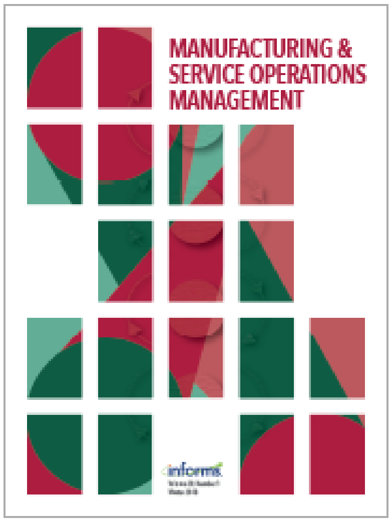


Manufacturing & Service Operations Management

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

Service Differentiation and Operating Segments: A Framework and an Application to After-Sales Services

Jose A. Guajardo, Morris A. Cohen

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 20, No. 3, Summer 2018, pp. 440–454
http://pubsonline.informs.org/journal/msom/
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

Service Differentiation and Operating Segments: A Framework and an Application to After-Sales Services

Jose A. Guajardo,[a] Morris A. Cohen[b]

[a] Haas School of Business, University of California Berkeley, Berkeley, California 94720; [b] The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Contact: jguajardo@berkeley.edu, http://orcid.org/0000-0003-0441-5353 (JAG); cohen@wharton.upenn.edu, http://orcid.org/0000-0002-8644-8820 (MAC)

Received: July 31, 2015
Revised: October 31, 2016; April 13, 2017
Accepted: April 14, 2017
Published Online in Articles in Advance: April 17, 2018

https://doi.org/10.1287/msom.2017.0645

Abstract. The decisions of whether and how to adopt service differentiation are at the core of a firm's service operations strategy. This paper proposes a framework for service differentiation that highlights the identification and use of *operating segments* as a central component in the delivery of differentiated services. The notion of operating segments and the general empirical methodology to identify them proposed in this paper integrally considers the consumer's preferences and operational capabilities required to fulfill the differentiated service offering. An application in the context of after-sales services for product-service bundles using data from a major manufacturer in the consumer electronics industry is presented, which illustrates how operational decisions need to be adjusted when multiple operating segments are defined to support a service differentiation strategy.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/msom.2017.0645.

Keywords: service differentiation • operating segments • after-sales services • service operations strategy

1. Introduction

As the world moves to a service-based economy, defining the right service strategy has become an increasingly important priority not only for firms that deliver traditional services (Frei and Morriss 2012) but also for original equipment manufacturers (OEMs) (Guajardo et al. 2016). The decisions of whether and how to adopt service differentiation are at the core of a firm's service operations strategy. Indeed, differentiating the provision of services could provide a more effective way to make supply meet demand for firms facing groups of customers with significantly heterogeneous service preferences. At the same time, adopting differentiation directly affects a firm's resources-management process, and it is part of its overall competitive strategy.

Service competition has been a major topic of interest in the field of service operations strategy. In traditional service industries, theoretical models have examined competition when consumer demand depends on price and service levels (Cachon and Harker 2002, Allon and Federgruen 2009), and empirical studies have tested related theories, e.g., in the fast-food sector (Allon et al. 2011) and banking industries (Buell et al. 2016a, b). In the context of product-service systems, both theory (e.g., Cohen and Whang 1997) and empirical applications (Guajardo et al. 2016) have highlighted the value of differentiated services in a competitive environment.

Similarly, product differentiation based on market segmentation has long been recognized as one of the most fundamental concepts in marketing (Wind 1978). Market segments are typically defined to divide a heterogeneous market into a set of homogenous submarkets, based on the variety of ways that customers can derive value from products (Wind and Bell 2008). Operations management (OM) also has a long history of dealing with segmentation. This is especially true in service operations and logistics environments where service standards are typically set to deliver targeted levels of availability, response time, and customer satisfaction. In many cases, these service standards vary across segments—by customer attributes, geography, product technology, and price. Such differentiation is especially important in after-sales service, wherein service standards are often based on the strategic value of product up-time.

