

- implications of shifts in strategic emphasis. *Journal of Marketing*, 67, 63-76.
- Olson, J. C., & Reynolds, T. J. (1983). Understanding consumer's cognitive structures: Implications for advertising strategy. In L. Percy & A. G. Woodside (Eds.), *Advertising and consumer psychology* (pp. 77-90). Lexington, MA: Lexington Books.
- Park, C. S., & Srinivasan, V. (1994). A survey-based method for measuring and understanding brand equity and its extendability. *Journal of Marketing Research*, 31, 271-288.
- Parkhurst, J. (2002). Leveraging brand to generate value. In B. Berman (Ed.), *From ideas to assets: Investing wisely in intellectual property* (pp. 395-420). New York: John Wiley.
- Perrier, R. (1997). *Brand valuation* (3rd ed.). London: (Interbrand), Premier Books.
- Pessemier, E. (1959). A new way to determine buying decisions. *Journal of Marketing*, 24, 41-46.
- Rao, V. R., Agrawal, M. K., & Dahlhoff, D. (2004). How is manifested branding strategy related to the intangible value of a corporation? *Journal of Marketing*, 68(4), 126-141.
- Reynolds, T. J., & Gutman, J. (1988, February-March). Laddering theory: Method, analysis, and interpretation. *Journal of Advertising Research*, pp. 11-31.
- Reynolds, T. J., & Whitlark, D. B. (1995, July-August). Applying laddering data to communications strategy and advertising practice. *Journal of Advertising Research*, pp. 9-17.
- Ritson, M., & Elliott, R. (1999). The social uses of advertising: An ethnographic study of adolescent advertising audiences. *Journal of Consumer Research*, 26, 260-277.
- Srinivasan, V. (1979, January). Network models for estimating brand-specific effects in multi-attribute marketing models. *Management Science*, 25, 11-21.
- Srinivasan, V., Park, C. S., & Chang, D. R. (2005, September). An approach to the measurement, analysis, and prediction of brand equity and its sources. *Management Science*, pp. 1433-1448.
- Srivastava, R. K., Shervani, T. A., & Fahey, L. (1998). Market-based assets and shareholder value: A framework for analysis. *Journal of Marketing*, 62(1), 2-18.
- Stull, T. K. (1984). Methodological techniques for the study of person memory and social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 2, pp. 1-72). Hillsdale, NJ: Lawrence Erlbaum.
- Thompson, C. J., Pollio, H. R., & Locander, W. B. (1994). The spoken and the unspoken: A hermeneutic approach to understanding the cultural viewpoints that underlie consumers' expressed meanings. *Journal of Consumer Research*, 21, 432-452.
- Vriens, M., & Frazier, C. (2003, Summer). The hard impact of the soft touch: How to use brand positioning attributes in conjoint. *Marketing Research*, pp. 23-27.
- Vriens, M., & Ter Hofstede, F. (2000, Fall). Linking attributes, benefits, and consumer values. *Marketing Research*, pp. 3-8.
- Wansink, B. (2003). Using laddering to understand and leverage a brand's equity. *Qualitative Market Research*, 6(2), 111-118.
- Zalman, G. (2003). *How customers think: Essential insights into the mind of the market*. Boston: Harvard Business School Press.
- Zalman, G., & Coulter, R. H. (1995, July-August). Seeing the voice of the customer: Metaphor-based advertising research. *Journal of Advertising Research*, pp. 35-51.

27

CUSTOMER SATISFACTION RESEARCH

RICHARD L. OLIVER
Vanderbilt University

This chapter entertains the general field of consumer (customer) satisfaction research in a manner expanding upon more general treatises. Whereas the field has been studied from many perspectives, the most common appearing in the areas of customer satisfaction surveys and customer satisfaction strategy; an alternative approach will be proposed. Most typically, works on customer satisfaction survey measurement take the form of "how to form scales and summarize results," and those on customer satisfaction strategy take the form of "how to generate satisfied customers and alleviate dissatisfaction." While these are worthy endeavors, it remains that little in-depth analysis has been performed on conceptually based managerially actionable strategies within the satisfaction response.

Unfortunately, these early perspectives have not significantly advanced the progress of satisfaction work for some time, and it was not until many and diverse perspectives were integrated, elaborated, and extended that others began the study of conceptual satisfaction in earnest. Here, the underlying mechanisms of how consumers construct, consciously or subconsciously, their satisfaction conclusions are explored so that a grander strategy of fostering satisfaction and

diminishing dissatisfaction can emerge. This will allow new actionable strategies resulting in more diverse practical implications. Readers interested in greater detail and elaboration, including discussion of topics not covered here, should consult the author's original work (Oliver, 1997).

CUSTOMER SATISFACTION: WHAT IS IT?

Recent interpretations in the consumer domain now couch satisfaction as a fulfillment response. Fulfillment implies that a consumption goal is known, as in basic motives of hunger, thirst, and safety. However, observers of human behavior understand that these and other goals can be and frequently are modified and updated in various ways. Thus, consumer researchers have moved away from the literal meaning of satisfaction and now pursue this concept as the consumer experiences and describes it.

In Oliver (1997), the following definition has been proposed as being consistent with the conceptual and empirical evidence to date:

Satisfaction is the consumer's fulfillment response. It is a judgment that a product or service feature, or the product or service itself, provided

(or is providing) a pleasurable level of consumption-related fulfillment, including levels of under- or over-fulfillment. (p. 13)

Here, pleasurable implies that *fulfillment* gives pleasure or reduces pain, as when a problem in life is solved. Thus, individuals can be satisfied just to get back to normalcy, as in the removal of an aversive state (e.g., pain relief). Moreover, fulfillment is not necessarily limited to the case of met needs. Over-fulfillment can be satisfying if it provides additional unexpected pleasure, and under-fulfillment can be satisfying if it gives greater pleasure than one anticipates in a given situation. Note that it has not been necessary to provide a separate discussion of dissatisfaction. If the word *displeasure* is substituted for *pleasure* in the satisfaction definition, dissatisfaction results. Thus, the displeasure of under-fulfillment typically is dissatisfying, and interestingly, over-fulfillment may be dissatisfying if it is unpleasant—the case of “too much of a good thing.”

Related Concepts

A number of related, but important, concepts are frequently used interchangeably with satisfaction, although closer inspection reveals that they are actually distinct from satisfaction despite the fact that they may be related to satisfaction in various ways. (Those discussed here are found in this chapter; a larger list is presented in Oliver, 1997.) Generally, they represent either the affective (liking/pleasure) or cognitive (thinking/judging) components of product and service experience, although some are hybrids of the two.

Within the antecedent categories, *moods* may play a role in satisfaction formation. These are positive or negative feelings of a largely non-thinking nature, although certain events may have preceded their appearance. *Quality*, in contrast, is a cognitive judgment that summarizes the exceptionally good (or bad) elements of the product, especially when compared to other direct alternatives or offerings (brands). In a similar vein, *value* is a judgment that compares the likely outcome of purchasing to the inputs forgone. Thus, value results when consumers compare what is to be received (e.g., performance) to the acquisition costs (e.g., financial, psychological, effort).

