GILBERT A. CHURCHILL, JR., and CAROL SURPRENANT*

The authors investigate whether it is necessary to include disconfirmation as an intervening variable affecting satisfaction as is commonly argued, or whether the effect of disconfirmation is adequately captured by expectation and perceived performance. Further, they model the process for two types of products, a durable and a nondurable good, using experimental procedures in which three levels of expectations and three levels of performance are manipulated for each product in a factorial design. Each subject's perceived expectations, performance evaluations, disconfirmation, and satisfaction are subsequently measured by using multiple measures for each construct. The results suggest the effects are different for the two products. For the nondurable good, the relationships are as typically hypothesized. The results for the durable good are different in important respects. First, neither the disconfirmation experience nor subjects' initial expectations affected subjects' satisfaction with it. Rather, their satisfaction was determined solely by the performance of the durable good. Expectations did combine with performance to affect disconfirmation, though the magnitude of the disconfirmation experience did not translate into an impact on satisfaction. Finally, the direct performance-satisfaction link accounts for most of the variation in satisfaction.

An Investigation Into the Determinants of Customer Satisfaction

The concept of consumer satisfaction occupies a central position in marketing thought and practice. Satisfaction is a major outcome of marketing activity and serves to link processes culminating in purchase and consumption with postpurchase phenomena such as attitude change, repeat purchase, and brand loyalty. The centrality of the concept is reflected by its inclusion in the marketing concept that profits are generated through the satisfaction of consumer needs and wants.

The need to translate the philosophical statement of the marketing concept into pragmatic operational guidelines has directed attention to the development and measurement of consumer satisfaction. In the early 1970s, consumer satisfaction began to emerge as a legitimate field of inquiry. The U.S. Department of Agriculture's Index of Consumer Satisfaction (Pfaff 1972) was the first study to report direct information on consumer satisfaction to policy makers. Both Olshavsky and Miller (1972) and Anderson (1973) examined disconfirmed expectancies and their influence on product performance ratings. These two studies along with Cardozo's (1964) experiment formed the foundation for much of the later theory testing and experimental research.

Since the early 1970s the volume of consumer satisfaction research has been impressive. Numerous theoretical structures have been proposed to examine the antecedents of satisfaction and develop meaningful measures of the construct. The vast majority of these studies have used some variant of the disconfirmation paradigm which holds that satisfaction is related to the size and direction of the disconfirmation experience, where disconfirmation is related to the person's initial expectations. More specifically, an individual's expectations

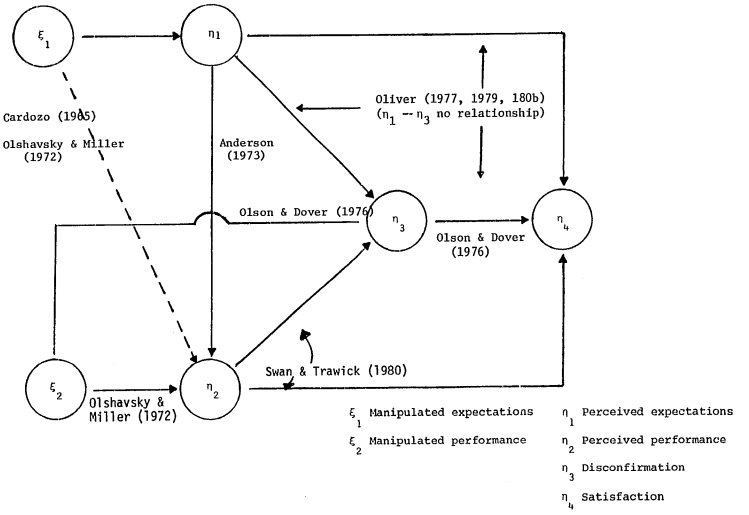
*Gilbert A. Churchill, Jr., is Donald C. Slichter Professor in Business Research, University of Wisconsin-Madison. Carol Surprenant is Assistant Professor, New York University.

The authors acknowledge the many helpful suggestions of three anonymous special issue reviewers and the special issue editor.

The AMA publications policy states: "No article(s) will be published in the *Journal of Marketing Research* written by the Editor or the Vice President of Publications." The inclusion of this article was approved because its review and the publication decision were completely handled by the special issue editor, Richard Bagozzi.

Journal of Marketing Research
Vol. XIX (November 1982), 491–504

Figure 1
LINKAGES TESTED IN PRIOR RESEARCH



are: (1) confirmed when a product performs as expected, (2) negatively disconfirmed when the product performs more poorly than expected, and (3) positively disconfirmed when the product performs better than expected. Dissatisfaction results when a subject's expectations are negatively disconfirmed.

The full disconfirmation paradigm encompasses four constructs: expectations, performance, disconfirmation, and satisfaction. Figure 1 depicts how the concepts are thought to be related and summarizes some of the major satisfaction studies to date.

Expectations. Expectations reflect *anticipated performance.* The satisfaction literature suggests consumers may use different "types" of expectations when forming opinions about a product's anticipated performance. Miller (1977) identified four types of expectations: ideal, expected, minimum tolerable, and desirable. Day (1977) distinguished among expectations about the nature of the product or service, expectations about the costs and efforts in obtaining benefits, and expectations of social benefits or costs.

Performance. The primary importance of performance in the satisfaction literature has been as a standard of comparison by which to assess disconfirmation. Olshavsky and Miller (1972) and Olson and Dover (1976) manipulated actual product performance, but their emphasis was on how performance ratings were influenced by expectations rather than on the impact of changes in performance level on satisfaction. Though it is reasonable to assume that increasing performance should increase satisfaction, the magnitude of the performance effect *vis-à-vis* expectation and disconfirmation effects has not been shown. If performance judgments are assimilated toward expectations, we might expect increases in performance to have relatively little impact on satisfaction if expectations remain constant.

Disconfirmation. In the satisfaction research literature, disconfirmation occupies a central position as a crucial intervening variable. Disconfirmation arises from *discrepancies* between prior expectations and actual performance. It is presumably the magnitude of the disconfirmation effect that generates satisfaction and dissatis-

faction. Oliver (1977) has stressed the importance of measuring disconfirmation apart from expectation, as he maintains the construct has an independent, additive effect on satisfaction.

In the traditional paradigm it is difficult to manipulate disconfirmation independently of expectation and performance precisely because it is *defined* as the *difference* between the two variables. That is, disconfirmation is determined jointly by the combination of the expectation and performance manipulations.

