Title
Dual Rules for Service Evaluation

Journal
Service Science, 5(4)

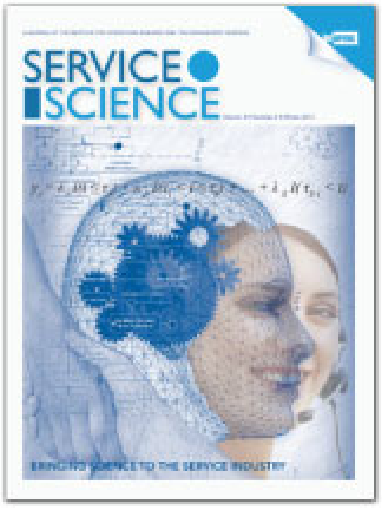
ISSN
2164-3962

Authors
Rolland, Erik
Patterson, Raymond A
Messinger, Paul R
et al.

Publication Date
2013-12-01

DOI
10.1287/serv.2013.0059

Peer reviewed



Service Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Dual Rules for Service Evaluation

Erik Rolland, Raymond A. Patterson, Paul R. Messinger, Keith F. Ward, Adam Finn

SERVICE SCIENCE
Vol. 5, No. 4, December 2013, pp. 279–295
ISSN 2164-3962 (print) | ISSN 2164-3970 (online)
http://dx.doi.org/10.1287/serv.2013.0059

Dual Rules for Service Evaluation

Erik Rolland
The Ernest and Julio Gallo Management Program, and School of Engineering, University of California, Merced, Merced, California 95343, erolland@ucmerced.edu

Raymond A. Patterson
Department of Accounting, Operations, and Information Systems, Alberta School of Business, University of Alberta, Edmonton, Alberta T6G 2R6, Canada, ray.patterson@ualberta.ca

Paul R. Messinger
Department of Marketing, Business Economics and Law, Alberta School of Business, University of Alberta, Edmonton, Alberta T6G 2R6, Canada, paul.messinger@ualberta.ca

Keith F. Ward
Department of Management, School of Management and Business, St. Edward's University, Austin, Texas 78704, keithfw@stedwards.edu

Adam Finn
Department of Marketing, Business Economics and Law, Alberta School of Business, University of Alberta, Edmonton, Alberta T6G 2R6, Canada, adam.finn@ualberta.ca

We find evidence that customers evaluate services according to different rules depending on whether the overall service experience impression is negative or positive. The functional form by which the underlying service attributes are mapped into an overall satisfaction evaluation is compensatory in the former (negative) case and has conjunctive properties in the latter (positive) case. These findings are based on an examination of a large data set of outpatient healthcare customer satisfaction surveys. We conjecture that different cognitive processes are at work for positive versus negative service encounters.

Key words: service analytics; service evaluation; customer satisfaction; dual processes; service management
History: Received January 18, 2013; Received in final revised form October 6, 2013; Accepted October 24, 2013 by Paul Maglio.

1. Introduction

A key to success for service organizations is sustainable relationships with customers. The quality of these relationships is often monitored with satisfaction surveys. Customer satisfaction, as measured by these surveys, has many ramifications for service organizations.

At the consumer level, customer satisfaction has been shown to help secure customer loyalty (Anderson and Sullivan 1993, Szymanski and Henard 2001) in various measurable dimensions, including positive behavioral intentions (e.g., Mittal and Kamakura 2001, Mittal et al. 1999), repeat purchases (Szymanski and Henard 2001), customer retention (Bolton 1998), high share of wallet (Cooil et al. 2007), and positive word of mouth (de Matos and Rossi 2008, Swan and Oliver 1989, Zeithaml et al. 1996).

At the firm level, customer satisfaction helps yield profitability (Anderson et al. 1994, Bernhardt et al. 2000), long-term financial performance (Mittal et al. 2005), and shareholder value through impact on future cash flows (Anderson et al. 2004) that exhibit increased growth and reduced variability (Gruca and Rego 2005). Given the importance of customer satisfaction at the consumer and firm levels, it is not surprising that service management researchers have explored how to measure it, as well as its antecedents and consequences.

The purpose of the current paper is to examine how people's perceptions of individual service features map into overall satisfaction assessments. In particular, we argue that a customer's general impression of a service experience, whether positive or negative, leads to different mappings from individual service features to overall customer satisfaction. Understanding this phenomenon, in context, is critical for management to correctly interpret the meaning of customer satisfaction survey results. Often, assessments of such results are a key input to important decisions involving commitment of resources to deliver service features.

Standard measures of customer satisfaction with a service consist of summary ratings of overall response, such as "overall satisfaction" or "overall quality of service" (e.g., from 1 to 5, where 1 represents an assessment of "poor," 3 represents "fair," and 5 represents "excellent"). In addition, ratings of relevant individual services features (on similar five- or seven-point scales) are also typically collected. The data set we analyze in this paper contains both types of measures. In addition, our data set includes answers to the question, "Would you recommend [this physician] to a friend?" with the physician's name being inserted for "[this physician]." We use this as an indicator of the patient's general impression of the service experience.

We estimate differing rules by which a service's individual features are combined to form a consumer's overall satisfaction perception. In particular, we identify two functional forms that map individual service feature assessments into overall satisfaction (or utility): one is compensatory and the other has conjunctive properties. As pointed out by Elrod et al. (2004), mathematical tractability in economic analysis is ensured by making simplifying assumptions related to the formation of the service utility. Often, the utilities of objects are modeled as being compensatory, where the valuation of every attribute affects the total utility of the object. However, compensatory utility forms may not be applicable for all situations.

Much research to date (that we discuss in the next section) has modeled a single rule (or functional form) that links overall customer satisfaction with the evaluations of individual service attributes. Instead, we demonstrate that two different evaluation rules may be triggered by the general impression of the service experience. If this can in fact be shown, then there are important and perhaps unexpected managerial implications for the strategic and tactical use of a firm's service evaluations.

In our statistical modeling, we estimate a two-equation recursive system. The first equation relates the indicator of a customer's general impression of the service experience with that customer's ratings of individual service features. The second equation relates the customer's assessment of overall quality of care with that customer's general impression of the service experience (the dependent variable in the first equation) and his or her ratings of individual service features (the same explanatory variables as in the first equation). To ascertain that there is no endogeneity problem that would bias our estimates, we conduct a conditional moment test that rules out correlation of the error terms in the two equations.

The rest of this paper is divided into six sections. Section 2 reviews relevant literature, §3 describes the hypotheses, §4 presents the data and analysis, and §5 discusses the potential mechanism underlying the divergence in evaluative rules for positive versus negative service encounters. Section 6 concludes and discusses the implications of this work.

2. Literature Review

The service management literature is extensive (for reviews, see Messinger et al. 2009, Rust and Chung 2006), and the topic has come to be recognized as one characterized by special management issues—what some authors have called "service-dominant logic," seen as distinct from the traditional "goods-dominant logic" (Vargo and Lusch 2004). Stemming from the work of Oliver (1980), one important development has been that of service quality scales, beginning with the prominent SERVQUAL scale (Parasuraman et al. 1988), followed by several similar scales (e.g., Cronin and Taylor 1992, Lytle et al. 1998, Parasuraman et al. 1991). Such scales have been applied in various areas, including physicians' services (Brown and Swartz 1989), information systems (Pitt et al. 1995), physical distribution (Bienstock et al. 1997), the banking industry (Bahia and Nantel 2000), and e-service (Parasuraman et al. 2005). To understand the determinants of overall service quality, a related stream of research has emerged linking overall customer satisfaction with measures of individual service attributes. We divide this research into three categories—linear models, nonlinear models, and models of customer heterogeneity—that use different functional forms for different customers. Our analysis goes a step beyond these approaches by describing how dual rules are applicable to a given customer's assessment depending on the circumstances (a positive versus a negative service encounter).

2.1. Linear Models

Examples of linear attribute formulations focusing on service performance dimensions are found in the work of Grönroos (1984) and Rust and Oliver (1994), and they can be expressed as

$$CS_{si} = \sum_{a=1}^{k} f_a(P_{sai}), \tag{1}$$

where CS_{si} is the customer satisfaction with service s for customer i, P_{sai} is the perceived performance with service s on dimension a for consumer i, and f_a is some functional form. This form is inherently linear, since the terms f_a (P_{sai}) are additive. Other examples are found across application areas such as software (Kekre et al. 1995), banking (Levesque and McDougall 1996), and online shopping (Szymanski and Hise 2000).

A second form of linear formulation, which originated in the expectancy disconfirmation model for consumer satisfaction (Oliver 1980), treated perceived performance and prior expectations or their discrepancies as a determinant of consumer satisfaction with products (Cardozo 1965). For services, a focus on the multivariate gap between the actual performance dimensions and the expected levels of performance on these dimensions was incorporated into a similar gap model (Parasuraman et al. 1985) that was operationalized as SERVQUAL (Parasuraman et al. 1988) and into subsequent work building on that formulation (e.g., Dabholkar et al. 2000). However, Cronin and Taylor (1992) found no evidence that an index that accounts for expectations provides a more valid measure; this conclusion was confirmed in a meta-analysis of 17 cross-sectional studies (Carrillat et al. 2007). Consistent with this evidence, the current paper builds on the simpler form that does not include expectations of performance.