Another antecedent concept that is actually a hybrid affective-cognitive judgment is *attitude*. It is a relatively stable judgment that a product or service has desirable or undesirable properties. The judgment takes the form of a liking or disliking and is based on many separate evaluations of product features. Whereas moods can emerge as pure feeling states, attitudes result from deliberate processing of (product- or service-related) information. Unlike satisfaction, consumers do not have to experience consumption to have formed an attitude; hence, attitudes can exist prior to purchase/usage. Satisfaction, in contrast, is a postusage phenomenon, is purely experiential, and results from comparative processes, discussed next; attitudes do not require comparisons.

The concepts discussed up to this point are shown as predisposing conditions in Figure 27.1. A discussion of other concepts in the figure follows.

To be discussed, *disconfirmation*, also known as expectancy disconfirmation, is the result of a comparison between what was expected and what was observed. In current satisfaction parlance, it more commonly refers to an expectation-performance discrepancy. Consumers would describe this concept in terms of performance being better than or worse than expected with regard to a product or service. The first component of disconfirmation, *expectation*, is a predisposing prediction—sometimes stated as a probability or likelihood—of attribute or product performance. *Performance* itself is the perceived amount of product or service attribute of outcomes received, usually reported on an objective scale bounded by good and bad levels of performance (e.g., courteous/discourteous service). This is frequently confused with *quality*, a judgment of performance excellence. Although frequently substituted for *satisfaction*, recent data suggest the concepts are separate and distinct.

In the domain of the consequences of satisfaction, a major concept is that of *loyalty*. Brand loyalty is a deeply rooted commitment to repurchase a product or repatronize a service in the future. It is not to be confused with repeat purchasing, which may involve constrained or happenstance (e.g., random) repetitive behavior. Although related to satisfaction, loyalty will be mentioned

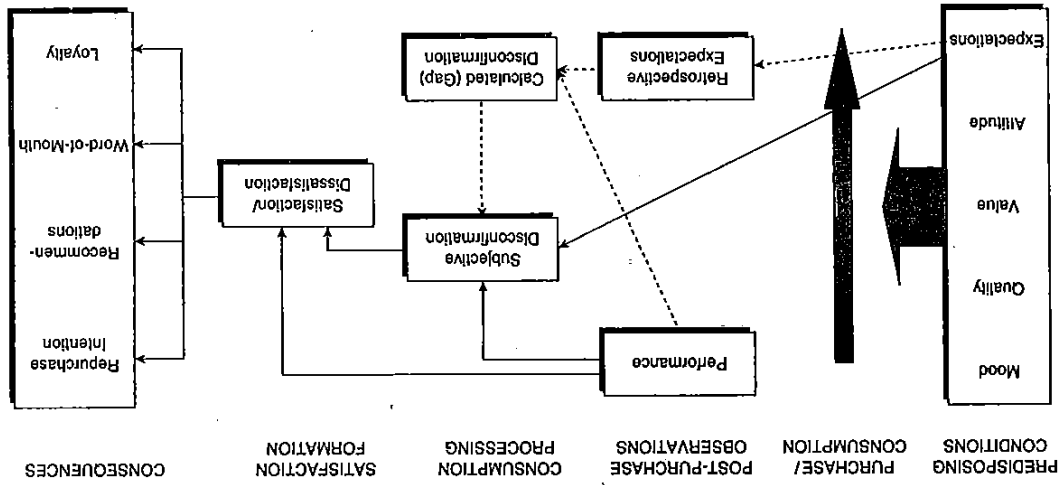


Figure 27.1 Antecedents and Consequences of Satisfaction Illustrating the Expectancy Disconfirmation Sequence

only briefly in the close of this chapter as research as to its meaning is ongoing. Other consequences are more fully discussed. These include repurchase intentions, recommendations, and word antecedents and consequences, as discussed here, appears in Figure 27.1.

Measurement Issues: Satisfaction Scales

Having defined and delineated satisfaction and related concepts, discussion proceeds to measurement more generally. This section will be necessarily brief because many reference sources are available (e.g., Chakrapani, 1998; Myers, 1999) and because measurement in general has been studied for some time. This section will also serve the purpose of a recapitulation of past research and as a prelude to the final section where measures are suggested based on the conceptual framework used here.

At this point, it will be helpful if the reader appreciates that, like attitude, satisfaction is a hybrid concept composed of both an affective component (pleasure) and a cognitive component (e.g., extent of need fulfillment) as in the previous definition—but after the purchase and use situation (i.e., after disconfirmation has been experienced). Thus, it is assumed that when a consumer affirms that "I am satisfied (dissatisfied)," both of these components are inherent in the reply. This understanding is useful if one wishes to expand a satisfaction scale with the use of affect-related or cognition-related "cousins."

Discussion begins with simple verbal (qualitative) statements of the form elicited when respondents are interviewed about their satisfaction or dissatisfaction with a product or service episode. Using a phone interview as an example, the respondent would be read the five categories of (1) very satisfied, (2) satisfied, (3) neither satisfied nor dissatisfied, (4) dissatisfied, and (5) very dissatisfied and asked to select one. This question would usually be asked once per attribute, product, and so on up to the point of respondent fatigue, a critical factor in phone surveys. For this reason, many prefer printed questionnaires, which can be completed at the respondent's leisure.

Whether measuring satisfaction with attributes or products, single-item scales of this

nature, while appearing to give discrete, unambiguous answers, have no inherent reliability unless they are readministered—an impracticality. Validity is also suspect if one were to test a single, five-point satisfaction response against many other factors thought to influence satisfaction. In the case of a negatively skewed distribution of satisfaction scores—the typical case with the bulk of data in the satisfied region—the variance in this variable would be very restrictive and validation would be compromised.

At a level more reliable than that of one-item scales, multi-item scales can be formed that include a satisfied/dissatisfied "anchor" along with items tapping related concepts. For example, the researcher could include feelings (pleasant/unpleasant), stronger emotions (anger/delight), and attitude-based (like/dislike) and cognitive (high/low performance) descriptors. (See the preceding discussion and Figure 27.1.) This method is frequently performed, but the scale produced (frequently through addition or weighted combination) is now one of a mixed nature. If carefully constructed, however, with the use of close relatives of satisfaction, the reliability will be very good. Validity is also considerably improved because each item in the multi-item scale adds to the available variance for testing.

At a still higher level, the companion descriptively based antecedents and consequences of satisfaction. This will require an understanding of the causal basis (i.e., expectancy disconfirmation in the present case) for making satisfaction judgments, the content of the remaining sections of this chapter. An example will be provided later after all concepts have been discussed.

Measurement Issues: Scaling

The Number of Scale Points

For predictive purposes, a minimum of three scale points is recommended, particularly for phone surveys. As the number of points becomes large—exceeding 10, for example—problems of interpretation are introduced. One reason is that consumers tend to use subintervals of very long scales, such as restricting responses to an interval of 5 points on a 10-point scale. Thus, one

respondent's "7" may be another's "9" for the same perceived performance level. In effect, these consumers are interpreting the scale's meaning in a manner not known to the researcher. An exception is the "chances in 10" scale. When percentages have clear meaning in the context used, this 11-point scale (0% to 100%) is admitted. In particular, "50-50"—representing a true midpoint—is easily understood.