Satisfaction. Conceptually, satisfaction is an outcome of purchase and use resulting from the buyer's comparison of the rewards and costs of the purchase in relation to the anticipated consequences. Operationally, satisfaction is similar to attitude in that it can be assessed as the sum of the satisfactions with the various attributes of the product or service. Pfaff (1977) suggests that both cognitive and affective models may be alternatives for describing satisfaction, whereas LaTour and Peat (1979) assert that the primary distinction between satisfaction and attitude derives from temporal positioning: attitude is positioned as a predecision construct and satisfaction is a postdecision construct.

The history of the consumer satisfaction research tradition is crudely captured in Figure 1. Most of the early research focused on the link between expectations and perceived product performance (ξ_1, η_2) although few investigators actually measured expectations, nor did they measure satisfaction (Cardozo 1964; Olshavsky and Miller 1972). The existence and direction of the disconfirmation experience were assumed to result from the expectation and/or performance manipulations and were not measured. Low expectations and/or high performance manipulations presumably produced positive disconfirmation whereas high expectations and/or low performance manipulations were thought to produce negative disconfirmation.

More recently, the focus of research has shifted to the relationship among perceived expectations, disconfirmation, and satisfaction (η_1, η_3, η_4). Oliver (1977, 1979, 1980b), for example, found independent additive effects of expectation and disconfirmation on satisfaction in field studies, although he did not manipulate performance. Olson and Dover (1976) examined the effects of expectation, performance, and disconfirmation on the consumer's belief elements. They found beliefs "displaced" toward expectation levels. However, they did not measure satisfaction, and thus could not examine the impact of these constructs on satisfaction. Swan and Trawick (1980) examined the influence of perceived product performance on disconfirmation and satisfaction. On the basis of correlational analysis, they found disconfirmation to be strongly related to performance, as was satisfaction. Taken together, these studies indicate the importance of disconfirmation as an intervening variable in the satisfaction process.

The preceding limited literature review reveals a body of empirical research examining one or more basic linkages in the satisfaction process. For a more extensive review of the consumer satisfaction literature, see Oliver (1980a) and Czepiel, Rosenberg, and Surprenant (1980). Much more work is needed, however, to conceptualize the constructs and their interrelationships adequately and to begin to integrate the results of prior research into more comprehensive models. Oliver's (1980b) study relating antecedents of satisfaction (expectation and disconfirmation) with consequences (postpurchase attitude and intention) is an important example of the direction this more extensive model development should take.

One of the crucial deficiencies in the extant literature is that no study has investigated the full set of interrelationships among expectations, perceived performance, disconfirmation, and satisfaction. Our study is an attempt to fill this gap in the literature. More specifically, we investigate whether it is necessary to include disconfirmation as an intervening variable affecting satisfaction as is commonly argued, or whether the effect of disconfirmation is adequately captured by expectation and perceived performance. Moreover, because the impact of expectations, performance, and disconfirmation on satisfaction may vary for different classes of products, we model the process for two types of products, a durable and a nondurable good.

METHOD

The model was investigated by experimental procedures. Three levels of expectations and three levels of performance were manipulated for two different products, each in a 3 × 3 factorial design.

Products

New products, a video disc player (VDP) and a hybrid plant (a chrysanthemum), were used in the experiment for several reasons. First, the use of new products is desirable to facilitate the manipulation of expectations. It is highly questionable whether a single communication can establish expectations about a product with which consumers may have had considerable prior experience. With new products, consumers would not have preconceived notions of what the product would deliver. Second, the proposed products were part of larger classes of products with which consumers did have some familiarity and thus it was possible to obtain judgments of salient attributes when expectations and performance were manipulated.

A final consideration affecting the choice of products was the ability to manipulate performance. Actual performance differences were necessary in the experiment because the research emphasis was on the relationship of satisfaction to the discrepancies between performance and expectations rather than the influence of expectations on perception. To some extent, the generalizability of the results may be limited by the characteristics of the products chosen. Both were new products for which performance variation was believable. As familiarity with

the products and believability vary, the findings of such experiments may differ.

Subjects

Subjects between the ages of 19 and 65 were recruited at a shopping mall and offered three dollars to participate in new product testing. All lived within a five-mile radius of the test city, a limitation necessary to allow followup for another part of the experiment.

Originally, 180 subjects were planned, a total of 20 in each cell. This number was ultimately used for the plant case but not for the VDP case, which was terminated because of equipment failure. Repair of the product would have taken a minimum of three weeks and we believed resumption of the experiment after such a delay might produce numerous history and maturation effects. Thus, the number of subjects for the video disc experiment was 126 and the cell sizes were slightly unequal.

Though all subjects were assigned randomly to treatment conditions, all subjects in the video disc product condition were chosen during the first three days of the experiment (Thursday through Saturday).

Manipulations

The proper levels for the expectation and performance manipulations were determined in extensive pretesting of the entire experimental process. Expectations were operationalized as a function of prior consumption experience and information delivered at the time of the experiment. This operationalization is similar to Miller's (1977) "expected" type and limited to Day's (1977) expectations about the nature and performance of the product. This limitation was necessary to ensure, insofar as possible, that subjects had a common standard of comparison.

For the expectation manipulations, subjects were shown one of three different printed messages giving information about the capabilities of the product, its general quality, and instructions on its use. To induce high source credibility, the source of the message was stated to be an independent testing lab.

The performance of the VDP itself was not manipulated. Instead a device was attached to the TV monitor to enable the experimenter to manipulate sound and picture quality. Performance level was controlled by two remote dials whereby the focus and hum could be varied independently. The high performance condition, with both dials in the "off" position for no distortion, represented the theoretical maximum level of performance that could be obtained.

For the plant, the physical characteristics in each performance condition were varied in the manner suggested by the pretest results. The high performance plants were, on the average, 24 inches tall and at the early stages of development, with 18 blossoms and at least five main stems. For moderate performance the plants averaged 15 inches in height, with eight blossoms on three main stems. For the poor performance condition the smallest plants, averaging seven inches, were selected. Excess buds and stems were removed so that these plants had only five blossoms and only two main stems. Some foliage was also removed so that the plants did not look full. Average flowering time was 26 days, 15 days, and 9 days for the best, average, and poor treatment conditions, respectively. The manipulation was refined further in an attempt to alleviate subjects' confusion about what an average chrysanthemum looked like. As part of the performance manipulation, subjects were shown an "ordinary mum chosen at random from a local florist's shop." The referent plant conformed to the average performance characteristics.

Disconfirmation levels were manipulated indirectly through the manipulation of expectation and performance. Pragmatic limitations dictated that disconfirmation could not be manipulated independently. It was possible, however, to obtain independent measures of disconfirmation which could then be used in the analysis.