2.2. Nonlinear Models

Many authors have suggested that nonlinear models may better describe customer satisfaction. Early work on nonlinear functional forms goes back to Einhorn (1970) (also see Einhorn 1971), who considered both parabolic and hyperbolic models used to detect nonlinear, noncompensatory judgment strategies. Other authors have investigated various discontinuous integrative hierarchical processes, such as differentiating between factors that contribute to individual satisfaction versus dissatisfaction (Herzberg et al. 1959), important factors versus nonimportant factors, basic versus higher needs, and various hybrid approaches (Kano et al. 1984, Swan and Combs 1976; for research overviews, see Yi 1990 and Chapter 5 of Oliver 2010).

An important class of discontinuous integrative decision rules includes conjunctive, disjunctive, and lexicographic rules (Coombs 1964, Coombs and Kao 1955, Dawes 1964; for overviews, see Einhorn 1970, Elrod et al. 2004). In general, a conjunctive rule requires an alternative to exceed acceptable thresholds on all relevant attributes for it to be chosen. A disjunctive rule requires an alternative to exceed a desired threshold on any attribute for it to be chosen. A lexicographic rule evaluates a set of alternatives by performance on attributes in order of the importance of the attributes. In contrast with the discontinuous rules mentioned above, a linear compensatory utility function is one where utility is a linear function of the service attributes (with constant coefficients on the attributes). Generally speaking, the indifference curves for such utility functions are downward sloping (if one is considering the trade-off among service attributes for which higher levels of the attributes are more attractive). By contrast, a fully conjunctive utility function is one where the overall utility of an alternative is equal to the smallest attribute level of all the service attributes of the alternative, and the resulting maps of the indifference curves are L-shaped. In this paper, we focus on evaluative rules with compensatory versus conjunctive features.

2.3. Heterogeneity Across Consumers in a Functional Form

Prior work has modeled consumer heterogeneity in the application of these rules. In the choice modeling literature, Gilbride and Allenby (2004) applied hierarchical Bayesian methods to empirically examine heterogeneity across consumers in their choice-set formation processes, allowing for conjunctive, disjunctive, and compensatory screening rules. Specifically for services marketing, Büschken et al. (2011) proposed that some consumers use compensatory rules when engaging in service evaluation, whereas other consumers use conjunctive rules. They call the first type of consumers “formators” and the second type “halos.” Formators first evaluate a service provider’s performance on the attributes of a service and then integrate these individual attribute evaluations in a compensatory fashion into a global evaluation, consistent with a formulation as shown in Equation (1). By contrast, halos first form a global evaluation of their satisfaction based on a limited number of attributes and then use this global evaluation as an anchor when generating (biased) answers to survey questions about particular individual attributes. For recent development of a related model that endogenously determines dimensionality of customer service drivers, see Büschken et al. (2013). In the context of halo effects, it has been observed that reverse causality may make the apparent relationships between the customers’ attribute perceptions and their overall satisfaction misleading to evaluators (Cooper 1981, Wirtz 2003).

Other work has considered different functional forms for different demographic groups or different circumstances. In the medical field, for example, Otani and Harris (2004) compared the applicability of conjunctive judgment rules versus other forms (disjunctive or compensatory) for four demographic groups (white males, white females, black males, and black females). Alternatively, as circumstances vary, Ganzach (1995, p. 497) noted that different nonlinear judgment rules may apply: “[I]t is important to note that this basic tendency [for performance evaluations to be conjunctive] may be mitigated by a number of factors, such as whether the evaluated object is human or nonhuman (e.g., professor or course) and whether the general evaluation of the object is positive or negative (e.g., good or bad professor; see Ganzach 1993 for other examples).” We will explicitly build on this last approach.

3. Hypotheses Construction

As with this last approach discussed above, we also argue that two types of evaluative rules apply. In particular, we suggest that the different evaluative rules apply for positive and negative service encounters. To develop this idea further, we define a customer's *evaluative mode* as a frame of reference (or perceptual perspective evoked by situational factors) that shapes how a customer perceives individual service features and how these perceptions combine to form an overall satisfaction assessment. We utilize the following notation to describe the customer's evaluative mode when answering a satisfaction survey:

Type P will refer to a positive evaluative mode arising from a positive service encounter.

Type N will refer to a negative evaluative mode arising from a negative service encounter.

3.1. Conjunctive vs. Compensatory Judgment Rules

We argue that if the customer has a positive evaluative mode, then there may be no need to seriously consider each of the attributes of the service when evaluating it, and thus low mental effort is required. This conceptualization is consistent with the Payne et al. (1993) discussion of choice based on utilization of a subset of the available alternatives. Decision making based on simple heuristics uses only a few attributes (often in a noncompensatory fashion) and involves less mental effort than full-information processing. Similar arguments are offered by Elrod et al. (2004), who suggested that low mental effort often invokes a noncompensatory decision-making process and that high mental effort invokes a compensatory decision-making process. Whether decision making is noncompensatory, compensatory, or something in between is governed by the shape of the applicable utility functions and associated indifference curves. Elrod et al. distinguished between three cases:

1. A "crisp" conjunctive utility function would be described by indifference curves of a discontinuous function that are L-shaped.
2. A "less than crisp," pervasive, two-dimensional conjunctive utility function would be described by indifference curves that are convex to the origin.
3. A linear compensatory assessment rule would form straight linear indifference curves of overall satisfaction.

We focus on the distinction between a linear compensatory functional form for utility (case 3) and some form of conjunctive utility (including cases 1 and 2).

In this context, our main contention is that a customer in a negative evaluative mode is unhappy with the service provider and more willing to expend mental effort to assess the trade-offs between the current service provider's strong and weak points—possibly as a first step toward finding a remedy for the situation. As noted above, a linear compensatory utility formulation involves explicit (and effortful) recognition of the trade-offs between various service features (and the linear form, in particular, is tractable). By contrast, a customer in a positive evaluative mode is happy with the current service provider, less in need of finding a new provider, and not particularly desirous of engaging in the mentally effortful process of understanding "what went wrong." One formulation consistent with this perspective of less discerning service–feature trade-offs, and less effortful evaluation, is a conjunctive assessment rule. This discussion motivates our first hypothesis.

Hypothesis 1A (H1A). *When in a type P evaluative mode, service consumers use a conjunctive judgment rule to form an overall satisfaction score.*

Hypothesis 1B (H1B). *When in a type N evaluative mode, service consumers use a linear compensatory judgment rule to form an overall satisfaction score.*

These hypotheses are premised on a view of consumers who rationally allocate greater cognitive effort (in assessing service satisfaction) to domains that can lead to greater benefits. Evidence of H1A would be established by utility contour maps that are L-shaped or convex to the origin. Evidence of H1B would be established by linear utility contour maps.

3.2. Halo Effects

Ganzach (1993, 1994, 1995) examined data that include objective attribute-specific measures and argued that the weights on each of the attributes are influenced by the halo effect in such a way as to be consistent with the overall satisfaction measure. This idea is consistent with the possibility that the functional form for utility may also change for an individual with a halo effect. Wirtz and Bateson (1995, p. 99), in a study of customer satisfaction with an online banking system, found that "halo effects can contaminate attribute-specific satisfaction measures" and that "attribute performance levels of a service can be obscured by halo effects." Büschken et al. (2011, 2013) made similar arguments. These findings suggest that there is a positive shift in consumer evaluations of individual service attributes when consumers are in a positive evaluative mode (type P process) relative to when consumers are in a negative evaluative mode (type N process).[1]

Various forms of evidence can support the existence of a halo effect. First, some support would be indicated if the mean individual attribute scores are consistently inflated for type P processes when compared with type N processes. Second, if H1 is true, and there is a halo effect, then one would expect both a relative shift downward in the utility function for a type N evaluative mode and a more linearly compensatory shape of the utility function.

3.3. Operationalizing Type P and Type N Evaluative Modes

As we mentioned in §1, our data set includes the following question: "Would you recommend [this physician] to a friend?" The use of this type of question as a predictor of customer repurchase intention or future firm performance has been debated in the marketing literature. Managers have been told that the aggregated response to this question in the form of a Net Promoter Score is the most important number they need to know to grow their businesses (Reichheld 2003). The academic literature has discredited this simplistic claim (e.g., Keiningham et al. 2007, Morgan and Rego 2006, Sharp 2008); however, at an individual respondent level, a positive response to this recommendation question does seem indicative of a positive experience with the service encounter. The satisfaction survey that generated the data used for the current paper included a dichotomous version of this recommendation question.

We operationalize a type P evaluative mode as present when a patient answers this question in the affirmative and a type N evaluative mode when a patient answers this question in the negative (not willing to recommend his or her physician to a friend). We acknowledge that this question is a proxy for the evaluative mode, which, more ideally, would have been measured during the service encounter, with further process measures regarding the customer's state of mind when answering the survey. Nevertheless, we emphasize that we do not rely on this variable as a predictor of repurchase intention or future firm performance, which is the interpretation of this measure that had been under contention in the literature. It is worth noting that the small proportion of respondents who would not recommend contains a disproportionately high share of the respondents who had a negative reaction to their service encounter. Because this willingness to recommend variable is only a binary proxy, it will misclassify some respondents. As a result, regression or similar models using this proxy as an independent variable will give parameters for the proxy and its interactions that are attenuated (i.e., biased toward zero) compared with those that would have been obtained for an error-free measure (Aigner 1973). Thus, tests using this willingness to recommend as an imperfect proxy for the evaluative mode are conservative. Furthermore, the worse the degree of misclassification, the more conservative these tests become. It is also important to note that this willingness to recommend variable is endogenous to the process of answering the satisfaction survey and should be accounted for as such in our modeling.

