The Asymmetric Impact of Negative and Positive Attribute-Level Performance on Overall Satisfaction and Repurchase Intentions

The relationship between attribute-level performance, overall satisfaction, and repurchase intentions is of critical importance to managers and generally has been conceptualized as linear and symmetric. The authors investigate the asymmetric and nonlinear nature of the relationship among these constructs. Predictions are developed and tested empirically using survey data from two different contexts: a service (health care, n = 4517) and a product (automobile, n = 9359 and n = 13,759). Results show that (1) negative performance on an attribute has a greater impact on overall satisfaction and repurchase intentions than positive performance has on that same attribute, and (2) overall satisfaction displays diminishing sensitivity to attribute-level performance. Surprisingly, results show that attribute performance has a direct impact on repurchase intentions in addition to its effect through satisfaction.

Consider the following scenarios:

- In an effort to increase customer satisfaction, an automotive manufacturer consistently has increased performance on engine power. However, results show that, after a while, increases in the ratings for performance on engine power do not yield corresponding changes in satisfaction ratings. Designers wonder if they should keep investing resources to enhance engine power.
- A physician historically has received high performance ratings on attributes such as politeness, timeliness, reliability of service, and correct diagnosis. In the current survey, she finds that performance ratings on timeliness are down, but are up on all the other attributes. However, she also finds a sharp decline in her overall satisfaction ratings. She is puzzled as to why negative performance on a single attribute is not offset by positive performance on a host of other attributes.

These scenarios highlight typical dilemmas that managers face at customer-driven organizations where the goals are to design products and services with attributes that maximize customer satisfaction. These goals generally are operationalized by multiple-regression models that identify key attributes into which managers should invest resources to enhance overall satisfaction (e.g., Bolton and Drew 1991; Hanson 1992; Wittink and Bayer 1994). The assumption underlying such “key driver” models is that there is a symmetric and linear relationship between attribute-level performance and dependent constructs such as overall satisfaction and purchase intentions. However, what if, at the attribute level, performance and satisfaction were linked asymmetrically? The answer to this question is important for academics and managers alike. Although academics (cf. Anderson and Sullivan 1993; Oliva, Oliver, and Bearden 1995) and practitioners (cf. Coyne 1989) recognize that the overall satisfaction function might not be linear and/or symmetric, no such conclusion can be offered at the attribute level. Yet, it is at the attribute level that managers must make decisions.

Therefore, our objective is to investigate—theoretically and empirically—the existence and nature of the asymmetric and nonlinear response of satisfaction to attribute-level performance. In the first section, we discuss the need for an attribute-level conceptualization of the performance/satisfaction relationship. That is followed by a discussion of the asymmetric and nonlinear nature of the relationship and the development of hypotheses. Then we present results from three studies that test the hypotheses.

Multi-Attribute Products and Satisfaction

There are several reasons—theoretical and managerial—to use multi-attribute models in the context of customer satisfaction (cf. LaTour and Peat 1979; Wilkie and Pessner 1973). First, consumers are more likely to render evaluations of their postpurchase experiences of satisfaction at an attribute level rather than at the product level (Gardial et al. 1994). Second, an attribute-based approach enables researchers to conceptualize commonly observed phenomena, such as consumers experiencing mixed feelings toward a product or service. A consumer can be both satisfied and dissatisfied with different aspects of the same product.
Although such phenomena are not easy to model in an overall satisfaction approach, the attribute-level approach provides a simple and elegant solution: Mixed feelings toward a product exist because a consumer may be satisfied with one attribute but dissatisfied with another. For example, in a restaurant, a customer may be highly satisfied with the food but highly dissatisfied with the service at the same time. Third, an attribute-level approach to satisfaction affords researchers a higher level of specificity and diagnostic usefulness compared with the product level or “overall” approach (LaTour and Peat 1979). For example, Parasuraman, Zeithaml, and Berry (1988) show that measuring various attributes or dimensions of a service provides a better understanding of global constructs, such as service quality. Similarly, Boulding and colleagues (1993) suggest that quality is multidimensional and different dimensions of quality are averaged together in some fashion to produce an overall assessment of quality. However, both works treat the relationship between attributes and overall quality as linear and symmetric. Fourth, the higher specificity and diagnostic usefulness of multi-attribute models also ensures that academic research will find more welcome applications among managers who generally work at the attribute level rather than at an overall level (e.g., Hanson 1992; Wittink and Bayer 1994). Finally, there is some evidence that attribute-level performance/disconfirmation and overall satisfaction are qualitatively different constructs (Oliva, Oliver, and Bearden 1995), and, if treated interchangeably, “global customer satisfaction responses may mask specific product issues” (Oliva, Oliver, and Bearden 1995, p. 26).

Thus, studying satisfaction at the attribute level can help extend both conceptual and empirical understanding of the phenomenon.

The Asymmetric Effect of Product Performance and Satisfaction: Theory and Hypotheses

When examining the theoretical and analytical importance of the link between attribute-level performance and overall satisfaction, it is important to recognize that the relationship could be asymmetric (cf. Anderson and Sullivan 1993; Oliva, Oliver, and Bearden 1995). One unit of negative performance on an attribute could have a greater effect on overall satisfaction or repurchase intentions than a corresponding unit of positive performance. Similarly, in a given set of attributes, negative performance on a single attribute could outweigh positive performance on many other attributes combined. Oliver (1993) finds that (1) attribute-level satisfaction and dissatisfaction significantly affect overall satisfaction with a product (automobile) and a service (an undergraduate course offering) and (2) attribute dissatisfaction has a larger weight than attribute satisfaction for the product but not for the service. However, no theoretical motivation for the observed disparity between the impact of attribute satisfaction and dissatisfaction is provided. In the next sections, the theoretical logic is developed along two lines of reasoning. One is based on prospect theory (Kahneman and Tversky 1979), and the other is rooted in cognitive research that examines the memorability of positive versus negative events.

Prospect Theory

Prospect theory (Kahneman and Tversky 1979) is a descriptive theory in which all of the alternatives that a person faces are reduced to a series of prospects that are evaluated independently on the basis of an S-shaped value function (see Figure 1).

As depicted in Figure 1, prospect theory postulates that peoples’ judgments display (1) reference dependence (carriers of value are gains and losses from a reference point) and (2) loss aversion (the function is steeper in the negative than in the positive domain). In addition, evaluations display diminishing sensitivity (marginal values of both gains and losses decrease with their size). The two key properties of the value function for our discussion are loss aversion and diminishing sensitivity.

The loss aversion built into prospect theory suggests that losses loom larger than gains (Einhorn and Hogarth 1981). Psychologically, a one-unit loss is weighted more than an equal amount of gain. In a satisfaction context, negative outcomes on attribute performance should carry more weight in the overall satisfaction judgment than equal amounts of positive outcomes on attribute performance. For example, if a car’s mileage were to decrease by 10 miles per gallon, it would have a greater impact on the overall satisfaction judgment than if the car’s mileage were to increase by 10 miles per gallon. Thus, negative performance on an attribute will loom larger than positive performance on the same attribute. On the basis of these theories, we propose

---

1Standardized loadings were as follows: attribute satisfaction and dissatisfaction for the automobile (.135 and −.175, respectively) and the course (.130 and −.059, respectively).
H_{1a}: Negative performance on an attribute will have a greater impact on overall satisfaction than positive performance on the same attribute.

H_{1b}: Negative disconfirmation on an attribute will have a greater impact on overall satisfaction than positive disconfirmation on the same attribute.

In addition, on the basis of prospect theory, overall satisfaction also should display diminishing sensitivity toward attribute performance. That is, at high (low) levels of performance, positive (negative) performance on an attribute should not affect satisfaction as dramatically as it does at lower levels of performance. This hypothesis is similar to the diminishing returns hypothesis in classical economics and also is depicted in Figure 1. Therefore,

H_2: Overall satisfaction will display diminishing sensitivity to changes in the magnitude of performance for a given attribute. In other words, at high levels of positive or negative performance on an attribute, overall satisfaction will be affected less than at intermediate levels of performance on that attribute.