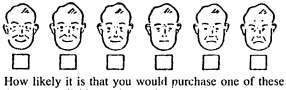
Procedure

A booth was set up in a shopping mall. The first part of the experiment involved the video disc player. Subjects were exposed to the expectation manipulation and were given ample time to read it. They were then given a manipulation check to assess their level of expectation. Next they were taken to another room where they were given a five-minute demonstration of the product and its capabilities. At the end of the demonstration, subjects were given the opportunity to examine the product and to ask questions. They then returned to the first room where they were provided with a scenario to facilitate role playing. They were asked to imagine that they had purchased the video disc player, had used the product in their home for one month, and that it is operating exactly like the one they had just seen. They were then asked to complete the performance, disconfirmation, and satisfaction measurements described in the following section.

One-half of the subjects were next given a similar task with a second new product, the plant. They were again given some information about the product to establish their initial expectations. These expectations were measured and subjects were shown the "treatment product" along with an ordinary chrysanthemum. Postexposure performance, disconfirmation, and satisfaction measures were taken as before.

The remaining one-half of the subjects were exposed to a somewhat different situation with the chrysanthemum. They were initially given the expectation manipulation and their expectations were assessed. Then they were told that as part of their compensation for participating in the experiment they would receive an experimental new product. They were given the "treatment" plant and were dismissed after their names and addresses had been obtained so that the rest of their payment could be mailed to them. Four weeks later, questionnaires

Table 1
DESCRIPTION OF MEASURES

Construct	Variable	Items: Type	Items: Number: Plant	Items: Number: VDP	Typical statement
Expectations	y_1	Attribute-specific	7	8	Good sound ___ ___ ___ ___ ___ ___ ___ Poor sound
	y_2	Global	1	1	Not very good ___ ___ ___ ___ ___ ___ ___ Excellent
Performance	y_3	Attribute-specific (line scale)	7	11	Visual resolution: Very inferior ___ ___ ___ ___ ___ ___ ___ Very superior
	y_4	Global (line scale)	1	1	Overall unit quality: Very inferior ___ ___ ___ ___ ___ ___ ___ Very superior
	y_5	Attribute-specific (rating scale)	0	11	The quality of color reproduction is: Terrible Very poor Poor Fair Good Very good Excellent
	y_6	Global (rating scale)	0	1	The overall quality of the unit is: Terrible Very poor Poor Fair Good Very good Excellent
Disconfirmation	y_7	Attribute-specific	10	11	My expectation of the number of flowers was: ___ ___ ___ ___ ___ ___ ___ Too high: It was worse than I thought / Accurate: It was just as I expected / Too low: It was better than I thought
	y_8	Global	1	1	My expectation regarding the performance of the product was: ___ ___ ___ ___ ___ ___ ___ Too high: It was worse than I thought / Accurate: It was just as I expected / Too low: It was better than I thought
Satisfaction	y_9	Attribute-specific (beliefs)	10	11	___ ___ ___ ___ ___ ___ ___ Strongly agree / Neutral / Strongly disagree
	y_{10}	Attribute-specific (affect)	10	11	___ ___ ___ ___ ___ ___ ___ Like / Dislike
	y_{11}	Global (verbal)	1	1	___ ___ ___ ___ ___ ___ ___ Completely satisfied / Neutral / Completely dissatisfied
	y_{12}	Global (faces)	1	1	
	y_{13}	Purchase probability	1	1	How likely it is that you would purchase one of these hybrids when they are available on the market? 0 (Very unlikely) / 50 (Don't know) / 100 (Almost certain)

were delivered to those subjects who received the plants. After they had completed the questionnaires, subjects were given the opportunity to exchange their plant for another.

At the conclusion of the experiment, subjects were fully debriefed to correct any misconceptions and to determine whether the experimental process was penetrated. None of the subjects interviewed had discovered the actual purpose of the experiment.

Measures

Several measurements were secured from each subject. Four constructs were assessed—expectation, performance, disconfirmation, and satisfaction—and multiple measures were used for each construct. At a minimum, each construct was assessed by using a single-item global measure and a multi-item, attribute-specific measure whereby the responses to the individual attributes were summed to generate the overall construct score.

The literature reflects controversy about whether satisfaction and dissatisfaction are on the same continuum. Aiello, Czepiel, and Rosenberg (1977) found some compression and upward skewing for a mixed scale. Leavitt (1977), however, found a strong negative correlation between satisfaction and dissatisfaction as one might expect in using separate satisfaction and dissatisfaction scales. We used a single satisfaction-dissatisfaction continuum. We believed the return from using separate satisfaction-dissatisfaction scales was not worth the burden imposed on subjects given the complexity of the study, and particularly given the equivocal nature of the evidence on whether satisfaction and dissatisfaction are at opposite ends of the same scale. Satisfaction with each product was assessed by using both belief and affect multi-item measures of satisfaction in accord with Pfaff's (1977) suggestion.

Subjects did not actually purchase and use the products. The elapsed time between the expectation manipulations and the collection of the dependent measures was limited. Subjects were asked basically to role-play the situation because, in an experimental setting, it is very difficult to provide extended usage of products, particularly durable goods. Nonetheless, satisfaction is conceptually distinguished as a postpurchase phenomenon. To the extent that the experimental setting did not allow for the necessary vicarious learning whereby subjects could experience satisfaction with use of the product, the satisfaction measures may be confounded with attitudinal measures.

The pretest also raised some concern about subjects' abilities to use the originally conceived line scale for recording their assessments of VDP performance. Consequently, a 7-point rating scale was also used for the VDP to capture subjects' performance evaluation. Table 1 describes each of the measures used.

RESULTS

Manipulation Checks

The success of the manipulations in producing the desired expectation and performance effects was assessed by analysis of variance procedures. The attribute-specific measures were used for this purpose because measurement theory suggests multi-item measures are typically much better than single-item measures (Churchill 1979; Nunnally 1978), although high correlations were found between the single-item global measures and the multi-item summed scores for both the expectation and performance measures.

Table 2
MANIPULATION CHECKS

			Analysis of variance table				
Manipulation	*Source*	*Degrees of freedom*	*Sum of squares*	*Mean square*	*F*	*Treatment level*	*95% confidence interval*
Plant							
Expectation	Between	2	288.24	144.12	161.28[a]	Low	2.94 to 3.55
	Within	177	158.17	.89		Moderate	4.81 to 5.25
	Total	179	446.41			High	6.14 to 6.53
Performance	Between	2	249.03	124.52	24.82[a]	Low	7.57 to 8.68
	Within	177	887.92	5.02		Average	9.20 to 10.40
	Total	179	1136.96			High	10.41 to 11.57
Video disc player							
Expectation	Between	2	135.26	67.63	94.52[a]	Low	3.43 to 4.03
	Within	123	88.01	.72		Moderate	4.95 to 5.45
	Total	125	223.27			High	6.03 to 6.51
Performance	Between	2	153.97	76.98	19.74[a]	Low	8.99 to 10.54
	Within	123	479.81	3.90		Average	10.67 to 11.69
	Total	125	633.78			High	11.94 to 13.00

[a] $p < .001$.

