Journal of Service Research

http://jsr.sagepub.com

An Empirical Investigation of Customer Satisfaction after Service Failure and Recovery

Michael A. McCollough, Leonard L. Berry and Manjit S. Yadav

Journal of Service Research 2000; 3; 121
DOI: 10.1177/109467050032002

The online version of this article can be found at:
http://jsr.sagepub.com/cgi/content/abstract/3/2/121

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Center for Excellence in Service, University of Maryland

Additional services and information for *Journal of Service Research* can be found at:

Email Alerts: http://jsr.sagepub.com/cgi/alerts

Subscriptions: http://jsr.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations http://jsr.sagepub.com/cgi/content/refs/3/2/121

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

An Empirical Investigation of Customer Satisfaction After Service Failure and Recovery

Michael A. McCollough
University of Idaho

Leonard L. Berry
Manjit S. Yadav
Texas A&M University

Relatively little research has addressed the nature and determinants of customer satisfaction following service failure and recovery. Two studies using scenario-based experiments reveal the impact of failure expectations, recovery expectations, recovery performance, and justice on customers' postrecovery satisfaction. Customer satisfaction was found to be lower after service failure and recovery (even given high-recovery performance) than in the case of error-free service. The research shows that, in general, companies fare better in the eyes of consumers by avoiding service failure than by responding to failure with superior recovery.

Service failure and recovery is a critical issue for both service managers and researchers. However, until recently, research on the nature and determinants of customer satisfaction following *service recovery* (i.e., the actions a service provider takes in response to service failure [Gronroos 1988]) has been limited. Therefore, recovery has been identified as a neglected area requiring additional research (Andreassen 1999; Fisk, Brown, and Bitner 1993; Singh and Widing 1991; Tax, Brown, and Chandrashekaran 1998). As a result of the limited attention given to recovery, little is known about how customers evaluate recovery efforts, what constitutes successful recovery, and the potential (and limits) of recovery to convert customer dissatisfaction into satisfaction.

Understanding recovery is important for managers. Service failure is one "pushing determinate" that drives customer switching behavior (Roos 1999), and successful recovery can mean the difference between customer retention and defection. In turn, customer retention is critical to profitability (Stauss and Friege 1999). Reicheld and Sasser (1990) maintain that, in certain circumstances, a service company can boost profits almost 100% by increasing customer retention just 5%. For service providers, recovery has special significance. Fisk, Brown, and Bitner (1993) argue that due to the unique nature of services (specifically, coproduction and the inseparability of production and consumption) it is impossible to ensure 100% error-free service.

This research addresses two broad research questions. First, the antecedents of postrecovery satisfaction are examined within the context of the disconfirmation and justice literature. Second, and of equal importance, this research explores the *recovery paradox* (McCollough and

The authors wish to thank the Center of Retailing Studies and the Graduate School of Business at Texas A&M University and the College of Business and Economics at the University of Idaho for providing partial support for this research.

Journal of Service Research, Volume 3, No. 2, November 2000 121-137
© 2000 Sage Publications, Inc.

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

Bharadwaj 1992) or the question of whether customers who experience a failure followed by superior recovery might rate their satisfaction as high as or even higher than they would have had no failure occurred. Marketing researchers have long wondered if truly superior recovery efforts actually can create greater satisfaction than if nothing had gone wrong (Etzel and Silverman 1981). Some researchers report results consistent with a recovery paradox effect (Bitner, Booms, and Tetreault 1990; Goodwin and Ross 1992; Hansen and Danaher 1999; Hart 1993; Kelly, Hoffman, and Davis 1993; Morris 1988; Smith and Bolton 1998; TARP [Technical Assistance Research Programs, Inc.] 1979, 1986; Triplett 1994). For instance, Hart, Heskett, and Sasser (1990) state, "A good recovery can turn angry, frustrated customers into loyal ones. It can, in fact, create more goodwill than if things had gone smoothly in the first place" (p. 148). However, other researchers have reported results inconsistent with a recovery paradox effect (Berry, Zeithaml, and Parasuraman 1990; Bolton and Drew 1991; Fornell 1992; Halstead and Page 1992; Smart and Martin 1993; Zeithaml, Berry, and Parasuraman 1996). In general, no theoretical explanation of why a recovery paradox effect is possible has been offered. In addition, very little research has directly compared postrecovery satisfaction with consumers who experienced error-free service while controlling for confounds that might affect satisfaction evaluations. Such a comparison is necessary if the conditions of the recovery paradox as laid out by McCollough and Bharadwaj (1992) for a recovery paradox are to be met. Therefore, our investigation of postrecovery satisfaction and the recovery paradox addresses a subject with contradictory findings.

CONCEPTUAL FRAMEWORK AND HYPOTHESES

Although, in general, the literature on postrecovery satisfaction is limited, some significant headway has been made recently (Smith and Bolton 1998; Tax, Brown, and Chandrashekaran 1998). This research adds to the growing literature on recovery by evaluating recovery from the perspective of disconfirmation (with emphasis on the role of failure and recovery expectations) and justice.

Disconfirmation

The most widely used model within the consumer satisfaction/dissatisfaction (CS/D) literature is the disconfirmation paradigm (Bearden and Teel 1983; Oliver 1980, 1981, 1989, 1993; Oliver and Bearden 1985; Oliver and Burke 1999; Swan and Trawick 1981). Disconfirmation also has been advanced as a model for understanding customers' reactions to recovery (Oliver 1981; Singh and Widing 1991). The disconfirmation paradigm holds that customers compare perceived product performance to expectations. Performance that exceeds expectations is *positively disconfirmed*, performance that meets expectations is *confirmed*, and performance that falls short of expectations is *negatively disconfirmed*. In general, the more negative the disconfirmation, the greater the dissatisfaction, whereas the more positive the disconfirmation, the greater the satisfaction.

In the left-hand portion of Figure 1, a disconfirmation model of recovery (which is empirically evaluated in Study 1) is presented. In this section, the disconfirmation portion of the model is explained, and concurrent with this discussion, a disconfirmation-based explanation of the recovery paradox will be advanced.

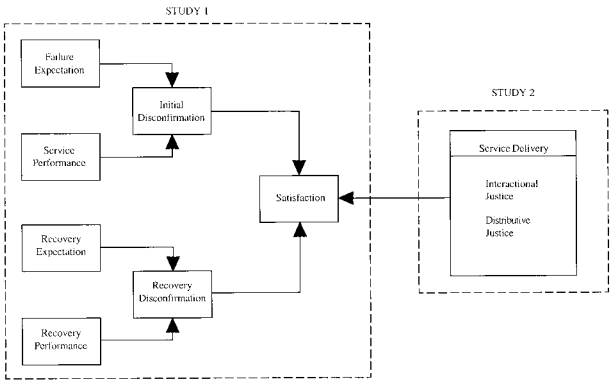
The disconfirmation portion of Figure 1 holds that *satisfaction* (modeled as a postrecovery transaction-specific judgment and not as a general attitude regarding the provider's overall service quality) is a function of *initial disconfirmation* and *recovery disconfirmation*. Initial disconfirmation is defined as the discrepancy between *failure expectations* (expectations that the service might fail) and *service performance* (initial service performance perceptions, broadly conceptualized in this context as either successful service performance or failure). Recovery disconfirmation is defined as the discrepancy between *recovery expectations* (expectations by the consumer regarding what the service provider will do given failure) and *recovery performance* (perceptions regarding steps taken by the service provider in response to failure).

Our model includes expectations given their importance in the disconfirmation, recovery, and complaining literature. In terms of understanding postrecovery satisfaction, recovery expectations have been held to be the standard against which recovery performance is judged (Kelly and Davis 1994; Oliver 1981; Singh and Widing 1991). From a motivational perspective, recovery expectations are also important. For instance, researchers investigating complaining behavior have found that many customers do not complain to the provider about a dissatisfying experience. However, those who do seek redress often are motivated by recovery expectations (referred to in the complaining literature as probability or likelihood of success) (see Blodgett, Granbois, and Walters 1993; Hirschman 1970; Richins 1983, 1987; Singh 1990).

An evaluation of the recovery paradox from a disconfirmation perspective would seem to indicate, all other things being equal, that because customers do not deliberately seek dissatisfaction, any transaction resulting in a problem (negative disconfirmation), even if successfully addressed, would lead to satisfaction lower than that resulting from error-free service. It appears reasonable that,

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

FIGURE 1
Organization Framework for Investigating Customer Satisfaction After Service Failure and Recovery



if a customer must first experience a problem and seek redress to achieve the level of satisfaction originally expected, this experience would be evaluated as inferior to one that had produced the desired result without any intervening problems. In essence, this perspective holds that the initial disconfirmation drives the final CS/D evaluation with recovery mitigating the dissatisfaction arising from the failure. To illustrate this situation, consider the following example:

> Customers Jones and Smith both have been satisfied customers of the same bank for many years. Customer Jones always has found the service of the bank to meet or exceed expectations and has never encountered a problem. Customer Smith also found the bank's service met or exceeded expectations until the bank made an error and returned a check written to a local grocery store marked "insufficient funds." When Customer Smith complained to the bank, a very courteous service representative promptly determined that the bank had made an error. The service representative apologized and outlined the steps the bank would take to correct the problem, which included sending a letter to the grocery store acknowledging that the bank was at fault and offering to cover any charges the store might assess. In addition, the service representative arranged to have Customer Smith's checking account credited for an additional $25.00 as compensation for any embarrassment and inconvenience. "After all," said the service representative, "if we expect our customers to pay for their errors, we should be prepared to pay for ours."