This paper will present a framework for services differentiation that builds upon prior research and practice drawn from both service OM and marketing. The proposed framework can be applied both to experiential services (such as hospitality, travel, and dining), as well as to product-service bundles, where the service component enhances the value derived by the customer from the use or consumption of the product (e.g., call centers for customer support, warranties). A key element in this proposed framework is the concept of *operating segments*, a term coined by Frei and Morriss (2012) to denote a list of service priorities shared by a meaningful group of customers. This paper outlines the basis for the notion of operating segments, develops an empirical strategy to define them in a general service setting, and highlights the operational implications of having multiple operating segments in the delivery of differentiated services. To illustrate the general empirical strategy, the authors apply it to a case of after-sales services (call centers, in-home repair) for high-definition televisions (HDTVs), using data from a leading consumer electronics company.

The rest of this paper is organized as follows. Section 2 reviews the literature that motivates the notion of operating segments and the proposed framework for differentiated services. The framework is developed in more detail in Section 3. Section 4 focuses on the notion of operating segments, detailing their most relevant aspects and proposing a general empirical methodology to identify them. Section 5 provides an application of this methodology to the case of an OEM of HDTVs, illustrating aspects and challenges that emerge when operating segments are defined in practice. Section 6 discusses managerial implications of operating segments, further illustrating how operational decision making needs to be modified when a company serves multiple operating segments. Section 7 reviews some limitations and extensions of the study, and Section 8 concludes.

2. Literature Review

The notion of customer heterogeneity has been analyzed extensively in the marketing literature in several contexts, including customer satisfaction (Mittal and Kamakura 2001) and brand loyalty (Fader and Lattin 1993). A major use of segmentation in marketing is to support the positioning of the firm's product portfolio. Methods such as conjoint analysis have been used extensively to capture variations in the desirability of different product attributes across different segments (e.g., Wind et al. 1989). Pricing and price discrimination also have been used in the definition of segments, as well as the definition of aggregate products which would be designed to serve multiple segments (e.g., Moorthy 1984, Moorthy and Png 1992). More complex multipart pricing schemes also have been analyzed to design and price service products, e.g., for telephone services (Iyengar et al. 2008). More recently, multicriteria optimization models have been developed for defining marketing segments (Liu et al. 2010).

An OM perspective on the topic is provided by Ho and Zheng (2004), who propose a model for dealing with the decision to provide a service delivery guarantee (e.g., for maximum delivery time). More generally, the view that product differentiation increases the level of product variety, adding complexity and potentially raising costs, has been highlighted in OM research. For example, this is seen in a series of classification schemes that are commonly used in practice to support inventory management, such as, e.g., the procedure developed by Ernst and Cohen (1990) in the context of spare parts used for automobile maintenance and repair. A recent stream of OM literature has looked at operational implications of service differentiation—e.g., analyzing the staffing and control decisions in call centers serving multiple customer classes (Gurvich et al. 2008, Bassamboo and Zeevi 2009, Gurvich and Whitt 2010, Mehrotra et al. 2012). However, this literature has not proposed a way to empirically identify such heterogeneous groups with differentiated service priorities based on the operational capabilities needed to deliver them. The notion of operating segments introduced in this paper and the general methodology proposed to identify them fills part of this gap by integrally considering the consumer's preferences and operational capabilities required to fulfill the differentiated service offering.

Some recent studies in OM have highlighted the role of customer heterogeneity in service settings, documenting significant heterogeneity in customer sensitivity to service attributes in industries such as banking (Campbell and Frei 2011; Buell et al. 2016a, b), purchases in a deli section of a super-center (Lu et al. 2013), and after-sales service support (Cohen et al. 2006a), or illustrating the important role of customer heterogeneity in the definition of business models in the sharing economy and peer-to-peer rental markets (Abhishek et al. 2016). Campbell and Frei (2011), in particular, consider geographic heterogeneity where local requirements differ; in such settings, local managers can resolve the capacity versus service trade-off by deviating from a uniform, central plan. The operational control of segmentation also leads to questions such as real-time matching of service resources with segment-specific service demand, prioritization of demand across segments, and rationing of resources. Deshpande et al. (2003) explored the issue of how to manage the delivery of differentiated service through prioritization and allocation of a commonly demanded resource. As this paper will illustrate, the consideration of these multiple operational decisions fundamentally distinguishes operating segments from marketing segments.

Customer demographic characteristics are usually an important factor in the implementation of segmentation. Past research provides a number of examples attempting to link customer demographic characteristics with preferences about products and services. Some researchers have documented gender differences in risk aversion for products such as insurance (Halek and Eisenhauer 2001) and extended warranties (Chen et al. 2009), age differences in brand loyalty and information search behavior (Ratchford 2001), and income differences in sensitivity to service times (Png and Reitman 1994, Propper 1995, Campbell and Frei 2011). These examples illustrate how customer demographic characteristics can be related to relevant aspects of service differentiation.