Table 3
SCALE RELIABILITIES FOR MULTI-ITEM MEASURES

Measurement	Coefficient alpha
Video disc	
Expectation	.88
Performance	
7-point	.91
Line	.91
Disconfirmation	.85
Satisfaction	
Belief	.87
Affect	.91
Plant	
Expectation	.95
Performance (line)	.93
Disconfirmation	.81
Satisfaction	
Belief	.95
Affect	.94

Table 2 suggests the manipulations had the intended effects. The overall F-tests are significant in each case and there is good separation in the mean response levels corresponding to each treatment condition.

Scale Reliability

Though investigating the reliability of all scales is desirable, it was not possible in our study. The nature of the contact precluded the opportunity to generate any stability estimates of reliability such as test-retest. This problem is greater for the single-item measures. For the multiple-item measures, one can calculate coefficient alpha which provides a good estimate of the reliability of a measure (Churchill 1979; Nunnally 1978; Peter 1979).

As Table 3 indicates, the alpha coefficients for each of the scales for each product are high and in general are well above the minimally acceptable level of .70 recommended for basic research (Nunnally 1978). Further, examination of the item-to-total score correlations (not shown) reveals some interesting patterns. Though all items for the VDP expectation measure correlate with the total score, the items that correlate most are those corresponding to the attributes manipulated in the expectation message. Items referring to appearance and disc quality, for example, are less highly correlated with overall performance than other attributes. Though these observations are not totally unexpected, they do provide indirect evidence that the scales were functioning properly and that response set bias was not a major problem. At the same time, factor analyses of the scale items suggest this scale and all the other multi-item scales for the VDP were unidimensional. The item-to-total correlations for the plant are uniformly high. This finding might be expected given the fact that more attributes were manipulated for the plant and the judgments seemed to be made in a more holistic manner. Again, factor analysis indicates that all of the plant multi-item scales were unidimensional.

Plant

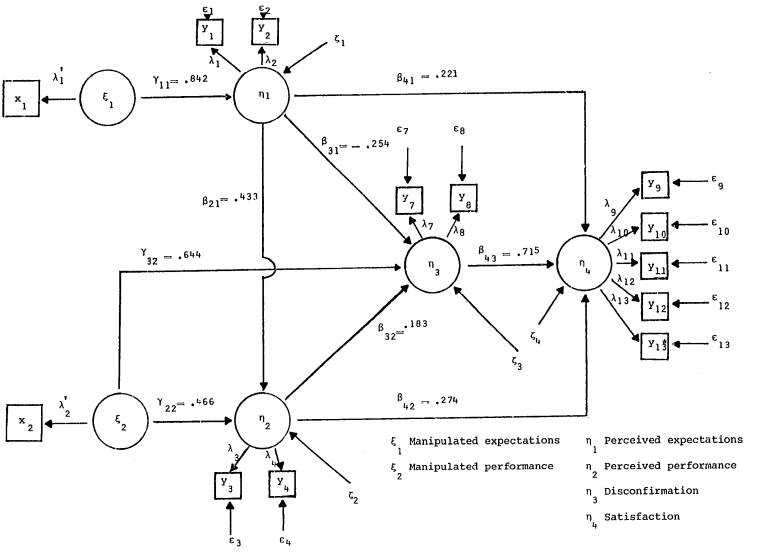
The lower triangular portion of Table 4 contains the pairwise correlations among the measures. Here x_1 and x_2 refer to the expectations and performance manipulations, respectively, and y_1 through y_{13} refer to the measured variables described in Table 1. Several aspects of these correlations are worth noting. First, most of the measures appear to have good convergent and discriminant validity. The measures basically correlate much higher with other measures assessing the same construct than with measures assessing different constructs. The single exception is y_{13}, the "purchase probability" measure of satisfaction, which has correlations with the other

Table 4
PAIRWISE CORRELATIONS

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	x_1	x_2
y_1		92	48	41	50	45	13	5	44	56	18	41	36	78	8
y_2	92		43	33	46	43	11	1	37	52	17	36	38	78	0
y_3	42	40		86	80	74	56	35	78	76	30	65	52	34	49
y_4	37	37	94		69	68	48	36	66	65	36	64	61	24	45
y_5						90	53	40	80	81	23	56	48	40	43
y_6							57	43	72	77	27	53	47	38	38
y_7	−18	−18	26	27				75	51	47	18	39	33	16	52
y_8	−10	−13	36	32			71		38	35	16	25	20	2	41
y_9	24	22	58	57			51	59		83	25	56	39	32	46
y_{10}	20	16	58	54			55	66	76		29	59	46	45	41
y_{11}	21	19	57	56			49	61	69	71		25	26	9	23
y_{12}	16	16	57	53			53	64	74	86	75		56	32	21
y_{13}	10	12	34	34			39	37	57	57	57	59		28	26
x_1	80	80	43	40			−26	−19	22	15	18	11	3		4
x_2	0	0	47	41			57	66	57	63	45	57	33	0	

The entries below the diagonal are the correlations for the plant data and those above the diagonal are the correlations for the video disc data. All decimals are omitted.

Figure 2
MAXIMUM-LIKELIHOOD PARAMETER ESTIMATES FOR THE PLANT DATA



*A measure that was dropped from the model because it did not display desirable measurement properties.

measures of satisfaction (y_9-y_{12}) that are equal to or lower than the correlations of these measures with the global disconfirmation measure (y_8). Further, the reliability of the behavioral intention measure when considered as part of the satisfaction construct is only .399 compared with .691 for y_9, .845 for y_{10}, .674 for y_{11}, and .844 for y_{12} (Bagozzi 1981). These results suggest y_{13} is not a good measure of satisfaction and therefore it was dropped from the analysis.[1] Otherwise, the pattern of correlations among the measures is impressive in terms of size and direction given the hypothesized model.

