduate students who came individually to a behavioral laboratory and were paid $5 for their participation in the study. After examining the same detailed instructions about judging relationships used in Study 1, subjects were taken to a room with an IBM Personal Computer. An assistant started the computer and inserted a diskette programmed to run the experiment for the between-subjects condition to which each subject had been randomly assigned. The diskette provided instructions for using the computer keyboard and for using the light pen to collect the rating data.

Each participant was asked first to use the light pen to rate his or her prior beliefs about price–quality relationships for grocery products in general on a -10 to $+10$ scale. Next, s/he was given a warm-up set of price–quality data in which quality data were in the left column and price data in the right. Once again, each set had 10 items presented simultaneously, denoted only as A through J, and with the ordered sets ordered on quality. Following this warm-up set, the computer presented the 14 experimental sets in an order created randomly for each subject. For those subjects in the random-format condition, the order of the 10 items within each set was randomized as well. Each set was presented on the monitor screen for 40 seconds. The subject then was asked to mark a -10 to $+10$ scale with the light pen to indicate his or her assessment of the degree of covariation present. The computer recorded the response and went on to the next set.

Selection of Stimulus Sets. A computer program was used to generate sets of rank orders of 10 items with values of Spearman's rho as close as possible to -0.9, -0.6, -0.3, 0, 0.3, 0.6, and 0.9. For each of these levels of correlation, an attempt was made to find two sets of ranks such that (1) the two sets had the same level of Spearman's rho and the same level of Kendall's tau, and (2) the levels of discrepancy at the extremes differed by as much as possible. Note that the levels of discrepancy that are possible to achieve vary by level of correlation. For example, one is much more constrained for sets with high correlations, whether positive or negative. This confounding is undesirable but cannot be avoided. The 14 sets presented in Table 4 best met these criteria.

Results

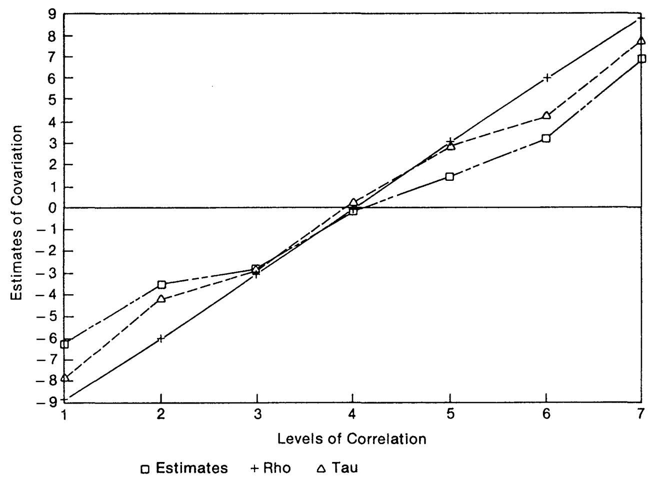
Prior belief groups were created by dividing the responses to the prior belief measure at the median (ratings less than 6 on the -10 to $+10$ scale were low, and ratings 6 or greater were high). This yielded cell sizes of 6, 10, 8, and 8 for ordered, low prior belief; ordered, high prior belief; random, low prior belief; and random, high prior belief cells, respectively. For the covariation estimate data, a two within-factor (seven levels, discrepancy low or high), two between-factor (format ordered or random, prior low or high) ANOVA was run. For accuracy, each subject's ratings for the 14 sets were correlated with the actual values of Spearman's rho and Kendall's tau for those 14 sets. This yielded two correlations per subject. These correlations were converted to z-values using Fisher's r to z transformation, and ANOVAs with format and prior beliefs as independent factors and z-values as dependent variables were run. Although any such measure of accuracy may have problems (Crocker 1981), these correlations provide a more direct measure of accuracy than the ratings.

Ratings Data. The results reveal main effects of format ($F(1,28) = 6.22$, $p < 0.03$, effect size = 0.43), level of correlation ($F(6,364) = 199.6$, $p < 0.001$, effect size = 0.88), and discrepancy ($F(1,364) = 14.9$, $p < 0.001$, effect size = 0.20). There were no significant main effects or interactions involving prior beliefs. Note that the effect size for level is stronger than that in Study 1.