Memory for Negative Versus Positive Instances

Researchers distinguish between transaction-based and cumulative satisfaction (Bittner and Hubbert 1994). Cumulative satisfaction typically is conceptualized as an overall judgment based on several transactions with a product or service. To the extent that all aspects of the transactions are not equally accessible (Gardial et al. 1994), the overall satisfaction judgment will vary on the basis of the accessibility of a given aspect. Memory accessibility is a function of stimulus salience, among other things (Taylor 1982). Evidence shows that negative information is more perceptually salient than positively valenced information, is given more weight than positive information, and elicits a stronger physiological response than positive information (Peeters and Czapinski 1990).

Similar psychological operations should occur for customer satisfaction because satisfaction is linked to memory-based processing (Yi 1990). To the extent that attributes with negative performance will be more perceptually salient than attributes with positive performance, attributes with negative performance should have a greater impact on the cumulative satisfaction judgment. Thus, within a given set of attributes, the relative impact of each attribute will be asymmetric. Consequently, when combined, attributes with negative performance should have a greater impact on overall satisfaction than their corresponding attributes with positive performance combined. On the basis of the previous discussion, we hypothesize that

H_{3a}: Cumulatively, attributes with negative performance will have greater impacts on overall satisfaction than attributes with positive performance.

Along similar lines, a diminishing sensitivity hypothesis between overall satisfaction and performance on various attributes can be proposed. In a given set of attributes, each additional instance of positive performance on an attribute will have a smaller impact than the other attributes. Conversely, each additional instance of negative performance should have a correspondingly smaller negative impact on overall satisfaction. Thus,

H_{3b}: Overall satisfaction will display diminishing sensitivity to additional instances of negative or positive performance. In other words, each additional instance of positive or negative performance should have a smaller impact on overall satisfaction.

Repurchase Intentions

Attribute-level performance should affect satisfaction and repurchase intentions differently (e.g., Ostrom and Iacobucci 1995). One reason could be that satisfaction and repurchase intentions are qualitatively different constructs. Satisfaction may be merely a judgment with cognitive and affective dimensions, whereas repurchase intentions also have a behavioral component. Based on the consumer’s goals (e.g., Mittal et al. 1993), performance on a certain attribute may become crucial for repurchase intentions but not satisfaction. For example, consider the case of a patient who is satisfied with all aspects of the service provided by his or her primary care physician (PCP), except that the patient has now relocated to another area far from the PCP’s office. When it comes time to renew and choose a PCP, the patient might indicate high overall satisfaction with the PCP but still might choose another PCP, because performance on a critical attribute has changed. In other words, though the performance on “distance of PCP’s office from home” had a negative impact on repurchase intentions, it has a small or virtually no impact on overall satisfaction. One reason for such a discrepancy can be found in the attribution literature (Folkes 1988). Here, the patient may attribute “poor” performance on “distance of PCP’s office from home” to him or herself or to causes other than the PCP and thus not change his or her overall satisfaction rating; nevertheless, the repurchase intentions of the patient change. Thus,

H_4: Overall satisfaction and attribute-level performance will have separate and distinct effects on repurchase intentions.

However, the relative magnitude of overall satisfaction and attribute performance should be driven contextually, and no specific predictions are made in that regard. H_4 simply states that attribute-level performance and overall satisfaction will have separate and distinct impacts on repurchase intentions.

Next, we propose an asymmetric effect of attribute-level performance on repurchase intentions. Previous research (e.g., Oliva, Oliver, and MacMillan 1992) shows that overall satisfaction and performance are related nonlinearly to repurchase intentions or loyalty. Feinberg and colleagues (1990, p. 113) found that across several product categories “the probability of repurchase was not isomorphic with either positive or negative service experiences.” Research also shows that satisfaction and dissatisfaction have different affective consequences (Oliver 1993), which may be related differentially to repurchase intentions. Furthermore, the same psychological principles that operate in the context of satisfaction also should operate in the context of repurchase intentions. Therefore,

H_5: Repurchase intentions will be affected asymmetrically by attribute-level performance. Negative performance on attributes will have a greater impact on repurchase intentions than positive performance.
In the next sections, we present results from three large-scale studies that test the preceding hypotheses. A unique feature of all three studies is that they use data from commercial studies conducted by firms measuring satisfaction with their product or service.

Study 1

Data for this study were provided by a large health maintenance organization (HMO) in the Northeastern United States. The HMO collected these data as part of its ongoing patient satisfaction measurement program. For this study, the entire data set, which contained information from 4517 telephone interviews conducted among patients enrolled with the HMO, was used. These interviews were conducted using Computer-Assisted-Telephone-Interview, which enabled the interviewer to probe, clarify, and follow up on responses to open-ended questions. Because responses to the open-ended questions were a vital part of the measurement process, interviewers were asked to exercise great caution in recording and following up on them.

The sampling frame for the data set consisted of patients who subscribed to 501 HMO offices affiliated with the HMO. A random sample was drawn from each HMO’s patient list. Thus, in effect, a proportionate-stratified sample, with each HMO as a stratum, was drawn. After contact with a household had been established, up to two members per household were interviewed as long as they evaluated different PCPs. Thus, each household provided only one evaluation per PCP. The qualified respondent was an adult from a family whose members had visited a PCP within the past 12 months. All 4517 interviews were used in the analysis. Although the exact response rate for this survey could not be ascertained, typical response rates for such surveys are more than 70%. Input from the managers at the HMO and informal comparisons with data from other waves of the study show that the sample is representative of the population served by the HMO.

Measures

The database contained measures for several constructs. There were two dependent measures for this analysis: overall satisfaction and repurchase intentions. “Overall satisfaction with the medical care at the PCP’s office” was operationalized as a five-point scale (1 = very dissatisfied, 5 = very satisfied). Repurchase intention or a patient’s intention to switch to a different PCP was operationalized as a yes or no response to the following question: “Do you intend to switch to a different primary care physician when you next have an opportunity?” Although both these measures are single-item scales, there is considerable precedent for using single-item measures in the context of large-scale satisfaction studies. LaBarbera and Mazursky (1983) discuss the issue of using single- versus multi-item scales for measuring overall satisfaction and conclude that, in large-scale survey research, the use of multi-item scales actually may decrease the quality of measurement rather than enhance it. Similarly, in their large-scale study of drivers of customer satisfaction in the computer industry, Kekre, Krishnan, and Srinivasan (1995) use a single-item measure for overall satisfaction.

Finally, Yi (1990), in his review of satisfaction research, compares the test/retest reliabilities of several multi- and single-item scales and finds that the test/retest reliability of single-item scales is acceptable (.55 to .84). For these reasons, we believe the measures are adequate.

Independent variables were created on the basis of open-ended data provided by patients. In a free-thought elicitation task during the telephone interview, respondents were asked about their experiences with various features of their visit to the PCP. The interviewer asked them to clarify and expand their thoughts.

These thought listings provided the basis for the independent variables for the first analysis. As shown in Table 1, each thought was coded in a distinct category. When queried, the project manager for this study indicated that the categories were determined by the managers at the HMO in conjunction with the project director. The process is to review all the open-ended data from a randomly chosen subset of interviews (usually 10%) and develop the categories so that they are mutually exclusive and collectively exhaustive. Then the coders code a portion of the data using these categories. After these data are coded, the coders and managers meet to see if the categories are “working,” that is, if the categories are descriptive and capture the “essence” of what respondents said. Furthermore, when making categories, the managers ensure that all attributes and thoughts are coded. In any given study, fewer than 2% of the thoughts are forced into the “other/not specified” category. The process of determining categories is done carefully, especially given the high cost of data collection and the strategic value of the information contained in these data.

For the purpose of this analysis, each category was coded as 1 if a respondent gave a response in that category and 0 otherwise. Three professional coders, each with more than 2–3 years of coding experience, coded the categories. However, each coder coded a separate set of respondents. Because the coding already had been done before the data set was made available for this study, no follow-up analyses could be done to assess the reliability and agreement of the coding. Therefore, to assess the agreement among the three coders, they were asked to code data from 337 respondents from a subsequent wave of a similar study conducted for another HMO in the same geographic area. For these 337 respondents of the pilot study, the three coders were in perfect agreement for 81% of the surveys. That is, coders coded every category identically for 81% of the respondents. Furthermore, Coders 1 and 2 match on 84.3% of the responses, Coders 2 and 3 match on 85.5%, and Coders 1 and 3 match on 85.2%. The associated Kendall’s tau correlation was as follows: .84 for Coders 1 and 2, .85 for Coders 1 and 3, and .85 for Coders 2 and 3 (all significant at p < .01).