Second, though the experimental manipulations had the intended impact, they had some other impacts as well. In particular, the correlation between expectation (x_1) and performance (x_2) as *manipulated* is zero and the correlations between *manipulated performance* (x_2) and *measured expectation* (y_1 and y_2) are zero, but the correlations between *manipulated expectation* (x_1) and *measured performance* (y_3 and y_4) are not zero. As a matter of fact, the correlations between manipulated expectation and measured performance are almost as large as those between manipulated and measured performance, a condition that has an important consequence in the fit-

[1]The caution is in order that whenever one drops measures from a model such as this the chance of finding a satisfactory model is increased, with the consequence that the research results may well be capitalizing on chance.

ting of a model to these data. This correlation seems to suggest that subjects' performance evaluations were affected not only by the performance treatment to which they were exposed, but also by the fact that they had been somewhat sensitized by the prior expectation manipulation and manipulation check. However, an analysis of variance of the responses as a function of the treatment levels (not shown) indicates no significant interaction effects between manipulated expectations and performance for either product.

LISREL (Jöreskog and Sörbom 1978) was used to estimate the parameters of the hypothesized model. This approach was in keeping with the idea expressed by Alwin and Tessler (1974) that if one is considering the effect of an experiment, the parameters of interest should be the structural parameters relating the latent constructs rather than any regression coefficients relating observed y's to observed x's. Bagozzi (1980), Blalock (1971), and Costner (1971) also advocate such a position.

The fit of the hypothesized model to the data is mediocre. The structural parameters are significant and in the expected directions. However, the chi square value indicates the model does not fit the data ($p = .000$) and there are several large residuals, particularly involving x_2, y_3, and y_4. A direct link from η_1 to η_2 was introduced which resolved the large y_3 and y_4 residuals. Those involving x_2 were a little more troublesome to treat. Their size and pattern seemed to suggest direct links from ξ_2 to both η_3 and η_4. The direct link from ξ_2 to η_4 proved to be not significant, though, and it was subsequently dropped. Figure 2 depicts the final fitted model but displays only the structural parameters for ease of reading. Table 5 lists the final parameter estimates, standard errors, and t-values for both the measurement and structural portions of the model;[7] Table 6 reports the reliability values for the individual measures and the composite reliability values for the constructs (Bagozzi 1981).

The evidence pertaining to the measures confirms the previous indications of their quality. All of the reliability values for the individual variables are high and the composite reliability values for the constructs are very high, the smallest being .834 for the disconfirmation measure. The evidence on structural relationships is also encouraging. The one disappointing set of evidence is the overall chi square value of 87.1 with 46 degrees of freedom which suggests the composite hypothesis that the model is true in the total population must be rejected as untenable. One must interpret this result with caution, though, because chi square is a valid test statistic only if the analysis is based on the sample covariance matrix and not the correlation matrix, which is certainly not true in this instance (Jöreskog and Sörbom 1982; Bentler and Bonett 1980). Further, other indicators suggest the model provides a reasonably good fit to the data. First, the average residual between the observed correlations and the theoretical correlations given the structural model is only −.0001. Second, the squared multiple correlations for each of the structural equations are reasonably high.

η_1—perceived expectations	.709
η_2—perceived performance	.406
η_3—disconfirmation	.581
η_4—satisfaction	.780

Finally, Tucker and Lewis' (1973) reliability coefficient indicates about 98% of the covariation to be explained is accounted for by the proposed structure (Bagozzi 1980, p. 106–12).

The t-values of 18.261 and 8.186 for γ_{11} and γ_{22}, respectively, indicate the experimental manipulations were clearly successful. Further, performance has a positive

Table 5
ESTIMATED PARAMETERS FOR PLANT EXPERIMENT

Parameter	LISREL estimate	Standard error	T value	Standardized value
β_{21}	.458	.062	7.379	.433
β_{31}	−.209	.057	−3.652	−.253
β_{32}	.143	.059	2.437	.183
β_{41}	.192	.052	3.671	.221
β_{42}	.225	.053	4.243	.274
β_{43}	.750	.084	8.909	.715
γ_{11}	.806	.044	18.261	.842
γ_{22}	.471	.058	8.186	.466
γ_{32}	.510	.059	8.613	.644
λ_1	1.000[a]			.957
λ_2	1.002	.037	27.216	.958
λ_3	1.000[a]			1.011
λ_4	.920	.036	25.473	.930
λ_7	1.000[a]			.791
λ_8	1.133	.089	12.785	.896
λ_9	1.000[a]			.829
λ_{10}	1.112	.069	16.062	.922
λ_{11}	.978	.075	13.061	.811
λ_{12}	1.106	.069	15.914	.917
Var(ζ_1)	.267	.034	7.798	.291
Var(ζ_2)	.608	.070	8.663	.595
Var(ζ_3)	.262	.047	5.547	.418
Var(ζ_4)	.151	.031	4.885	.220
θ_{ε_1}	.084	.020	4.189	.084
θ_{ε_2}	.081	.020	4.066	.081
θ_{ε_3}	−.020	.028	−.726	−.020
θ_{ε_4}	.137	.028	4.926	.137
θ_{ε_7}	.373	.050	7.527	.373
θ_{ε_8}	.195	.041	4.766	.195
θ_{ε_9}	.313	.038	8.249	.313
$\theta_{\varepsilon_{10}}$	.151	.024	6.257	.151
$\theta_{\varepsilon_{11}}$	.343	.041	8.403	.343
$\theta_{\varepsilon_{12}}$	.160	.025	6.474	.160

$\chi^2_{46} = 87.1$, $p = .000$.
[a]Constrained.

[2]Note that Table 1 indicates a negative error variance in y_3, which is impossible. A condition like this often can be traced to extremely high initial correlations between measures designed to assess the same construct. In this case, $r_{y3y4} = .939$; y_3 was subsequently dropped and the model reestimated. All parameter estimates came out exactly the same to one decimal place (many were the same to three decimal places) and thus we decided to use the model with y_4 included because of its greater degrees of freedom.