The OM literature also includes some research about the fundamental question of whether a firm should offer differentiated services. A recent paper by Sainathan (2015) deals with differentiation by prioritizing segments according to the delay sensitivity of customers in the context of a queueing model. Related questions also have been of interest in the specific context of after-sales services. For example, Wang et al. (2002) analyzed an environment where two classes of customer service (based on delivery lead time) can be provided. Their analysis quantified the economic value of providing differentiated service quality (i.e., a shorter lead time) based on inventory reduction that could be attained by introducing a second class of customer service. The existing OM literature on service differentiation focuses heavily on single-metric approaches (such as customer wait time) and does not deal explicitly with the impact of product attributes or customer heterogeneity, unlike the framework that will be proposed in this paper.

An important OM concept related to service differentiation is that of pooling of demand for a common product across multiple segments, which usually implies a trade-off between pooling efficiencies and the demand of product differentiation across segments. A recent illustration of the impact of demand pooling in the context of service parts inventory management was developed by Cohen and Cohen (2017), who explored the impact of demand pooling by centralizing repairs of failed parts at a regional depot versus at a local forward base. More generally, a stream of OM research in queueing (especially in the context of call centers) has analyzed deviations from the conventional operational view that pooling resources leads to operational efficiencies (Eppen 1979) and has explored scenarios in which the benefits of having "dedicated services" may outweigh cost inefficiencies (Gilbert and Weng 1998; Mandelbaum and Reiman 1998; van Dijk and van der Sluis 2008, 2009). Similarly, Song et al. (2015) provided empirical evidence for the advantages of dedicated services in a healthcare setting, and Shunko et al. (2018) showed that behavioral considerations can also play a role in the analysis of dedicated service versus pooling.

Overall, although the efficiencies obtained from pooling and one-size-fits-all services are well known in OM, the operational implications of offering differentiated services are less well understood. In what follows, the authors build upon the aforementioned streams of research to present a framework for service differentiation and to develop the construct of operating segments.

3. A Framework for Service Differentiation

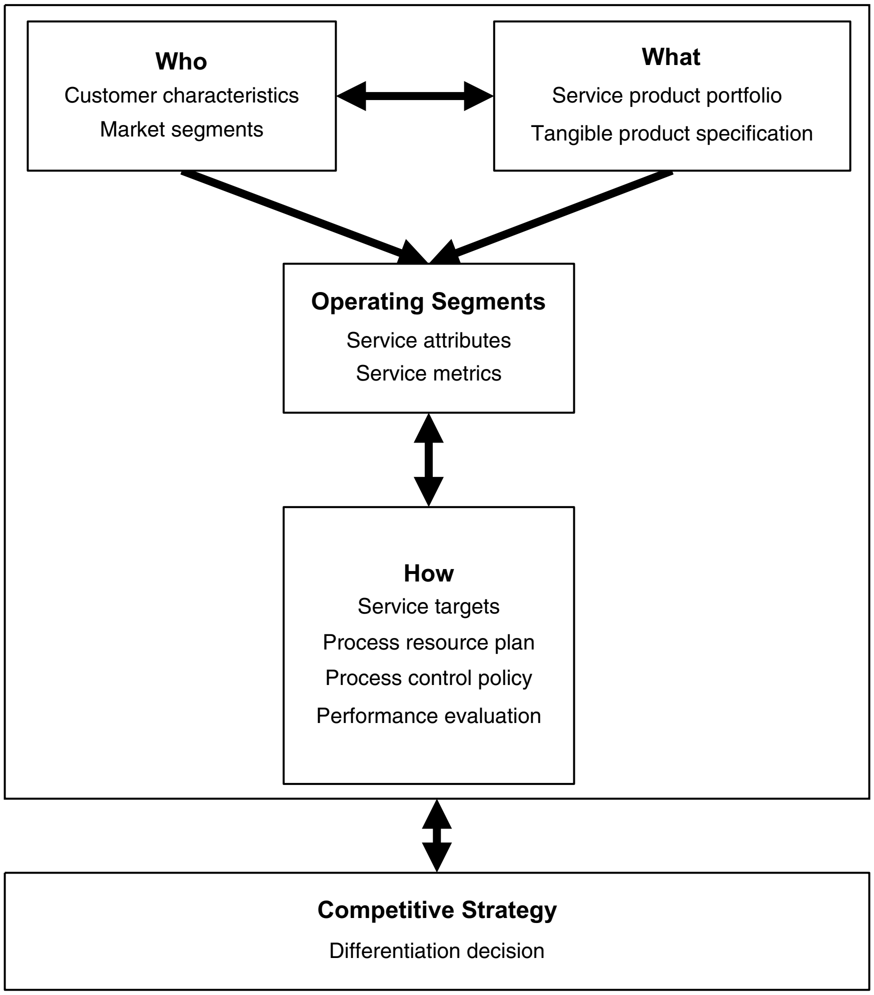
Figure 1 provides a visualization of the proposed framework. Customers, products, and processes form the basic inputs needed to define the operating segments: *who* (which customers) are being served, *what* service products are associated with or targeted to the segments, and *how* these service products should be produced and delivered through management and control of the underlying resources and operations processes.