Odd or Even Number of Items: The Issue of a Midpoint

The previous discussion addresses the debate over whether there should be an even or odd number of items in a scale. Consider the following two alternatives:

- (1) Bad 1 2 3 4 5 6 Good
- (2) Bad 1 2 3 4 5 6 7 Good

In (1), there is no midpoint, and respondents who believe that the feature is neither bad nor good are forced to commit themselves to answering with a "3" or with a "4," the first suggesting modestly negative feelings, the second modestly positive. In (2), this respondent can comfortably respond with a "4." Opponents of the second scale argue that there is no information in "fence sitting" and that few, if any, consumers are really neutral toward features and products. This author disagrees in that neutrality is information, and the percentage of respondents who feel this way may provide the researcher with additional insight. For the above and other reasons, use 5 to 7 points. Three is permissible if attention spans, brevity, or respondent fatigue operate, while 11-point scales work nicely for chances-in-10 responses.

Because consumers tend to rate products positively, particularly with regard to satisfaction, most ratings are skewed, with the bulk of responses in the positive half of the scale. This tendency is more pronounced for scales with a smaller number of points (e.g., three and four—the latter not recommended). In fact, this issue also answers the question of "what a good satisfaction score is." The shape of the score distribution is a very revealing quality. Poor scores are represented by a "normal" distribution

peaking at the scale midpoint. Good scores peak near the upper (positive) extreme and then decline before reaching the extreme. Excellent scores, in contrast, rise continuously up to the upper extreme of the scale without falling.

A SATISFACTION MODEL USEFUL FOR CURRENT RESEARCH

At this point, it would be helpful to envision the framework upon which this chapter is based. If one can construct the antecedents of the satisfaction response—that is, how a consumer consciously or unconsciously constructs satisfaction—many conceptual and measurement issues will fall into place. It is this author's position that the most useful model of this process is the expectancy disconfirmation framework, as shown in the middle (unboxed) section of Figure 27.1. Omitted from the following discussion are the related antecedents and shorter term consequences of satisfaction posed earlier.

As portrayed in the middle area of Figure 27.1, expectations, whether measured before or after consumption (predicted or recalled retrospective expectations), and performance are compared to form an "objective" (or gap) disconfirmation level; objective disconfirmation provides the basis for a subjective interpretation of this expectation-performance difference, and subjective disconfirmation is direct cause of satisfaction. In addition, the performance-satisfaction link represents the direct effect of performance on satisfaction not channeled through disconfirmation. Dashed arrows represent links to and from recalled expectations and gaps. These variables may or may not be collected by the researcher (to be discussed).

This middle graphic represents the "expectancy disconfirmation with performance model." For discussion purposes, we begin with the most common treatment of satisfaction, that of performance influences. This is followed by discussion of the expectation and disconfirmation variables. As noted, the chapter concludes with brief discussions of the short-term consequences of satisfaction, namely, repurchase intention and communications to the firm and other consumers.

THE PERFORMANCE OF FEATURES (ATTRIBUTES)

In an all-too-familiar research scenario, much of satisfaction research is conducted according to the following format: A list of key product or service features is generated that, it is hoped, contains an exhaustive set of factors thought to cause satisfaction and dissatisfaction. In a fairly direct manner, consumers are asked to retrospectively rate the product or service on the degree to which each of the features was delivered. At the same time, the consumer may be asked to rate the product on an overall basis or on satisfaction. An assumption of this technique is that those features that are more aligned with the overall score across consumers (high feature scores corresponding to satisfaction; low feature scores to dissatisfaction) are those that have the greatest impact.

Despite the ubiquity of this method, problems are inherent in its implementation. One problem is that the list of features cannot be exhaustive for all consumers. A second problem arises from the disparate goals of consumers (need fulfillment) versus those of designers, engineers, and manufacturing personnel. In part, this problem can be addressed by using the proper level of feature abstraction when preparing the attribute list (to be discussed). Another problem, discussed next, is that of the relevance of features at different stages of decision making.

Satisfaction Drivers Versus Choice Criteria

In pursuing the reasons behind the consumer's satisfaction response, it should be borne in mind that the researcher's goal is to determine the correct feature list of *satisfaction drivers*, as opposed to product or service *choice criteria*. A common mistake is that of assuming the features consumers use in selecting a product from a list of alternatives are identical to the set of features that play into satisfaction and dissatisfaction judgments. Although many of the choice features will also be those used in forming satisfaction judgments, the *assumption* that the choice feature set and the satisfaction set are identical or even similar has been shown to be incorrect.

Consider summer vacation travel. Some features, such as the availability of travel and lodging, are used by the consumer only in the choice phase of decision making. The combined cost of the travel package could be an example. A consumer may shop for the lowest total monetary outlay. Once found, this cost ceases to be a factor in satisfaction with the trip, which may be months away.

As an example of a "pure" satisfaction driver, certain destination amenities such as air-conditioning would provide an example of this situation. Brownouts and equipment failures could not have been predicted previously, and any hotel stay in the destination locale is randomly subject to this effect.

Other features are examples of mixed or dual influence, affecting both choice and satisfaction to the same degree. Prevailing weather patterns in the chosen destination would be an example of this phenomenon. Although the vacation planner has no control over the actual weather encountered, destinations are chosen for their *expected* weather; mother nature is fickle in this regard.

Thus, for all these reasons, researchers are advised to determine satisfiers and dissatisfiers independently of choice determinants. In addition, the prior discussion illustrates why a satisfaction measure is preferred to one of attitude or quality if consumer satisfaction is, in fact, the goal of the firm. This is so because both attitude and quality judgments are used in choice and thus may give a distorted picture of the features most strongly related to satisfaction.

Levels of Feature Abstraction

An additional concern early in the satisfaction measurement process is determination of how *specific* the feature list should be. This issue addresses whether the list is to contain detailed "micro"-dimensions of a product or micro-behaviors of a service provider or if it is to contain more general dimensions of higher abstraction. As noted, consumers may think at higher levels of consumption goals, whereas management must think in terms of the details of product or service design. It is also the case that many intricacies of product engineering are not understood by the lay consumer.

An obvious drawback to the micro-dimension approach is that the list of feature experiences presented to consumers becomes overly long despite the fact that it can contain maximum diagnostic value—value deriving from the fact that specific details of a (dis)satisfying situation will be singled out for study. In an alternative approach operating at a higher level of abstraction, the researcher can simply have consumers rate the consumable on, for example, quality, value, and need satisfaction (met my needs). Note, however, that this simple list of three overarching criteria of the consumption experience does not provide design details required by management. This leaves management with a bare minimum of real actionable conclusions.

The answer to the level of abstraction problem can be found in two areas. The first relates to progress made in determining key features; that is, whether management is in the early exploratory phases of discovery or whether analysis is closer to identifying critical problem dimensions. Greater specificity should be more appropriate at the exploratory study phases. The second area reflects the nature of data collection. Greater specificity requires longer lists of features. If issues of questionnaire length, respondent apathy or fatigue, and inadequate incentives are in play, then shorter, more abstract feature lists are required. The lack of diagnosticity will have to be compensated for in other ways, such as the inclusion of open-ended sections of the survey.