Table 6
INDIVIDUAL MEASURE AND COMPOSITE CONSTRUCT RELIABILITIES

	Plant		Video disc	
Construct/measure	*Measure*	*Composite*	*Measure*	*Composite*
Expectations		.957		.962
y_1 —Attribute-specific	.916		.921	
y_2 —Global	.919		.926	
Performance		.970		.932
y_3 —Attribute-specific (line)	1.020*		.771	
y_4 —Global (line)	.873		.620	
y_4 —Attribute-specific (rating)			.887	
y_6 —Global (rating)			.820	
Disconfirmation		.834		.872
y_7 —Attribute-specific	.627		.982	
y_8 —Global	.805		.591	
Satisfaction		.926		.904
y_9 —Attribute-specific (beliefs)	.687		.811	
y_{10} —Attribute-specific (affect)	.849		.840	
y_{11} —Global (verbal)	.657			
y_{12} —Global (faces)	.840			

*The reliability cannot exceed one. An impossible condition like this often can be traced to extremely high initial correlations between measures designed to assess the same construct. In this case $r_{3/4} = .939$; y_4 was subsequently dropped and the model reestimated. As all parameter estimates came out approximately the same, we decided to use the model with y_4 included because of its greater degrees of freedom.

impact on disconfirmation, with higher product performance leading to subjects rating the product as performing better than expected ($\beta_{32} = 0.183$). Expectations have the opposite impact in that high expectations produced disappointments among subjects about the performance of the plant ($\beta_{31} = -0.253$). Satisfaction is related positively to disconfirmation as is commonly suggested ($\beta_{43} = 0.715$). Participants who felt that the plant performed better than originally anticipated were satisfied with it. Note, however, that the level of satisfaction is also affected directly by subjects' expectations ($\beta_{41} = .221$) and performance perceptions ($\beta_{42} = .274$); when either of these are high, subjects are more satisfied with the plant. Expectations and performance thus have a direct impact on satisfaction in addition to their indirect impact through disconfirmation, although the impact of disconfirmation on satisfaction is the largest of the three. Note finally that the unexplained variation with respect to satisfaction is low (variance $\zeta_4 = .220$). As in regression analysis, one can divide this unexplained variation by the total variance in η_4 to produce a measurement of the proportion of the total variation left unexplained. This in turn can be subtracted from one to produce an assessment of the proportion of the total variation in satisfaction accounted for by the model. These calculations indicate that 78% of the variation in satisfaction is accounted for, suggesting the model does very well in "explaining" respondent satisfaction with the plant.

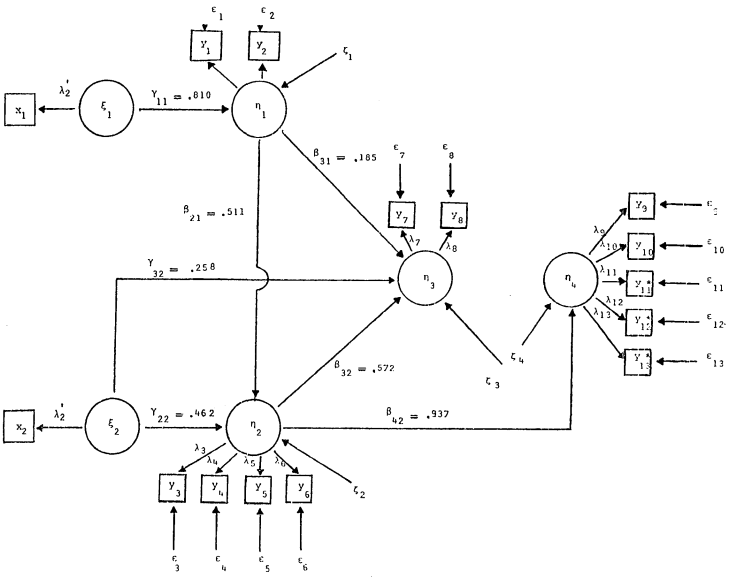
Video Disc Player

The upper triangular portion of Table 4 contains the correlations for the measures in the video disc experiment. The evidence on the convergent and discriminant validity of the measures is not as strong as it is with the plant data. The satisfaction measures are particularly troublesome. Neither y_{11}, y_{12}, nor y_{13} satisfies the criteria for convergent and discriminant validity. All have correlations with the other measures of satisfaction that are less than their correlations with other measures of different constructs; consequently, y_{11}, y_{12}, and y_{13} were dropped from the analysis. Another feature to note in the pairwise correlations is that those involving expectations and disconfirmation have a different pattern than before. In the plant experiment, all the pairwise correlations between the two expectation measures and the two disconfirmation measures are negative; in the VDP experiment they are positive though small.

The final model used to account for the plant data was fit to the VDP data. The chi square value indicated the model did not fit the data very well. Further, the *t*-values indicated the parameters β_{41} and β_{43} were not significant. They were dropped and the model was reestimated. An incremental chi square analysis indicated they indeed should be omitted. The $\Delta\chi^2_2$ value of 2.2 was not significant, indicating that including the two parameters did not improve the fit.

Although the chi square test indicates that the overall model must be rejected, other rules of thumb on reliability, validity, and fit indicate a reasonably good model overall. The constructs are adequately captured by the measures. The composite reliability values in Table 6 range from .872 for the disconfirmation construct to .962 for the expectation construct. The residuals are also small; the average residual between the observed cor-

Figure 3
MAXIMUM-LIKELIHOOD PARAMETER ESTIMATES FOR THE VIDEO DISC DATA



*Measures that were dropped from the model because they did not display desirable measurement properties.

relations and the theoretical correlations produced by the proposed structure is only .0014. Tucker and Lewis' (1973) reliability coefficient suggests that 92% of the covariation available to be explained is accounted for by the proposed structure. Finally, the squared multiple correlations for each of the structural equations are reasonably high, indicating that each of the constructs is captured rather well by the model.

η_1—perceived expectations	.657
η_2—perceived performance	.488
η_3—disconfirmation	.455
η_4—satisfaction	.877

Satisfaction is particularly well accounted for, as almost 88% of the variation in the satisfaction construct is explained by the fitted model.

Several parallels can be seen between the VDP results (Figure 3, Table 7) and the results of the plant experiment. The experimental manipulations were very successful. Subjects' performance evaluations were affected by their perceived expectations. Disconfirmation is negatively related to expectations and positively related to performance, and the performance manipulation also has a positive direct impact on disconfirmation. The main differences in the results are related to the satisfaction construct. Whereas expectations, disconfirmation, and performance all affected satisfaction in the plant experiment, only performance had an impact in the VDP ex-

Table 7
ESTIMATED PARAMETERS FOR VIDEO DISC EXPERIMENT

Parameter	LISREL estimate	Standard error	T value	Standardized value
β_{21}	.466	.067	6.924	.511
β_{31}	−.190	.091	−2.102	−.185
β_{32}	.644	.118	5.467	.572
β_{41}				
β_{42}	.961	.074	12.914	.937
β_{43}				
γ_{11}	.777	.056	13.871	.810
γ_{22}	.405	.062	6.523	.462
γ_{32}	.254	.082	3.108	.258
λ_1	1.000[a]			.959
λ_2	1.003	.044	23.042	.962
λ_3	1.000[a]			.876
λ_4	.897	.079	11.373	.785
λ_5	1.073	.065	16.454	.939
λ_6	1.031	.069	15.015	.903
λ_7	1.000[a]			.986
λ_8	.773	.090	8.567	.762
λ_9	1.000[a]			.899
λ_{10}	1.018	.065	15.651	.915
Var (ξ_1)	.316	.047	6.718	.344
Var (ξ_2)	.393	.066	5.937	.512
Var (ξ_3)	.529	.111	4.809	.544
Var (ξ_4)	.099	.026	3.206	.123
θ_{11}	.079	.025	3.205	.079
θ_{22}	.074	.025	3.016	.074
θ_{33}	.229	.032	6.703	.229
θ_{44}	.380	.049	7.332	.380
θ_{55}	.113	.024	5.081	.113
θ_{66}	.180	.031	6.259	.180
θ_{77}	.028	.086	.321	.028
θ_{88}	.419	.074	5.685	.419
θ_{99}	.189	.036	5.345	.189
θ_{1010}	.160	.033	4.740	.160