The "who" question is associated with the definition of market segments. Market segments can be defined based on a variety of customer characteristics. These attributes can include demographic, behavioral, economic, and social variables.

The "what" question is answered by defining the portfolio of service products and the supporting tangible products in terms of their attributes. Design standards for each service product, in particular, determine the service product portfolio. These product designs and their positioning with respect to the preferences of market segments define the service attributes that lead to operating segments. A typical design standard for call centers, for example, would be the waiting time of customers calling in. Because the delivery of the services associated with operating segments requires operationally relevant service attributes, it is necessary also to consider the attributes of the tangible resources that are used to produce the services of interest. Typically, these resources affect the environment where customers consume services and hence are relevant to the perception of service quality, e.g., leaving the facilities in clean condition after an in-home repair service. In the case of product-service bundles, such as after-sales service, the service support requirements of the tangible product are peculiar to its design and function.

The answer to the question of "how" is reflected in the design and control policies for the various operational processes that are required to produce and deliver the products. Indeed, the definition of *operating segments* leads to the specification of group-specific operational *service metrics* and *targets*. From a managerial perspective, there are several challenges associated with resource management for a company wanting to deliver service differentiation in a cost-effective manner, both at strategic and tactical levels. At a strategic level, possibly the most salient problem is capacity management, which is an important component of the *process resource plan*. The tactical *process control policy*, in turn, relates to how firms determine mechanisms for the allocation of "servers" capable of meeting the requirements of the customers in the different operating segments and prioritizing these requirements across segments. *Performance evaluation* is also an integral part of the service delivery, for which is useful the definition of key performance indicators (KPIs) that can be tracked and monitored. The definition of a service strategy requires the firm to carry out a complete analysis of the economic implications. This evaluation would ultimately indicate whether adopting a differentiation strategy would result in increased profits. Firms base this decision on multiple factors, including

Figure 1. Framework for Service Differentiation



not only internal factors, but also considerations about the firm's competitive environment. Its detailed assessment will depend on the specifics of each application.

The general framework depicted in Figure 1 provides a useful background for the role of operating segments in the definition of a service differentiation strategy. As illustrated in the figure, the various aspects of the framework for the delivery of differentiated service products are highly interrelated.

4. Operating Segments

4.1. Defining Operating Segments

An operating segment is defined as a *list of service priorities shared by a meaningful group of customers*. Customers with similar service priorities are part of the same operating segment. Importantly, not only are the service priorities different across operating segments, but so are the operational capabilities needed to deliver them. For example, a call center may hire and train two groups of agents to serve its multiple operating segments.

Let $C = (c_1, \ldots, c_m)$ be the set of relevant observable customer characteristics, $P = (p_1, \ldots, p_n)$ be the set of relevant product characteristics based on service priorities, and $S = (s_1, \ldots, s_o)$ be the set of service process policies that can be used to produce and deliver a differentiated set of service products. Note that C can include both demographic characteristics (such as gender, age, and income) as well as other customer characteristics (e.g., price sensitivity). On the other hand, P refers to attributes of the services to be delivered that matter to the customers and that are operationally relevant (e.g., response time, ability to resolve problems, technical knowledge). It is important to emphasize that the attributes that are part of P entail differences in the operational aspects of service delivery. For example, "fast service" is an attribute that clearly has operational implications (e.g., larger capacity, better-trained service force); hence, it is considered as a possible attribute of P. The hierarchy of decisions (S) associated with the design and management of the service delivery process, which (as will be detailed in Section 6) can include setting capacity levels for different classes of resources, as well as scheduling and control policies that govern the utilization of these resources.