In this situation, why should Customer Smith's postrecovery satisfaction be greater than Customer Jones's if Customer Smith had to contact the bank to have the problem corrected and possibly experienced embarrassment and anxiety because of the returned check? Customer Smith might be pleased with the prompt and efficient recovery effort, but Smith would probably have preferred error-free service.

One possible answer to the previous question regarding Smith's satisfaction versus Jones's is that customers have expectations regarding appropriate recovery efforts in the event of failure. Extending this argument, some customers may not only have recovery expectations, but some may also have failure expectations. For instance, many customers recognize that consumption entails some potential for dissatisfaction (Murray and Schlacter 1990). Therefore, to determine what will be done in the event of a failure, they inquire about warranties, exchange, and refund policies.

If service failure is not totally unexpected, satisfaction judgments may be suspended until the recovery efforts can be evaluated. If recovery efforts meet expectations, then the evaluation of the overall transaction would be one of confirmation. (The customer's expectation of a possible problem has been confirmed as has the customer's expectations of appropriate recovery efforts.) However, this evaluation should be lower than that of a customer who recognized the potential for a problem but did not encounter one, resulting in a state of positive disconfirmation.

Let us imagine another situation in which a customer felt that sooner or later all banks would make an error and that it would be very difficult to have the problem corrected. The check returned in error would, therefore, result in a state of confirmation, but having the problem quickly and agreeably resolved would be a positively disconfirming experience. In this case, the prompt correction of the problem and the monetary credit might so exceed Customer Smith's original expectations of problem resolution that high postrecovery satisfaction would result. Therefore, within the framework of disconfirmation, the recovery paradox would best be explained by recovery performance that exceeds expectations given failure expectations.

Our conceptualization of recovery modeled in Figure 1 can be viewed as a second disconfirmation path (for recovery judgments) that can be incorporated within the standard disconfirmation model. This conceptualization is potentially consistent with Halstead and Page (1992), who state that the customer's ultimate, postrecovery satisfaction is driven by the initial satisfaction judgment with recovery mediating the final, postrecovery satisfaction judgment. The treatment of recovery performance is also harmonious with the recovery model proposed and evaluated by Smith and Bolton (1998), who found an increase in

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

cumulative satisfaction when the customer was very satisfied with an organization's recovery effort. In many aspects, this perspective also is consistent with the disconfirmation-based recovery models proposed (but which were not empirically evaluated) by Oliver (1981) and Singh and Widing (1991). On the basis of the previous discussion, the following hypotheses are offered (see Figure 1):

Initial Disconfirmation Hypotheses

Hypothesis 1: The greater (lower) failure expectations, the less (more) negative initial disconfirmation.

Hypothesis 2: The higher (lower) perceived service performance, the more (less) positive initial disconfirmation.

Hypothesis 3: The more negative (positive) initial disconfirmation, the greater dissatisfaction (satisfaction).

Recovery Disconfirmation Hypotheses

Hypothesis 4: The greater (lower) recovery expectations, the less (more) positive recovery disconfirmation.

Hypothesis 5: The higher (lower) recovery performance, the more positive (negative) recovery disconfirmation.

Hypothesis 6: The more positive (negative) recovery disconfirmation, the greater satisfaction (dissatisfaction).

Recovery Paradox Hypothesis

Hypothesis 7: Postrecovery satisfaction will be equal to or greater than the satisfaction when no-service failure occurs given (a) high failure expectations, (b) low recovery expectations, and (c) high recovery performance.

Distributive and Interactional Justice

An emerging literature has examined recovery evaluations from the perspective of justice (Goodwin and Ross 1992; Hocutt, Chakraborty, and Mowen 1997b; Tax, Brown, and Chandrashekaran 1998). This research adds to this work by exploring the role played by distributive and interactional justice in recovery situations (see Figure 1).

Distributive justice (DJ) specifies that individuals evaluate the fairness of an exchange by comparing costs with the gains received (Greenberg 1987, 1990b). Although DJ can be conceptualized as customers' evaluations of whether they got "their money's worth," it also can include nonmonetary inputs and outputs involving such intangibles as emotions (anger and embarrassment), complaining costs (time and effort), and ego benefits. From a DJ perspective, classic inequity arises when an individual believes the outcome is inadequate given inputs. The most common recoveries are those in which the customer responds to "negative" inequity by seeking redress and involve some combination of additional rewards or lowered costs, for example, a refund, adjustment of costs, or an exchange.

Feelings of preference or advantageous inequity can occur when an individual believes the outcome is greater than deserved given the inputs (Oliver and Swan 1989a, 1989b). Given such positive inequity, it is possible for feelings of guilt or indebtedness to arise. However, justice research suggests that individuals have an egocentric bias (Greenberg 1987) wherein they are more tolerant of advantageous than disadvantageous inequity (Oliver and Swan 1989b).

Because neither inputs nor outputs need to be economic and are viewed from the subjective viewpoints of the exchange principles, it is possible for differences in perspective to lead to differences in perceived DJ. In a redress situation, a problem of inequity could result in dissatisfaction if a service provider does not recognize the same costs as the customer (the customer's time and effort to seek redress). Alternatively, if the customer's norm is that only direct monetary costs will be reimbursed, then compensating nonmonetary costs would result in feelings of advantageous inequity. Although feelings of guilt are possible, the most likely outcome of reimbursing both monetary and nonmonetary costs (given the customer's egocentric bias) should be postrecovery satisfaction superior to that experienced when only monetary costs are reimbursed. Therefore,

Hypothesis 8: The greater (lower) perceptions of distributive justice, the greater postrecovery satisfaction (dissatisfaction).

Interactional justice (IJ) concerns the fairness of the recovery process itself or the interactional aspects of the encounter (Bies and Moag 1986; Goodwin and Ross 1992). Interactional justice is generally considered to be a subelement of the more global justice construct of procedural justice (PJ), which concerns the fairness or the processes used to determine the distribution of outputs (Greenberg 1990b). In a recovery situation, PJ would concern the perceived fairness of the organization's recovery policies, whereas IJ would be concerned with how the organization's policies would be carried out. An example of IJ would be an apology to the customer for a service failure. This research will focus specifically on the IJ aspects of PJ.

Customers may judge the recovery effort on the basis of both the recovery process (IJ) and outcome (DJ). For example, consider two different recovery situations in which the same financial outcome is achieved by the customer

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

(identical DJ). If the service provider is perceived as apologetic, empathetic, and responsive in one situation and is curt and indifferent in the other, the customer's postrecovery satisfaction almost certainly would be higher in the first situation even though the same tangible outcomes are achieved in both cases.

Hypothesis 9: The greater (lower) perceptions of IJ, the greater postrecovery satisfaction (dissatisfaction).

DJ and IJ are not independent constructs but aspects of the same overriding construct, justice. To some degree, DJ is a necessary, but not sufficient, condition for IJ and PJ; without a truly fair outcome, the procedures and interaction very well may be judged as flawed (McFarlin and Sweeney 1992). Likewise, DJ does not ensure favorable perceptions of IJ. (Distribution might be fair even if the interaction is not.) Therefore, organizations seeking to provide high satisfaction must be perceived as offering both DJ and IJ. Indeed, Greenberg (1990a) labeled situations in which individuals attempt to be recognized as being fair without actually behaving fairly as "hollow justice" and noted it may backfire if individuals suspect insincerity. Likewise, it has been argued that in some situations, PJ efforts that do not lead to appropriate changes in DJ may create a frustration effect (Folger et al. 1979). Such situations have been labeled as "sham" participation or as "pseudoparticipation" (Cohen 1985; Goodwin and Ross 1992) and may result in greater dissatisfaction and exit behavior by employees and customers. Likewise, Tax, Brown, and Chadrashekaran (1998) found evidence of an interactive effect between DJ and IJ.

Hypothesis 10: DJ and IJ will interact with each other as they affect postrecovery satisfaction.

RESEARCH METHOD

Research Plan

Scenario-based experiments were conducted to investigate service recovery. Bitner (1990) notes that role-playing experiments (scenarios) allow expensive or difficult manipulations to be more easily operationalized, provide the researcher with control over otherwise unmanageable variables, and facilitate the compression of time by summarizing events that might otherwise unfold over days or weeks. Furthermore, the use of scenarios avoids the expense and ethical consideration associated with observing or enacting actual service failure while avoiding the response bias due to memory lapses and rationalization likely to be present in surveys that rely on recall (Smith and Bolton 1998). The key drawbacks of role-playing are a greater likelihood of demand effects and the possible inability of participants to project their behavior and to respond as they actually would in a real situation. To minimize these problems, we recruited individuals who were in the midst of the actual service encounter being studied. This procedure would ensure that participants were familiar with the service offering and minimize concerns that participants in a laboratory setting might respond differently from those actually involved in the service setting.

Research Setting

Airline travel was chosen as the research setting. Airline travel represents a service for which failure is common (Andreeva 1998). As a result, it was anticipated that most airline travelers would find manipulations regarding recovery expectations, recovery performance, and justice realistic and believable. Using an airport served by numerous airlines also eliminated any possible brand bias. In addition, most passengers, when facing a flight delay or cancellation, will have no choice but to seek redress as canceling the trip is not an option for most passengers. Therefore, in this research setting, no bias is created by imposing redress-seeking behavior on respondents who might not ordinarily voice complaints. Finally, the administration of scenario-based experimental manipulations by intercepting airline passengers was established by Bitner (1990) as a reliable and valid research methodology.