Scaling Feature Performance

In this section, it is assumed that an "optimal for the purpose" list of critical features has been generated. This list can be constructed in many ways, including "intuition" (not recommended) and from inputs provided internally within the firm, from channel middlemen, and from consumers. The latter source can be tapped in many ways, including focus groups, correspondence with the firm (including complaining data), and from various forms of survey research. Readers are directed to other works within this volume of readings.

Because performance is measured ex post in satisfaction surveys, the questions are typically worded in the past tense. Other, more pressing,

measurement issues also require attention. These are the polarity of the performance items and whether or not one wishes to also include the valence (positivity or negativity) of the performance rating. Issues concerning the number of scale points and, secondarily, a preference for whether the number should be odd or even have been discussed previously in the context of measurement issues.

Polarity

All performance dimensions have some ability to underperform or to negatively perform. Take the case of a stain remover, and presume that "cleansing stains" is the basic performance dimension. Consider the following scales:

- (1) Doesn't cleanse _____ Cleanses
 (2) Partially cleanses _____ Cleanses

In practice, researchers would put numbers under the horizontal lines and between the endpoints. These endpoints are referred to as poles, and the issue is whether the negative pole should be truly negative or something less. Scale (1) assumes that the worst the stain remover can do is not lift stains. Scale (2) presumes only that the cleanser partially, but not fully, lifted stains.

Which scale assumption should be followed? The best answer is that the scale should encompass the experience of the users responding to the survey. If consumers have access to the better oxy-based cleansers on the market, Scale (2) may be best. However, if one is sampling consumers using home remedies, possibly Scale (1) may be necessary.

Handling the Case of an Ideal Point

Often, the desired performance level exists at some level less than the maximum limits of performance. The carbonation level of a soft drink is an oft-cited example. Consider, again, alternative scales:

- (1) Not carbonated _____ xxxxxx
 Highly carbonated
 (2) Not carbonated _____ Ideally carbonated

In Scale (1), the "positive" pole has been labeled as some unspecified maximum level of performance. The "X" marks an individual consumer's preferred level of carbonation, at least subjectively, while the lowercase *x*s represent the range of other consumers' preferences. Scale (1) has little value if the range of the *x*s is not known to the researcher. In fact, a rightmost score on this scale may be interpreted as maximally fulfilling when, in fact, it represents a level of carbonation that is excessive to most consumers. Scale (2) overcomes this problem if the researcher wishes high scores to represent maximum performance as viewed by the consumer but unfortunately does not inform as to what "ideal" is. Experimentation is necessary to unravel this dilemma. Another solution is provided next.

Valence

Note that no mention has been made of how good (or bad) the cleansing of stains is to the consumer. In this and many cases, the feature itself connotes a goodness or badness, with removing stains connoting a desired feature and, in a worst-case example, "setting stains" connoting an undesired feature. For this reason, most surveys do not separately measure the valence of performance. However, there are many other situations where a valence must also be measured. The preceding carbonation example illustrates one such case. If a consumer finds carbonation undesirable, then the only favorable rating for this consumer is the "not carbonated" pole on the scale. The researcher, however, may interpret this as a negative response. Thus, the consumer must also be asked if the degree of carbonation he or she believes the beverage contains is good or bad or likable or not as follows:

This level of carbonation is: _____ Desirable
Bad _____ Good
or: Undesirable _____

This separate measurement of valence can be avoided if the researcher is able to include valence in the feature description. Consider the following:

Low gas mileage _____ High gas mileage
Poor gas mileage _____ Good gas mileage

While these two alternatives appear almost identical and in practice may yield equivalent results, there is a subtle distinction between them. In the first, one is not able to determine if high gas mileage is believed to be desirable (or even important!) to the respondent. In the second, one has no idea what range of mileage is considered "good."

Temporal Survey Issues

A last issue in performance measurement vis-à-vis satisfaction is the issue of whether the satisfaction measure(s) should precede or follow attribute performance measurement. This issue becomes very critical because the logic of the survey comes into play. If satisfaction measurement comes first, a general halo effect may occur whereby attribute performance is seen to conform to the overall judgment. If performance measures precede satisfaction, then satisfaction may be colored by the first few (primacy) attributes or the last few (recency) attributes.

One solution is to randomly split the surveys into those with satisfaction first, then second. Logic issues can then be measured and resolved. Note that this does not solve the primacy/recency effects in the attribute list. Randomization of the list will partially address this. However, this author's experience is that randomization promotes greater correlations (multicollinearity) in the attribute list. There are problems here that defy solution.

EXPECTATIONS AND THEIR MEASUREMENT

Expectations Defined

Generally, an expectation is an anticipation of future consequences based on prior experience and other many and varied sources of information. Expectations can also be described as a *comparative referent* for performance. The reason is that performance alone is an unreferenced concept. Meaning is attached only when

performance can be compared to some standard. In an elementary sense, the adjectives *good* and *bad* suggest the operation of a "goodness" standard, although they provide little diagnostic value to management as the terms have highly variable meaning across consumers. Diagnosticity is increased as the standard becomes less ambiguous and more objective to the point that it can be measured.

In fact, any number of referents can be used in later satisfaction assessments, but all become channeled into expectations when the product or service is purchased. The concept of needs, discussed previously, is one of the many referents available to consumers. These same consumers, however, will pursue only those products that they expect to fulfill their needs. Thus, the expectation and not the need is what the consumer brings into the purchase. Oftentimes, the expectation and need will overlap exactly, becoming interchangeable.

Multiple Expectation Referents (Standards)

Frequently, consumers express different variations of what they would prefer a product deliver. At one level, they may have an ideal perception of a product offering, something they wish they could receive in a perfect world. Others, or the same consumer at a different time, expect only what they believe the firm's product can or will deliver. Researchers have referred to these two different perceptions as *ideal* and *predicted* expectations or, alternatively, as *should* and *will* or *desired* and *likely* outcomes. By measuring both of these expectation levels, the firm will be in a better position to understand how their offering compares to what consumers really want and how their offering stacks up against those of their competitors. If both ideal and predictive expectations are actively processed, the researcher may have to allow for the measurement of multiple levels.

A number of research investigations have established that consumers do indeed recognize and use multiple levels of expectations or standards. Among the most common of these are studies investigating the influence of normative

(should) and predicted (will) expectations or, alternatively, ideal and expected referents. The results of all studies tend to be similar. When the ideal or should level of expectations was the referent, satisfaction was lower than when actual expected or predicted expectations were used. This implies (but may not — to be discussed) that high expectations can frustrate satisfaction attainment. Moreover, these studies generally conclude that consumers do entertain multiple standards and that inclusion of more than just the predicted level may improve a model's ability to understand satisfaction. Researchers are encouraged to measure both if the study design permits as the expectation measure must be taken twice.

Expectations: Examples of Measures

The measurement of expectations essentially asks consumers to place likelihood estimates on specific occurrences of product or service performances in the future. Consumers may also be asked to place likelihood estimates on events that have already occurred (to be discussed), a practice necessitated in many customer environments, such as emergency rooms, where the identity of clientele cannot be known a priori. Discussion centers, first, on the measurement of actual predictive expectations, those that are estimated before the occurrence of an event. In this sense, these expectations are *proactive*, or proactively obtained by the researcher.