With these limitations in mind, it can be said that such an approach to understanding satisfaction has been used by other consumer satisfaction researchers (Bitner, Booms, and Tetreault 1990; Gardial et al. 1994). This approach, though more unstructured than rating scales, is advantageous because it captures attribute perceptions that are not contaminated by the preconceived notions that a researcher could impose on the respondents by cuing them regarding specific attributes. The thought categories then were classified fur-
ther as positively or negatively valenced. The positively valenced items for each respondent were summed to create a new variable, POSITIVE. Similarly, a variable called NEGATIVE was created by summing all the negative items for each respondent. The positive/negative classification of each thought category is indicated in Table 1.

**Analysis and Results**

The general strategy employed for testing the proposed asymmetry consists of computing separate estimates for positive and negative performance and statistically comparing the absolute magnitude of the positive and negative estimates. The absolute magnitude is compared because negative performance should be related negatively to the dependent constructs, and positive performance related positively to the dependent constructs. Therefore, to get a proper comparison, the “signs” of the estimates must be ignored. As an illustration, a test of the asymmetry between the composite variables POSITIVE and NEGATIVE is done, in which the coefficient for variable POSITIVE is expected to have a positive sign, and the coefficient for the variable NEGATIVE is expected to have a negative sign. Therefore, testing whether \( |b_{\text{NEGATIVE}}| > |b_{\text{POSITIVE}}| \) suggests the following hypothesis:

\[ (1) \quad H_A: |b_{\text{NEGATIVE}}| > |b_{\text{POSITIVE}}| \]

The corresponding null hypothesis can be stated as follows:

\[ (2) \quad H_0: |b_{\text{NEGATIVE}}| \leq |b_{\text{POSITIVE}}| \]

To test \( H_A \), two rival models are compared to see whether \( H_0 \) can be rejected statistically in favor of \( H_A \). One model formally constrains \( |b_{\text{NEGATIVE}}| \leq |b_{\text{POSITIVE}}| \), as suggested in \( H_0 \), whereas in the other, the coefficients are free to vary as suggested in \( H_A \). Comparing the performance of the constrained model to that of the unconstrained model provides a test of the asymmetry. If the model with the constraint is rejected in favor of the alternative model (i.e., \( H_0 \) is rejected in favor of \( H_A \)), it can be concluded that the

**TABLE 1**

<table>
<thead>
<tr>
<th>Thought Categories and Their Relative Impact on Overall Satisfaction with Medical Care</th>
<th>Coefficient</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive Mentions</strong></td>
<td><strong>Matched Categories</strong></td>
<td><strong>Negative Mentions</strong></td>
</tr>
<tr>
<td>Thorough/attentive</td>
<td>.11***</td>
<td>Not thorough</td>
</tr>
<tr>
<td>Spends time with me/personable</td>
<td>.17****</td>
<td>Doesn’t spend enough time/hurries you in and out</td>
</tr>
<tr>
<td>Not afraid to refer to another doctor</td>
<td>.07</td>
<td>Won’t give referrals/trouble getting referrals</td>
</tr>
<tr>
<td>Listens to his/her patients</td>
<td>.03</td>
<td>Not interested in patients/doesn’t listen</td>
</tr>
<tr>
<td>Convenient office hours</td>
<td>.07</td>
<td>Office hours are not convenient</td>
</tr>
<tr>
<td>Convenient office location</td>
<td>.02</td>
<td>Distance of office is inconvenient</td>
</tr>
<tr>
<td>Always available/availability of doctor</td>
<td>.15***</td>
<td>Not available/isn’t always in office</td>
</tr>
<tr>
<td>No wait/sees patients right away</td>
<td>.14**</td>
<td>Wait too long/office too crowded</td>
</tr>
<tr>
<td><strong>Unmatched Categories</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good, excellent doctor care</td>
<td>.26****</td>
<td>No follow up/don’t tell you results</td>
</tr>
<tr>
<td>Knowledge/educated/competent</td>
<td>.02</td>
<td>Doctor/staff overbooks appointments</td>
</tr>
<tr>
<td>Explains everything well/answers questions</td>
<td>.12**</td>
<td>Poor staff/problem with staff</td>
</tr>
<tr>
<td>Helpful</td>
<td>.13*</td>
<td>Problems with HMO</td>
</tr>
<tr>
<td>Good bedside manner</td>
<td>.13</td>
<td>Incorrect diagnosis/unnecessary testing</td>
</tr>
<tr>
<td>Professional</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>Friendly/nice/courteous</td>
<td>.11***</td>
<td></td>
</tr>
<tr>
<td>Concerned/caring/sympathetic</td>
<td>.22****</td>
<td></td>
</tr>
<tr>
<td>Good with kids</td>
<td>.19****</td>
<td></td>
</tr>
<tr>
<td>Makes me comfortable/easy to talk to</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Very patient</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>Sincere/honest/straightforward</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>Outgoing/good personality/sense of humor</td>
<td>.25*</td>
<td></td>
</tr>
<tr>
<td>Is warm/kind</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>Is an old-fashioned type of doctor</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>Calls back right away/keeps in contact/follows thorough</td>
<td>.17***</td>
<td></td>
</tr>
<tr>
<td>Always available during emergency</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>Promptness of service</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Has been our doctor for a long time</td>
<td>.33****</td>
<td></td>
</tr>
<tr>
<td>I like him/her</td>
<td>.20***</td>
<td></td>
</tr>
<tr>
<td>Staff is very nice/good</td>
<td>.14**</td>
<td></td>
</tr>
<tr>
<td>Cleanliness of office/building</td>
<td>.05</td>
<td></td>
</tr>
</tbody>
</table>

*\( p < .10; \) *** \( p < .05; \) **** \( p < .01; \) ***** \( p < .001; \) Model \( R^2 = .23, F = 30.9, \) error \( df = 4481, \) \( p < .0001. \)
absolute magnitude of the coefficient for NEGATIVE is greater than POSITIVE, or that |b_{NEGATIVE}| > |b_{POSITIVE}|. Next, the results are described.

Negative versus positive attribute performance (H_{1a}). A dummy-variable regression was conducted using all of the thought categories as independent variables and overall satisfaction as the dependent variable. As shown in Table 1, there were 30 positively valenced thought categories and 13 negatively valenced thought categories. On the basis of these, the following model was estimated:

\[
\text{Overall Satisfaction} = \text{Intercept} + Attnpos_1 + \ldots + Attnpos_{30} + Attnneg_1 + \ldots + Attnneg_{13}
\]

In this model, the impact of each attribute on overall satisfaction is calculated individually. Furthermore, positively valenced attribute mentions are expected to have positive signs, whereas negatively valenced attribute mentions are expected to have negative signs. To test H_{1a}, we must test whether the absolute impact of a negative mention for an attribute is greater than the absolute impact of a positive mention for the same attribute. The model is significant (F = 30.9, p < .0001) and is reported in Table 1.

Several findings of interest emerge from this table. First, as shown in the first eight rows of Table 1, there were eight attributes for which there were both negative and positive categories. Therefore, for these eight attributes, H_{1a} could be tested. For all eight pairs, the negatively valenced attributes have a larger (in the absolute sense) coefficient than the corresponding positively valenced attributes. We used the statistical test explained previously to test whether the negative coefficient for each pair of attributes is larger than the corresponding positive coefficient for the same attribute; the null hypothesis in Equation 2 was rejected for seven of the eight pairs at p < .001. For example, “always available/ availability of doctor” (.15, p < .01) has a smaller impact on patient satisfaction than “not available/ isn’t always in his office” (−.67, p < .0001). Similar results can be noted for (1) “not afraid to refer to another doctor” (.07, p < .26) versus “won’t give referrals/trouble getting referrals” (−.77, p < .0001) and (2) “spends time with me/personable” (.17, p < .0001) versus “doesn’t spend enough time/rushes you in and out” (−.40, p < .0001). Only for “convenient office location,” for which both the positive and negative coefficients were nonsignificant, was the hypothesis not supported. These results support H_{1a}. However, because this study is only able to provide a sense of the direction of a person’s evaluation on each attribute, a full test of H_{1a} cannot be done. In other words, because there are no data on the magnitude of positive or negative performance on each attribute, additional testing for this hypothesis is required. Recall that though H_{1b} is stated in terms of varying levels of performance on each attribute, there are no data on varying levels of performance in this study. Nevertheless, these results are encouraging and point toward the type of investigation that is carried out in Study 3.