Research Design

To test the research hypotheses, two separate studies were developed. In Study 1, the disconfirmation-based hypotheses (Hypotheses 1-7) were investigated by the manipulation of recovery performance and recovery expectations. Study 2 evaluated the justice hypotheses (Hypotheses 8-10) by the manipulation of DJ and IJ. Two separate studies were conducted because the excessive length and complexity of a single research instrument to investigate both disconfirmation and justice explanations of recovery were not deemed practical.

Data Collection

Pretests. Three pretests were conducted. The first two employed student participants and assisted in determining the realism of the scenarios and in the development and refinement of the measures. After modifying the research instrument, concurrent verbal protocols were conducted to verify that respondents were interpreting the scenarios as intended.

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

The final pretest ($N = 240$) was intended to mirror the final data collection plan as closely as possible. A small airport used exclusively by commuter airlines was the location for the final pretest. Previous research has shown that the consumers' attributions[1] regarding a failure can influence consumer evaluations (Folkes 1984; Folkes and Kotsos 1986). To establish firm responsibility for the failure (and avoid possible attribution confounds), the cause of the service failure was initially attributed to a mechanical problem. However, this seemed to create confound by triggering negative beliefs regarding the airline's safety, possibly due to the smaller aircraft used by the airlines at the pretest airport and a well-publicized crash of a commuter aircraft just before the sampling period. Therefore, for the final data collection, the cause of the service problem was changed from a mechanical problem to crew unavailability.[2] This change maintained the airline's responsibility for the service failure but eliminated the safety confound.

Main study. The data were collected at a medium-sized regional airport in the southwestern United States. This airport was viewed as ideal for the final data collection because it was served by the top nine domestic carriers, was geographically accessible, and was served primarily by jet service. After securing the cooperation of the airport manager, the principal researcher met with the managers of the carriers to review the research objectives and methodology and to address any concerns. The airport's one commuter airline was not included in the study, and all passengers surveyed were preparing to board jet aircraft. All participants were randomly assigned to treatment conditions. Passengers were approached in the waiting areas (most were seated) and asked to participate. In total, passengers from 50 departures were included in the sample, and all major carriers were represented by several departures. Observations were obtained for all hours of operation and all days of the week during a 2-week period. To minimize disruption to the operation of the airport, and in consultations with airport management, data for both studies were collected concurrently. Therefore, the population of both studies is the same.

For passengers unable to complete and return the survey before boarding, a postage-paid, preaddressed envelope was attached to the research instrument. The primary purpose of this option was to overcome travelers' concerns that they would not have time to complete the survey before boarding. This approach also allowed the inclusion of late arrivals in the research, precluding any systematic bias. The vast majority of participants (75.7%) returned the survey in the terminal. The method of survey return (in terminal or by mail) was entered into all subsequent data analyses, and in no case did it have a material effect on the research conclusions.

A total of 1,005 airline passengers were approached while waiting to board flights and asked to participate in the study. Of those, 727 (72.3%) agreed to cooperate. A total of 550 passengers returned the surveys before boarding, whereas 177 took surveys with them. Of these 177 sample members, 65 mailed back their surveys within 3 weeks of the completion of sampling. Therefore, the final sample was 615 (61.2% of those approached and 84.6% of those who agreed to participate).

Study 1: Failure Expectations, Recovery Expectations, and Recovery Performance

Study 1, a 2×3 between-subjects design, evaluated the recovery disconfirmation-based hypotheses (Hypotheses 1-7) by manipulating recovery expectations (high/low) and *service performance* (service failure with either high or low recovery or the control condition of no service failure).[3] Recovery expectations were manipulated by including in the scenarios a sign posted at the airline ticket counter. In the high-recovery expectation condition, the sign promised a $150 ticket voucher for a delay of 2 hours or more that was the airline's responsibility. In the low-recovery expectation condition, the sign consisted of a disclaimer in which the airline refused to accept responsibility for delays and cancellations beyond that required by law. Failure expectations were not manipulated but were measured before the participants read the experimental scenario. This approach simplified the research while allowing the impact of failure expectations to be assessed.

The high- and low-recovery performance conditions were based on a review of the recovery literature and feedback from consumers. Many of the aspects of recovery that were manipulated were consistent with those identified subsequent to the data collection by Boshoff (1999) as important determinants of recovery satisfaction. High recovery was designed to be a prototypical, superb recovery that would significantly exceed the typical airline response to a cancellation. In the high-recovery situation the airline treated the passenger courteously, made several apologies, and anticipated the traveler's needs. In addition, the passenger was given a $150 ticket voucher, meal vouchers, and offered use of the phone. The passenger was rebooked on the next available flight.

1. Attributions regarding both the service failure and the recovery are potentially important factors for understanding postrecovery satisfaction judgments. However, due to the limitations inherent in any one study, they were not the focus of this research.

2. Examples of crew unavailability might be when crew members are delayed in arriving at the airport due to a previous flight that is late.

3. A summary of all manipulations for both Studies 1 and 2 is included in the appendix.

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

TABLE 1
Scale Items and Measurement Properties

Measurement Item	*Study 1*[a] *SL*	*IR*	*CR*
Failure Expectations			.79
In general, I am not surprised if I encounter some kind of problem when I fly.	.55	.32	
I would consider myself lucky if I did not experience some kind of problem with my flight today.	.85	.69	
I consider the odds of running into a problem when I fly as being pretty high.	.81	.67	
Service Performance			.55
The airline's on-time performance was very poor.[b]	.48	.22	
The airline's reliability was very high.	.74	.51	
Recovery Expectations			.73
My expectations were high that I would receive compensation if I encountered a long delay.	.63	.39	
For the situation described, I had very high expectations concerning actions the airline would take to deal with a lengthy delay.	.65	.44	
After reading the posted sign, I expected the airline to do whatever it took to guarantee my satisfaction.	.76	.58	
I didn't expect this airline to do much for me if I encountered a long delay.[b]	–.48	.24	
Recovery Performance			.83
I would rate the performance of the airline in dealing with the cancellation as exceptional.	.81	.64	
For the situation described, I would rate the efforts of the airline to deal with my problem as superior.	.88	.78	
Initial Disconfirmation			.56
The airline's reliability was about what I expected.	.47	.12	
This airline's on-time performance was much better than I expected.	.77	.28	
Recovery Disconfirmation			.75
I expected the airline would do more in response to the canceled flight.[b]	.72	.52	
The compensation for my problem was much better than I expected the airline would provide.	–.72	.53	
After reading the airline's posted policy, I expected the airline would do more for me.[b]	.68	.48	
Final Satisfaction/Postrecovery Satisfaction			.90
Overall, how satisfied or dissatisfied did this experience leave you feeling?[c]	.88	.77	
How well did this service experience meet your needs?[d]	.84	.72	
Overall, I am very satisfied with this experience.	.87	.76	

NOTE: Unless noted, all items used a 7-point, *strongly disagree/strongly agree*, Likert-type format. See Notes c and d below.
a. SL = standardized loadings; IR = item reliability; CR = composite reliability; VE = variance extracted. VE and CR calculated consistent with Fornell and Larcker (1981).
b. Reverse coded.
c. Used a 9-point, *extremely dissatisfied/extremely satisfied*, Likert-type format.
d. Used a 9-point, *extremely poor/extremely well*, Likert-type format.

In the low-recovery condition the agent was polite but not as courteous as under the high-recovery condition and responded to, but did not anticipate, the traveler's needs. The passenger was rebooked on the next available flight. The passenger was given a meal voucher and offered the use of a phone but was not offered the $150 ticket voucher. When, in the high-recovery expectation condition, the passengers inquired as to whether the $150 ticket voucher would be issued, the airline agent stated, "No, I am sorry, but that policy applies to our full-fare passengers, and you have a discounted ticket." In the case of both high- and low-recovery performance, the total length of the delay (3 hours) was the same.

Method of data analysis. The Study 1 disconfirmation model for the case of service failure was examined using LISREL-VIII (Jöreskog and Sörbom 1993). The recovery paradox was evaluated by using analysis of covariance (ANCOVA).

Measures. The scales employed in Study 1 are given in Table 1. All questions used a 7-point (*strongly disagree-strongly agree*) Likert-type scale, except for two of the Satisfaction scale questions, which were measured using a 9-point scale to limit skewness (Fornell 1992) and which employed different anchors (see Table 1 and the following discussion of the specific satisfaction measures). During the final data collection, respondents reported very little confusion regarding the questionnaire items. Because of the uniqueness of the research setting (airline service failure and recovery) and the presence of many related and distinct constructs such as failure and recovery expectations, initial and recovery performance, and initial and recovery disconfirmation, many measures are substantially original to this research.

Satisfaction. The Satisfaction scale employed three items. One question, adapted from Westbrook (1980), assessed how well the service experience met the consum-

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

ers' needs and was anchored by *extremely poor/extremely well*. The other satisfaction items were consistent with common satisfaction measures reported in the literature. One item treated satisfaction/dissatisfaction as a bipolar construct, anchored by *extremely dissatisfied/extremely satisfied* (with the midpoint labeled *neither*). This measure was similar to that employed by Oliver and Bearden (1985). Consistent with Fornell (1992), the two preceding satisfaction measures used a 9-point scale to limit skewness. Another satisfaction measure (embedded in a different battery of questions) treated satisfaction as a unidimensional construct (similar to Oliver and Bearden 1985; Westbrook 1980, 1981) and employed a 7-point (*strongly agree/strongly disagree*) Likert-type scale. For all Satisfaction scales, care was taken to ensure that the measures tapped satisfaction with the service encounter portrayed in the scenarios and not a generalized attitude regarding the service provider by explicitly asking participants to rate their satisfaction with the service experience given in the scenarios (see Table 1).