Predictive (Proactive) Expectations

In most consumption situations, consumers will hold *valenced* expectations—for the basis of purchase and acquisition is to obtain products with pleasant consequences and to avoid those with potentially unpleasant outcomes. Thus, "high" expectations reflect the possibility that desirable events will occur and/or that undesirable events will not occur, while "low" expectations suggest that desirable events will not occur and/or that undesirable events will occur.

Note that valence is clearly indicated in these statements, phrased as desirabilities or undesirabilities. In practice, the valence component of the expectation generally will be implied in the

wording of the expectation item on a survey. For example, one might refer to "good" gas mileage or use words with a positive or negative connotation such as "praise from friends," "courteous servers," or "delays on takeoff." In working with expectations in this manner, it is assumed that the performance referent is also phrased in properly valenced fashion.

In measuring expectations before purchase or patronization, one must be careful to ensure that the consumer sample has some basis for forming an expectation set. Consumers probably do go into many purchases without complete expectation knowledge, but it would be rare for consumers not to have any awareness of the product at all. At the minimum, the brand and price will be known to the consumer, and some individuals will use these to infer the remaining performance characteristics. Because of the possibility that some elements of the expectation set will not be known to all consumers, however, researchers should include a "don't know" category in the expectation section of the survey.

The types of expectation measures available to the researcher are straightforward. A number of variations are shown here, including the familiar Likert version in agree-disagree format

If I complain, I will get a response:

0 1 2 3 4 5 6 7 8 9 10 (times in 10 tries)

Brand X has (possesses) feature Y:

Unlikely _____ Likely

Brand X will satisfy my needs:

Disagree _____ Neither _____ Agree

There is no evidence that any one measure is better than another. Rather, the context of the investigation should dictate the question format. For example, a "chances-in-10" scale is better used when the consumer has had a chance to sample the product over multiple purchases. A household staple and a familiar restaurant both qualify as examples. In contrast, a likelihood format could be used for a first-time purchase, as could the agree-disagree scale.

Note that no example of "raw" product performance has been suggested, as in "The car will average 35 miles per gallon." There are two problems with the use of such measures. First,

Medicine X: Hard to swallow:

Should be
Hard 1 2 3 4 5 Easy

Will be

Hard 1 2 3 4 5 Easy

Courier X: Speed of delivery:

Desired
Slow 1 2 3 4 5 Fast

Predicted

Slow 1 2 3 4 5 Fast

While this format provides some space-saving economies, the close juxtaposition of the two forms of expectation questions may introduce a fairly obvious compound. The researcher's goal, under this format, is somewhat more transparent than if the two sections were widely separated. This is so because consumers may adopt a halo response set whereby the "should" questions are always scored high and the "will" questions are scored, for example, two scale units below.

Expectations After the Fact: Retrospective Expectations

Very frequently in satisfaction studies, the researcher cannot identify purchasers before consumption and is therefore constrained to measure expectations at the same time that the levels of performance and satisfaction are assessed. In this event, the expectation section of the survey should appear before any other material, especially satisfaction. The wording of the question would conform to the following general format: "Referring back to the time you first purchased the product (engaged the service), what were your expectations at that point?" The subsequent list of attribute expectations would then be worded in the following tense: "At the time I purchased the car, I believed that it would give me good gas mileage."

Although these are not predictive expectations and raise some fairly important issues, three reasons account for the majority of cases when the researcher is forced to rely on expectations ex post. The first is simply neglect; the researcher was not aware of or saw no need for expectation measurement prior to purchase and usage and was content to make interpretations from the performance ratings alone. A second reason is that some firms may not wish to suggest possible product experience to consumers before usage. An example of this is the probability of breakdown and repair or of complaining outcomes.

The third reason is due to necessity; most firms cannot predict and do not have access to their customers before they purchase the product. Virtually all tangibles sold at retail fit this description. Producers first learn who their buyers are if and when warranty cards are returned. Services provide greater latitude because the consumer can be intercepted before the service episode. However, it is still true that the consumer, in most cases, has made an earlier decision to patronize the service, and this decision may affect expectation measurement in subtle ways. In both cases, the researcher must rely on the ability of the consumer to reflect back on what he or she thought the product/service would deliver. A number of problems now become evident.

The most troubling is that the consumer will have already experienced the product's performance. In this case, what the consumer has experienced and what he or she recalls will be confounded. Most probably, the recalled expectations will be biased toward the experienced performance. One explanation for this finding is that consumers with ill-defined expectations, perhaps because they are unfamiliar with the product or service category, will construct "data-driven" or bottom-up strategies for responding to expectations ex post. This may be especially true if the consumer feels that ill-formed expectations reflect poorly on his or her judgment or decision-making skills or if the survey appears to assume that consumers are "supposed to have" expectations. Nonetheless, if proactive measurement cannot or has not been taken, the researcher will still be better served than if no expectation data had been collected at all.

DISCONFIRMATION AND ITS MEASUREMENT

When consumers compare performance to their expectations, the response of *disconfirmation*—more specifically, disconfirmation of performance standards—results. Because the early

work in consumer satisfaction was conducted with predictive expectations as a standard, the phrase *disconfirmation of expectations* or "expectancy disconfirmation" has come to apply to this concept. Many standards consumers bring to the consumption experience can be disconfirmed, so an alternative phrase to describe the discrepancy from a standard could be simply *disconfirmation*. In this section, the phrases *expectancy disconfirmation* and *disconfirmation* will be used interchangeably.

In contrast to the lay interpretation of a disconfirmation, which usually connotes a negative outcome, performance can also be favorably compared to expectations. This permits disconfirmation to take on a positive as well as a negative value. In the same way that product performance can be worse than expected, it can also be better than expected. Because the phrase *disconfirmation* without the valence qualifier is ambiguous as to direction, the phrase *negative disconfirmation* will be used to refer to the negative discrepancy that occurs when performance is below standard, and *positive disconfirmation* will be used to refer to the positive discrepancy that occurs when performance is above standard. When performance is equal to standards or expectations, a zero disconfirmation or, simply, a confirmation of expectations exists. Measurement of this concept is straightforward, as in the two examples that follow:

(Regarding a product or service) My expectations were:

Too high:	Accurate:	Too low:
It was poorer than I thought	It was just as I thought	It was better than I thought
----- ----- -----		

Overall, this product/service feature/attribute was:

Much worse than expected	As expected	Much better than expected
----- ----- -----		

These scales are "generic" and have numerous applications. For example, the entire consumption experience can be judged on the degree to

which it was better or worse than expected, as can individual attributes. In air travel, for example, the entire trip can be assessed as can each minute element of the trip from the ticket agent's speed, to the waiting area, to the food in flight, to the landing, and so on. Other possibilities include dimensions (attribute groups) of performance and the benefits (good aspects) reaped from consumption as separate from the problems (bad aspects) encountered.

Objective (Calculated) Versus Subjective Disconfirmation

Early attempts to measure disconfirmation used a discrepancy or "gap" approach. That is, separate survey sections were used to capture, first, attribute expectations and, later, attribute performance perceptions. Then the expectation scores were subtracted from their respective performance scores, and these "gaps" were added. The logic is direct. When attribute performance was higher than its respective expectation, the gap is positive and is considered favorable. Similarly, when expectations were higher, the gap was negative and unfavorable. Satisfaction should increase as the positivity of the gap score increased and should decrease (contributing to dissatisfaction) with the negativity of the gap score.