4. Description of Data and Analysis

The above hypotheses are examined by analyzing satisfaction data for primary care physicians. The data were collected from outpatients who received treatment during an eight-year period, covering 222 doctors representing 5 medical practices distributed over 12 different clinics in California. The data set contains 41,085 usable and complete responses. Of the 29 original questions, 15 were used in this paper. The questionnaire included measures of patients' assessment of service attributes on a scale from 1 to 5, where 1 represents "poor," 3 represents "fair," and 5 represents "excellent." Other questionnaire items included, "Would you recommend [this physician] to a friend?" (coded as $RF = 1$ as a proxy for type P and $RF = 0$ as a proxy for type N as an indicator of a positive or negative evaluative mode); "Overall quality of care," used to measure overall customer satisfaction (CS); and "Health worse now than a year ago," used to measure the health condition (HC).

4.1. Overview of Data

As a background, Table 1 reports the mean score for each variable while controlling for RF. We observe that the averages for customer satisfaction are significantly higher (4.53) when patients report that they would recommend the physician to a friend but are lower (3.45) when they would not recommend the physician to a friend. The table also demonstrates that patients who would provide positive recommendations also report significantly higher scores on all service attributes measured than the patients who would not recommend. This effect can be described as a halo effect, and it is correlational, not causal.

[1] In this case, a positive shift (or bias) refers to an evaluation where response values are positively inflated with respect to another value that could be said to have a negative shift. This is a relative comparison, as both assessments may be individually accurate but different.

Table 1. Responses from Patients Who "Recommend" and "Would Not Recommend"

Question	*RF*	Mean	SE	*t*-Statistic
Overall quality of care (i.e., customer satisfaction; *CS*)	0	3.45	0.027	−62.91
	1	4.53	0.003	
Ability to get appointment in nonemergency (*A*)	0	3.03	0.032	−46.75
	1	4.21	0.005	
Time between appointment and visit (*B*)	0	3.01	0.031	−42.95
	1	4.11	0.005	
Time spent waiting in the reception area (*C*)	0	3.22	0.031	−39.76
	1	4.17	0.005	
Time spent waiting in the exam area (*D*)	0	3.20	0.030	−47.41
	1	4.25	0.004	
Doctor's personal interest in you and your medical problems (*E*)	0	2.73	0.031	−134.09
	1	4.73	0.003	
Thoroughness of your examination (*F*)	0	2.85	0.030	−116.57
	1	4.69	0.003	
Doctor's explanation of treatment options (*G*)	0	2.73	0.032	−121.31
	1	4.69	0.003	
Explanation of tests and procedures (*H*)	0	2.83	0.031	−109.78
	1	4.66	0.003	
Doctor's explanation of prescribed medicine (*I*)	0	2.82	0.032	−107.65
	1	4.64	0.003	
Accuracy of the diagnosis (*J*)	0	2.98	0.034	−96.52
	1	4.64	0.003	
Physician's explanation for referrals to other physicians and/or practitioners (*K*)	0	2.92	0.033	−98.40
	1	4.63	0.003	
Time spent with doctor during visit (*L*)	0	2.69	0.031	−100.05
	1	4.56	0.004	
Health worse now than a year ago (i.e., health condition; *HC*)[a]	0	3.11	0.027	2.77
	1	3.03	0.006	

Notes. The sample size when $RF = 0$ is 1,606 for all variables (except *HC*, which is 1,549); when $RF = 1$, the sample size is 39,479 (except *HC*, which is 38,550).

[a]Note that *HC* is coded based on responses to the following question: "Compared to one year ago, how would you rate your health in general now?" (1: much better than one year ago; 2: somewhat better than one year ago, 3: about the same as one year ago, 4: somewhat worse than one year ago, 5: much worse than one year ago).

Note that the mean difference for *HC* in Table 1 ("Health worse now than a year ago") between the two referral groups is greater than zero, with a *t*-statistic of 2.77. In particular, the average health condition score is slightly better for patients who recommend (3.03) versus those who do not (3.11).

4.2. Constructing Two Principal Components

In pursuing the objective of investigating nonlinearity in the satisfaction data, it will be useful to be parsimonious with the number of parameters and variables in our formulations. In particular, note that the 12 service attributes reported in Table 1 can be accounted for by using principal components associated with the entire sample (see Table 2).

Principal component analysis (PCA) is suitable as a dimensionality reduction technique in cases such as this, where many service attributes are correlated with others along a particular axis (see Suhr 2005 for support that PCA applies in cases such as ours). This is distinct from typical applications of factor analysis, where there might be many variables that are not so clearly correlated with each other along particular axes but can be explained theoretically by a limited number of latent factors, with a range of different loadings on these latent factors. This latter scenario does not appear to apply in our data set.

The data set exhibits strong loadings on the first two orthogonal components: the first is related to the provider (i.e., physician) and the second to other experiences in the outpatient office. Accordingly, we refer to the first

Table 2. Principal Component Analysis: Component Matrix

Question	Component	
	C1 weights	*C2* weights
Ability to get appointment in nonemergency (A)	0.221	**0.799**
Time between appointment and visit (B)	0.200	**0.822**
Time spent waiting in the reception area (C)	0.213	**0.783**
Time spent waiting in the exam area (D)	0.299	**0.753**
Doctor's personal interest in you and your medical problems (E)	**0.862**	0.228
Thoroughness of your examination (F)	**0.866**	0.253
Doctor's explanation of treatment options (G)	**0.899**	0.219
Explanation of tests and procedures (H)	**0.883**	0.243
Doctor's explanation of prescribed medicine (I)	**0.866**	0.246
Accuracy of the diagnosis (J)	**0.791**	0.221
Physician's explanation for referrals to other physicians and/or practitioners (K)	**0.809**	0.260
Time spent with doctor during visit (L)	**0.783**	0.340

Notes. Extraction method: PCA. Rotation method: Equamax with Kaiser normalization. *C1*, physician quality component; *C2*, office quality component. Weights above 0.750 are shown in bold.

Table 3. Cronbach's α for Implied Construct Scales for the Entire Data Set

Sample	N	Cronbach's α	
		Physician quality	Office quality
Entire data set	41,085	0.958	0.841
Would recommend ($RF = 1$)	39,479	0.942	0.830
Would not recommend ($RF = 0$)	1,606	0.938	0.825

component as physician quality (*C1*) and the second component as office quality (*C2*). These two components were selected using the Kaiser criterion, where only components with eigenvalues greater than 1 are selected. Two additional principal component analyses were examined, one for the portion of the data set containing positive recommendations ($RF = 1$) and another for the portion containing negative recommendations ($RF = 0$). Approximately the same loadings for both parts as for the whole sample (shown in Table 2) were found. Values are not reported because they essentially replicate those in Table 2.

If the first four items loading heavily on the office quality component and the last eight items loading heavily on the physician quality component were treated as scales measuring two constructs, the Cronbach's α for both components in the entire data set would exceed 0.80, the widely accepted threshold for a desirable level of scale reliability. This is shown in Table 3. The same is also true for those respondents making a positive recommendation ($RF = 1$) and a negative recommendation ($RF = 0$).

For parsimony, we focus on models that relate overall customer satisfaction to the two principal components above. Accordingly, two component scores were created using the loadings from Table 2 as weights with the associated variables A–L (in Table 1). To maintain interpretability of the component scores as ranging from 1 to 5, we normalized the weights for a given principal component by dividing each loading by the sum of the loadings for that component.[2] These two underlying service components, *C1* and *C2*, descriptive of the care that patients perceive, will be used in our analysis.

[2] To be concrete, we accordingly calculated the two components as follows:

$$\begin{aligned} C1 &= (0.221 \times A + 0.200 \times B + 0.213 \times C + 0.299 \times D + 0.862 \times E + 0.866 \times F + 0.899 \times G + 0.883 \times H \\ &\quad + 0.866 \times I + 0.791 \times J + 0.809 \times K + 0.783 \times L)/7.692; \\ C2 &= (0.799 \times A + 0.822 \times B + 0.783 \times C + 0.753 \times D + 0.228 \times E + 0.253 \times F + 0.219 \times G + 0.243 \times H \\ &\quad + 0.246 \times I + 0.221 \times J + 0.260 \times K + 0.340 \times L)/5.167. \end{aligned}$$

4.3. Estimating Dual Rules: Methodology

We examine the relationship between underlying service components, $C1_{ijt}$ and $C2_{ijt}$, and overall customer satisfaction, CS_{ijt}, for doctor i, patient j, and time t. Our goal is to understand whether different rules apply when customers are in a type P or type N evaluative mode as proxied by the patient's willingness to recommend the physician, RF_{ijt} ("Would you recommend [this physician] to a friend?"). (One limitation of these data is that the patients' responses are anonymous and not identified by patient across time, so that we cannot use some panel methods or even patient-specific intercepts; this is a property of a number of satisfaction surveys, as with, for example, academic instructor evaluations.)