The results for level show a remarkable degree of accuracy (see the Figure). Even with seven different levels, subjects discriminated quite well, although there is a tendency to underestimate higher levels (see Jennings et al. 1980 and Lipe 1982 for similar "conservatism" results).

FIGURE

COVARIATION ESTIMATES AND ACTUAL VALUES OF RHO AND TAU



NOTE: The actual values of rho and tau have been multiplied by 10 to make them consistent with the −10 to +10 scale used for the estimates of covariation.

For discrepancy, high discrepancy at the extremes led to higher estimates ($\bar{X}$ = 0.34 vs. −0.65). As noted earlier, higher discrepancy at the extremes generally means a lower total sum of the absolute differences across ranks for the stimulus sets used here. Thus, this result shows that subjects give higher covariation estimates when there is a lower total sum of the absolute differences across ranks, and hence higher discrepancy at the extremes. Therefore, these data provide evidence that subjects may use a total sum of the absolute deviations heuristic rather than one that entails looking at such deviations only at the extremes.

There was no format × discrepancy interaction. The format main effect shows merely that subjects give higher ratings for the random than for the ordered list ($\bar{X}$ = 0.24 vs. −0.56). The reason for this effect is not entirely clear.

Accuracy Data. Accuracy was generally extremely high. The average correlation between ratings and actual values is approximately 0.9. The ANOVA on accuracy yields very similar results whether Spearman's rho or Kendall's tau is used as the measure of actual correlation. Thus, only the former results are reported. Although accuracy is lower for the random format ($\bar{X}$ = 0.88 vs. 0.91 after reconverting z to r), the effect is not significant ($F(1,28) = 1.88$, n.s.). There is, however, a main effect of prior belief ($F(1,28) = 5.28$, $p < 0.03$, effect size = 0.40), but no format × prior belief interaction ($F(1,28) = 0.32$, n.s.). The direction of the prior belief effect is that subjects with lower prior beliefs are more accurate (after reconverting z to r, $\bar{X}$ = 0.93 vs. 0.86).

Discussion

The results of Study 2 replicate those of Study 1 in important ways. Even with 14 sets of data and seven different levels of actual correlation, participants once again discriminated well between levels of covariation. Although quite different from the sample of individuals participating in Study 1, the sample in Study 2 provided remarkably similar estimates for comparable covariation levels. For data sets with a correlation near 0.6, the mean estimates were 3.78 and 3.28 for Study 1 and 2, respectively. For sets of correlation near 0, the mean ratings were 0.3 and −0.14. Subjects were also found to be fairly accurate estimators of covariation when the correlation of their estimates to a statistical standard was the criterion.

In addition, Study 2 investigated the use of two heuristics for dealing with rank-order data that could perhaps account for the high level of accuracy observed in both studies. Subjects were found to provide estimates of covariation in line with the total sum of the absolute differences between ranks, and not with the discrepancy only at the extreme ranks. As will be discussed, the apparent use of this heuristic may have contributed to consumers' ability to maintain accuracy in covariation assessment regardless of the way the rank-order data were presented.

GENERAL DISCUSSION

The most striking feature of the results of both studies is the documentation of at least one situation in which people can be accurate assessors of covariation. Although there were a few effects of prior beliefs and format, the general pattern of results does not support the Alloy and Tabachnik (1984) predictions. Several potential explanations for these results can be identified and investigated in future research.

Perhaps the simplest explanation is that these studies included more detailed instructions than did many previous ones and included explicit requests to be accurate. Thus, previous results may have been due to misunderstandings of the task and the concept of covariation. Alloy and Abramson (1979), for example, obtained accuracy in some of their conditions when detailed instructions were given, and Shaklee and Tucker (1980) also reported improved performance after instruction in the concept of covariation. The current studies may thus test what consumers are capable of, and not what individuals do naturally without special instructions and motivation.