$\chi^2_{28} = 160.8$, $p = .000$.
[a]Constrained.

periment. The impact was positive and very pronounced, the proportion of variation left unexplained being only 12%.

DISCUSSION

We sought to investigate the determinants of customer satisfaction. Is satisfaction simply affected by the extent of the disconfirmation experience as is often suggested, or do expectations and performance exert independent effects on satisfaction in addition to their impact via disconfirmation? In an experiment involving two products, a chrysanthemum and video disc player, both expectations and performance were manipulated independently for each product, and each subject's perceived expectations, performance evaluations, disconfirmation, and satisfaction were subsequently measured by using multiple measures for each construct.

The analysis indicated the effects were different for the two products. For the plant, the relationships were as hypothesized. Initial expectations had a negative effect on the disconfirmation experience whereas performance had a positive effect. Disconfirmation positively affected satisfaction as is commonly held; when subjects perceived the product performing better than expected, they were more satisfied with it and vice versa. Expectations and performance also affected satisfaction directly, and the three variables in combination explained 78% of the total variation in satisfaction.

The results for the VDP were different in important respects. First, neither the disconfirmation experience nor subjects' initial expectations affected subjects' satisfaction with the product. Rather, their satisfaction was determined solely by the performance of the VDP. When it performed well they were satisfied with it and when it performed poorly they were dissatisfied with it, regardless of their initial expectations. Expectations did combine with performance to affect disconfirmation, though the magnitude of the disconfirmation experience did not translate into an impact on satisfaction. Finally, the direct performance-satisfaction link accounted for most of the variation in satisfaction, 88%.

Several results warrant additional explanation. Contrary to the findings of Oliver (1977, 1980b), who found no correlation between perceived expectations and disconfirmation, we found a statistically significant negative, but small, correlation between the constructs in the plant experiment. The correlation between the two constructs was positive, but not significant, in the VDP experiment. Two explanations for this apparent contradiction are possible. First, Oliver (1977) has argued that there is no necessary association between expectations and disconfirmation when judgments are *subjective*. The nature of the expectation manipulations may have produced judgments about the plant that were more objective than the judgments about the VDP. Subjects were given objective levels on attributes for the plants ("size will be about 10 inches"), whereas more subjective descriptors were used for the video disc ("picture quality is excellent"). Second, expectations were not manipulated in Oliver's studies, but in our study they were. Thus, some of the association between expectations and disconfirmation may reflect demand characteristics.

The most crucial issue needing additional clarification is why the discrepancy is found between the models for the two products. Why are the results for the plant consistent with prior research findings in that disconfirmation had the greatest effect on satisfaction whereas those for the VDP are inconsistent? Not only does disconfirmation not have the major impact on satisfaction in the latter case, but its impact is not even statistically significant.

Several explanations are possible. First, the VDP was an innovative, technologically complex, durable product. Most previous satisfaction research has focused on nondurable, frequently purchased, and relatively inexpensive products like our alternative product, the plant. Two exceptions should be noted. Olshavsky and Miller (1972) manipulated performance of a tape-recorder and examined the influence of expectations on performance

ratings. They found that the ratings assimilated toward expectations. They did not, however, measure disconfirmation or satisfaction. Oliver (1977) examined the influence of expectation and disconfirmation on the performance ratings of cars. He found both variables to be significantly related to performance. In that study, no manipulations were used and product performance was held constant. To our knowledge, in all other satisfaction studies in which performance was manipulated, small, nondurable products were used. Thus, one explanation for the differences in the two models is that for durable products performance differences (if present) are the major determinant of satisfaction, and conversely that the disconfirmation of initial expectations has little impact. This argument suggests previous researchers have found evidence for the disconfirmation experience because they have focused almost exclusively on nondurable products.

Another possibility is that the experiment does not allow satisfaction as it is typically conceived to operate. Consumer satisfaction is a postpurchase phenomenon. It reflects how much the consumer likes or dislikes the product after using it. Respondents did not actually use the products in our experiment, except for the subset of respondents who were allowed to take a plant home. Rather they were asked to imagine that they had purchased each product, had used each for a month, and that each had performed exactly as seen in the demonstration. The validity of the model specifications depends on how well these imagined role-playing situations actually produced the vicarious learning needed for satisfaction to operate. Preferably each respondent would have been allowed to use the plant and VDP for some time before their satisfaction was assessed, but resource limitations precluded that procedure. Because of the temporal constraint imposed in securing the measurements, the validity of our conclusions about what determines satisfaction are questionable, although the same criticism could be directed at a number of other studies as well.

A third alternative explanation for the conflicting results involves the measures. The performance and satisfaction measures for the VDP were highly correlated, with some between-construct correlations slightly higher than the within-construct correlations. This is *one* of the reasons why the three satisfaction measures—y_{11}, y_{12}, and y_{13}—had to be eliminated. A natural question is whether performance and satisfaction measures capture the same thing. In spite of conceptual differences between the constructs, are our operationalizations likely to share method variance? Swan and Trawick (1980), for example, also found high correlations between the two constructs. Or is it possible, indeed likely, that satisfaction with durable products would be more dependent on actual product performance than would satisfaction with nondurable or convenience goods, and that researchers have failed to note this before because their designs have not allowed them to address the issue? After all, the measures of performance and satisfaction used in *each* analysis did have convergent and discriminant validity. Yet the results *were different*!

IMPLICATIONS

Our results indicate that the effects of expectation, disconfirmation, and performance on satisfaction may differ for durable and nondurable products. Because most of the research on consumer satisfaction has used nondurables, this finding has important implications for future research. Day (1977) has suggested that the processes consumers use to assess satisfaction may differ for durables and nondurables. To the authors' knowledge, no other study has directly compared the two product classes. To verify and extend our finding, much more research is needed on the satisfaction process for durable products. Despite the well-known pragmatic difficulties of doing research with durable products, researchers must extend their findings beyond the realm of consumer package goods.