A challenge for our estimation is that our proxy for evaluative mode RF_{ijt} is endogenous to the survey response process, and may itself be influenced by the underlying service components $C1_{ijt}$ and $C2_{ijt}$. To recognize this endogeneity (and avoid introducing bias into our estimation), we explicitly model both RF_{ijt} and CS_{ijt} within the following system:

$$RF_{ijt} = \begin{cases} 1 & \text{if} \quad y^*_{1,\,ijt} \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$y^*_{1,\,ijt} = \alpha_0 + \alpha_1 C1_{ijt} + \alpha_2 C2_{ijt} + \alpha_3 C1_{ijt} C2_{ijt} + u^1_{ijt}; \tag{2}$$

$$CS_{ijt} = k \quad \text{if} \qquad l_k \geq y^*_{2,\,ijt} > l_{k-1},$$

$$y^*_{2,\,ijt} = \beta_1 C1_{ijt} + \beta_2 C2_{ijt} + \beta_3 C1_{ijt} C2_{ijt} + (1 - RF_{ijt})(\beta_4 + \beta_5 C1_{ijt} + \beta_6 C2_{ijt} + \beta_7 C1_{ijt} C2_{ijt}) + u^2_{ijt}. \tag{3}$$

Here, $y^*_{1,\,ijt}$ and $y^*_{2,\,ijt}$ are latent dependent variables determined by their respective predictive equations, and the associated observed (limited) dependent variables are RF_{ijt} and CS_{ijt}. In this system, the errors are assumed identically, independently normally distributed:

$$\begin{pmatrix} u^1_{ijt} \\ u^2_{ijt} \end{pmatrix} \sim \text{IIDN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right). \tag{4}$$

Note that the threshold parameters, l_k, $k = 1, \ldots, 4$, are also to be estimated, and there is no intercept β_0 in Equation (2) because it is subsumed in these threshold parameters (for completeness, we set $l_0 \equiv -\infty$ and $l_5 \equiv \infty$). Given these assumptions, (2) describes a probit model and (3) describes an ordered probit model.

We observe that the two-equation system (2) and (3) constitutes a recursive probit system. This system is recursive, by definition, because RF_{ijt}, the dependent variable in (2), enters (3) as an independent variable, but CS_{ijt}, the dependent variable in (3), does not enter (2) as an independent variable. It is well established that linear recursive systems (with continuous dependent variables) can be consistently estimated with ordinary least squares, applied equation by equation, provided that $\rho = 0$ (Fisher 1966, Wold 1960; for a simple explanation, see Steward and Wallis 1981, pp. 266–299). Such an approach would have been applicable for estimation if the (continuous) dependent variables $y^*_{1,\,ijt}$ and $y^*_{2,\,ijt}$ were directly observed. When such continuous dependent variables are latent (as in Equations (2) and (3)), it has come to be realized that a recursive probit model can also now be estimated, equation by equation. This fact was first observed by Heckman (1978), formally proven for a simple linear case by Wilde (2000), and further demonstrated by Monfardini and Radice (2008) for the case similar to our model, where there is an interaction between the endogenous dummy from the first equation and exogenous regressors in the second equation.

These latter authors, however, provide a critical qualification to this approach of estimating the two probit models separately. After estimation, a test must be run to ensure that the errors in the two equations are not correlated (i.e., in our case, that $\rho = 0$); otherwise, endogeneity bias in estimation is induced through the error terms. Therefore, after estimating (2) and (3), we will test whether $\rho = 0$ to check for the applicability of our estimation approach.[3]

[3] As a modeling strategy, we find it convenient to use probit rather than logit models (i.e., we assume normally distributed error terms in (4)) because this allows us to parameterize the covariance between the error terms in (2) and (3) simply as ρ in (4)—which is a familiar formulation. This allows us to set up a relatively straightforward test of $\rho = 0$. And establishing that (we cannot reject) $\rho = 0$ is required for a recursive system to be estimable equation by equation without biased estimates for the coefficients. Being able to estimate the model equation by equation allows us to use available software for estimating probit and ordered probit models, rather than having to adapt software to estimate the two-equation system (2) and (3) jointly. Incidentally, we did estimate logit and ordered logit models for equations analogous to (2) and (3) (but with standard Type I extreme value distributions for the error terms). The relative coefficient estimates were very similar to the probit and ordered probit models that we discuss in this paper.

4.4. Estimates of the Two-Equation System

We experimented with linear predictors of $y^*_{1,ijt}$ and $y^*_{2,ijt}$ and other functional forms on a calibration sample before arriving at our final specifications. For brevity, we report the final model using Equations (2) and (3) estimated on the whole sample in Table 4. The appendix provides a verification of the results for the whole sample and also shows the results for our calibration and holdout samples (which were determined by a random split of the whole sample into two parts of roughly equal size). The estimates for the calibration and holdout samples matched each other, as well as the entire sample, very closely, with nearly identical test statistics. Table A.1 in the appendix provides results for Equation (2), and Table A.2 in provides results for Equation (3).

In the context of our model, knowing whether the first equation is above the threshold indicates whether patients are in a type P or type N evaluative mode. The only significant driver of evaluative mode in the estimated Equation (2) is *C1*, which we described as a component descriptive of the physician quality (arrived at using principal component analysis; see Table 2). *C2*, a component descriptive of the office quality, is not statistically significant (at the 0.05 level). In the estimated Equation (3), we see that all the estimated coefficients that describe the difference between the two rules for service evaluation, $\hat{\beta}_4$ (coefficient of $(1-RF)$), $\hat{\beta}_5$ (coefficient of $(1-RF)\times C1$), $\hat{\beta}_6$ (coefficient of $(1-RF)\times C2$), and $\hat{\beta}_7$ (coefficient of $(1-RF)\times C1\times C2$), are significantly different from 0. This indicates that there are indeed two significantly different rules for service evaluation. We will discuss this further after testing for exogeneity.

4.5. Testing for Exogeneity

The conditional moment test has been shown to be a test of $\rho = 0$ with good performance characteristics in simulation studies, relative to several other possible tests (Monfardini and Radice 2008). This test has the practical advantage that it can be carried out from separate probit and ordered probit estimates of (2) and (3). We accordingly estimated Equations (2) and (3) separately using available software. Then we carried out a conditional moment test as described below.

The conditional moment test statistic is

$$\hat{\tau} = \frac{1}{N}\sum_{i,j,t}\hat{u}_{1,ijt}\hat{u}_{2,ijt}, \tag{5}$$

which is asymptotically distributed as N(0, 1). This test is constructed from pseudo-residuals defined generally as $\hat{u}_{m,ijt} \equiv \hat{E}(u_{m,ijt} \mid \text{estimated model parameters})$. Although past applications of this statistic of which we are

Table 4. The Estimated System of Equations

Equation (2)			Equation (3)		
Parameter	Estimate	Sig.	Parameter	Estimate	Sig.
Threshold parameters			Threshold parameters		
$[-\hat{\alpha}_0]$	3.1670	0.000	$[\hat{l}_1 = 1]$	−3.2738	0.000
			$[\hat{l}_2 = 2]$	−2.8543	0.000
			$[\hat{l}_3 = 3]$	−1.6347	0.000
			$[\hat{l}_4 = 4]$	−0.0286	0.903
Location parameters			Location parameters		
			$(1-RF)$	−3.4290	0.000
C1	1.4463	0.000	*C1*	−0.8350	0.000
C2	−0.1484	0.086	*C2*	−0.8240	0.000
			$(1-RF)\times C1$	0.7181	0.000
			$(1-RF)\times C2$	1.2976	0.000
$C1\times C2$	−0.0184	0.359	$C1\times C2$	0.3818	0.000
			$(1-RF)\times C1\times C2$	−0.2876	0.000
Pseudo-R^2					
Nagelkerke	0.515			0.431	
McFadden	0.473			0.243	
Observations	41,085			41,085	
CM statistic				−0.0724	0.7958

Note. Equation (2) is estimated using the standard (dichotomous) probit model.

aware have been applied to contexts of two simple probit models in a system, we derive and write the pseudo-residuals for the more general probit case, which can include ordinal probit in one or both equations. We have

$$\hat{u}_{m,ijt} \equiv \hat{E}(u_{m,ijt} \mid \hat{l}_{y_{m,ijt}}, \hat{f}_{m,ijt}) = \frac{\phi(\hat{l}_{y_{m,ijt}-1} - \hat{f}_{m,ijt}) - \phi(\hat{l}_{y_{m,ijt}} - \hat{f}_{m,ijt})}{\Phi(\hat{l}_{y_{m,ijt}} - \hat{f}_{m,ijt}) - \Phi(\hat{l}_{y_{m,ijt}-1} - \hat{f}_{m,ijt})}. \tag{6}$$

Here, ϕ and Φ are the probability density and cumulative distribution functions, respectively, for the N(0, 1) distribution. Note that this formulation reduces to the simpler formulation (given in Monfardini and Radice 2008) applicable when both equations are dichotomous (not-ordered) probit relationships.