Second, different attributes of the service encounter contribute differently to overall satisfaction. For example, “cleanliness of office/building” (.05, p = .78) is not consequential, whereas “no follow-up/doesn’t tell you results” (−.65, p < .0001) has a greater impact on overall satisfaction. Several other comparisons can be made with similar results. These echo earlier findings (Bolton and Drew 1991) that service attributes may be core, facilitating, or supporting and provide confidence in our results.

Third, with an R^2 of 23%, the overall variation explained by the model appears to be low. However, this could have resulted partially from loss of information resulting from the dichotomous nature of the independent variable (Cohen 1983). Furthermore, because the main concern in this study is to test the asymmetry hypothesis rather than explain variation in the dependent variable, the low R^2 is not particularly troublesome.

Cumulative impact and diminishing sensitivity (H_{3a} and H_{3b}). To test H_{3a} and H_{3b}, the following equation was estimated:

\[
\text{Overall satisfaction} = b_1(\text{POSITIVE}) + b_2(\text{POSITIVE})^2 + b_3(\text{NEGATIVE}) + b_4(\text{NEGATIVE})^2
\]

In Equation 4, POSITIVE and NEGATIVE are the sums of positively and negatively valenced mentions of attributes, respectively. Therefore, b_1 and b_2 model the asymmetric cumulative impact of positive and negative performance on overall satisfaction. To test for diminishing sensitivity, the squared terms are introduced (cf. Hamilton 1992). Thus, for positive performance, b_1 and b_2 model diminishing sensitivity, whereas for negative performance, b_3 and b_4 model diminishing sensitivity. These tests (for H_{3a} and H_{3b}) are described in detail next.

Cumulative impact of number of positive or negative attribute mentions (H_{3c}). As stated previously, b_1 and b_2 can be used to test the asymmetry for the cumulative impact of negative and positive mentions of attributes on overall satisfaction stated in H_{3a}. Because negative mentions of attribute performance are hypothesized to have greater impacts on overall satisfaction than are positive mentions, |b_3| should be greater than |b_2|. Similar to the test of H_{1a}, two rival models are compared to test whether the model with the constraint |b_3| ≤ |b_2| performs better or worse than the unconstrained model. If the constrained model performs worse than the unconstrained model, the constraint is rejected, which supports |b_2| > |b_3|. Results of the analysis are shown in Table 2. The coefficients for POSITIVE and NEGATIVE are .29 and −.72, respectively. Both are in the expected direction and significant at p < .0001. Moreover, their absolute magnitude is as hypothesized; the absolute value of the coefficient for NEGATIVE is greater than the absolute value of the coefficient for POSITIVE. This was confirmed by comparing the constrained and unconstrained models as explained previously. The model with the constraint |b_3| ≤ |b_2| was rejected in favor of the unconstrained model, where |b_2| > |b_1| (F_{1,4477} = 50.37, p < .0001). Thus, H_{3a} is supported.

These models also were estimated using standardized coefficients. The results for the models with standardized coefficients were practically identical to the results reported here.

38 / Journal of Marketing, January 1998
negative performance on attributes has a significantly greater impact on overall satisfaction than positive performance on attributes. The regression coefficients show that for this particular service, the deleterious effect of an additional instance of negative performance on attributes outweighs the beneficial effect of positive performance by a margin of two to one (−.72 versus .29).

**Diminishing sensitivity (H3b).** This hypothesis suggests a nonlinear relationship between overall satisfaction and performance on additional attributes. This relationship is modeled using the polynomial function described in Equation 4. In Equation 4, b1 and b2 test for diminishing sensitivity in the domain of positive performance. The linear positive coefficient b1 models the positive impact of performance, whereas b2, the coefficient for the squared term, models diminishing sensitivity by creating a “dampening” effect on the slope of b1. Instead of being a straight line as when b1 was used alone, the function becomes a curve that “tapers off” when the squared term (b2) is included in the equation. This occurs because every increase in b1 is countered by a smaller (because of the squared term) decrease in b2. For example, the estimate for b1 is .28, whereas the estimate for b2 is −.04. Consider the effect of increasing positive performance for one and two attributes. In the absence of the squared term, the beneficial impact of one attribute is .29 and of two attributes is .58; however, with b2, the beneficial impact increases but at a decreasing rate. Using both b1 and b2, the beneficial impact of one attribute is .25, calculated as .29(1) − .04(1)2, whereas the impact of two attributes is .42, calculated as .29(2) − .04(2)2. Thus, inclusion of the squared term captures diminishing sensitivity. Similar results apply for b3 and b4 in the domain of negative performance. Results based on this analysis are reported in Table 2.

As seen in Table 2, all four coefficients are statistically significant and in the predicted direction, which thereby supports H3b. However, a key concern when using these data to test H3b is that support for H3b may be an artifact of the way the variables POSITIVE and NEGATIVE were constructed. Some of the attributes that went into the construction of the variable may be inherently less important or salient than others. Thus, the test for H3b may be simply an artifact of the lower importance or salience of some attributes used in the construction of the cumulative variable than others. Additional tests, described in the Appendix, were used to test for this possibility. In these tests, the variables POSITIVE and NEGATIVE were constructed in eight different ways, and H3b was tested eight different times. Therefore, including the base case described in Table 3, a total of nine tests were conducted for H3b. All nine tests supported the diminishing sensitivity in the positive performance domain, but only five supported diminishing sensitivity for negative attribute performance. Therefore, erring on the side of caution, the results reported in Table 2 are believed to be valid.

### Table 2
**Regression Results for Satisfaction With Medical Care**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intercept</th>
<th>POSITIVE</th>
<th>NEGATIVE</th>
<th>POSITIVE²</th>
<th>NEGATIVE²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>4.24</td>
<td>.29</td>
<td>−.72</td>
<td>−.04</td>
<td>.07</td>
</tr>
<tr>
<td>Standard Error</td>
<td>.02</td>
<td>.02</td>
<td>.05</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>t-statistic</td>
<td>199.48</td>
<td>13.87</td>
<td>−13.48</td>
<td>−7.78</td>
<td>3.01</td>
</tr>
<tr>
<td>p-value</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0030</td>
</tr>
</tbody>
</table>

R² = .23; F₄,₄₄₇₇ = 330.65, p < .0001.

Test for constraint for H₀ in Equation 2: F₁,₄₄₇₇ = 50.37, p < .0001 (one-tailed test).

### Table 3
**Logistic Regression for Intention to Switch to a Different PCP**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1 Attribute</th>
<th>Model 2 Overall Satisfaction</th>
<th>Model 3 Attribute and Overall Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−1.50 (.07)</td>
<td>4.17 (.23)</td>
<td>3.14 (.24)</td>
</tr>
<tr>
<td>POSITIVE</td>
<td>−.66 (.05)</td>
<td>—</td>
<td>−.43 (.05)</td>
</tr>
<tr>
<td>NEGATIVE</td>
<td>1.21 (.08)</td>
<td>—</td>
<td>.76 (.08)</td>
</tr>
<tr>
<td>Overall satisfaction</td>
<td>—</td>
<td>−1.45 (.06)</td>
<td>−1.15 (.06)</td>
</tr>
<tr>
<td>Chi-square (d.f.)</td>
<td>708.7 (2)</td>
<td>989.9 (1)</td>
<td>1182.2 (3)</td>
</tr>
<tr>
<td>(p &lt; .0001)</td>
<td>(p &lt; .0001)</td>
<td>(p &lt; .0001)</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>2905.22</td>
<td>2606.10</td>
<td>2417.71</td>
</tr>
<tr>
<td>% correctly classified</td>
<td>86.0%</td>
<td>88.1%</td>
<td>89.5%</td>
</tr>
</tbody>
</table>

All coefficient estimates are significant at p < .001; standard errors are in parentheses. It could be argued that POSITIVE and NEGATIVE are not 2 but 43 variables because they are composed of multiple binary variables. Therefore, the AIC also was recalculated after penalizing Models 1 and 3 for 43 rather than 2 variables. The results for model fit were virtually identical to the one reported in the table.

Negative and Positive Attribute-Level Performance / 39
of conservatism, the conclusion is that diminishing sensitivity holds for positive performance but not for negative performance on attributes.