Second, rank-order data may be easier to handle than either nominal or continuous data. That is, it may be that people in the present study were able to find a simple heuristic for rank-ordered data that was robust under the manipulations used. In particular, if participants used a sum of the absolute differences between ranks heuristic, as the evidence from Study 2 suggests, that heuristic is easy to employ whether the data are random or ordered, regardless of the timing of product category information, or even if data were presented sequentially rather than simultaneously. One simply computes the absolute difference in ranks for each pair and keeps a running total. Thus, subjects using such a heuristic could maintain accuracy fairly easily. It may be, then, that one characteristic of an incoming data stream that would make it "strong" relative to prior beliefs is the ready availability of a reasonably accurate simple heuristic. Thus, presentation of continuous data might lead people to be less accurate.

Bettman et al. (1986) reported evidence consistent with these notions in conditions where no prior beliefs are relevant. They obtained further support for the use of the total sum of absolute differences heuristic and examined the boundary conditions for accuracy of covariation assessment for rank-order data. In particular, subjects were able to maintain high levels of accuracy when data were presented sequentially and when the ranks were presented as words (e.g., first, second) rather than numbers, regardless of format. Accuracy deteriorated only when one characteristic was presented as ranks and the other as actual values (e.g., 159) *and* these rank-value pairs were presented randomly. Hence, there is some support for the contention that use of rank-order data is at least partially responsible for some of the present high level of performance.

Third, prior beliefs about price and quality may not have been particularly salient in the studies conducted here. Although attempts were made to invoke prior beliefs by (1) asking consumers about their price–quality beliefs before they rendered covariation judgments, (2) clearly labeling the data as price–quality data for a specific product category, and (3) including a series of questions about price and quality variation for several product categories, it is possible that these procedures may not have succeeded. It is also possible that prior beliefs about price and quality, whether salient or not, may not exert the same influence as the prior beliefs studied in social cognition. Beliefs or stereotypes about individuals and groups (Hamilton and Rose 1980) may simply be more powerful than those about products, thereby producing more powerful effects on covariation assessment. There may also be a different dynamic for social perception and consumer perception. In the social realm, covariation may be formed after many episodic encounters, where each encounter may be very salient and have relatively high affect. However, consumer covariation beliefs may be based on semantic knowledge that one is told (e.g., you get what you pay for), and not on integration of episodic experiences. Thus, consumers may not relate prior covariation beliefs to current experiences in any systematic fashion.[2] The possible explanations above argue for further inquiry into consumer learning about covariation. It would be particularly useful to do more direct process-tracing studies of the assessment process, using verbal protocols or other means.

A final explanation for these results may be that prior beliefs may influence the covariation assessment process at some point other than the data integration stage examined here. As Crocker (1981) has suggested, prior beliefs may affect judgments at much earlier stages, such as those of obtaining and interpreting data. Indeed, as noted earlier, John et al. (1986) have recently shown that searches conducted to assess the degree of price–quality relationship for specific product categories are systematically affected by individuals' prior price–quality beliefs. Prior expectations have also been shown to bias the interpretation of evidence and quality judgments for product categories where quality is ambiguous or difficult to judge (e.g., Allison and Uhl 1964). Combined with the current results, these studies suggest that the stages of obtaining and interpreting data, and not necessarily the combining of data, may be the locus of covariation assessment problems to the extent that such problems exist. Further research isolating the influence of these stages could add to our understanding of the covariation assessment process.

[*Received December 1985. Revised June 1986.*]

[2]We are grateful to an anonymous reviewer for suggesting this argument.

REFERENCES

Allison, Ralph I. and Kenneth P. Uhl (1964), "Brand Identification and Perception," *Journal of Marketing Research,* 1 (August), 36–39.

Alloy, Lauren B. and Lyn Y. Abramson (1979), "Judgment of Contingency in Depressed and Nondepressed Students: Sadder but Wiser?" *Journal of Experimental Psychology: General,* 108 (December), 441–485.

——— and Naomi Tabachnik (1984), "Assessment of Covariation by Humans and Animals: The Joint Influence of Prior Expectations and Current Situational Information," *Psychological Review,* 91 (January), 112–149.