A reasonable question at this point is, Why is a subjective interpretation needed if the gap calculation is available to the consumer and, thus, to the researcher? A first answer is that the "true" expectation level, the "true" performance level, and the consumer's calculation (and even its accuracy) are typically *not* available to the researcher. In fact, many product and service attributes are not subject to this type of evaluation at all. How, for example, do consumers quantify the comfort level of the car's seating or the richness of the upholstery? If the consumer cannot quantify or scale expectations and performance, then disconfirmation can only exist at the subjective level.

A second answer to this question is because *only the consumer* can attach the proper amount of valence (goodness/badness) to the difference he or she assumes. This valence, then, determines *how much* better or worse performance is perceived to be; the consumer then scales the

amount of perceived positive or negative disconfirmation on the "better-than/worse-than" scale.

Now, assume that the consumer omits one of the numeric steps in two different ways. First, what if consumers are less than diligent and fail to maintain the proper records for calculating gas mileage, a step that requires a log of distance traveled and a measure of the volume of gasoline used? Can the consumer still respond to the *subjective* disconfirmation query? Most certainly, as long as a perception of "better/same/worse than" is sensed. This "sense" has been shown to exist in many studies and can exist for purchases having no objective performance dimensions at all (e.g., art, insurance in the absence of claims, nutritional supplements). In a second example, what if the researcher was able and willing to collect only expectation and performance data and form gap scores? Could assumptions be made regarding the likely satisfaction of the respondents? Perhaps, but not necessarily. The researcher would be able to calculate objective differences between these values. Unfortunately, this difference exists only as a *managerial* calculation and may not be as accurate as the consumer would necessarily find it. Three reasons for this disparity exist.

The first is lack of precision in the gap calculations. Gap scores are difference scores, which are known to be notoriously unreliable. The reason is that the inevitable measurement error in each of the gap components is compounded when the two are combined. While it is true that even the subjective score contains measurement error, it is not compounded as it is not used in a calculation. A second reason why a difference score may not reflect a consumer-generated subjective score is that the consumer may implicitly weight either expectations or performance more highly than the other. For example, expectations may be only vaguely recalled because of the time interval between purchase and usage. Unequal weightings could also arise because of differences between consumers who place greater weight on expectations versus those who place greater weight on performance in making their judgment. Last, as noted, the raw difference score does not contain the consumer's valence toward the discrepancy. Using a gas mileage example, a 5-mpg positive difference from expectation may be only "slightly" better

than expected to one consumer but may be "much-much" better to another. The subjective difference score clearly accounts for this difference in interpretation. For all these reasons, researchers are strongly advised to measure and use subjective as opposed to calculated disconfirmation estimates if satisfaction prediction is the firm's goal.

Predicting Satisfaction

Is there evidence for the superiority of subjective disconfirmation over calculated disconfirmation in the prediction of satisfaction? Yes. A number of studies have examined both the calculated and single-score varieties of disconfirmation, most using rating scale scores. The results of all studies were similar, with the majority of the evidence suggesting that the subjective version of disconfirmation correlates more highly with satisfaction scales than do the discrepancy scores. Moreover, when analyzed in an ordering of cause and effect, the following configuration of concepts is consistently found to best fit the data:

Calculated disconfirmation → Subjective disconfirmation → Satisfaction

This sequence of events forms the basis for the expectancy disconfirmation model of consumer satisfaction discussed here and displayed in Figure 27.1. As shown, this sequence portrays a calculated expectation-performance discrepancy (if performed) as input to the consumer's subjective interpretation of this difference. The subjective interpretation then becomes the most immediate antecedent of satisfaction. If no "objective" score is available, then a subjective judgment is "sensed." Expectations and performance are implicitly incorporated in the disconfirmation judgment in this sequence.

Survey Placement Issues

The better than/worse than expected concept is a fairly recent development and requires careful thought as to its placement. If disconfirmation is to be obtained in stand-alone fashion, the disconfirmation attribute list would be placed in a separate section of a survey or personal or phone

interview apart from *and after* the expectation and performance attribute responses to maintain the logical flow of decision making portrayed in Figure 27.1. Note that this subjective measure involves no actual numeric comparison made to the expectation level. This means that, to fully test the concepts in the expectancy disconfirmation model, three survey sections are needed: expectation, performance, and subjective disconfirmation. It could be argued that three attribute lists are too onerous to the respondent. This, in part, explains why calculated disconfirmation is popular as it requires only two lists. But, for all the reasons cited here, subjective disconfirmation captures information not available from a gap score.

The solution to this "three lists" problem is to creatively design the survey. There are variations on this strategy. First, one could measure expectations before purchase; this mentally separates expectations from the other concepts because of the passage of time as the product is used. A second solution is to use a two-column combined expectation-performance section. Here, the attribute list has two answer columns headed by phrases similar to "I desired" and "I received." This gives the appearance of one list instead of two. Then, the disconfirmation section would follow, not necessarily immediately, but perhaps after a set of other questions seeking answers needed by the researcher.

Operation of Disconfirmation in the Satisfaction Model

Reviews of studies measuring disconfirmation in various forms are now available to suggest that it is a powerful predictor of satisfaction, even when combined with expectation and performance in the manner shown in Figure 27.1 (Oliver, 1997; Szymanski & Henard, 2001; Yi, 1990). In fact, disconfirmation typically dominates expectation and frequently dominates performance in terms of the strength of effect. There are times, however, when both performance and disconfirmation are input to the same regression that the disconfirmation effect is obscured (becomes nonsignificant). The reason is multicollinearity as disconfirmation is a performance-based concept, and both the

performance and disconfirmation variables may be highly correlated. When this happens, two regressions must be run separately—the first containing performance and the second containing disconfirmation in its stead. While there are more sophisticated statistical techniques for ferreting out multicollinearity, this simple technique will demonstrate the separate effects of performance and disconfirmation. Some attributes will demonstrate both effects while others will show divergent influences on satisfaction.

While the separate performance effect is most well-known and more thoroughly researched historically, especially in practice, the interplay of expectation and disconfirmation is particularly interesting. Here, two scenarios are entertained, one where expectations are thought to dominate disconfirmation in the prediction of satisfaction and a second where disconfirmation is known to dominate expectations.

When Expectations Dominate

For expectations to dominate a satisfaction decision, the processing of expectations must be more salient to the consumer than is the processing of performance or of comparing expectations to performance. For products of this nature (to be discussed), it is imperative that the researcher measure expectations. Without estimates of expectations, satisfaction cannot be predicted or it will be poorly predicted in the situations described here.

It has been said that it is not necessary that the consumer have objective measurements of performance for a disconfirmation judgment to occur. If the dominant expectations scenario is to be useful, it must also be assumed that a subjective performance judgment that would give rise to a strong or salient disconfirmation perception is not or cannot be made. Two possible reasons may explain why consumers do not attend to performance and, hence, disconfirmation. The first is that they are unable to judge performance; the second is that they do not do this as a practical matter.

When are consumers unable to judge performance? Three answers are (a) whenever no objective performance can be observed, (b) whenever performance is an ambiguous concept, or (c) whenever measurement is so

technically involved that the consumer would not be aware of or have access to the procedures. The first two cases could be represented by "health" foods and aesthetics, while the third could pertain to any of a number of "high-tech" items, including computers, water treatments, and catalytic converters.