These results suggest that overall satisfaction displays diminishing sensitivity to each additional instance of positive performance but not necessarily to negative performance. In other words, though the relative benefit of positive performance on additional attributes may decrease, each additional instance of negative performance may be equally deleterious.

Repurchase intentions (H₄ and H₅). Because the measure for switching intentions was a dichotomous variable, a logistic regression analysis was conducted to test H₄ and H₅. Three models predicting intention to switch to a different PCP were estimated. Note that though the hypotheses are stated in terms of repurchase intentions, results are discussed in terms of intention to switch to maintain consistency with the data. The models used for testing H₄ and H₅, which are shown in Table 3, are as follows:

(5) Model 1: intention to switch = f(POSITIVE, NEGATIVE);
(6) Model 2: intention to switch = f(Overall satisfaction);
(7) Model 3: intention to switch = f(POSITIVE, NEGATIVE, Overall satisfaction).

In Model 1, only POSITIVE and NEGATIVE were used as predictor variables. In Model 2, only overall satisfaction was used as a predictor variable. The third model used POSITIVE, NEGATIVE, and overall satisfaction as predictor variables. Note that these models are concerned only with asymmetry, not diminishing sensitivity; therefore, only the cumulative variables POSITIVE and NEGATIVE are used in these models. Results from these models are reported in Table 3.

Results for Model 1, the attributes-only model, confirm the hypothesis that negative attributes are more consequential for intention to switch than are positive attributes of the service. Thus, H₅ is supported. Next, all three models in Table 3 were examined to test H₄. The fit of the models was examined using the Akaike Information Criterion (AIC statistic) of model evaluation (Bollen 1989). The AIC evaluates alternative models while accounting for the number of parameters in rival models; it penalizes models with many rather than few free parameters. Models with smaller AIC values provide a better fit. The AIC for each of the three models is shown in Table 3.

The overall satisfaction model (Model 2) provides a better fit in predicting repurchase intentions than does the attribute performance model (Model 1); the AIC fit for Model 2 is smaller than that of Model 1. However, Model 3, which includes both attribute performance and overall satisfaction, outperforms either model. The value of the AIC for Model 3 is smaller than either the overall satisfaction or the attribute performance model. Thus, H₄ is supported.²

²These results also were verified by dividing the data set in two parts: one with 3517 and the other with 1000 people. The three models were estimated with the first data set and then used to predict the value for intention to switch in the holdout sample of 1000 people. The coefficient estimates were virtually identical to the ones reported in Table 4, and the percentages correctly classified for the three models were 65%, 69.9%, and 78.9%, respectively. These results again support the hypotheses.

Attribute performance and overall satisfaction have separate and distinct effects on repurchase intentions.

Summary of results and limitations. The results from this study support the hypothesized asymmetry and provide a strong indication of the existence of diminishing sensitivity, at least on a cumulative basis. Negative performance on seven of eight pairs of attributes has a greater impact on overall satisfaction than does positive performance. This is also true of the cumulative impact of negative versus positive attributes. Similarly, intentions to switch also are affected asymmetrically by positive and negative performance. Overall satisfaction appears to display diminishing sensitivity to attribute performance in the domain of positive performance but not negative performance; however, this conclusion cannot be offered for different levels of performance within each attribute. Finally, both overall satisfaction and attribute performance have separate and distinct impacts on repurchase intentions.

These results should be viewed in light of the study’s limitations. First, there were no data on expectation disconfirmation at the attribute level; this precluded testing H₁b, a limitation that is addressed in Study 2. Second, attribute-level performance was measured using 0,1 categories, which precluded measurement of different levels of performance for an individual attribute. Thus, H₁a and H₂, which are stated for different levels of performance within an individual attribute, could not be tested. Study 3, in which performance levels are measured for each attribute, fully examines H₁a and H₂. Third, Study 1 was conducted in a service setting; therefore, it could be argued that the results of this study would not necessarily generalize to a product-related setting. Although there is no theoretical reason to expect this, testing these results in a product setting should increase confidence in the generalizability.

Study 2

The purpose of this study is twofold: (1) to corroborate the findings of Study 1 in a different setting—automobiles—and (2) to test H₁b, which postulates an asymmetric impact of positive and negative disconfirmation on attributes on overall satisfaction. It is important to provide a separate test of H₁b, because both performance and disconfirmation have been found to have distinct impacts on overall satisfaction (Yi 1990).

Sample

Data for this study were obtained from a mail survey of new car buyers throughout the United States. This survey is part of an ongoing tracking study conducted by a domestic automotive manufacturer. This study uses a mail survey, and the response rate for a typical wave of this study ranges from 36% to 40%, which is similar to the industry standard for satisfaction surveys in general. For example, Mittal and colleagues (1993) report results from a large-scale automotive study with a response rate of 33.5%, Hall (1995) reports results from a patient satisfaction context with a response rate of 41.1%, and LaBarbera and Mazursky (1983) report results from grocery item shoppers with a response rate of 25%.
The database included responses from 9359 new car buyers in one wave of the larger study. These respondents were selected randomly from the larger database of 250,000 respondents that the organization has accumulated. A chi-square test on key demographics and t-tests on the dependent variable showed no difference between the larger database and the random subsample. Furthermore, informal comparisons with other industry databases and inspection by managers in the automotive research industry confirmed that the sample was nationally representative of automobile buyers.

### Measures

Overall satisfaction was measured on a ten-point scale with the following anchors: 1 = very dissatisfied and 10 = completely satisfied. Furthermore, disconfirmation measures were available on ten attributes that had been selected by the firm on the basis of qualitative research. Attribute-level disconfirmation was measured on a three-point scale in which respondents were asked to indicate whether performance on each attribute was "better than what you expected," "about what you expected," or "worse than what you expected."

Operationalizing disconfirmation in this way facilitates a comparison of the asymmetric impact of positive versus negative disconfirmation in a manner that is consistent with previous studies (cf. Oliver and DeSarbo 1988). This measure was coded as a dummy variable such that "better than expected" was coded as (1,0), "worse than expected" was coded as (0,1), and "about as expected" was coded as (0,0). With this coding scheme, the asymmetric effects between positive and negative disconfirmation on each attribute can be compared; the neutral level serves as the "baseline" against which the coefficients for "above" or "below" expectations are computed (cf. Pedhazur 1982). In other words, the estimates for the neutral level (about as expected) for each attribute are confounded with the intercept term.

### Analysis and Results

A dummy-variable regression was conducted to test H1b. The specific equation that was estimated is

(8) Overall Satisfaction = Intercept + \(b_{1i}\) Attribute\(_i\) (above expect.) + \(b_{2i}\) Attribute\(_i\) (below expect.) + \(...\) \(b_{10}\) Attribute\(_{10}\) (above expect.) + \(b_{20}\) Attribute\(_{10}\) (below expect.)

In Equation 8, \(b_{1i}\) represents the relative impact of positive disconfirmation on attribute \(i\), whereas \(b_{2i}\) represents the relative impact of negative disconfirmation on the same attribute on overall satisfaction. Using this formulation, the impact of each attribute on overall satisfaction was estimated. This data set had expectation/disconfirmation data on ten separate attributes, which resulted in 20 specific coefficients, one for positive and one for negative disconfirmation on each of the ten attributes. Results of the analysis are shown in Table 4 and are discussed subsequently in two parts. First, the substantive results related to H1b are discussed; second, the overall model, in terms of the variance explained and the different attributes included in the model, is discussed.

Recall that H1b postulates that negative disconfirmation on an attribute will have a greater impact than positive disconfirmation on the same attribute. Table 4 supports this hypothesis. For all ten attributes, the absolute value of the coefficients is greater for negative disconfirmation than for positive disconfirmation. For example, the impact of negative disconfirmation on "comfort" is three times larger than the impact of positive disconfirmation on the same attribute (−0.86 versus .24). Negative disconfirmation on "quality of vehicle" affects overall satisfaction three times more than positive disconfirmation (−1.43 versus .42).