Based on the customer characteristics, it is possible to define "meaningful" groups of customers $CL = (CL_1, \ldots, CL_j)$, according to their sensitivity to the service attributes in P. Groups will therefore be defined by a membership function that maps C into CL, where the members of a group share common priorities for different aspects of the service. Moreover, for each CL_j the set of attributes $P_j = (p_{[1]}, \ldots, p_{[j]})$ characterizing the group's service preferences needs to be determined. Each relevant (CL_j, P_j) defines a different operating segment; the collection of $\{(CL_1, P_1), (CL_2, P_2), \ldots, (CL_j, P_j)\}$ thus defines all potential operating segments. The collection of operating segments can be reduced to a subset of "meaningful" segments if a given operational policy can deliver to more than one segment, i.e., multiple segments are served by a common process or policy. The final collection of operating segments is denoted by $\{(CL_k, P_k)\}$ with $k = 1, 2, \ldots, k'$ where $k' \leq j$, which represents the service product portfolio that the firm has undertaken to deliver to the market.

As noted previously, an operational policy s_k corresponds to the set of resources and management decisions associated with the delivery of services to operating segment k. It could, for example, entail use of a *first-come, first-served* (FCFS) rule for responding to incoming calls in a call center, or the use of multiple classes of servers. Thus, a service strategy can be defined by the triple (CL_k, P_k, s_k), which defines the grouping of customers, the service performance attributes, and the operational policy required to deliver this service product to this collection of customers.

It is important to highlight the differences between operating segments and market segments, for which Figure 1 may serve as a useful reference. Market segments classify customers in terms of marketing needs and exploit customer heterogeneity for that purpose. Operating segments, in contrast, emanate from heterogeneity in the *service preferences and operational requirements for* (*optimally*) *serving different customer groups*. The definition of operating segments will have direct consequences for defining the mechanisms and processes required to design, produce, and deliver the products in the portfolio to the collection of operating segments. Note that there will be operating segments defined for every product in the portfolio. Moreover, every product could be associated with one or more market segments, and within a market segment, there may be several operating segments. Occasionally, operating segments and market segments may coincide, although that is not generally the case (see Frei and Morriss 2012 for further examples).

In short, the operations perspective considers how relative attribute performance can be realized through the allocation and management of resources associated with the delivery of the service. In particular, it is necessary to consider the mechanisms and processes required to design, deliver, and sustain the products in the portfolio for the collection of operating segments.

4.2. An Empirical Strategy for Defining Operating Segments

Multiple strategies can be used to identify operating segments. This paper proposes a two-step procedure, wherein the first step uses customers' heterogeneity in their general valuation of services for defining meaningful groups of customers, and the second step identifies heterogeneity in specific service attributes for the groups of customers defined in the first step. The inputs required by the procedure are data on customer characteristics, their preferences regarding service attributes, and guiding performance metrics. The required data can (and routinely are) collected by firms through surveys or from transactional databases.

Step 1: Identification of Meaningful Groups of Customers. The first objective is to identify customer heterogeneity in the valuation of services, with the goal of defining "meaningful" customer groups as a function of their characteristics $C = (c_1, \ldots, c_m)$. This step will make use of two additional inputs. Generally speaking, let M be a performance metric of interest (e.g., customers' likelihood to recommend the brand, satisfaction) and Q be an intermediate metric(s) capturing some aspect(s) of the customers' overall service experience with a firm. For example, Q can be the overall perception of service quality in the case of pure services, or the preferences of product quality versus service quality in the case of product-service bundles. The first step focuses on characterizing the relationship between customer groups and the overall metrics of interest. In particular, one can make use of regression analysis to study the relationship between M, C, and Q, through some function f:

$$M = f(C, Q, C \times Q)$$

The proposed approach includes the interaction $C \times Q$ as a way to identify the groups of customers with meaningful variation based on the relationship between the overall service experience and the performance metric of interest. This gives a broad sense of the customer groups who are more sensitive to services, or for whom service perceptions are a more important factor in the overall performance metric M. The definition of the group categories that form part of C will depend on the data available in a given application. As illustrated in Section 5, the categories that form part of C can rely on arbitrary definitions (e.g., solely based on business rules) or on additional statistical analysis (e.g., clustering techniques).