Performance, expectations, and disconfirmation measures. The research setting and the need to carefully distinguish between two types of expectations (failure and recovery), two sets of performance (initial service performance and recovery performance), and two types of disconfirmation (initial and recovery) necessitated the development of original scales. The disconfirmation literature was examined carefully in constructing these scales. For instance, both Disconfirmation scales included a question related to "much better," which was adapted from Swan and Trawick (1981). Consistent with the needs of this study and the advice of Yi (1990), perceived disconfirmation measures were employed.

Manipulation checks. Manipulation checks indicated that the two recovery performance conditions (high/low [HI/LO]) differed significantly in perceived recovery performance, $\overline{X}_{HI/LO} = 5.44/3.54$, $F(1, 113) = 55.4$, $p < .0001$, $\eta^2 = .33$. The manipulation of recovery performance had a small effect (based on effect size) on recovery expectations, $\overline{X}_{HI/LO/NO\text{-}FAIL} = 4.28/4.85/4.48$, $F(2, 184) = 3.5$, $p = .03$, $\eta^2 = .04$. Manipulation checks also indicated that the high- and low-expectation treatments differed significantly in perceived recovery expectations, $\overline{X}_{HI/LO} = 5.04/4.04$, $F(1, 184) = 31.6$, $p < .0001$, $\eta^2 = .15$. The manipulation recovery expectations had a slight effect on perceived recovery performance, $\overline{X}_{HI/LO} = 4.27/4.72$, $F(1, 113) = 3.2$, $p = .08$, $\eta^2 = .03$. Using the guidelines proposed by Perdue and Summers (1986), these results indicate that the manipulations worked as intended.

Measurement model. The analysis of the disconfirmation model of recovery given in Figure 1 followed the two-step approach recommended by Anderson and Gerbing (1988) with the measurement model examined followed by the structural equations model. The standardized loadings, item reliabilities, and composite reliability for the items and scales used to measure the latent constructs in the disconfirmation model of recovery are given in Table 1. The total sample size available for the structural model was $N = 117$. Standardized loadings (all significant at $p < .001$) and the item reliabilities suggest that the constructs' items cohere reasonably well. Bagozzi and Yi (1988) suggest as a guideline that composite reliability should be equal to or greater than about .60. All scales with the exception of Service Performance (.55) and Initial Disconfirmation (.56) meet this criterion.

Discriminant validity was assessed using Anderson's (1987) criterion that the correlation between two latent constructs plus or minus two standard errors does not include one. This criterion is satisfied for all construct pairs except for between Recovery Disconfirmation and Recovery Performance [0.93 ± 2 (.05) = 1.03, .83]. Thus, even here the rule is violated only marginally. The discriminant validity of the scales was also assessed using confirmatory factor analysis (Anderson and Gerbing 1988), focusing specifically on the conceptually similar constructs that could potentially overlap: initial disconfirmation, recovery disconfirmation, and satisfaction. The results of this analysis suggest that the three-factor solution is better than a one-factor solution (three-construct solution: $\chi^2[17] = 26.16$; one-construct solution: $\chi^2[20] = 65.76$; $\chi^2[3]$ difference = 39.59, $p < .001$). Thus, the measures of these conceptually similar constructs are empirically distinct.

The overall fit of the structural model (i.e., the Study 1 disconfirmation portion of the model portrayed in Figure 1) was first assessed by examining the chi-square statistic, which is significant, $\chi^2(140) = 262.3$, $p < .001$. However, as this statistic is sensitive to sample size (Bagozzi and Yi 1988), additional measures of fit were examined. Bentler and Bonett's (1980) Normed Fit Index (NFI) (D) is .79, whereas Tucker and Lewis's (1973) Non-Normed Fit Index (NNFI) is .86. The Goodness-of-Fit Index (GFI) is .82, whereas the Adjusted Goodness-of-Fit Index (AGFI) is .75. The Comparative Fit Index (CFI) is .89, and the Incremental Fit Index (IFI) is .89. The standardized root mean square residual (RMR) is .077. The root mean square error of approximation (RMSEA) was .08. The modification indexes did not suggest any significant modifications of either the measurement or structural model.

Thus, as was the case for the measurement model, the structural model demonstrates modest fit. When assessing the overall fit of the model, it is important to take into consideration the number of related constructs that were modeled, the use of new scales for many of the constructs, and the nature of the research setting. For instance, two different types of expectations (failure and recovery), two differ-

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

TABLE 2
Standardized Estimates of Hypothesized Relationships

Relationship (Hypothesis)	*Standardized Estimate*	t-*Value*
Effect of failure expectation on Initial Disconfirmation (Hypothesis 1: +)	.17	1.34**
Effect of service performance on Initial Disconfirmation (Hypothesis 2: +)	.96	3.60*
Effect of initial disconfirmation on Satisfaction (Hypothesis 3: +)	.70	4.56*
Effect of recovery expectation on Recovery Disconfirmation (Hypothesis 4: –)	–.51	–4.82*
Effect of recovery performance on Recovery Disconfirmation (Hypothesis 5: +)	.71	7.02*
Effect of recovery disconfirmation on Satisfaction (Hypothesis 6: +)	.43	4.68*

*$p < .001$ (one-tailed). **$p < .10$ (one-tailed).

ent types of performance (initial service performance and recovery), and two different types of disconfirmation (initial and recovery) were included in the model. Given the overall adequate fit of the model, an examination of the hypotheses tests was deemed appropriate.

Hypotheses tests. Table 2 provides the standardized estimates and *t*-values for the hypothesized relationships involving service failure and recovery. The direction of all paths is consistent with that hypothesized and is significant at $p < .001$, with the exception of the effect of failure expectations on initial disconfirmation ($p < .1$).

Failure expectations exert a relatively weak, positive effect on initial disconfirmation. Those with low (high) failure expectations felt the airlines' reliability was worse (better) than expected. However, Hypothesis 1 is not supported at $p < .05$. Service performance exerts a significant positive influence on initial disconfirmation, supporting Hypothesis 2. In turn, initial disconfirmation has a significant positive effect on satisfaction. Thus, consistent with Hypothesis 3, those who reported the airlines' reliability was worse (better) than expected reported greater dissatisfaction (satisfaction).

The recovery disconfirmation relationships mirrored the initial disconfirmation findings. Those with high (low) recovery expectations perceived the recovery performance as worse (better) than expected, supporting Hypothesis 4. Also, consistent with Hypothesis 5, perceived recovery performance had a significant effect on recovery disconfirmation. Finally, recovery disconfirmation exerted a positive effect on satisfaction (Hypothesis 6). Those who perceived that the recovery was better (worse) than expected were more satisfied (dissatisfied).

Examining the relative magnitudes of the standardized estimates sheds some light on the strength of the relationships in the model. In general, the impact of performance (initial or recovery) on disconfirmation is greater than the effect of expectations (either failure or recovery) on disconfirmation. This finding mirrors the consistent findings from the disconfirmation literature that performance exerts a greater impact on disconfirmation than expectations. Interestingly, initial disconfirmation exerts a much stronger impact on satisfaction than does recovery disconfirmation. This implies that recovery can only mitigate the impact of failure on satisfaction and that initial service performance is the primary driver of final satisfaction.

As stated in Hypothesis 7, the recovery paradox was proposed to be most likely given (a) high failure expectations, (b) low recovery expectations, and (c) high recovery performance. Under these circumstances, failure should be the least disconfirming and recovery should have the greatest positive disconfirmation effect, potentially resulting in satisfaction equal to or greater than the control case of no-service failure. To evaluate this hypothesis, an ANCOVA model (using the general linear model) was run with satisfaction as the dependent variable and the treatment conditions entered as the main effects and failure expectations[4] treated as a covariate. Figure 2 graphically portrays the mean satisfaction ratings in the failure and no-failure conditions.

To test the recovery paradox, the mean of the low-recovery expectations and high-recovery performance group was compared with the means of the no-failure conditions. The mean of the low-recovery expectations and high-recovery performance group is 5.69, significantly lower than the mean of either of the no-failure groups (the mean of the high-recovery expectations and no-failure condition was 7.15, $p < .001$; the mean of the low-recovery expectations and no-failure condition was 6.96, $p < .001$). These results do not support a recovery paradox effect, and Hypothesis 7 is rejected.

Study 2: Distributive and Interactional Justice

Study 2, a 3×3 between-subjects design, evaluated Hypotheses 8 through 10 by manipulating three levels of *distributive justice* (high, moderate, and low) and three levels of *interactional justice* (high, moderate, and low). In addition, a control condition of no-service failure was included. The failure was, as in the case of Study 1, depicted as "crew unavailability," resulting in a 3-hour delay to the passenger. Under high DJ, the passenger received a $150

4. The recovery paradox was evaluated on a sample composed of those that responded to scenarios involving failure and recovery (the same as the path analysis data analysis) as well as the control condition of no-service failure.