Equivalent results can be noted for other attributes, though the magnitude of the asymmetry between positive and negative disconfirmation varies somewhat. For example, the magnitude of the asymmetry is smaller for "ease of getting in and out of vehicle" (−0.39 versus .34) than for "power and pickup" (−0.62 versus .08). Overall, the results support H1b. More interesting, for many attributes, positive disconfirmation has a nonsignificant impact (e.g., "ride quality," "quietness of engine"). For these attributes, it seems that there is no additional benefit in exceeding customer expectations, though the negative consequences of not meeting expectations are relatively high. These results also confirm the findings of Study 1 that different attributes have varying degrees of impact on overall satisfaction. For example, the impact of negative disconfirmation on "quality of vehicle" (−1.43, \(p < .0001\)) is much larger than negative disconfirmation on "driver's seating comfort" (−.29, \(p < .05\)) or "ride quality" (−.42, \(p < .01\)). Again, this suggests that attributes could be classified as core or facilitating attributes on the basis of the magnitude of the asymmetry.

However, the results of both Studies 1 and 2 must be scrutinized further in light of the fact that the models used in both studies explain a small amount of variation (23% and 18%, respectively) in overall satisfaction. There are several possible reasons for this. One explanation is that when dummy variables instead of regularly scaled items are used to operationalize attribute-level performance or disconfirmation, the overall variance explained by the model generally decreases because of loss of information for the inde-
dendent variable (Cohen 1983; Cox 1980). In the context of satisfaction literature, only one study has examined the issue of scale categories and explained variation. Wittink and Bayer (1994) conducted a quasi-experimental study to investigate different measurement systems for satisfaction research. They found that the system using dichotomized variables for measuring attribute-level performance (e.g., high versus low) had a lower $R^2$ than the one using a five-point scale for measuring attribute-level performance. Thus, it can be argued that the lower $R^2$ is simply an artifact of the measurement system adopted here; still, we cannot rule out the alternative explanation that all possible attributes that explain overall satisfaction were not included. This alternative explanation will be examined by comparing the results from Studies 2 and 3, because both pertain to overall satisfaction with an automobile and measure similar attributes.

Study 3

Study 3 included data from 13,759 respondents who filled out a satisfaction survey in the context of the automotive industry. Before describing the study, it is important to understand the specific rationale behind including Study 3. Although the results of the previous studies indicate some support of the central thesis—namely, the asymmetry and diminishing returns—if this work, they do not test adequately for both asymmetry and diminishing returns for each individual attribute ($H_{1a}$ and $H_2$). For example, Study 1 shows diminishing sensitivity but only for all attributes combined, whereas Study 2 is unable to test for diminishing sensitivity. Therefore, to provide direct evidence for $H_{1a}$ and $H_2$, this third study is reported.

Sample

Analysis for Study 3 was based on a customer satisfaction data set of a major automotive manufacturer in the United States. This data set was assembled using a mail survey methodology. Similar to the data in Study 2, these data were collected as part of an ongoing tracking study to assess satisfaction with various aspects of the vehicle ownership experience. A survey was mailed to people who owned their vehicle for two or more years. The response rate for this survey varied from wave to wave but ranged between 38% and 42% for the last three waves.

In the survey, respondents rated their overall satisfaction with the vehicle and evaluated performance on six different attributes: brakes, transmission, power and pickup, vehicle quality, quietness, and interior roominess. Data from 13,759 respondents who answered this section of the survey were made available for this analysis. Similar to Study 2, the demographic profile of these respondents was compared with the demographic profile of the larger database from which they were extracted. No statistically significant differences were found in terms of age, sex, income, race, or geographic location (for all, $p > .5$).

Measures

There were two main constructs: overall satisfaction and attribute-level performance. Overall satisfaction was measured on a ten-point scale (10 = completely satisfied, 1 = completely dissatisfied). Performance on each of the six attributes was measured on an eight-point scale (4/3 = very satisfied; 2/1 = somewhat satisfied; -1/-2 = somewhat dissatisfied; and -3/-4 = very dissatisfied with performance on the attribute). Note that this scale is a satisfaction scale and therefore does not measure performance directly at the attribute level. However, recent studies in the area of customer satisfaction show that though attribute-level satisfaction and performance are conceptually distinct constructs, their measures are highly correlated. For example, Spreng, MacKenzie, and Olshavsky (1996) find high correlation between attribute-level satisfaction and performance for camcorders. In their study, the correlation between performance and satisfaction for “versatility” was .72 ($p < .01$), and “picture” was .80 ($p < .01$). Furthermore, both performance and satisfaction ratings on each attribute had similar correlations with overall satisfaction. Bittner and Hubbert (1994, p. 85) find a high interfactor correlation (.94) between satisfaction and quality/performance and conclude that “this finding provides evidence for the notion that overall service satisfaction and service quality may be tapping elements of the same construct.” Thus, for the sake of measurement, the use of the satisfaction scale as a surrogate of attribute-level performance is reasonable for the current study. However, this is in no way designed to suggest that the conceptual distinction between performance and satisfaction at the attribute level should not be maintained.

Recall that the scale being used to measure attribute-level performance is an eight-point scale; on this scale, positive ratings on an attribute reflect positive performance on that attribute, whereas negative ratings indicate negative performance on that attribute. For each attribute, there are four levels or grades of positive performance (1, 2, 3, 4) and four grades of negative performance (-1, -2, -3, -4). Thus, the effect of different levels of performance and diminishing sensitivity can be modeled for each individual attribute.

Analysis and Results

The analytic strategy for this study was adapted from Anderson and Sullivan (1993). Accordingly, the asymmetric and diminishing impact of each attribute on overall satisfaction is modeled as follows:

$$\text{Overall Satisfaction} = \text{Intercept} + \text{LN_PERF}_1 + \text{LP_PERF}_1 \ldots \text{LN_PERF}_6 + \text{LP_PERF}_6. \tag{9}$$

In Equation 9, the measurement variable for each attribute is decomposed into $\text{LP_PERF}$ and $\text{LN_PERF}$, where $L$ denotes the natural logarithm of each measure, $P_{\text{PERF}}$ or $N_{\text{PERF}}$ (for positive and negative performance, respectively). Thus, performance on each attribute is decomposed into positive and negative performance as follows: If performance on the first attribute is negative, the measurement variable for negative attribute performance, $\text{LN_PERF}_1$, is equal to $-\ln (-\text{PERF}_1)$,\(^5\) and

\(^5\)The negative sign for $\text{PERF}_1$ for negative performance makes the final outcome positive and allows the use of the natural logarithm of the number. For example, if there is a negative performance rating of $-4$, then setting the variable to $-(-4)$ yields $+4$, for which it is possible to take the natural logarithm. Recall that natural logarithms do not exist for negative numbers.
LP_PERF is equal to zero. If performance on an attribute is positive, then LP_PERF is equal to ln (PERF_i), and LN_PERF is equal to zero. Thus, if the first attribute received a rating of -4, LN_PERF = ln {(-4)}, and LP_PERF = 0. Conversely, if an attribute was rated 3, LP_PERF = ln(3), and LN_PERF = 0. Note that in this analysis plan, two coefficients are estimated for each attribute for a total of 12 coefficients.

This attribute-level measurement plan accomplishes two goals. First, it ensures that all of the coefficients will be positive; thus, for example, the coefficients for positive and negative performance on a given attribute will be positive. This makes the interpretation of results more convenient and therefore managerially useful. Thus, if the coefficient for negative performance is larger than the coefficient for positive performance on an attribute, the hypothesized asymmetry would be supported. Second, the natural logarithm transformation captures diminishing return or sensitivity (Anderson and Sullivan 1993). For example, if the coefficient for positive performance on an attribute is significant, it can be interpreted as supporting the diminishing sensitivity hypothesis for positive performance on that attribute.

At this point, it is necessary to explain why this plan of analysis rather than the one used in Equation 4 was adopted to model the asymmetry and diminishing sensitivity. First, to use an analytic plan similar to Equation 4, four coefficients would have been required to model the asymmetry and diminishing returns on each of the six attributes (positive, negative, positive^2, and negative^2 for each attribute); thus, the use of a plan similar to Equation 4 would necessitate the estimation of 24 coefficients. Equation 9, conversely, requires the estimation of only two coefficients per attribute for a total of 12 coefficients. Second, the use of squared terms in polynomial equations such as Equation 4 introduces severe multicollinearity (e.g., positive and positive^2 for an attribute are likely to be highly correlated). Although such multicollinearity may not be a serious issue for a single attribute, when modeling several attributes, it becomes difficult to detect and control for (e.g., Wittink and Bayer 1994). Thus, Equation 9 not only provides a parsimonious representation of the model (12 instead of 24 terms), but also circumvents issues related to multicollinearity that would arise because of the use of squared terms along with the base terms (e.g., positive and positive^2) for several attributes. Finally, the use of Equation 9 not only complements the results of Study 1 by testing the asymmetry and diminishing sensitivity for each individual attribute, but also compares attribute-level results with those reported by Anderson and Sullivan (1993).