The role of performance in determining satisfaction with the VDP also has important consequences for both research and management. Most satisfaction research has ignored variations in product performance except as it affects disconfirmation. The major variables of interest have been expectations and disconfirmation. This bias is perhaps natural for marketers because those variables pertain to communications (advertising and product information, for example) which are more directly under the control of marketing managers. The models we estimated, however, indicate that performance levels *per se* have a direct impact on satisfaction judgments in addition to their impact on disconfirmation. In fact, for the VDP this link is the *only* significant influence on satisfaction. This finding suggests that satisfaction with the VDP could be increased only by increasing performance, not by minimizing negative disconfirmation. Further, if this result were generalizable to some other products, it would indicate that both researchers and managers must direct much more attention to the impact of performance levels. It is unrealistic to ignore the impact of performance on satisfaction, as the U.S. auto industry has learned. Managers often have a choice of what level of performance should be delivered to satisfy the consumer and require research evidence to aid them in that decision.

REFERENCES

Aiello, Albert, John A. Czepiel, and Larry J. Rosenberg (1977), "Scaling the Heights of Consumer Satisfaction: An Evaluation of Alternative Measures," in *Consumer Satisfaction, Dissatisfaction and Complaining Behavior,* Ralph Day, ed. Bloomington: Indiana University (April), 43–50.

Alwin, D. F. and R. C. Tessler (1974), "Causal Models, Unobserved Variables, and Experimental Data," *American Journal of Sociology*, 80 (No. 1), 58–86.

Anderson, Rolph E. (1973), "Consumer Dissatisfaction: The Effect of Disconfirmed Expectancy on Perceived Product

Performance," *Journal of Marketing Research*, 10 (February), 38–44.

Bagozzi, Richard P. (1980), *Causal Models in Marketing*. New York: John Wiley & Sons, Inc.

——— (1981), "An Examination of the Validity of Two Models of Attitude," *Multivariate Behavioral Research*, 16 (July), 323–59.

Bentler, Peter M. and Douglas G. Bonett (1980), "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures," *Psychological Bulletin*, 88 (No. 3), 588–606.

Blalock, H. M., Jr., (1971), *Causal Models in the Social Sciences*. Chicago: Aldine Publishing Co.

Cardozo, Richard (1964), "Customer Satisfaction: Laboratory Study and Marketing Action," *Journal of Marketing Research*, 2 (August), 244–9.

Churchill, Gilbert A., Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (February), 64–73.

Costner, H. L. (1971), "Utilizing Causal Models to Discover Flows in Experiments," *Sociometry*, 34 (September), 398–410.

Czepiel, John A., Larry J. Rosenberg, and Carol F. Surprenant (1980), "The Development of Thought, Theory and Research in Consumer Satisfaction," paper presented at the American Marketing Association Special Educators' Conference.

Day, Ralph (1977), "Toward a Process Model of Consumer Satisfaction," in *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*, H. Keith Hunt, ed. Cambridge, MA: Marketing Science Institute (May), 153–86.

Jöreskog, Karl G. and Dag Sörbom (1978), *Analysis of Linear Structural Relationships by the Method of Maximum Likelihood: Lisrel IV*. Chicago: National Education Resources.

——— and ——— (1982), "Recent Developments in Structural Equation Modeling," *Journal of Marketing Research*, 19 (November).

Kurth, Rudolf (1965), "Testing the Significance of Consumer Complaints," *Journal of Marketing Research*, 2 (August), 283–4.

LaTour, Stephen A. and Nancy C. Peat (1979), "Conceptual and Methodological Issues in Consumer Satisfaction Research," in *Advances in Consumer Research*, William L. Wilkie, ed. Association for Consumer Research, 432–7.

Leavitt, Clark (1977), "Consumer Satisfaction and Dissatisfaction: Bipolar or Independent," in *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*, H. Keith Hunt, ed. Cambridge, MA: Marketing Science Institute (May), 132–52.

Miller, John A. (1977), "Exploring Satisfaction, Modifying Models, Eliciting Expectations, Posing Problems, and Making Meaningful Measurements," in *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*, H. Keith Hunt, ed. Cambridge, MA: Marketing Science Institute (May), 72–91.

Nunnally, Jum C. (1978), *Psychometric Theory*. New York: McGraw-Hill Book Company.

Oliver, Richard (1977), "A Theoretical Reinterpretation of Expectation and Disconfirmation Effects on Posterior Product Evaluation: Experiences in the Field," in *Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, Ralph Day, ed. Bloomington: Indiana University (April), 2–9.

——— (1979), "Product Satisfaction as a Function of Prior Expectation and Subsequent Disconfirmation: New Evidence," in *New Dimensions of Consumer Satisfaction and Complaining Behavior*, Ralph Day and H. Keith Hunt, eds. Bloomington: Indiana University (October), 66–71.

——— (1980a), "Theoretical Bases of Consumer Satisfaction Research: Review, Critique, and Future Direction," paper presented at the American Marketing Association Second Special Conference on Marketing Theory.

——— (1980b), "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," *Journal of Marketing Research*, 17 (November), 460–9.

Olshavsky, Richard and John A. Miller (1972), "Consumer Expectations, Product Performance, and Perceived Product Quality," *Journal of Marketing Research*, 9 (February), 19–21.

Olson, Jerry C. and Philip Dover (1976), "Effects of Expectations, Product Performance, and Disconfirmation on Belief Elements of Cognitive Structures," in *Advances in Consumer Research*, Association for Consumer Research.

Peter, J. Paul (1979), "Reliability: A Review of Psychometric Basis and Recent Marketing Practices," *Journal of Marketing Research*, 16 (February), 6–17.

Pfaff, Anita B. (1972), "An Index of Consumer Satisfaction," *Proceedings of the Third Annual Conference*, Association for Consumer Research (November), 713–37.

Pfaff, Martin (1977), "The Index of Consumer Satisfaction: Measurement Problems and Opportunities," in *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*. H. Keith Hunt, ed. Cambridge, MA: Marketing Science Institute (May), 36–71.

Swan, John E. and I. Fredrick Trawick (1980), "Inferred and Perceived Disconfirmation in Consumer Satisfaction," in *Marketing in the 80's*, Proceedings of the AMA Educators' Conference, Chicago, 97–101.

Tucker, Ledyard R. and Charles Lewis (1973), "A Reliability Coefficient for Maximum Likelihood Factor Analysis," *Psychometrika*, 38 (March), 1–10.