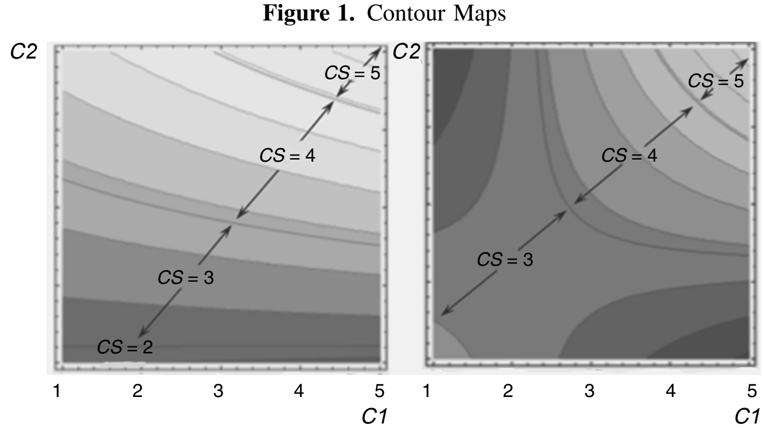
For our application of this test, for equation $m = 1$, the estimated thresholds reduce to l_k, $k = -1, 0, 1$, where $\hat{l}_0 = -\hat{\alpha}_0$ (and $\hat{l}_{-1} \equiv -\infty$ and $\hat{l}_1 \equiv \infty$). The estimated predictors are $\hat{f}_{1,ijt} \equiv \hat{f}_1(C1_{ijt}, C2_{ijt}) = \hat{\alpha}_0 + \hat{\alpha}_1 C1_{ijt} + \hat{\alpha}_2 C2_{ijt} + \hat{\alpha}_3 C1_{ijt} C2_{ijt}$. And for equation $m = 2$, the estimated thresholds are $\hat{l}_k$, $k = 0, \ldots, 5$ (where $\hat{l}_0 \equiv -\infty$ and $\hat{l}_5 \equiv \infty$ for completeness). Here, the estimated predictors are $\hat{f}_{2,ijt} \equiv \hat{f}_2(C1, C2, RF) = \hat{\beta}_1 C1_{ijt} + \hat{\beta}_2 C2_{ijt} + \hat{\beta}_3 C1_{ijt} C2_{ijt} + (1 - RF_{ijt})(\hat{\beta}_4 + \hat{\beta}_5 C1_{ijt} + \hat{\beta}_6 C2_{ijt} + \hat{\beta}_7 C1_{ijt} C2_{ijt})$. Using the estimated Equations (2) and (3) shown in Table 4, we calculate the conditional moment test to be $\hat{\tau} = -0.0724$. This statistic is approximately N(0, 1) (with a large sample size of 41,085 observations). This test is not significant at the 0.1 or 0.05 level (or even the 0.2 level). Therefore we do not reject the hypothesis of $\rho = 0$, and there is no endogeneity bias introduced by the inclusion of RF_{ijt} on the right-hand side of Equation (3).

4.6. Discussion of Our Results and Hypotheses

We now provide graphical representations and further interpretation of the results of Table 4. Figure 1 depicts the contour map for the negative ($RF = 0$) and positive ($RF = 1$) recommendation groups from Table 4. For respondents who would recommend their physician to a friend, the functional form is strongly conjunctive. However, for respondents who would not recommend, the functional form is a compensatory model that tends to be only somewhat conjunctive. Overall, Table 4 provides statistical evidence that different functional forms are at work, and Figure 1 provides visual evidence of the same. In particular, the type *P* (positive) evaluative model (when $RF = 1$) is characterized by a larger interaction term and a higher intercept than that of the type *N* (negative) evaluative model (when $RF = 0$).

Related to H1A and H1B, the larger interaction term in Table 4 is consistent with a conjunctive rule for combining two components such as *C1* and *C2*. Conjunctive rules are often applied strictly to combining two dichotomous variables, whereby the two variables are joined by a logical "and" operator. The analog of this for two continuous variables would be a multiplicative interaction term. In our case, *C1* and *C2* (with the myriad ordinal combinations of the individual attributes) are approximately continuous. Thus, the interaction term for type *P* is evidence of a conjunctive utility function. This corresponds with indifference curves characterized by the convex shape toward the origin (as seen in Figure 1). This is strong support for H1A. The indifference curves in the left panel (type *N*) are mostly linear and characterize a linear compensatory utility function. In terms of measured statistics, there is a very little interaction effect between *C1* and *C2* for type *N* (see Table 4). In addition, the significance is high for all variables. Therefore, based on the above statistical tests, support for H1B is also found.

Figure 1. Contour Maps



Would not recommend ($RF = 0$)
compensatory functional form

Would recommend ($RF = 1$)
conjunctive functional form

Table 5. Responses According to Straight-lining and Process Type

	Total count		Straight-lining (all attributes $= CS$)		Not straight-lining (not all attributes $= CS$)	
CS score	Type *P*	Type *N*	Type *P*	Type *N*	Type *P*	Type *N*
5	24,104	290	8,674 (36.0%)	34 (11.7%)	15,430 (64.0%)	256 (88.3%)
4	12,673	499	725 (5.7%)	4 (0.8%)	11,948 (94.3%)	495 (99.2%)
3	2,425	573	90 (3.7%)	11 (1.9%)	2,335 (96.3%)	562 (98.1%)
2	174	126	1 (0.6%)	0 (0.0%)	173 (99.4%)	126 (100%)
1	103	118	5 (4.9%)	13 (11.0%)	98 (95.1%)	105 (89.0%)
Total	39,479	1,606	9,495 (24.1%)	62 (3.9%)	29,984 (75.9%)	1,544 (96.1%)

A higher intercept term is found to be applicable when patients recommend their physician to a friend (the intercept is 0 when $RF = 1$ and -3.4290 when $RF = 0$ in Table 4). This is consistent with a halo effect and accounts for the generally high overall satisfaction measures ($CS = 3$, 4, or 5) for all values of *C1* and *C2* in the right panel of Figure 1. The lower intercept applicable when patients do not recommend accounts for generally lower overall satisfaction measures ($CS = 2$, 3, 4, or 5) for all values of *C1* and *C2* in the left panel of Figure 1.

Overall, Table 4 supports H1A and H1B in that there are two functional forms for type *P* and *N* processes. In particular, Table 4 and Figure 1 are supportive of a "pervasive" conjunctive utility function (Elrod et al. 2004) for the type *P* process and a linear compensatory utility function for the type *N* process. As we discuss in §5, this leads to our conjecture that there may be two mental processes in play when an individual is asked to evaluate a service.

4.7. Additional Evidence of Dual Mental Processes

To further explore the implications of the evaluative modes (types *P* and *N*), we construct a contingency table (see Table 5). In this table, we interpret the cases where patients score all attributes (and the overall satisfaction score) at the same level as indicative that few cognitive resources are employed. These patients appear to be responding in an automatic fashion by marking all attributes the same as their overall satisfaction score. If customers were diligent in using deliberative cognitive resources, it would be very unlikely for a subject to register all attributes to be the same—for example, as all 3's ("fair"), or all 2's or all 4's for that matter. The process of providing the same answer to every question is referred to as *straight-lining*.

A survey respondent who gives all 5's (all "excellent" scores) might be more likely to use deliberative cognitive resources than another respondent who gives all 3's (all "fair" scores), because a score of all 5's might arise from true customer delight with all aspects of the service, together with truncation effects (Oliver et al. 1997). If truncation effects were the only factors giving rise to all 5's, a similarly high relative percentage incidence of all 1's (all "poor" scores) would be expected. However, that is not the case,[4] suggesting that two different levels of mental engagement are discernible (when $RF = 1$ versus $RF = 0$).

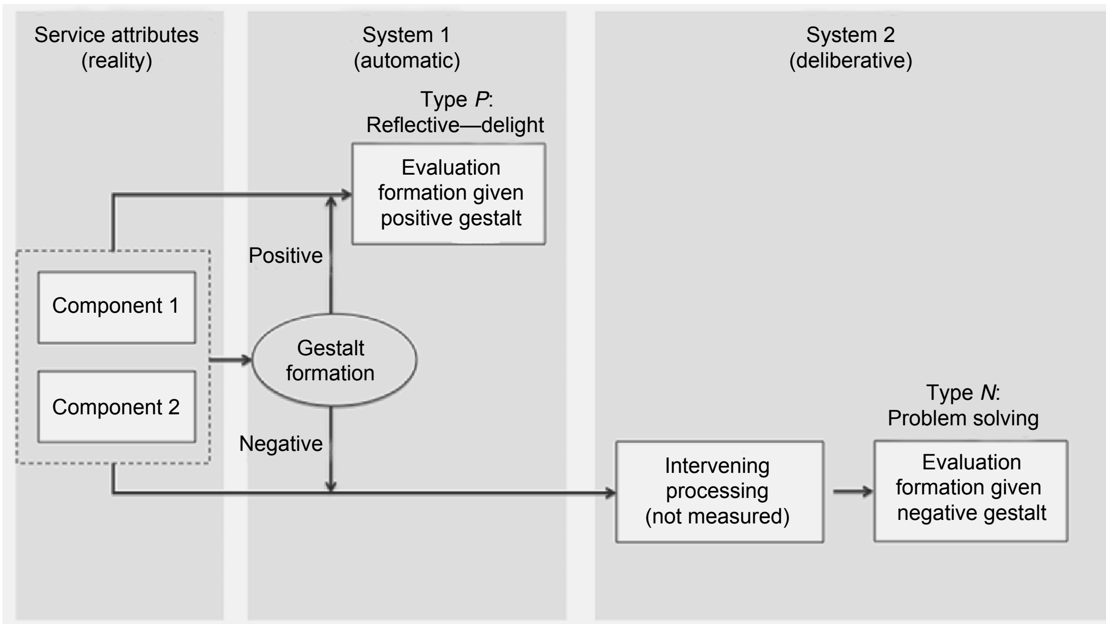
Table 5 addresses whether the relative proportion of cases of straight-lining varies systematically according to whether customers would or would not recommend their physician to a friend. Of the customers that would recommend, 24.1% straight-line for all attributes. By contrast, of the customers that would not recommend, 3.9% straight-line for all the attributes. This is a statistically significant difference between the two groups when just comparing the respondents in the bottom row as a single contingency table ($\chi^2 = 394.75$, $p < 0.0001$). This provides prima facie evidence that many customers in type *P* processes are responding in an automatic fashion.