Given this measurement and analysis plan, each coefficient in Equation 9 should be positive. The asymmetry is tested by constraining the positive and negative coefficients for each attribute to be equal (LN_PERF_i = LP_PERF_i), determining whether the constraint can or cannot be rejected, and reporting whether LN_PERF > LP_PERF. If the constraint is rejected and the absolute size of the coefficient for negative performance is greater than the coefficient for positive performance, then the asymmetry is supported. Furthermore, the natural logarithm transformation captures diminishing sensitivity, and a statistically significant coefficient indicates support for the diminishing hypothesis. Results for these analyses are reported in Table 5 and discussed next.

**Asymmetry between positive and negative performance (H1a)**. The results in Table 5 show that all estimates are statistically significant. For five of the six attributes, H1a is supported because LN_PERF > LP_PERF and the constraint LN_PERF = LP_PERF is rejected. A comparison of the coefficients shows that the magnitude of the asymmetry is different for each attribute. Thus, the magnitude of the asymmetry is much larger for “transmission” (.04 versus .35) than for “quietness” (.11 versus .23). Similar to the results of Study 2, “vehicle quality” has the largest impact on overall satisfaction (even after two years of ownership).

**Diminishing sensitivity**. The logarithm transformation facilitates testing the diminishing sensitivity hypothesis. As the values get more extreme, a logarithmic function tapers off and thus resembles the diminishing sensitivity curve. Thus, significant coefficients that are based on the natural logarithm indicate support for the diminishing sensitivity hypothesis (cf. Anderson and Sullivan 1993). However,

<table>
<thead>
<tr>
<th>Attribute</th>
<th>LP_PERF</th>
<th>LN_PERF</th>
<th>Reject Constraint?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Regression coefficients for positive performance on an attribute)</td>
<td>(Regression coefficients for negative performance on an attribute)</td>
<td></td>
</tr>
<tr>
<td>Brakes</td>
<td>0.00*</td>
<td>0.24***</td>
<td>Yes***</td>
</tr>
<tr>
<td>Transmission</td>
<td>.04*</td>
<td>.35***</td>
<td>Yes***</td>
</tr>
<tr>
<td>Power and pickup</td>
<td>.06**</td>
<td>.21***</td>
<td>Yes**</td>
</tr>
<tr>
<td>Vehicle quality</td>
<td>1.41***</td>
<td>1.66***</td>
<td>Yes***</td>
</tr>
<tr>
<td>Quietness</td>
<td>.11***</td>
<td>.23***</td>
<td>Yes*</td>
</tr>
<tr>
<td>Interior roominess</td>
<td>.49***</td>
<td>.17**</td>
<td>Yes***</td>
</tr>
<tr>
<td>R^2 = .65; F_{12,13758} = 2210.1, p &lt; .0001.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***p < .0001; **p < .01; *p < .05; n.s > .10.
whether the logarithmic transformation added any additional explanatory power to that of a model without diminishing returns also was checked. Another baseline model in which attribute performance had not undergone a logarithmic transformation also was estimated. Results from both models were virtually identical. For the baseline model—that is, the model without the log-transformed variables—\( R^2 = .65 \) and \( F_{12,13758} = 2210.9 \) show equally good fit. A chow-test confirmed that the model with diminishing returns did not fit significantly better than the model without diminishing returns. Therefore, even though the model with diminishing returns is statistically significant, it can be concluded that it does not add much to the explanatory power of the model.

**Summary of results.** The results support the asymmetry between different levels of negative and positive attribute performance but are inconclusive with regard to diminishing returns. One important part of the results is that for “roominess” a reversal of the hypothesized asymmetry is observed. The implications of this asymmetry for developing a theoretical typology of attributes as they relate to overall satisfaction are discussed in a subsequent section.

In addition, Study 3, which has a higher \( R^2 \) than Study 2, is based largely on the same attributes as Study 2. “Interior roominess” is the only attribute that is not measured directly in Study 2. However, it can be argued that in Study 2, attributes such as “comfort,” “ease of getting in and out of vehicle,” “driver’s seating comfort,” and “ride quality” capture most of the aspects of “interior roominess” from Study 3. Similarly, though Study 1 has four times the attributes of Study 3, the variance explained by Study 3 is much higher. Thus, it seems reasonable to infer that the lower \( R^2 \) obtained in Studies 1 and 2 is an artifact of information loss due to coarse measurement scales rather than exclusion of relevant attributes.

**Discussion**

We investigate the link between attribute-level performance/disconfirmation, satisfaction, and repurchase intentions. Using three large-scale data sets, several hypotheses about these relationships were tested. The results are summarized as follows:

- Both overall satisfaction and repurchase intentions are affected asymmetrically by attribute-level performance and disconfirmation. That is, negative performance/disconfirmation on an attribute has a greater impact than positive performance/disconfirmation.
- Regarding diminishing returns, the results are mixed. For all attributes combined, there is support for diminishing returns in the domain of positive performance but not negative performance (Study 1). When tested for diminishing returns for each individual attribute (Study 3), there is support for diminishing sensitivity for both positive and negative performance. However, data also show that a linear conceptualization fits the data as well as the diminishing return conceptualization.
- Finally, results show that, in addition to the impact mediated by satisfaction, attribute-level performance has a direct impact on repurchase intentions. This result calls into question previous models that assume that the impact of performance on repurchase intentions is mediated by overall evaluations (e.g., satisfaction).

These results show that the relationship between attribute-level outcomes and overall evaluations is complex. These results have several implications for research and practice that are discussed next.

**Implications for Research**

Satisfaction research at an attribute level typically has viewed attributes as unidimensional and classified them accordingly. That is, most typologies classify attributes on the basis of their weight or importance in determining overall satisfaction (cf. Bolton and Drew 1991; LaTour and Peat 1979). These results show that typologies that classify attributes on the basis of their weight alone might not be enough. More specifically, there is a need to develop a typology that enables researchers to understand why the magnitude of the asymmetry is different for different attributes. That is, is the observed asymmetry related to the utility-preserving or enhancing qualities of an attribute (cf. Kahn and Meyer 1991)? By their very nature, utility-preserving attributes (e.g., correct diagnosis by a doctor, safety in air travel) seem to have a higher potential for negative disconfirmation, whereas utility-enhancing attributes (e.g., doctor is humorous, entertainment aboard a flight) have a higher potential for positive disconfirmation. Accordingly, in Study 3, “interior roominess” would be a utility-enhancing attribute, whereas the other attributes might classify as utility-preserving attributes. Yet another reason could be the level of performance and disconfirmation observed in the past. Thus, for airline safety, consumers typically observe high positive performance/disconfirmation (i.e., air travel is relatively safe). If this expectation suddenly were disconfirmed negatively, then it would become salient and affect overall satisfaction. If a customer frequents a restaurant where the service is consistently below expectations (i.e., negative performance) and receives bad service again on his or her next visit, there would be little further impact. However, if the customer suddenly received top-quality service, it would have a great impact on overall satisfaction. This line of logic also helps explain why there was a reversal of the hypothesized asymmetry for “interior roominess.” It could be that in the past, automobile customers always have expected low performance on interior roominess; thus, encountering high performance on “interior roominess” causes a higher level of disconfirmation (in the positive direction). However, these ideas are speculative and their resolution is beyond the scope of this research. The development of a theoretically driven typology that enables a resolution of these findings should be a welcome addition to satisfaction research.

Analytically, these results call into question previous conceptualizations that treat the relationship between attribute-level performance and overall satisfaction as symmetric and linear (cf. LaTour and Peat 1979). More broadly, these results have implications for multi-attribute models of consumer decision making and choice (cf. Kahn and Meyer 1991; Wilkie and Pessemer 1973). Typical models of decision making do not distinguish between attributes on which a consumer has experienced positive or negative performance. Depending on this past experience, the weights of
the attributes in subsequent decisions could shift dramatically. For example, a consumer who has experienced positive performance on comfort in the decision to purchase a car may weight it less than another who has experienced negative performance on comfort. Previous experiences of attribute-level performance thus serve as important contextual cues for subsequent decisions and choices. Applying these findings to multi-attribute models of consumer choice is a key direction for further research.