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

FIGURE 2
Satisfaction in the Failure and No-Failure Conditions

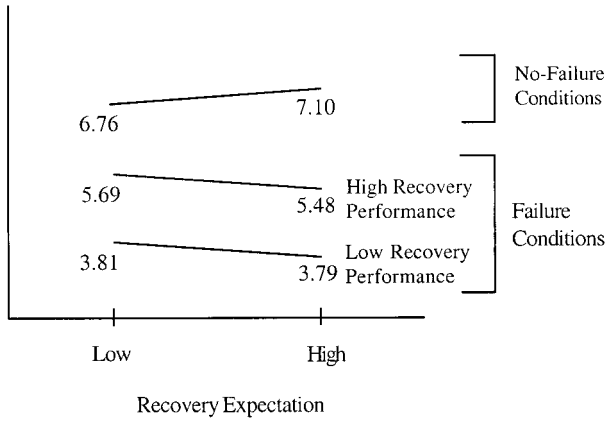


TABLE 3
Scale Items and Coefficient Alpha for Study 2

Scale Items	*Cronbach's Alpha*
Interactional Justice	.79
The agent demonstrated a poor understanding of my needs.[a]	
In dealing with me, the agent treated me in a courteous manner.	
I was offered an apology.	
Distributive Justice	.83
This trip resulted in a very positive outcome for me.	
I was more than compensated for any out-of-pocket expenses I might have incurred.	
I was more than compensated for any frustration.	
I got more out of this transaction than the airline.	

NOTE: All scales used a 7-point *strongly disagree/strongly agree* Likert-type format.
a. Reverse coded.

ticket voucher, a meal voucher, and the use of a phone. Under moderate DJ, the passenger did not receive a ticket voucher but did receive meal vouchers and use of a phone. Under low DJ, the passengers did not receive a ticket voucher, meal voucher, or use of the phone (see the appendix for an overview of the manipulations).

Under high IJ, the agent was portrayed as anticipating the passengers' needs and as being very professional, courteous, and apologetic. Under moderate IJ, the agent was less apologetic and responded to, but did not anticipate, passengers' needs. Under low IJ, the agent was rude and indifferent to passengers' plights and did not offer an apology. In addition, under low IJ, and unlike the high and moderate conditions, the passenger was placed on standby and not immediately rebooked on another flight. However, to keep the objective level of harm caused by the failure the same, the total length of the delay (3 hours) was the same in all conditions.

Method of data analysis. For the Study 2 justice model, analysis of variance (ANOVA) using the general linear model was employed.

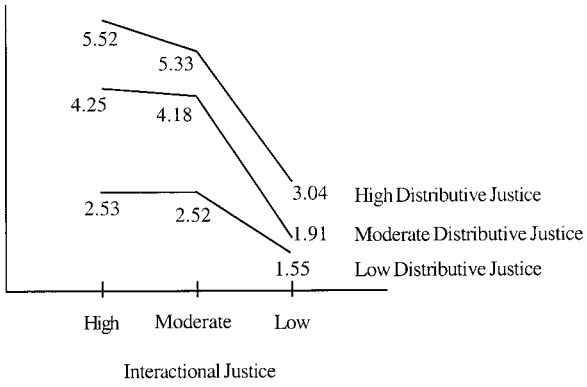
Measures. The items composing the *interactional* and *distributive justice* scales are given in Table 3. Some items were adapted from Oliver and Swan (1989a). Others were developed to assess aspects of justice that have been cited as important in the recovery literature such as compensation for out-of-pocket expenses and frustration and the presence of an apology (Goodwin and Ross 1992). The items composing the Satisfaction scale were identical to Study 1. The Study 2 coefficient alpha for satisfaction is .92 ($N = 376$). Thus, all scales demonstrate adequate reliability.

Manipulation checks. Manipulation checks using the perceived-IJ scale indicated that the IJ manipulation was successful, $\bar{X}_{HI/MOD/LO} = 5.56/5.19/2.37$, $F(2, 314) = 223.1$, $p < .0001$, $\eta^2 = .59$. The difference between the high- and moderate-IJ treatment is significant at $p = .03$, whereas the difference between all other mean comparisons is significant at $p < .0001$. The manipulation of IJ had a slight but significant effect on DJ, $\bar{X}_{HI/MOD/LO} = 3.26/3.28/2.40$, $F(2, 302) = 18.9$, $p < .0001$, $\eta^2 = .11$. The check of the manipulation of DJ indicates that this treatment also was successful. The respective perceived-DJ means for the DJ treatment groups are $\bar{X}_{HI/MOD/LO} = 4.19/2.87/1.87$, $F(2, 302) = 92.4$, $p < .0001$, $\eta^2 = .39$. The difference between all means is significant at $p < .0001$. Manipulating DJ had a slight, significant effect on IJ, $\bar{X}_{HI/MOD/LO} = 4.92/4.55/3.64$, $F(2, 314) = 31.7$, $p < .0001$, $\eta^2 = .17$. A comparison of the η^2 indicates that the manipulations were successful (see Perdue and Summers 1986).

Hypotheses tests. To evaluate the justice hypotheses, an ANOVA model was run with satisfaction as the dependent variable and the main effect treatments and interaction as the dependent variables. Figure 3 graphically presents the effect of interactional and distributive justice on postrecovery satisfaction. In general, the higher the distributive justice, the higher the satisfaction. The impact of interactional justice is most pronounced when comparing either high or moderate IJ with low IJ. DJ had a significant effect on postrecovery satisfaction, $\bar{X}_{HI/MOD/LO} = 4.63/3.44/2.20$, $F(2, 319) = 50.7$, $p < .0001$, $\eta^2 = .24$. The difference between all means is significant at $p \leq .0001$. Therefore, Hypothesis 8 is supported. IJ also had a significant effect on postrecovery satisfaction, $\bar{X}_{HI/MOD/LO} =$

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

FIGURE 3
The Effect of Interactional and Distributive Justice on Postrecovery Satisfaction



4.10/4.01/2.16, $F(2, 319) = 44.5$, $p < .0001$, $\eta^2 = .22$. The difference between high and moderate IJ is not significant ($p = .7005$), whereas all other mean differences are significantly different at $p \leq .0001$. Therefore, Hypothesis 9 is supported. The interaction between IJ and DJ is also significant, supporting Hypothesis 10, $F(4, 319) = 2.8$, $p = .0265$, $\eta^2 = .03$.

To shed further light on the effect of IJ and DJ on postrecovery satisfaction, the interaction effect was isolated from the main effects for each cell (interaction effect = cell mean – grand mean – row effect – column effect) where the grand mean is the overall mean, the row mean effect = row mean – grand mean, and the column mean = column mean – grand mean (see Ross and Creyer 1993). The interaction effects, presented in Table 4, are illuminating. When the relative effects of IJ and DJ are similar (the diagonal elements), the interaction is "positive." In these cases, IJ and DJ act together to produce greater postrecovery satisfaction than average. However, when the main effects are mismatched (the off-diagonal elements), the interaction approaches zero and becomes negative the more "mismatched" the respective justice elements are. For instance, the most negative values occur in the case of high DJ and low IJ or high IJ and low DJ. These negative interactions for cases of high or moderate IJ and low DJ are consistent with a "sham" effect (Cohen 1985; Goodwin and Ross 1992) or "hollow" justice (Greenberg 1990a). These findings also mirror those of Tax, Brown, and Chandrashekaran (1998), who found significant interactions among DJ and PJ as well as DJ and IJ. Likewise, Hirschman (1970) notes that voice that leads to frustration may actually increase exiting behavior. In the case of high IJ and low DJ, it is possible the respondents doubted the sincerity of the gate agents' apologies when even the most basic of compensation was not provided. Likewise, when gate agents provided high levels of compensation without corresponding levels of empathy and understanding (high DJ and low IJ), passengers may have felt they were being "bought off."

No specific recovery paradox hypothesis was proposed concerning Study 2. However, in an attempt to gain a greater understanding of the recovery paradox, the high-IJ and high-DJ cell was compared to the no-failure control group for satisfaction. The mean satisfaction of the no-failure condition (7.43) is significantly higher ($p < .0001$) than the mean satisfaction for the high-IJ and high-DJ failure group (5.52). Therefore, replicating the findings of Study 1, no recovery paradox emerges on the strength of recovery performance alone.

DISCUSSION AND IMPLICATIONS

In general, the research hypotheses were well supported, with evidence confirming both the disconfirmation model as well as the important role of justice. Among the most intriguing findings of this research are (a) a lack of support for a recovery paradox effect, (b) the impact of both initial and recovery disconfirmation on final postrecovery satisfaction, (c) the relatively greater impact of initial disconfirmation versus recovery disconfirmation on satisfaction, and (d) the finding of an interaction between DJ and IJ, which underscores the importance of providing consistent levels of these two types of justice in recovery efforts. Following is a review and discussion of the implications of these findings.

Disconfirmation

Ultimately, final satisfaction provides the litmus test against which the impact of service failure and recovery should be judged. In Study 1, satisfaction is greatest in the case of no failure. Given failure, the path analysis shows that initial disconfirmation has a greater impact on satisfaction than recovery disconfirmation. In turn, initial service performance is the primary predictor of initial disconfirmation, whereas recovery performance is the primary predictor of recovery disconfirmation. Therefore, consistent with Halstead and Page (1992), it appears that satisfaction is primarily driven by the initial service failure with recovery performance acting to mitigate the damage to satisfaction caused by the failure. However, given the impact of recovery performance, recovery expectations, and recovery disconfirmation on postrecovery satisfaction, failure service providers should strive to offer high-recovery performance that exceeds customer expectations.