5. Proposed Synthesis

Our analysis shows the applicability of two different decision rules depending on whether patients were in a type *P* or type *N* evaluative mode, which is proxied by the measure "Would you recommend [this physician] to

[4] That is, of patients that chose 5 for *CS* and were also type *P*, 36% chose 5 for all other attributes; of patients that chose 1 for *CS* and were also type *N*, only 11.0% chose 1 for all the other attributes.

Figure 2. Conjectured Dual Process Model



a friend." In this section, we discuss how such results may reflect the presence of dual cognitive processes being operative, depending on the circumstances. We begin this section by reviewing past work on dual cognitive process models.

Multiple authors have suggested that consumer decision making is generally governed by two systems: system 1 is unconscious, rapid, automatic, and high capacity; and system 2 is conscious, slow, and deliberative (see Bond et al. 2009, Evans 2008). These two processes are distinct cognitive resources that can be called upon (separately or jointly) in different decision-making situations. These two cognitive resources have different performance characteristics, including the quality of the judgment and the cognitive load required to reach that judgment. Similar thinking is exemplified in Petty et al. (1983), who introduced the idea of peripheral and central processing of ad messages. The cognitive resources of systems 1 and 2 are also relevant for understanding consumers' responses to service encounters.[5]

As a way of interpreting the empirical results of this paper, we speculate that system 1 could govern the "lens" by which consumers view and assess service encounters, thus applying the approaches of Bargh (1997), Loewenstein (2001), and Smith and DeCoster (2000) to the service science domain. This lens has two valences, depending on whether the consumer's evaluation of a service encounter is positive or negative, triggering an instinctual gestalt response. The positive gestalt response would initiate what is defined above as the *type P service assessing process*, and the negative gestalt response would initiate what is defined as the *type N service assessing process*. In this way, system 1 provides the automatic switching mechanism that determines the use of type *P* or type *N* processes.

We conjecture that systems 1 and 2 may form the basis of a dual process theory of service evaluation. In Figure 2, we show that the evaluation process begins with the service attributes being experienced by the customer. A gestalt impression of the service experience is then formed with system 1 mental processes. If the impression is positive, typically no deliberative thinking is invoked, and few additional mental resources are consumed. Conjunctive (noncompensatory) functional forms result, governed by system 1 processing. If the impression of the service experience is negative, additional deliberative thinking is likely required, and system 2 mental processes are engaged. Compensatory functional forms result.

It is worth noting that there is a close correspondence between the evaluative processes represented in Figure 2 and the evaluative processes described in our two-equation system (Equations (2) and (3)). Equation (2) describes the relationship between the service components/attributes, *C1* and *C2*, and the gestalt (evaluative) formation, as

[5] The beginnings of a dual process theory can be found in the concepts of customer delight (Finn 2005, 2012; Oliver et al. 1997; Schneider and Bowen 1999) and customer outrage. Customer delight was identified as a direct determinant of behavioral intention, quite separate from the long-established direct effect of customer satisfaction (Oliver et al. 1997). No parallel models yet exist for outrage, although Schneider and Bowen (1999) speculated on a similar relationship in the negative direction; they attributed an emotive response to service failures arising from a perceived violation of a basic human need for security, justice, or self-esteem.

proxied by the recommend to a friend variable. Equation (3) describes the system 1 and system 2 evaluations, contingent on the gestalt evaluative formation (i.e., as moderated by the recommend to a friend variable). This moderator adjusts the functional form on the right-hand side of the overall satisfaction equation (Equation (3)). In our estimated model, the weights when $RF = 0$ (which we proxy as a negative, type N, evaluative mode) are consistent with a nearly linearly compensatory model (left panel of Figure 1), and the weights when $RF = 1$ (which we proxy as a positive, type P, evaluative mode) are consistent with a more conjunctive functional form. Furthermore, our analysis of the contingency table (Table 5) indicates that customers make more use of the different levels of the five-point scales when they are in a type N evaluative mode, as proxied by the willingness to recommend variable, which is also more effortful.

We might even suggest that when in a type N evaluative mode, service users are in a more problem-solving mode precisely because they are in a negative evaluative mode about the service experience. Perhaps they might be close to contemplating how to improve their service experience by switching service providers. In the case where the initial gestalt assessment is clearly positive, we would suggest that service users never leave the automatic mode of cognitive processing. This is our synthesis from the conclusion of our study and our conjecture about the underlying cognitive processes that may be at work.

We look forward to future work that examines such a dual process theory of service evaluation. Such an approach can both explain the results of this paper and also form the basis for better understanding other aspects of human choices to remain loyal to service providers and for better interpreting customer satisfaction surveys.

6. Conclusions and Implications for Future Work

A set of 41,085 outpatient satisfaction surveys was examined to determine whether or not dual rules apply for service evaluation. We examined how the underlying service attributes are combined to form an overall service evaluation, depending on whether the customers are in a positive or negative evaluative mode. The presence of a positive or negative (type N or type P) evaluative mode was proxied by the response to the question, "Would you recommend [this physician] to a friend?" We found that customers in a type N evaluative mode use a compensatory model by which individual attribute assessments are combined to form an overall satisfaction assessment, whereas respondents in a type P evaluative mode use a noncompensatory model. Also, there is a halo shift in the intercepts between type P and N respondents. We conjecture that type P customers are not engaging the more mentally taxing system 2 cognitive decision-making processes. In contrast, type N customers appear to fully engage the system 2 cognitive evaluation processes (i.e., their service experience primes them to carefully assess the trade-offs between different attributes in a compensatory fashion). Since 96% of our respondents are of type P—which we believe is typical for highly rated service providers—our conclusion is that much of the current customer satisfaction survey data are not useful for strategic decisions that require compensatory trade-off information. Thus, caution must be taken when making conclusions using data from type P respondents.

6.1. Limitations and Implications for Future Research

We acknowledge limitations of this research. For the healthcare domain, certain service experiences cannot be undone, which adds salience to negative evaluations. In particular, we expect that the functional form governing customer satisfaction could be different across application areas or different types of services.

A question for future research would be to examine demographic factors (e.g., education, gender, age, income) to determine whether they influence an individual's preference for one type of mental processing over another. We realize that activation of dual or multiple rules for service evaluation may not apply in a "one-size-fits-all" manner. Thus, one might speculate a greater proclivity toward the more deliberative approach of system 2 than the automatic approach of system 1 when high-ticket or high-risk items are being evaluated (as opposed to evaluations of low-ticket or low-risk items, which often are routinized decisions). In addition, the experiences of family members may also play a role: perhaps a family member's experiences create a prior expectation for the patient; this precedent might then be discarded as the patient has his or her own personal experiences with a healthcare provider. Limitations of our data preclude exploration of these issues, constituting a limitation of the current work that we hope future research can address. Situational factors (e.g., imposed deadlines, elapsed time between the service encounter and when the survey is given) may also exert an influence on the type of processing, independently or in combination with other factors. For example, these healthcare data involve a retrospective evaluation that may color patient responses. Determining how the evaluative mode was formed most likely requires factoring in the elapsed time between the service encounter and when the survey is taken. Such data were not available for this study. Longitudinal testing will ameliorate this issue by assessing the consumer's evaluative mode at multiple points in the service encounter and the decision-making process. This may permit more precise identification and estimation of different evaluative functional forms of customers.

Because our proxy for an evaluative mode is imperfect, we suggest that future work should investigate better proxies for revealing the customer's evaluative mode.

The context of a service (e.g., healthcare, banking) may influence the observed phenomena. If the service encounter is highly personal or frequently repeated, the impact of the service attributes on the customer evaluative mode (type P or N) might be different from that of interactions that are more generic or infrequent. In particular, the engagement of the customer could be linked to the domain, and engagement may vary in terms of what a customer has previously experienced. In addition, there may be interaction effects between consumer characteristics and the situation. This could include different value cocreation roles that customers perform in a service dominant encounter, affecting prevailing emotions and prejudices. Furthermore, self-relevance will probably differ with involvement in the service, and an individual with more frequent and diverse experiences with one type of service may be more able to evaluate his or her service experience.[6]

It would also be useful to know how and when during a service encounter different functional forms get triggered and whether this can be reversed if the service provider responds quickly enough after a service breakdown. More generally, are there ways managers can detect and influence a shift in mental processes indicative of a service breakdown? To help consider these questions, multiple methodologies may be useful, including field or laboratory experiments (which also can measure ongoing consumer decision processing).