Beach, Lee Roy and Thomas S. Scopp (1966), "Inferences about Correlations," *Psychonomic Science,* 6, 253–254.

Bettman, James R., Elizabeth H. Creyer, Deborah Roedder John, and Carol A. Scott (1986), "Effects of Presentation Format and Data Properties on Covariation Assessment in Rank Order Data," unpublished paper, Fuqua School of Business, Duke University, Durham, NC 27706.

———, Deborah Roedder John, and Carol A. Scott (1984), "Consumers' Assessment of Covariation," in *Advances in Consumer Research,* Vol. 11, ed. Thomas C. Kinnear, Provo, UT: Association for Consumer Research, 466–471.

Chapman, Loren J. and Jean P. Chapman (1967), "Genesis of Popular but Erroneous Psychodiagnostic Observations," *Journal of Abnormal Psychology,* 72 (June), 193–204.

Crocker, Jennifer (1981), "Judgment of Covariation by Social Perceivers," *Psychological Bulletin,* 90 (September), 272–292.

——— and Shelley E. Taylor (1978), "Theory-driven Processing and the Use of Complex Evidence," paper presented at the American Psychological Association Annual Conference, Toronto, Canada.

Duncan, Calvin P. and Richard Olshavsky (1982), "External Search: The Role of Consumer Beliefs," *Journal of Marketing Research,* 19 (February), 32–43.

Erlick, Dwight E. and Robert G. Mills (1967), "Perceptual Quantification of Conditional Dependency," *Journal of Experimental Psychology,* 73 (January), 9–14.

Hamilton, David L. and Terrence L. Rose (1980), "Illusory Correlation and the Maintenance of Stereotypic Beliefs," *Journal of Personality and Social Psychology,* 39 (November), 832–845.

Jenkins, Herbert M. and William C. Ward (1965), "Judgment of Contingency between Responses and Outcomes," *Psychological Monographs,* 79 (1, Whole No. 594).

Jennings, Dennis L., Teresa M. Amabile, and Lee Ross (1980), "Informal Covariation Assessment: Data-based versus Theory-based Judgments," in *Judgment Under Uncertainty: Heuristics and Biases,* eds. Daniel Kahneman, Paul Slovic, and Amos Tversky, Cambridge, England: Cambridge University Press, 211–230.

John, Deborah Roedder, Carol A. Scott, and James R. Bettman (1986), "Sampling Data for Covariation Assessment: The Effect of Prior Beliefs on Search Patterns," *Journal of Consumer Research,* 13 (June), 38–47.

Lipe, Marlys G. (1982), "A Cross-Study Analysis of Covariation Judgments", working paper, Center for Decision Research, University of Chicago, Chicago, IL: 60637.

Monroe, Kent B. and Susan M. Petroshius (1981), "Buyers' Perceptions of Price: An Update of the Evidence," in *Perspectives in Consumer Behavior, Third Edition,* eds. Harold H. Kassarjian and Thomas S. Robertson, Glenview, IL: Scott, Foresman, 43–55.

Nisbett, Richard E. and Lee Ross (1980), *Human Inference: Strategies and Shortcomings of Social Judgment,* Englewood Cliffs, NJ: Prentice-Hall.

Olson, Jerry C. (1977), "Price as an Informational Cue: Effects on Product Evaluations," in *Consumer and Industrial Buying Behavior,* eds. Arch G. Woodside, Jagdish N. Sheth, and Peter D. Bennett, New York: North-Holland, 267–286.

Shaklee, Hariett and Diane Tucker (1980), "A Rule Analysis of Judgments of Covariation between Events," *Memory and Cognition,* 8 (September), 459–467.

Smedslund, Jan (1963), "The Concept of Correlation in Adults," *Scandinavian Journal of Psychology,* 4 (3), 165–173.

Ward, William C. and Herbert M. Jenkins (1965), "The Display of Information and the Judgement of Contingency," *Canadian Journal of Psychology,* 19 (September), 231–241.