As a result of this step, a set of "meaningful" customer groups $CL_1, \ldots, CL_j$ is identified, and, broadly speaking, each has a different sensitivity to services. In the case of product-service bundles, the distinction can refer to the relative sensitivity to services versus products.

Step 2: Definition of Operating Segments. Having identified the customer groups $CL_1, \ldots, CL_j$, one can make use of the operational service attributes $P = (p_1, \ldots, p_n)$ to identify a set of relevant attributes $P_k = (p_{[1]}, \ldots, p_{[j]})$ for each group k. The metric Q can be used as a guiding criterion for this purpose. In other words, for each customer group, the goal is to use its heterogeneity with respect to optimal service attributes, to identify the set of service attributes that have the highest explanatory power in the overall service evaluation by customers in the group. There are multiple ways to accomplish this goal. Methodologically, this is a standard problem in statistics, and in principle most feature selection algorithms (Guyon and Elisseeff 2003) can be used to select the attributes that are most relevant for each group. For example, wrapper methods (Kohavi and John 1997; see Guajardo et al. 2010 for an application), filter methods (e.g., Koller and Sahami 1996), and embedded methods (e.g., Lal et al. 2006) have been used for feature selection.

One particular approach that can be used for this purpose is stepwise estimation, which is widely available in most statistical packages for a variety of methods. In this context, for each meaningful customer group $CL_1, \ldots, CL_j$, a stepwise regression of Q on $P = (p_1, \ldots, p_n)$ can be performed to identify the subset of service attributes $P_k = (p_{[1]}, \ldots, p_{[j]})$ that are most relevant to explain the service preferences of the group. This procedure leads to the selection of the subset of features that have the most explanatory power for a certain dependent variable and statistical relationship. It allows using both backward (starting with the full subset of potential variables, iterate to sequentially eliminate variables with low explanatory power), forward (starting with a model that uses only an intercept, add variables that have the most explanatory power sequentially), and other search algorithms, as well as defining different criteria for entry/exit of variables (Lindsey and Sheather 2010). It also requires defining the guiding information criteria (e.g., Akaike's information criterion, Bayesian information criterion, adjusted R-square, among others; see Hastie et al. 2009 for additional details). Overall, stepwise regression is a flexible method and is the tool that will be used in this application to identify the set of attributes that are most relevant for each customer group.

As an outcome of this two-step process, the complete set $\{(CL_k, P_k)\}_{k:1,\ldots,k'}$ of operating segments is obtained. The resulting segments are linked to operational policies s_k, a discussion that Section 6 expands upon. The application in Section 5 will provide an illustration of how this proposed strategy can be applied to a specific case. In general, such applications must be closely tied to the strategic objective of implementing a differentiation strategy.

5. An Application to the Case of HDTVs

5.1. Context and Data

This application focuses on the after-sales services of a major consumer electronics OEM, more specifically, on the OEM's U.S. market for the HDTV segment. In consumer electronics, TVs represent the product segment with the largest market share. Supporting services for customers whose products are under warranty consist of call centers for customer service (owned and operated by the OEM) and in-home repair services (provided by exclusive service providers or by authorized service centers). The OEM has considerable influence on the ultimate quality of service delivered to customers by either retailers or service providers. For example, it can set specific operational targets associated with the service delivery—e.g., customer wait time at the call center or the probability of a first-time fix for home visits.

Collaborating with the company, a survey that is intended to capture consumer perceptions of different variables of interest was designed and analyzed. The survey was run by a third-party market research company using a web-based interface. See the online appendix for a description of the survey questions. The survey captured consumer perceptions about product quality and service quality and likelihood to recommend the brand. A seven-point scale was used, and consumers were asked to rank their evaluation of the quality of the product and supporting services (1 = poor to 7 = excellent), as well as their likelihood to recommend the brand based on their overall experience (1 = not at all likely to 7 = extremely likely). The survey also collected information on customer characteristics such as gender, age, education, and income, as well as customer perceptions with regard to specific attributes of the service that was delivered. Overall, the sample consisted of 345 owners of the HDTV brand who had recently experienced a service interaction with the company. Table 1 provides some descriptive statistics for various subgroups in the sample.

The statistics in Table 1 indicate some apparent differences for distinct groups in the sample. For example, the relative magnitude of the association between product–service quality and likelihood to recommend the brand seems to vary across the different groups, i.e., for women, the correlation between service quality and likelihood to recommend is bigger than the correlation between product quality and likelihood to recommend. For men, exactly the opposite is observed. Similar differences are observed for the distinct income groups.