Cronin et al. (2000) showed that the customer perceptions of quality, value, and satisfaction are highly correlated yet independent concepts. They concluded that service quality not only affects consumer perceptions of value and satisfaction but also has a direct influence on behavioral intentions. Consistent with findings from Olsen (2002) and Szymanski and Henard (2001), in this paper we assumed that the relationship between quality performance and satisfaction is high. Future work could attempt to disentangle the relationships between the evaluative mode, service quality, and satisfaction.

Finally, in this study, the consumers were evaluating a type of service where ratings are historically high (i.e., outpatient services). It remains to be seen if this holds for services that do not enjoy such high evaluation ratings.

6.2. Implications for Managers

Service organizations collect and analyze internal and external performance data to facilitate strategic and operational planning. In strategic planning, it is necessary to understand the value or quality of the organization's services relative to competitors in order to establish and/or maintain competitive advantage. In operational planning, management seeks to better allocate resources and improve processes based on analysis of the collected customer satisfaction data. Service organizations primarily derive their success from interactions that are perceived as useful to the customer. Therefore, managers of service organizations must focus on facilitating successful customer interactions, but they must carefully analyze all interactions so as to improve their services. This is crucial to attracting and retaining customers—the lifeblood of the service organization.

To make strategic decisions, service managers require information on customer trade-offs between perceived service attributes. Type N respondents, who tend to be more deliberative (and compensatory), thus provide greater insight for the purpose of redesigning services and improving organizational processes. It is these individuals that provide the most useful feedback for the purposes of understanding trade-offs between individual services features (and for identifying problem areas).

This does not entirely devalue the service evaluations of type P respondents. For the purpose of evaluating the effectiveness of different service delivery units (and for determining rewards), it remains reasonable to compare the percentage of type P versus type N responses across different delivery units.

In summary, when a satisfaction driver changes, the "Would you recommend [this physician] to a friend?" proxy may change. This may alter the customer's evaluative mode, which then changes the functional form used to evaluate overall satisfaction. Hence, the level of overall satisfaction may change. A practical implication is that the service provider should attempt to prevent customers from having a negative experience that would lead them to engage in a negative evaluative mode. Although this appears to be a statement of the obvious, firms may neglect elements in the service delivery process that they deem to be superficial but yet are contributors to the customer's formation of a negative impression. This means that there needs to be an active search of the service process to fix problems that would be triggers to entering a negative evaluative mode. Internal service processes should therefore be agile and reflective, allowing managers to learn from and correct service component weaknesses before the service is finalized.

[6] We thank an anonymous reviewer for suggesting several of these points.

Another practical implication for managers conducting service studies is to recognize that there may be a need to collect sufficient customer evaluations to have statistically valid samples of type N respondents. For example, in our data set, only 3.8% of the responses exhibited a type N evaluative mode (i.e., these respondents were not willing to recommend the physician to a friend and did not straight-line responses). For a small survey sample, this issue can have a detrimental effect on the manager's ability to draw specific remedial conclusions from the data beyond overall satisfaction. A challenge in practice is how to engage more respondents to use deliberative mental processes or how to achieve sample sizes where the respondents using deliberative processes are sufficiently large. Alternatively, the question of how a service manager ensures that survey participants are engaged looms large. Indeed, it might be useful for managers to ask more probing follow-up questions after learning that a respondent would not recommend his or her physician to a friend in order to get more informative data.

Appendix

Table A.1. Split Sample Verification for Equation (2)

	Complete		Calibration		Holdout	
Parameter	Estimate	Sig.	Estimate	Sig.	Estimate	Sig.
Threshold parameters						
$[-\hat{\alpha}_0]$	3.1670	0.000	3.4159	0.000	2.9396	0.000
Location parameters						
C1	1.4463	0.000	1.5271	0.000	1.3743	0.000
C2	−0.1484	0.086	−0.1241	0.323	−0.1712	0.149
C1 × *C2*	−0.0184	0.359	−0.0283	0.342	−0.0097	0.722
Pseudo-R^2						
Nagelkerke	0.515		0.530		0.499	
McFadden	0.473		0.489		0.458	
Observations	41,085		20,548		20,537	

Note. Equation (2) is estimated using the standard (dichotomous) probit model.

Table A.2. Split Sample Verification for Equation (3)

	Complete		Calibration		Holdout	
Parameter	Estimate	Sig.	Estimate	Sig.	Estimate	Sig.
Threshold parameters						
$[\hat{l}_1 = 1]$	−3.2738	0.000	−3.3384	0.000	−3.1963	0.000
$[\hat{l}_2 = 2]$	−2.8543	0.000	−2.9434	0.000	−2.7484	0.000
$[\hat{l}_3 = 3]$	−1.6347	0.000	−1.7228	0.000	−1.5270	0.000
$[\hat{l}_4 = 4]$	−0.0286	0.903	−0.1441	0.728	0.1109	0.753
Location parameters						
$(1 - RF)$	−3.4290	0.000	−3.7124	0.000	−3.1613	0.000
C1	−0.8350	0.000	−0.8473	0.000	−0.8183	0.000
C2	−0.8240	0.000	−0.8584	0.000	−0.7846	0.000
$(1-RF)$ × *C1*	0.7181	0.000	0.7478	0.000	0.7024	0.000
$(1-RF)$ × *C2*	1.2976	0.000	1.4635	0.000	1.1312	0.000
C1 × *C2*	0.3818	0.000	0.3865	0.000	0.3761	0.000
$(1-RF)$ × *C1* × *C2*	−0.2876	0.000	−0.3198	0.000	−0.2580	0.000
Pseudo-R^2						
Nagelkerke	0.431		0.426		0.436	
McFadden	0.243		0.240		0.247	
Observation	41,085		20,548		20,537	
CM statistic	−0.0724	0.7958	−0.0698	0.7959	−0.0715	0.7958

Note. Equation (3) is estimated using ordinal probit.

References

Aigner DJ (1973) Regression with a binary independent variable subject to errors of observation. *J. Econometrics* 1(1):49–59.

Anderson EW, Sullivan MW (1993) The antecedents and consequences of customer satisfaction for firms. *Marketing Sci.* 12(2):125–143.

Anderson EW, Fornell C, Lehmann DR (1994) Customer satisfaction, market share, and profitability: Findings from Sweden. *J. Marketing* 58(3):53–66.

Anderson EW, Fornell C, Mazvancheryl SK (2004) Customer satisfaction and shareholder value. *J. Marketing* 68(4):172–185.

Bahia K, Nantel J (2000) A reliable and valid measurement scale for the perceived service quality of banks. *Internat. J. Bank Marketing* 18(2):84–91.

Bargh JA (1997) The automaticity of everyday life. Wyer RS Jr, ed. *The Automaticity of Everyday Life: Advances in Social Cognition*, Vol. 10 (Lawrence Erlbaum Associates, Mahwah, NJ), 1–61.

Bernhardt KL, Donthu N, Kennett PA (2000) A longitudinal analysis of satisfaction and profitability. *J. Bus. Res.* 47(2):161–171.

Bienstock CC, Mentzer JT, Bird MM (1997) Measuring physical distribution service quality. *J. Acad. Marketing Sci.* 25(1):31–44.

Bolton RN (1998) A dynamic model of duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Sci.* 17(1):45–65.

Bond SD, Bettman JR, Luce MF (2009) Consumer judgment from a dual-systems perspective: Recent evidence and emerging issues. Malhotra NK, ed. *Review of Marketing Research*, Vol. 5 (M.E. Sharpe, Armonk, NY), 3–37.

Brown SW, Swartz TA (1989) A gap analysis of professional service quality. *J. Marketing* 53(2):92–98.

Büschken J, Otter T, Allenby GM (2011) Do we halo or form? A Bayesian mixture model for customer satisfaction data. Fisher College of Business Working Paper 1620863, Ohio State University, Columbus. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1620863.

Büschken J, Otter T, Allenby GM (2013) The dimensionality of customer satisfaction survey responses and implications for driver analysis. *Marketing Sci.* 32(4):533–553.

Cardozo RN (1965) An experimental study of customer effort, expectation, and satisfaction. *J. Marketing Res.* 2(3):244–249.

Carrillat FA, Jaramillo F, Mulki JP (2007) The validity of the SERVQUAL and SERVPERF scales: A meta-analytic view of 17 years of research across five continents. *Internat. J. Service Indust. Management* 18(5):472–490.

Cooil B, Keiningham TL, Aksoy L, Hsu M (2007) A longitudinal analysis of customer satisfaction and share of wallet: Investigating the moderating effect of customer characteristics. *J. Marketing* 71(January):67–83.

Coombs CH (1964) *A Theory of Data* (John Wiley & Sons, New York).

Coombs CH, Kao RC (1955) Nonmetric factor analysis. Engineering Research Bulletin 38, University of Michigan Press, Ann Arbor.