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

Interactional and Distributive Justice

Both DJ and IJ are important predictors of postrecovery satisfaction. Research reported in the organizational behavior literature has found that when the initially desired outcome is blocked, evaluations can still be relatively high given high IJ or PJ (Folger and Konovsky 1989; Greenberg 1990b; McFarlin and Sweeney 1992). In this case, the originally desired outcome (arriving at their destination on time) was unobtainable; therefore, IJ (the how) may have added importance in determining postrecovery satisfaction.

To reap the maximum impact on postrecovery satisfaction, the relative levels of IJ and DJ must be consistent. A synergistic effect seems to emerge when the levels of IJ and DJ are similar. In this case, the sum of the effect of IJ and DJ on postrecovery satisfaction is greater than the individual effects. Conversely, situations in which IJ and DJ are mismatched appear to create a situation where DJ and IJ work against one another. For instance, high IJ and low DJ may be perceived by consumers as indicating that the service provider's apologies are insincere. In the extreme, a "sham effect" appears to emerge. Conversely, high DJ and low IJ may leave consumers feeling "bought off" by a service provider who is not sincerely interested in their welfare.

For airlines (and service providers, in general), the finding that IJ and DJ are both necessary is noteworthy. Although expensive, giving monetary compensation is a straightforward response. Delivering IJ may be more of a challenge. Even the best gate agent might find it a challenge to be truly empathetic and responsive when facing a long line of upset customers following a flight cancellation. Service managers should consider special training and support services that help frontline service providers respond effectively to service failures.

The Recovery Paradox

By using a control condition of no-service failure, this research was able to directly evaluate the recovery paradox, whereas most previous studies that lacked such a condition have only been able to report results that indirectly addressed the recovery paradox. In this study, no support was found for a recovery paradox. In neither Study 1 nor Study 2 were the mean satisfaction ratings of participants who experienced service failure as great or greater than those who did not experience service failure. One possible explanation for the lack of a recovery paradox in this research is that for the service failure portrayed, a 3-hour delay, no recovery effort can completely mitigate the harm caused by the failure. Although travelers might find the $150 ticket voucher and the smooth and professional recovery portrayed as a pleasant surprise, the recovery effort cannot completely erase the harm caused by the failure. For instance, if the delay causes a business traveler to miss an important meeting, no realistic compensation is apt to totally erase the harm caused by the failure.

TABLE 4
Interactional and Distributive Justice Interaction Effects on Postrecovery Satisfaction (Study 2)

	Interactional Justice		
	High	*Moderate*	*Low*
Distributive justice			
High	.1463	.0535	–.3966
Moderate	.0629	.1827	–.3423
Low	–.4058	–.3329	.5422

NOTE: Interaction effect = cell mean – grand mean – row effect – column effect. Grand mean = overall mean; row effect = row mean – grand mean; column mean = column mean – grand mean. See Ross and Creyer (1993).

Data were collected in both studies on the perceived harm caused by the service failure. To explore further the role of harm in determining satisfaction, an ANCOVA model was run for both studies with satisfaction as the dependent variable and the experimental main effects and the perceived harm covariate[5] as the independent variables. In Study 1, the harm variable is significant. The mean of the high-recovery performance/low-recovery expectation condition (adjusted for harm) is 5.85. This is not statistically greater than the mean of either the no-failure/high-recovery expectation condition of 6.26 (p = .28) or the mean of the no-failure/low-recovery expectation condition of 6.34 (p = .14). This supports the possibility of a recovery paradox effect when the harm caused by the failure is taken into account. The results are similar for Study 2. The mean for the control condition of no failure is 5.97, whereas the mean of the high-IJ and high-DJ condition is 5.24 (p = .1006). Therefore, when full or near-full recovery is possible, superior recovery efforts might be able to produce a recovery paradox effect. This finding is consistent in some respect with the finding of Smith and Bolton (1998) that postrecovery cumulative satisfaction could increase if the customer was very satisfied with the recovery effort and the magnitude of the failure was considered. Likewise, Webster and Sundaram (1998) and Sundaram, Jurowski, and Webster (1997) report that service consumption criticality, similar to the construct of harm discussed here, interacts with recovery effort to influence customer satisfaction.

5. The following three items composed the perceived harm scale: (a) This service encounter created a major problem for me; (b) This service experience caused me a great deal of inconvenience; and (c) Overall, this service performance did not result in any real harm (reverse coded). Coefficient alpha Study 1 = .84, Study 2 = .74.

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

The finding that a recovery paradox effect may be possible when the recovery effort can completely mitigate the harm caused by failure may explain why some research has indicated that a recovery paradox effect is possible, whereas others have not. Hocutt, Chakraborty, and Mowen (1997a) found that under conditions of high redress, responsiveness, empathy, and courtesy, postrecovery satisfaction can be higher than the control condition of no-service failure. However, this recovery paradox effect may be explained by the failure, in this case an improperly prepared steak in a restaurant. Under the high-recovery performance condition, the steak was replaced, the customer was not charged, and the server was portrayed as responsive and empathetic. This would appear to be a case of relatively low harm in which the recovery effort could completely mitigate the harm caused by the failure. Would the results have been different had preparing another steak required 20 additional minutes and made the diners late for a show?

Conversely, Kelly, Hoffman, and Davis (1993), in a study of retail failure and recovery, reported that "correction plus recovery" (which involved additional compensation beyond the correction of the failure) was rated slightly less favorably than recovery strategies that simply corrected the problem. In explaining this counterintuitive finding, they note that most failures that resulted in correction plus recoveries were generally more severe in nature. Thus, the lower rating of the correction plus recoveries might have been due to the severity of the initial failure.

Recovery paradox effects are limited by the requirement that in most situations, customers must seek redress for recovery to occur. In the case of airline delays and cancellations, most consumers seek redress or correction of the problem. However, for many service failures, customers choose not to complain. In fact, under low-harm failure conditions, the customer is less likely to complain, and redress seeking may occur less often than under high-harm failure conditions (Richins 1983, 1987). Conversely, in a study of postcomplaint satisfaction, Tax, Brown, and Chandrashekaran (1998) found that most complaints resulted from problems judged by the consumer to be highly important.

It also should be remembered that harm is specific to the individual and the context. One customer's low-harm failure is another's high-harm failure. Consider the difference in harm incurred between a dine-in customer and a drive-through customer who receive the wrong sandwiches at a fast-food restaurant. Likewise, consider the difference in harm caused by a service failure between an important business meeting and informal dinner with casual acquaintances. Clearly, the issue of the harm caused by the failure is worthy of future study.

Is Service Recovery an Opportunity?

Does service recovery present an organization with an opportunity to improve customer satisfaction as argued by some (Abrams and Paese 1993; Hart, Heskett, and Sasser 1990)? Our finding that the higher the recovery performance, the higher the postrecovery satisfaction supports the importance of superior service recovery. Inferior recovery performance can lead to what Bitner, Booms, and Tetreault (1990) termed a *double deviation* from customer expectations: The firm fails to deliver on the initial service and the recovery service. Therefore, superior recovery could be viewed as an opportunity when compared to inferior recovery.

A consistent finding from the literature is that most dissatisfied customers never bother to complain. If superior service recoveries encourage customers who would not ordinarily complain to seek redress, then overall postservice failure customer satisfaction could be enhanced following outstanding service recovery. Therefore, postrecovery satisfaction following superior recovery is an opportunity when viewed against the dissatisfaction of those who do not seek redress. In addition, organizations may capitalize on the information gained from such complaints to design more reliable service delivery systems (Tax and Brown 1998). In this respect, service recovery can be viewed as an opportunity to gain access to superior market intelligence on the cause of customer dissatisfaction, which can, in turn, lead to more reliable service offerings.

However, our research indicates that excellent recovery is not an opportunity when compared to the satisfaction resulting from error-free service delivery. Error-free service is the better option for a variety of reasons, including customer confidence in the company's reliability and the absence, by and large, of a recovery paradox. Smith and Bolton (1998), in reporting evidence supporting a recovery paradox effect, observed that such an effect depended on achieving consistently high-service recovery performance. Uneven- or poor-service recovery risks alienating and losing customers. Our findings support the work of quality researchers (e.g., Crosby 1979; Deming 1986) who conclude that organizations should strive to identify and eliminate all potential sources of failure before the consumption experience to maximize both firm financial performance and customer satisfaction.

LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

This study relied on scenario-based experiments, an approach with strong precedent, and specific steps increased

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

realism and generalizability. Nevertheless, different methodologies should be employed to confirm and possibly extend the conclusions of this research. A survey approach could replicate many of the findings here and increase generalizability. The Critical Incident Technique (Bitner, Booms, and Tetreault 1990) could be useful in identifying specific actions service providers can take to ensure successful recoveries.

The empirical conclusions in this study are based on findings from one service industry. Caution must be exercised in extending the conclusions of this study to other services. For instance, it is possible that customer reactions to service failure and recovery differ based on their involvement in the service. As an example, customers rarely notice electrical services unless something goes wrong. Failure for such a service would result in elevating a "low-involvement" service to a "high-involvement" one. If recovery is superior, a recovery paradox effect might emerge when comparing the "unequal" satisfaction of the no-failure low-involvement group with that of the high-involvement postrecovery satisfaction group. Therefore, researchers should investigate customer satisfaction after low-involvement service failure and recovery.