Table 1. Descriptive Statistics and Correlations

		Average scores (1–7 scale)			Pairwise correlations		
Customer characteristic	No. obs. (%)[a]	Product quality (*PQ*)	Service quality (*SQ*)	Likelihood to recommend (*LtR*)	*PQ*, *LtR*	*SQ*, *LtR*	*PQ*, *SQ*
Gender							
Man	80.3	5.78	5.35	5.17	0.82	0.78	0.60
Woman	19.7	4.94	4.76	4.46	0.75	0.84	0.58
Age							
*Age*1 (15–34)	18.2	4.95	4.74	4.87	0.68	0.71	0.46
*Age*2 (35–54)	36.7	4.78	4.87	4.61	0.78	0.78	0.56
*Age*3 (55 or more)	45.1	5.48	5.73	5.46	0.85	0.83	0.63
Education							
No college	33.7	5.40	5.30	5.25	0.82	0.83	0.64
College or more	66.3	5.00	5.25	5.03	0.81	0.78	0.58
Income							
Less than \$100,000	63.3	5.16	5.19	5.01	0.85	0.78	0.67
\$100,000 *or more*	36.7	4.98	5.31	5.06	0.74	0.83	0.52
Full sample	100	5.13	5.23	5.03	0.80	0.80	0.59

Note. Sample size in each case is as follows: gender = 345 obs., age = 335 obs., education = 312 obs., and income = 259 obs.
[a]For each customer characteristic, only valid responses are considered in the calculations.

5.2. Operating Segments in the HDTV Application

Because this application involves product–service bundles, one natural way to think about sensitivity to services is by contrasting customers' perceptions about service quality and product quality. Indeed, a given customer class may be more sensitive to service quality perceptions, whereas others may be more sensitive to product quality perceptions. In this scenario, a firm could offer different service levels to different operating segments, e.g., prioritizing service delivery for customer groups with higher sensitivity to a given service dimension. The following discussion focuses on how to use the proposed strategy to identify relevant operating segments in the context of the HDTV application.

Step 1: Identification of Meaningful Groups of Customers. The main performance metric M available in the data is the likelihood to recommend the brand. Let the subscript i index consumers, such that LtR_i denotes consumer i's likelihood to recommend the brand. Because the data include the consumers' perceptions of product quality (PQ_i) and service quality (SQ_i), one can define $Q_i = (PQ_i, SQ_i)$ including both components. The vector C_i reflects customer i's characteristics, which in the available data corresponds to demographic characteristics (gender, age, education, and income). As noted earlier, one can define the group categories that form part of C arbitrarily (e.g., based on business use) or through, for example, clustering techniques. This case illustrates the first approach by making direct use of the categories for each customer variable in Table 1. This paper will first consider the methodology using these categories and then will discuss how C can be alternatively defined using clustering techniques. Finally, let $Q_i \times C_i$ represent the interactions between the quality vector and the consumer characteristics vector. The first-step regression can be thus represented as follows:

$$LtR_i = \alpha + Q_i\beta + C_i\gamma + (Q_i \times C_i)\delta + \varepsilon_i. \qquad (1)$$

Table 2 displays the results for model specifications that gradually incorporate the interaction effects. All models are constructed with 258 observations; bootstrap standard errors are displayed in parentheses.

The results in Table 2 suggest that gender and income moderate the relative importance of the associations between product–service quality and the likelihood to recommend the brand. Service quality exhibits a greater association with the likelihood to recommend the brand for women than for men, and for high-income customers than for low-income customers. Conversely, the association between product quality and likelihood to recommend the brand is higher for men than it is for women, and for low-income customers than for high-income customers. These effects are present in both the single-variable interaction models (2 and 5) and the model with the full set of interactions (6 and 8). The models suggest that there are no important differences according to age and education.