This latter case of the consumer's disinterest or inability in testing performance is an intriguing phenomenon. Some products have performance dimensions that are measurable, but the actual measurement procedures may be too intrusive, cumbersome, or inconvenient. Take the example of a long-life light bulb. Because consumers do not keep logs of the time durations of lights, it is unlikely that the actual life is known for a particular bulb. In this case, only guesses will be obtained that, more than likely, will track the manufacturer's claims. In conclusion, researchers investigating consumables for which performance is nebulous or for which expectations play a major role must focus on expectations effects. Even when there are lesser performance and/or disconfirmation effects, the role played by expectations will reveal major strategic directions for raising or lowering expectations held by consumers.

When Disconfirmation Dominates

When can one expect to find that disconfirmation dominates the expectation effect? One answer from the literature pertains to the role of involvement. More involved consumers have been found to report stronger levels of disconfirmation and satisfaction. Note that involvement enhances the effect of both positive and negative disconfirmation so that positive (negative) disconfirmation under high involvement has a more favorable (unfavorable) effect on satisfaction than it would if involvement were not operating. Generally, any concept that affects the salience of performance to the consumer should increase the degree to which disconfirmation, more so than expectation, affects satisfaction.

A second explanation relates to the degree to which performance clearly and unambiguously refutes expectations. Individuals who recognize discrepancies from expectations and are willing to accept them regardless of the potential damage to one's ego should demonstrate

disconfirmation influences. One example of such an effect occurs with investments where the final (not interim) results are known. Here, performance is unambiguous, regardless of the expectations of the investor. A gain, of course, can be very gratifying. A loss, however, will require soul-searching or blame toward the broker's or management's recommendation.

A third possibility may result from the temporal interval between expectation formation and eventual performance observation. Lengthy time frames before attempts to recall expectations may cause expectations to decay. Thus, a declining memory for expectations may cause satisfaction to be dominated by disconfirmation only. Note from the previous discussion that it is not necessary to know precise expectation levels to form a "better/worse than expected" judgment.

It is not necessary to impress the reader with the necessity of measuring performance and its companion variable, disconfirmation. Performance measures are typically included in surveys, and disconfirmation measurement is becoming much more common, as evidenced by surveys received by this author. They are separate concepts, however, and require separate measures for reasons discussed next.

Why Measure Disconfirmation?

In simple terms, because it is becoming apparent that disconfirmation is a critical variable in the satisfaction equation. References to meeting or exceeding expectations are now commonly found in various promotions used by firms. Early forms of this type of communication appeared as promises to "meet" one's expectations. It is now clear that this is not sufficient as exceeding expectations will always improve satisfaction over that of meeting expectations. It is dangerous, however, to overpromise by stating that a firm will exceed all potential expectations because this may not be feasible, particularly when human performance is a large part of the performance delivery as it is in services. Such firms now use a mix of promises by stating that they will meet or exceed expectations. More and more, comparisons to expectations are observed to drive satisfaction. Ironically, this phenomenon could not be known until expectations and disconfirmation were measured.

SHORT-TERM CONSEQUENCES OF SATISFACTION

Intentions and Their Measurement

One of the most common results of satisfaction/dissatisfaction is a stated intention to repurchase (or not) in the future. Sometimes, this is posed in surveys in a hypothetical sense as in, "If you were in the market for a (generic product), how likely would you buy a (specific brand)?" using a likely/unlikely bipolar scale. In other cases, the measure would be more specific, as in a scale bounded by certain to buy/not buy or even chances in 10. For an example of the latter, a consumer might be asked, "If a complaint were made to a firm, how many times in 10 would a satisfactory response be obtained?"

These scales are ubiquitous in the academic literature; in particular, it is unusual not to see them used because the researcher rarely can observe repetitive behavior in a cross-sectional one-shot survey. In commercial research, however, intention scales may not need to be used because actual repeat behavior is more easily obtained. This could occur in panel data, scanner data, e-shopping data, and in TV viewing when the set is connected to a media observation device such as ITV.

Still another measure of intention is the degree to which the consumer splits purchasing between alternatives. Similar to multibrand loyalty, consumers may intend to repatronize one of a set of acceptable alternatives, such as in restaurant dining. In this case, an intention to repatronize a particular establishment is more akin to a probability across choices as opposed to a probability within a choice. The difference is that, in the latter case, one alternative may be purchased or not, while in the former, one alternative will be purchased with some degree of certainty, but only one. To this author's knowledge, this issue has not been thrashed out. In any event, a chances-in-10 scale can be used in both instances.

It is known, however, that stated intentions without behavioral validation are very unreliable. Consumers frequently overstate their intentions due to a positivity bias in consumer responding and because of unforeseen obstacles in predicting the future, such as product unavailability and/or

insufficient funds. In lieu of other measures of satisfaction validation, however, intention data may be among the best measurement modalities one can achieve.

Complaining/Praising and Their Measurement

It is this author's experience that the most neglected, infrequently found satisfaction-related concepts in satisfaction surveys are complaining and its polar opposite, praising. This is surprising as the complaining literature is vast and, in fact, was the first of the satisfaction concepts to be extensively studied because of its relation to the consumerism movement. Complaining is important because, unlike dissatisfaction, complaining is a behavior. While dissatisfaction and complaining are related, they have been found to be imperfectly correlated. Not all dissatisfied consumers complain, so that those who do are very disaffected. This is critical from the standpoint of the firm because problems cannot be addressed unless they have been brought to the firm's attention.

One might argue that low performance ratings in the attribute list will pinpoint the sources of dissatisfaction. This assumes that the majority of consumers will agree on the troublesome attribute(s). However, be reminded that there may be subsets of dissatisfied consumers, however small, that will not greatly affect an otherwise positive attribute score. Knowing who these individuals are, in terms of their demographics and other determining variables, may assist the firm in ferreting out problems faced by specific categories of customers.

In the same vein, the extent of complimenting or praising—the related bipolar concept of complaining—may also be of value to the firm. Praising is not as frequent as complaining but does occur. One might view it in the context of an extreme expression of satisfaction, having information value to the firm beyond high performance and satisfaction ratings. This could be particularly important in the service industry where many and varied service providers are involved. Typically, performance differs across service personnel, and it would be important to recognize those who have exceeded usual standards.

Measurement of both these concepts is not all that difficult and takes up only a small portion of a survey. For example, the following complaining/complimenting scales were used in a study by this author in the context of automobile purchasing and may prove useful to researchers:

	NO	YES	
I complained to the dealership about the car	N	Y	1 2 3 4 5+
I praised the car to the dealership	N	Y	1 2 3 4 5+

Scoring is straightforward. Those who respond with a no are given a zero, while those who respond with a yes are given the numeric number of reported complaints. The scale ends at 5+, although this number was arbitrary. Alternatively, one could simply use a binary score whereby any number of individual complaints is interpreted as "complaining."

(WOM)

About how many people have you talked to about the good things or bad things about your car?

Did you tell these people mostly positive or mostly negative things about the car?