Furthermore, although adequate, the fit indexes for the disconfirmation model of service failure and recovery are modest. Future research should seek improved measures of the constructs investigated here. In addition, the structural model might be enhanced by the incorporation of additional or alternative constructs such as justice or the incorporation of service failure harm.

The role of attributions regarding both the service failure and the recovery effort deserves study. The locus for the service failure was fixed in this research on the service provider, and no effort was made to manipulate stability and controllability attributions. However, Hocutt, Chakraborty, and Mowen (1997b) report that postrecovery satisfaction following service failure due to the customers' actions equals the control condition of no-service failure. Therefore, more attention should be paid to service failures contributed to by the consumers' actions or an act of God. Likewise, the role of harm in determining postrecovery satisfaction and a recovery paradox deserves additional research.

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

APPENDIX
Summary of Manipulations

	Study 1 Manipulations
High-recovery expectations	• Sign in scenario text promises a voucher for $150 off the purchase price of a round-trip ticket for delays of 2 hours or more if the airline is responsible for the delay
Low-recovery expectations	• Sign in scenario text declaims any responsibility for delayed or canceled flights
High-recovery performance	• Agent checks to see if passenger can be rebooked on another airline without prompting • Three apologies are offered • Passenger is offered a $150 ticket voucher, a meal voucher, use of a phone for local or long-distance calls, and use of lounge • Agent asks if there is anything else the passenger needs • Rebooks passenger on a later flight
Low-recovery performance	• Agent checks to see if passenger can be rebooked on another airline after being prompted • Two apologies are offered • With prompting offers use of phone and meal voucher, refuses $150 ticket voucher • Rebooks passenger on a later flight
	Study 2 Manipulations
High distributive justice	• Passenger is offered a $150 ticket voucher, a meal voucher, and use of phone for local or long-distance calls
Moderate distributive justice	• Passenger is offered meal voucher and use of phone for local and long-distance calls
Low distributive justice	• Passenger is refused meal voucher and use of phone
High interactional justice	• Agent checks to see if passenger can be rebooked on another airline without prompting • Three apologies are offered • Agent asks if there is anything else the passenger needs • Rebooks passenger on a later flight
Moderate interactional justice	• Agent checks to see if passenger can be rebooked on another airline after being prompted • Two apologies are offered • Rebooks passenger on a later flight
Low interactional justice	• Agent refuses to check to see if passenger can be rebooked on another flight, noting all future flights are booked • No apology is offered • Passenger is placed on standby (total length of the delay is the same as in the case of high and moderate interactional justice) • Agent is portrayed as in a hurry to move on to the next customer • The customer must go to the customer service desk for meal voucher and use of the phone (high and moderate distributive-justice conditions only)
	Studies 1 and 2 No-Failure Conditions
No-failure condition	• Scenario portrays an uneventful flight with no problems • Agent is polite and friendly • Customer arrives at final destination on time

REFERENCES

Abrams, Michael and Matthew Paese (1993), "Wining and Dining the Whiners: Treating Complaints as Opportunities Is One Customer Strategy That Really Works," *Sales and Marketing Management*, 145 (February), 72-75.

Anderson, James C. (1987), "An Approach for Confirmatory Measurement and Structural Equation Modeling of Organizational Properties," *Management Science*, 33 (4), 525-41.

———and David W. Gerbing (1988), "Structural Equations Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (3), 411-23.

Andreassen, Thor Wallin (1999), "What Drives Customer Loyalty with Complaint Resolution?" *Journal of Service Research*, 1 (May), 324-32.

Andreeva, Nellie (1998), "Unsnarling Traffic Jams at U.S. Airports," *Business Week*, August 10, 84.

Bagozzi, Richard P. and Youjae Yi (1988), "On the Evaluation of Structural Equation Models," *Journal of the Academy of Marketing Science*, 16 (Spring), 74-94.

Bearden, William O. and Jesse E. Teel (1983), "Selected Determinants of Consumer Satisfaction and Complaint Reports," *Journal of Marketing Research*, 20 (February), 21-28.

Bentler, Peter M. and Douglas G. Bonett (1980), "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures," *Psychological Bulletin*, 8 (3), 588-606.

Berry, Leonard L., Valarie A. Zeithaml, and A. Parasuraman (1990), "Five Imperatives for Improving Service Quality," *Sloan Management Review*, 31 (Summer), 29-39.

Bies, Robert J. and Joseph S. Moag (1986), "Interactional Justice: Communication Criteria of Fairness," in *Research on Negotiation in Organizations*, Vol. 1, R. Lewicki, M. Bazerman, and B. Sheppard, eds. Greenwich, CT: JAI, 57-79.

Bitner, Mary Jo (1990), "Evaluating Service Encounters: The Effects of Physical Surroundings and Employee Responses," *Journal of Marketing*, 54 (April), 69-82.

———, Bernard H. Booms, and Mary Stanfield Tetreault (1990), "The Service Encounter Diagnosing Favorable and Unfavorable Incidents," *Journal of Marketing*, 54 (January), 71-84.

Blodgett, Jeffrey G., Donald H. Granbois, and Rockney G. Walters (1993), "The Effect of Perceived Justice on Complainants' Negative

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

Word-of-Mouth Behavior and Repatronage Intentions," *Journal of Retailing*, 69 (Winter), 399-428.

Bolton, Ruth N. and James H. Drew (1991), "A Multistage Model of Customers' Assessments of Service Quality and Value," *Journal of Consumer Research*, 17 (March), 375-84.

Boshoff, Christo (1999), "RECOVSAT: An Instrument to Measure Satisfaction with Transaction-Specific Service Recovery," *Journal of Service Research*, 1 (February), 236-49.

Cohen, Ronald L. (1985), "Procedural Justice and Participation," *Human Relations*, 38 (July), 643-63.

Crosby, Phillip B. (1979), *Quality Is Free: The Art of Making Quality Certain*. New York: McGraw-Hill.

Deming, W. Edwards (1986), *Out of Crisis*. Cambridge: MIT Press.

Etzel, Michael J. and Bernard I. Silverman (1981), "A Managerial Perspective on Directions for Retail Customer Dissatisfaction Research," *Journal of Retailing*, 57 (Fall), 124-36.

Fisk, Raymond P., Stephen W. Brown, and Mary Jo Bitner (1993), "Tracking the Evolution of the Services Marketing Literature," *Journal of Retailing*, 69 (Spring), 61-103.

Folger, Robert and Mary A. Konovsky (1989), "Effects of Procedural and Distributive Justice on Reactions to Pay Raise Decisions," *Academy of Management Journal*, 32, 115-30.

———, David Rosenfield, Janet Grove, and Louise Corkran (1979), "Effects of 'Voice' and Peer Opinions on Responses to Inequity," *Journal of Personality and Social Psychology*, 37 (December), 2253-61.

Folkes, Valerie S. (1984), "Consumer Reactions to Product Failure: An Attributional Approach," *Journal of Consumer Research*, 10 (March), 398-409.

——— and Barbara Kotsos (1986), "Buyers' and Sellers' Explanations for Product Failure: Who Done It?" *Journal of Marketing*, 50 (April), 74-80.

Fornell, Claus (1992), "A National Customer Satisfaction Barometer: The Swedish Experience," *Journal of Marketing*, 56 (January), 6-21.

——— and David F. Larcker (1981), "Evaluating Structural Equation Models with Unobserved Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Goodwin, Cathy and Ivan Ross (1992), "Consumer Responses to Service Failures: Influence of Procedural and Interactional Fairness Perceptions," *Journal of Business Research*, 25 (September), 149-63.

Greenberg, Jerald (1987), "Reactions to Procedural Injustice in Payment Distributions: Do the Means Justify the Ends?" *Journal of Applied Psychology*, 72 (February), 55-61.

———(1990a), "Looking Fair vs. Being Fair: Managing Impressions of Organizational Justice," *Research in Organizational Behavior*, 12, 111-57.

———(1990b), "Organizational Justice: Yesterday, Today, and Tomorrow," *Journal of Management*, 16 (June), 399-432.

———and Robert Folger (1983), "Procedural Justice, Participation, and the Fair Process Effect in Groups and Organizations," in *Basic Group Processes*, P. B. Paulus, ed. New York: Springer-Verlag, 235-56.

Gronroos, Christian (1988), "Service Quality: The Six Criteria of Good Perceived Service Quality," *Review of Business*, 9 (Winter), 10-13.

Halstead, Diane and Thomas J. Page, Jr. (1992), "The Effects of Satisfaction and Complaining Behavior on Consumer Repurchase Intentions," *Journal of Consumer Satisfaction, Dissatisfaction, and Complaining Behavior*, 5, 1-11.

Hansen, David E. and Peter J. Danaher (1999), "Inconsistent Performance during the Service Encounter: What's a Good Start Worth?" *Journal of Service Research*, 1 (February), 227-35.

Hart, Christopher W. L. (1993), *Extraordinary Guarantees*. New York: American Management Association.

———, James L. Heskett, and W. Earl Sasser, Jr. (1990), "The Profitable Art of Service Recovery," *Harvard Business Review*, 68 (July-August), 148-56.

Hirschman, Albert O. (1970), *Exit, Voice, and Loyalty-Responses to Decline in Firms, Organizations, and States*. Cambridge, MA: Harvard University Press.