Similarly, these results also point toward a careful reexamination of satisfaction and its consequent behaviors, such as retention and word-of-mouth activity. Previous conceptualizations have assumed these relationships to be linear and symmetric (Yi 1990). However, satisfaction’s impact on these behaviors could be asymmetric and nonlinear. High levels of satisfaction might not increase retention, but high levels of dissatisfaction might have a large and deleterious impact on retention. An asymmetric conceptualization also could help explain the curious finding that most firms are experiencing, that is, high rates of customer defection despite high rates of customer satisfaction (Reichheld 1996). High dissatisfaction with a product might prompt high levels of information search for alternative brands, whereas high levels of satisfaction might not decrease the amount of information search substantially; this asymmetry has implications for consideration set formation and choice. Given the increased emphasis on linking satisfaction research to the bottom line (Reichheld 1996), these issues warrant careful attention.

Finally, the finding that attribute-level performance also has a direct impact on repurchase intentions not only is surprising but also calls into question previous models of satisfaction. These models (for a review, see Iacobucci et al. 1996) presume that attribute-level evaluations affect behavioral intentions but only through an overall satisfaction judgment. This finding raises several important questions that need more research. First, the conditions under which the direct impact of attribute-level outcomes is larger (smaller) than the one mediated through satisfaction should be outlined. Second, it could be that the mediating role of overall satisfaction is different for different attributes and dependent on the magnitude and direction of the asymmetry. Developing a typology of attributes should help resolve these issues.

**Implications for Practice**

The key implication for managers is that they no longer should view attribute performance with a neutral eye. Although positive and negative performance on an attribute are two sides of the same coin, each side of the coin buys a different amount of overall satisfaction or repurchase intentions. This asymmetry should be kept in mind in trying to manage attribute-level performance. The strategic implication of this result is clear: To maximize overall satisfaction and repurchase intentions, managers should optimize and not maximize attribute-level performance. For example, for any given attribute it is more important to eliminate negative performance first and then focus on increasing performance in the positive direction. Similarly, in a given set of attributes, focus on attributes for which customers are experiencing negative performance and then allocate resources to maximizing performance on attributes for which consumers are experiencing positive performance.

**Concluding Comments**

Despite the initial wave of skepticism with satisfaction research among practitioners and academics, it now is acknowledged that customer satisfaction and retention are key strategic imperatives and not fads (e.g., Honomichl 1993; Reichheld 1996). Increasingly, researchers also are finding that the relationships between satisfaction and its antecedents and consequences are complex (cf. Anderson and Sullivan 1993; Coyne 1989; Oliva, Oliver, and Bearden 1995). Therefore, understanding the complexity of these interrelationships has become a key strategic imperative for most firms (e.g., Jones and Sasser 1995). We hope that this work represents a small step in that direction and will encourage further research aimed at understanding these complexities.

**Appendix**

Follow-up tests for $H_5$ were conducted to ensure that the reported results are not due to the differing salience or importance of some of the attributes used in the study. Salience is operationalized as the frequency with which an attribute is mentioned (higher number of mentions = higher salience), whereas importance is operationalized as the magnitude of the regression weight (larger regression coefficient = higher importance). To check for this possibility, the cumulative variables were reconstructed, excluding those attributes that were not important or salient.

To check for the salience bias, the first set of variables was constructed from attributes on the basis of the number of mentions. With number of mentions as a proxy for salience, only those attributes that received the top 3, 4, 5, or 6 mentions were included. To clarify, POSFRQ3 and NEGFREQ3 used only those attributes that received the first, second, and third highest number of mentions. POSFRQ4 and NEGFREQ4 also included the attribute that received the fourth highest number of mentions, and so on. To check for the importance bias, the cumulative variables were constructed on the basis of their imputed importance. With the regression weight in Table 2 as a proxy for an attribute’s importance, variables called POSREG5 and NEGREG5 were constructed using those attributes that had the five highest regression weights for the positive and negative categories. Similar variables also were constructed on the basis of the top six and seven regression weights.

These strategies of integration are consistent with prior theoretical and empirical research (e.g., Dawes 1979). Furthermore, satisfaction researchers (e.g., Oliver 1993) have used additive-linear combinations to integrate performance on several attributes. To summarize, eight additional sets of variables were constructed: POSFRQ or NEGFREQ3–6, POSREG or NEGREG5–7, and POSWTD or NEGWTDS. Next, the following model was estimated for each set of variables:

**Negative and Positive Attribute-Level Performance / 45**
TABLE A1
Summary of Various Tests for Asymmetry and Diminishing Sensitivity

<table>
<thead>
<tr>
<th>Variable</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
<th>R²</th>
<th>(F-statistic, significance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POS3FRQ (salience)</td>
<td>.58***</td>
<td>-.16***</td>
<td>-.98***</td>
<td>.15*</td>
<td>.14</td>
<td>9.8, .0020</td>
</tr>
<tr>
<td>POS4FRQ (salience)</td>
<td>.53***</td>
<td>-.13***</td>
<td>-.78***</td>
<td>.07n.s.</td>
<td>.15</td>
<td>4.9, .0300</td>
</tr>
<tr>
<td>POS5FRQ (salience)</td>
<td>.49***</td>
<td>-.11***</td>
<td>-.73***</td>
<td>.02n.s.</td>
<td>.17</td>
<td>6.4, .0100</td>
</tr>
<tr>
<td>POS5RG (importance)</td>
<td>.54***</td>
<td>-.14***</td>
<td>-.38***</td>
<td>.28***</td>
<td>.18</td>
<td>50.9, .0001</td>
</tr>
<tr>
<td>POS6FRQ (salience)</td>
<td>.44***</td>
<td>-.13***</td>
<td>-.78***</td>
<td>.08n.s.</td>
<td>.18</td>
<td>16.9, .0001</td>
</tr>
<tr>
<td>POS6RG (importance)</td>
<td>.52***</td>
<td>-.13***</td>
<td>-.78***</td>
<td>.08n.s.</td>
<td>.18</td>
<td>16.9, .0001</td>
</tr>
<tr>
<td>POS7RG (importance)</td>
<td>.46***</td>
<td>-.09***</td>
<td>-.21***</td>
<td>.22***</td>
<td>.21</td>
<td>66.6, .0001</td>
</tr>
<tr>
<td>POSMATCH (importance)</td>
<td>.35***</td>
<td>-.08**</td>
<td>-.80***</td>
<td>.05n.s.</td>
<td>.14</td>
<td>24.9, .0001</td>
</tr>
<tr>
<td>POSALL (all attributes/initial case)</td>
<td>.29***</td>
<td>-.04***</td>
<td>-.72***</td>
<td>.07**</td>
<td>.23</td>
<td>50.4, .0001</td>
</tr>
</tbody>
</table>

***p < .0001; **p < .001; *p < .05; n.s.p > .10.

Overall satisfaction = $b_1(\text{POSITIVE}) + b_2(\text{POSITIVE})^2 + b_3(\text{NEGATIVE}) + b_4(\text{NEGATIVE})^2$.

The results are described in Table A1. In all of the eight additional cases plus the initial case, $|b_4| > |b_2|$, which indicates that negative performance has a greater impact than positive performance as shown in the columns labeled $b_1$ and $b_2$ in Table A1. This result was significant at $p < .03$ in all nine cases (see the column labeled “Restrictions Test” in Table A1). These results show that the conclusion about $H_3$ is not a methodological artifact.

Regarding diminishing sensitivity, both $b_2$ and $b_4$ have the expected sign in all nine cases. Furthermore, $b_2$ was significant for all of the nine models ($p < .001$), but $b_4$ was significant only for five of the nine models. To ascertain the “joint” statistical significance on the basis of these nine models, a combined test such as Fisher’s combined test would be appropriate (Wolf 1986). However, because the nine comparisons are not independent, such a test of significance cannot be done. Therefore, erring on the side of conservatism, we conclude that over the nine tests, $b_4$ is not significant, whereas $b_2$ is. Thus, diminishing sensitivity holds for positive performance but not for negative performance on attributes.

REFERENCES