	Mostly Negative	Half and Half	Mostly Positive
1	1	2	3
2	2	3	4
3	3	4	5
4	4	5	6
5	5	6	7

(Recommendations)

Did you recommend that these people:

	1 <th>2 <th>3 <th>4 <th>5 <th>6 <th>7</th> </th></th></th></th></th>	2 <th>3 <th>4 <th>5 <th>6 <th>7</th> </th></th></th></th>	3 <th>4 <th>5 <th>6 <th>7</th> </th></th></th>	4 <th>5 <th>6 <th>7</th> </th></th>	5 <th>6 <th>7</th> </th>	6 <th>7</th>	7
Buy the car	1	2	3	4	5	6	7
Neither							

Word-of-Mouth, Recommendations, and Their Measurement

Word-of-mouth (WOM) is the third of the postsatisfaction concepts discussed here. In fact, it is such an important concept that all of author (Reichheld, 2003) has suggested that one of satisfaction can be tapped by simply asking for an answer to this one concept. Although this would certainly simplify survey administration, it is not good practice because no supplemental diagnostic information is available. Attribute performance data are not measured, nor are any of the related concepts in the expectancy disconfirmation model. However, if one wishes only to gauge the sentiments of one's customer base, the one-measure strategy may suffice perhaps on an interim basis between less frequent large-scale surveys.

The nature of recommendations is very closely intertwined with WOM. While WOM can consist of praising or damning (to other consumers as opposed to the firm or its representatives), recommendations are targeted communications to potential purchasers. Note that *recommendations* is a general term and can be either positive (to buy) or negative (to not buy). Because of their close correspondence, they can be measured in the same survey section, also shown from the automobile study, as follows:

As one would surmise, satisfaction was correlated with all of these measures in the expected direction. In the automobile study, the extent of complaining correlated negatively, complimenting correlated positively, the positivity of WOM correlated positively, and the favorableness of recommendations correlated positively. All of these relations were significant, and some were substantial. Researchers would be well served if items similar to these were included in satisfaction surveys. It has been noted that very few companies examine the incidence of WOM resulting from different levels of satisfaction.

A CONSUMPTION SATISFACTION SCALE BASED ON THE CHAPTER CONCEPTS

It is now time to illustrate a multi-item satisfaction scale based on the conceptual work presented here. This is not similar to the brief discussion of satisfaction scales presented in the beginning of this chapter because these were not integrated within a conceptual framework. As noted earlier, it is recommended that one begin with a structured framework, such as that of the expectancy disconfirmation model, containing concepts known to affect satisfaction—namely, expectations, performance, disconfirmation, satisfaction, intention, complaining/complimenting, and word-of-mouth recommendations. These potential satisfaction scale items are presented next (R stands for reverse-scored) and displayed at the product level. Adjustments would need to be made if measurement were at the attribute-specific level.

Expectations (+): I expected this product to be excellent.

Expectations (-): My expectations of this product were very low (R).

Performance (+): This product has performed very well.

Performance (-): This product has performed very poorly (R).

Disconfirmation (+): This product has exceeded my expectations.

Disconfirmation (0): This product has performed exactly as I had expected.

Disconfirmation (-): This product has fallen short of my expectations (R).

Satisfaction (the anchor): I am satisfied with my choice of this product.

Dissatisfaction: I am dissatisfied with my choice of this product (R).

Intention (+): I intend to repurchase this product.

Intention (-): I do not intend to repurchase this product (R).

Complimenting: I have complimented (management) about this product.

Complaining: I have complained (to management) about this product (R).

WOM (+): I have told people many good things about this product.

WOM (-): I have told people many unfavorable things about this product.

Note that each of these concepts can be converted into bipolar scales so that only half will be necessary. (Disconfirmation can also be bipolar as shown previously.)

Within each variable category (e.g., performance), items can be expanded to flesh out a richer set of descriptors for the variable of interest. As an example, satisfaction can be expanded to include satisfaction at different arousal levels—delight for high arousal, happiness for moderate arousal, and contentment for low arousal. In this manner, the framework in Figure 27.1 can be tested using various forms of modeling (regression, structural equation) with greater confidence in the reliability of the model's components.

CONCLUSION

This ends the present discussion of satisfaction research and satisfaction measurement focusing primarily on the expectancy disconfirmation model. For those interested in further study of the concept, please be referred to the author's book on the subject (Oliver, 1997). Of note is

the fact that the expectancy model dates from the early 1980s (Oliver, 1980) and is now just appearing in satisfaction surveys. While this framework is not a panacea and it is acknowledged that others are available, particularly those that incorporate value (e.g., Woodruff & Gardial, 1996), it is now well supported by extensive research in academia.

For those who have not yet used concepts from this model, be aware that not all the linkages in Figure 27.1 are supported in all contexts. It would be rare indeed if performance were not related to satisfaction, and oftentimes a high correlation (multicollinearity) between performance and disconfirmation obscures the effect of this latter variable. If this should occur, one would regress satisfaction on disconfirmation alone, and the disconfirmation effect will assuredly appear. Depending on whether expectations are measured prior to or after performance is observed, the effect of this variable may or may not emerge. In a tracking study of restaurant patronage (Oliver & Burke, 1999), the effect of "predictive" expectations was observed initially and then decayed as the experience progressed. However, retrospective expectations continued to have an effect because performance had been observed and respondents very likely used performance to guide them in their recollections of their expectations. Nonetheless, there is value in measuring expectations because they set the stage for the operation of later variables in the model.

Last, space did not permit an examination of customer loyalty. Readers are referred to many good practitioner-oriented works. A more academic discussion can be found in Oliver (1997), and an extended version appears in Oliver (1999). Be aware, however, that *psychological* loyalty has not been *researched* as extensively as satisfaction, so works that exist treat loyalty largely as repeat purchasing—a behavior—and not as psychological loyalty, which exists as

one's mental state. Repeat purchase data do reveal insights, but many times repurchase results from happenstance or constrained purchasing. In contrast, satisfaction is largely mental, which is why one is advised to measure it along the lines suggested here.

REFERENCES

- Chakrapani, C. (1998). *How to measure service quality & customer satisfaction*. Chicago: American Marketing Association.
- Myers, J. H. (1999). *Measuring customer satisfaction: Hot buttons and other measurement issues*. Chicago: American Marketing Association.
- Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17, 460-469.
- Oliver, R. L. (1997). *Satisfaction: A behavioral perspective on the consumer*. New York: Irwin/McGraw-Hill.
- Oliver, R. L. (1999). Whence consumer loyalty? *Journal of Marketing*, 63, 33-44.
- Oliver, R. L., & Burke, R. R. (1999). Expectation processes in satisfaction formation: A field study. *Journal of Service Research*, 1, 196-214.
- Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*, 81, 46-54.
- Szymanski, D. M., & Henard, D. H. (2001). Customer satisfaction: A meta-analysis of the empirical evidence. *Journal of the Academy of Marketing Science*, 29, 16-35.
- Woodruff, R. B., & Gardial, S. F. (1996). *Know your customer: New approaches to understanding customer value and satisfaction*. Cambridge, MA: Blackwell.
- Yi, Y. (1990). A critical review of consumer satisfaction. In V. A. Zeithaml (Ed.), *Review of marketing 1990* (pp. 68-123). Chicago: American Marketing Association.