Hocutt, Mary Ann, Goutam Chakraborty, and John C. Mowen (1997a), "The Art of Service Recovery: Fact or Fiction? An Empirical Study of the Effects of Service Recovery," in *Marketing Theory and Applications*, Debbie Thorne LeClair and Michael Hartline, eds. Chicago: American Marketing Association, 50-51.

———, ———, and ———(1997b), "The Impact of Perceived Justice on Customer Satisfaction and Intention to Complain in a Service Recovery," in *Advances in Consumer Research*, Merrie Brucks, ed. Ann Arbor, MI: Association for Consumer Research, 457-63.

Jöreskog, Karl G. and Dag Sörbom (1993), *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Chicago: Scientific Software International.

Kelly, Scott W. and Mark A. Davis (1994), "Antecedents to Customer Expectations for Service Recovery," *Journal of Academy of Marketing Science*, 22 (Winter), 52-61.

———, K. Douglas Hoffman, and Mark A. Davis (1993), "A Typology of Retail Failures and Recoveries," *Journal of Retailing*, 69 (Winter), 429-52.

McCollough, Michael A. and Sundar G. Bharadwaj (1992), "The Recovery Paradox: An Examination of Consumer Satisfaction in Relation to Disconfirmation, Service Quality, and Attribution Based Theories," in *Marketing Theory and Applications*, Chris T. Allen, Thomas J. Madden, Terence A. Shimp, Roy D. Howell, George M. Zinkhan, Deborah D. Heisley, Richard J. Semenik, Peter Dickson, Valarie Zeithaml, and Roger L. Jenkins, eds. Chicago: American Marketing Association, 119.

McFarlin, Dean B. and Paul D. Sweeney (1992), "Distributive and Procedural Justice as Predictors of Satisfaction with Personal and Organizational Outcomes," *Academy of Management Journal*, 35 (August), 626-37.

Morris, Susan V. (1988), "How Many Lost Customers Have You Won Back Today? An Aggressive Approach to Complaint Handling in the Hotel Industry," *Journal of Consumer Satisfaction, Dissatisfaction, and Complaining Behavior*, 1, 86-92.

Murray, Keith B. and John L. Schlacter (1990), "The Impact of Services versus Goods on Consumers' Assessment of Perceived Risk and Variability," *Journal of the Academy of Marketing Science*, 18 (Winter), 51-65.

Oliver, Richard L. (1980), "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," *Journal of Marketing Research*, 17 (September), 46-49.

———(1981), "Measurement and Evaluation of Satisfaction Processes in Retail Settings," *Journal of Retailing*, 57 (Fall), 25-48.

——— (1989), "Processing of the Satisfaction Response in Consumption: A Suggested Framework and Research Propositions," *Journal of Consumer Satisfaction, Dissatisfaction, and Complaining Behavior*, 2, 1-16.

——— (1993), "Cognitive, Affective, and Attribute Bases of the Satisfaction Response," *Journal of Consumer Research*, 20 (December), 418-30.

———and William O. Bearden (1985), "Disconfirmation Processes and Consumer Evaluations in Product Usage," *Journal of Business Research*, 13 (June), 235-46.

———and Raymond R. Burke (1999), "Expectation Processes in Satisfaction Formation," *Journal of Service Research*, 1 (February), 196-214.

——— and John E. Swan (1989a), "Consumer Perceptions of Interpersonal Equity and Satisfaction in Transactions: A Field Survey Approach," *Journal of Marketing*, 53 (April), 21-35.

——— and ———(1989b), "Equity and Disconfirmation Perceptions as Influences on Merchant and Product Satisfaction, *Journal of Consumer Research*, 16 (December), 372-83.

Perdue, Barbara C. and John O. Summers (1986), "Checking the Success of Manipulations in Marketing Experiments," *Journal of Marketing Research*, 23 (November), 317-26.

Reicheld, Frederick F. and W. Earl Sasser, Jr. (1990), "Zero Defections: Quality Comes to Services," *Harvard Business Review*, 68 (September-October), 105-11.

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009

Richins, Marsha L. (1983), "Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study," *Journal of Marketing*, 47 (Winter), 68-78.

——— (1987), "A Multivariate Analysis of Responses to Dissatisfaction," *Journal of the Academy of Marketing Science*, 15 (Fall), 24-31.

Roos, Inger (1999), "Switching Processes in Customer Relationships," *Journal of Service Research*, 1 (August), 68-85.

Ross, William T., Jr. and Elizabeth H. Creyer (1993), "Interpreting Interactions: Raw Means or Residual Means?" *Journal of Consumer Research*, 20 (September), 330-38.

Singh, Jagdip (1990), "Voice, Exit, and Negative Word-of-Mouth Behaviors: An Investigation Across Three Service Categories," *Journal of the Academy of Marketing Science*, 18 (Winter), 1-15.

——— and Robert Widing II (1991), "What Occurs Once Consumers Complain? A Theoretical Model for Understanding Satisfaction/Dissatisfaction Outcomes of Complaint Responses," *European Journal of Marketing*, 25 (5), 30-46.

Smart, Denise T. and Charles L. Martin (1993), "Consumers Who Correspond with Business: A Profile and Measure of Satisfaction with Response," *Journal of Applied Business Research*, 9 (Spring), 30-42.

Smith, Amy K. and Ruth N. Bolton (1998), "An Experimental Investigation of Customer Reactions to Service Failure and Recovery Encounter: Paradox or Peril?"*Journal of Service Research*, 1 (August), 65-81.

Stauss, Bernd and Christian Friege (1999), "Regaining Service Customers," *Journal of Service Research*, 1 (May), 347-61.

Sundaram, D. S., Claudia Jurowski, and Cynthia Webster (1997), "Service Failure Recovery Efforts in Restaurant Dining: The Roles of Criticality of Service Consumption," *Hospitality Research Journal*, 20 (3), 137-49.

Swan, John E. and Frederick I. Trawick (1981), "Disconfirmation of Expectations and Satisfaction with a Retail Service," *Journal of Retailing*, 57 (Fall), 49-67.

TARP (1979), *Consumer Complaint Handling in America: Summary of Findings and Recommendations*. Washington, DC: U.S. Office of Consumer Affairs.

———(1986), *Consumer Complaint Handling in America: An Updated Study Part II*. Washington, DC: U.S. Office of Consumer Affairs.

Tax, Stephen S. and Stephen W. Brown (1998), "Recovering and Learning from Service Failure," *Sloan Management Review*, 39 (Fall), 75-88.

———, ———, and Murali Chandrashekaran (1998), "Customer Evaluations of Service Complaint Experiences: Implications for Relationship Marketing," *Journal of Marketing*, 62 (April), 60-76.

Triplett, Tim (1994), "Product Recall Spurs Company to Improve Customer Satisfaction," *Marketing News*, 28 (April 11), 6.

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

Webster, Cynthia and D. S. Sundaram (1998), "Service Consumption Criticality in Failure Recovery," *Journal of Business Research*, 41, 153-59.

Westbrook, Robert A. (1980), "Intrapersonal Affective Influences on Consumer Satisfaction with Products," *Journal of Consumer Research*, 7 (June), 49-54.

———(1981), "Sources of Consumer Satisfaction with Retail Outlets," *Journal of Retailing*, 57 (Fall), 68-85.

Yi, Youjae (1990), "A Critical Review of Consumer Satisfaction," in *Review of Marketing*, Valarie A. Zeithaml, ed. Chicago: American Marketing Association, 68-123.

Zeithaml, Valarie A., Leonard L. Berry, and A. Parasuraman (1996), "The Behavioral Consequences of Service Quality," *Journal of Marketing*, 60 (April), 31-46.

Michael A. McCollough is an assistant professor of marketing at the University of Idaho. He obtained his Ph.D. in marketing from Texas A&M University. His research interests are services marketing, services quality, service failure and recovery, service guarantees, and customer satisfaction and dissatisfaction. He has published in the *Journal of Hospitality and Tourism Research*, the *Journal of Marketing Education*, *Marketing Education Review*, the *Journal of Professional Services Marketing*, and *International Review of Retail, Distribution and Consumer Research*.

Leonard L. Berry holds the JCPenney Chair of Retailing Studies and is a distinguished professor of marketing at Texas A&M University. He is a former national president of the American Marketing Association and founder of Texas A&M University's Center for Retailing Studies. His research interests are services marketing, service quality, and retailing strategy. He has published numerous journal articles and books, including *Discovering the Soul of Service—The Nine Drivers of Sustainable Business Success* (Free Press, 1999), *On Great Service: A Framework for Action* (Free Press, 1995), *Marketing Services: Competing through Quality* (Free Press, 1991), and *Delivering Quality Service: Balancing Customer Perceptions and Expectations* (Free Press, 1990).

Manjit S. Yadav is an associate professor of marketing and Center for Retailing Studies Faculty Fellow, Texas A&M University. His current research focuses on strategic issues related to electronic commerce. His work has been published in a number of leading journals, including the *Journal of Marketing Research,* the *Journal of Consumer Research,* and *Sloan Management Review.* He is a member of the Editorial Review Board of the *Journal of the Academy of Marketing Science*. He developed and currently teaches a graduate course (Strategic Foundations of E-Commerce) and an undergraduate course (Fundamentals of Internet Marketing) dealing with the strategic challenges and opportunities in the emerging electronic marketplace.

Downloaded from http://jsr.sagepub.com at SAGE Publications on December 2, 2009