Cooper WH (1981) Ubiquitous halo. *Psych. Bull.* 90(2):218–244.

Cronin JJ Jr, Taylor SA (1992) Measuring service quality: A reexamination and extension. *J. Marketing* 56(3):55–68.

Cronin JJ Jr, Brady MK, Hult GTM (2000) Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. *J. Retailing* 76(2):193–218.

Dabholkar PA, Shepherd CD, Thorpe DI (2000) A comprehensive framework for service quality: An investigation of critical conceptual and measurement issues through a longitudinal study. *J. Retailing* 76(2):139–173.

Dawes RM (1964) Social selection based on multi-dimensional criteria. *J. Abnormal Soc. Psych.* 68(1):104–109.

de Matos CA, Rossi CAV (2008) Word-of-mouth communications in marketing: A meta-analytic review of the antecedents and moderators. *J. Acad. Marketing Sci.* 36(4):578–596.

Einhorn HJ (1970) The use of nonlinear, noncompensatory models in decision making. *Psych. Bull.* 73(3):221–230.

Einhorn HJ (1971) Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organ. Behav. Human Performance* 6(1):1–27.

Elrod T, Johnson RD, White J (2004) A new integrated model of noncompensatory and compensatory decision strategies. *Organ. Behav. Human Decision Processes* 95:1–19.

Evans JSBT (2008) Dual-processing accounts of reasoning, judgment and social cognition. *Annual Rev. Psych.* 59:255–278.

Finn A (2005) Reassessing the foundations of customer delight. *J. Service Res.* 8(2):103–116.

Finn A (2012) Customer delight: Distinct construct or zone of nonlinear response to customer satisfaction? *J. Service Res.* 15(1):99–110.

Fisher FM (1966) *The Identification Problem in Econometrics* (McGraw-Hill, New York).

Ganzach Y (1993) Goals as determinants of nonlinear noncompensatory judgment strategies: Leniency vs strictness. *Organ. Behav. Human Decision Processes* 56(3):422–440.

Ganzach Y (1994) Inconsistency and uncertainty in multi-attribute judgment of human performance. *J. Behav. Decision Making* 7(3):193–211.

Ganzach Y (1995) Negativity (and positivity) in performance evaluation: Three field studies. *J. Appl. Psych.* 80(4):491–499.

Gilbride TJ, Allenby GM (2004) A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Sci.* 23(3):391–406.

Grönroos C (1984) A service quality model and its marketing implications. *Eur. J. Marketing* 18(4):36–44.

Gruca TS, Rego LL (2005) Customer satisfaction, cash flow, and shareholder value. *J. Marketing* 69(3):115–130.

Heckman JJ (1978) Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46(4):931–959.

Herzberg F, Mausner B, Snyderman BB (1959) *The Motivation to Work*, 2nd ed. (John Wiley & Sons, New York).

Kano N, Seraku N, Takahashi F, Tsuji S (1984) Attractive quality and must-be quality. [In Japanese.] *J. Japanese Soc. Quality Control* 14(2):39–48.

Keiningham TL, Cooil B, Andreassen TW, Aksoy L (2007) A longitudinal examination of net promoter and firm revenue growth. *J. Marketing* 71(3):39–51.

Kekre S, Krishnan MS, Srinivasan K (1995) Drivers of customer satisfaction for software products: Implications for design and service support. *Management Sci.* 41(9):1456–1470.

Levesque T, McDougall GHG (1996) Determinants of customer satisfaction in retail banking. *Internat. J. Bank Marketing* 14(7):12–20.

Loewenstein G (2001) The creative destruction of decision research. *J. Consumer Res.* 28(3):499–505.

Lytle RS, Hom PW, Mokwa MP (1998) SERV∗OR: A managerial measure of organizational service-orientation. *J. Retailing* 74(4):455–489.

Messinger PR, Li J, Stroulia E, Galletta D, Ge X, Choi S (2009) Seven challenges to combining human and automated service. *Canad. J. Admin. Sci.* 26(4):267–285.

Mittal V, Kamakura WA (2001) Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics. *J. Marketing Res.* 38(1):131–142.

Mittal V, Kumar P, Tsiros M (1999) Attribute-level performance satisfaction, and behavioral intentions over time: A consumption-system approach. *J. Marketing* 63(2):88–101.
Mittal V, Anderson EW, Sayrak A, Tadikamalla P (2005) Dual emphasis and the long-term financial impact of customer satisfaction. *Marketing Sci.* 24(4):544–555.
Monfardini C, Radice R (2008) Testing exogeneity in the bivariate probit model: A Monte Carlo study. *Oxford Bull. Econom. Statist.* 70(2):271–282.
Morgan NA, Rego LL (2006) The value of different customer satisfaction and loyalty metrics in predicting business performance. *Marketing Sci.* 25(5):426–439.
Oliver RL (1980) A cognitive model of the antecedents and consequences of satisfaction decisions. *J. Marketing Res.* 17(4):460–469.
Oliver RL (2010) *Satisfaction: A Behavioral Perspective on the Consumer*, 2nd ed. (M.E. Sharpe, Armonk, NY).
Oliver RL, Rust RT, Varki S (1997) Customer delight: Foundations, findings, and managerial insight. *J. Retailing* 73(3):311–336.
Olsen SO (2002) Comparative evaluation and the relationship between quality, satisfaction, and repurchase loyalty. *J. Acad. Marketing Sci.* 30(3):240–249.
Otani K, Harris LE (2004) Different integration processes of patient satisfaction among four groups. *Health Care Management Rev.* 29(3):188–195.
Parasuraman A, Berry LL, Zeithaml VA (1991) Refinement and reassessment of the SERVQUAL scale. *J. Retailing* 67(4):420–450.
Parasuraman A, Zeithaml VA, Berry LL (1985) A conceptual model of service quality and its implications for future research. *J. Marketing* 49(4):41–50.
Parasuraman A, Zeithaml VA, Berry LL (1988) SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *J. Retailing* 64(1):12–40.
Parasuraman A, Zeithaml VA, Malhotra A (2005) E-S-QUAL: A multiple-item scale for assessing electronic service quality. *J. Service Res.* 7(3):213–233.
Payne JW, Bettman JR, Johnson EJ (1993) *The Adaptive Decision Maker* (Cambridge University Press, New York).
Petty RE, Cacioppo JT, Schumann D (1983) Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *J. Consumer Res.* 10(2):135–146.
Pitt LF, Watson RT, Kavan CB (1995) Service quality: A measure of information-systems effectiveness. *MIS Quart.* 19(2):173–187.
Reichheld FF (2003) The one number you need to grow. *Harvard Bus. Rev.* 81(12):46–54.
Rust RT, Chung TS (2006) Marketing models of service and relationships. *Marketing Sci.* 25(6):560–580.
Rust RT, Oliver RL (1994) Service quality: Insights and managerial implications from the frontier. Roland TR, Oliver RL, eds. *Service Quality: New Directions in Theory and Practice* (Sage, Thousand Oaks, CA), 1–19.
Schneider B, Bowen DE (1999) Understanding customer delight and outrage. *Sloan Management Rev.* 41(1):35–45.
Sharp B (2008) Net Promoter Score fails the test. *Marketing Res.* 20(4):28–30.
Smith ER, DeCoster J (2000) Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality Soc. Psych. Rev.* 4(2):108–131.
Steward MB, Wallis KF (1981) *Introductory Econometrics* (John Wiley & Sons, New York).
Suhr DD (2005) Principal component analysis vs. exploratory factor analysis. *SUGI 30 (30th Annual SAS Users Group Internat. Conf.) Proc.* (SAS Institute, Cary, NC), Paper 203-30.
Swan JE, Combs LJ (1976) Product performance and consumer satisfaction: A new concept. *J. Marketing* 40(2):25–33.
Swan JE, Oliver RL (1989) Postpurchase communications by consumers. *J. Retailing* 65(4):516–533.
Szymanski DM, Hise RT (2000) E-satisfaction: An initial examination. *J. Retailing* 76(3):309–322.
Szymanski DM, Henard DH (2001) Customer satisfaction: A meta-analysis of the empirical evidence. *J. Acad. Marketing Sci.* 29(1):16–35.
Vargo SL, Lusch RF (2004) Evolving to a new dominant logic for marketing. *J. Marketing* 68(1):1–17.
Wilde J (2000) Identification of multiple equation probit models with endogenous dummy regressors. *Econom. Lett.* 69(3):309–312.
Wirtz J (2003) Halo in customer satisfaction measures: The role of purpose of rating, number of attributes and involvement. *Internat. J. Service Indust. Management* 14(1):96–119.
Wirtz J, Bateson JEG (1995) An experimental investigation of halo effects in satisfaction measures of service attributes. *Internat. J. Service Indust. Management* 6(3):84–102.
Wold HOA (1960) A generalization of causal chain model (part III of a triptych on causal chain systems). *Econometrica* 28(2):443–463.
Yi Y (1990) A critical review of consumer satisfaction. Zeithaml VA, ed. *Review of Marketing*, Vol. 4 (American Marketing Association, Chicago), 68–123.
Zeithaml VA, Berry LL, Parasuraman A (1996) The behavioral consequences of service quality. *J. Marketing* 60(2):31–46.