Note that in these models, C is directly obtained from the customer characteristics, based on the categories reflected in Table 1. Alternatively, clustering can be used to define the customer groups that are part of C, to then interact the clusters with the components of Q to analyze which clusters of customers are more sensitive to product quality and service quality. A hierarchical clustering analysis led to a partition of the space of customer characteristics into 10 different clusters that best characterize the customers' information. The regression analysis in Step 1 was

Table 2. Step 1 Regression of M on Q, C, and $C \times Q$

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PQ	0.602***	0.661***	0.540***	0.632***	0.707***	0.642***	0.638***	0.733***
	(0.064)	(0.073)	(0.149)	(0.119)	(0.078)	(0.167)	(0.064)	(0.166)
SQ	0.511***	0.454***	0.483***	0.532***	0.396***	0.434***	0.553***	0.429***
	(0.059)	(0.067)	(0.109)	(0.108)	(0.073)	(0.120)	(0.058)	(0.120)
Woman	−0.102	−0.009	−0.101	−0.101	−0.112	−0.027	−0.128	−0.069
	(0.153)	(0.145)	(0.157)	(0.155)	(0.153)	(0.148)	(0.149)	(0.144)
*Age*2	−0.340*	−0.337	−0.310	−0.337	−0.271	−0.244	−0.410**	−0.307
	(0.204)	(0.198)	(0.195)	(0.203)	(0.202)	(0.190)	(0.199)	(0.185)
*Age*3	−0.215	−0.220	−0.191	−0.213	−0.130	−0.126	−0.324*	−0.214
	(0.190)	(0.185)	(0.192)	(0.192)	(0.187)	(0.192)	(0.186)	(0.183)
College	0.0472	0.0621	0.035	0.0543	0.0315	0.049	0.058	0.069
	(0.128)	(0.123)	(0.124)	(0.132)	(0.123)	(0.130)	(0.122)	(0.124)
Income_high	0.116	0.130	0.113	0.114	0.103	0.099	0.150	0.118
	(0.135)	(0.130)	(0.135)	(0.133)	(0.128)	(0.128)	(0.129)	(0.126)
PQ × *Woman*		−0.289**				−0.304**		−0.191
		(0.133)				(0.143)		(0.135)
SQ × *Woman*		0.261**				0.281**		0.230*
		(0.120)				(0.130)		(0.126)
PQ × *Age*2			0.044			0.180		0.168
			(0.179)			(0.165)		(0.157)
PQ × *Age*3			0.098			0.138		0.053
			(0.185)			(0.171)		(0.166)
SQ × *Age*2			0.069			−0.065		−0.053
			(0.144)			(0.132)		(0.127)
SQ × *Age*3			0.004			−0.046		0.003
			(0.151)			(0.137)		(0.132)
PQ × *College*				−0.039		0.016		−0.031
				(0.141)		(0.119)		(0.113)
SQ × *College*				−0.036		−0.118		−0.074
				(0.127)		(0.103)		(0.102)
Q × *Income_high*					−0.214*	−0.247**		−0.261**
					(0.125)	(0.132)		(0.126)
SQ × *Income_high*					0.260**	0.303***		0.307***
					(0.111)	(0.112)		(0.106)
PQ × *SQ*							0.057***	0.0558***
							(0.020)	(0.018)
Constant	5.206***	5.198***	5.193***	5.196***	5.143***	5.133***	5.118***	5.031***
	(0.192)	(0.190)	(0.192)	(0.199)	(0.190)	(0.197)	(0.197)	(0.197)
Adj. *R*-squared	0.804	0.810	0.802	0.803	0.812	0.818	0.814	0.825

*, **, and *** indicate significance at the 0.1, 0.05, and 0.01 confidence levels, respectively.

based on these 10 groups instead of based on predefined categories, i.e., customer demographic variables (*gender*, *age*, *income*, *education*) were replaced by nine dummy variables based on the categories obtained from the clustering analysis, and the interaction terms were adjusted in Equation (1) accordingly. Using a group of men with low income as a reference category, similar results to the core regression analysis were obtained (see the online appendix for details). Indeed, the only significant interactions were related to a group of women (covering most income categories, except the "Less than $25,000 annual income" category), showing predominantly a gender effect. This illustrates how clustering techniques can be used in the context of defining meaningful groups of customers, leading to similar conclusions in this case.

As a result of Step 1, two groups of customers were identified as more sensitive to services than the rest of the population: women and high-income customers.

Step 2: Definition of Operating Segments. The results in Step 1 indicate that meaningful customer groups in the application can be defined by gender and income attributes. Indeed, this categorization could lead to as many as four relevant groups: $CL_1, \ldots, CL_4$. For simplicity, and as a result of sample size limitations, this section will use a two-group segmentation, i.e., either (i) women versus men, or (ii) low income versus high income as the meaningful customer groups. The goal is to identify the service attributes that are considered more relevant for each of these customer groups. As noted, the survey included several questions related to

Table 3. Stepwise Regression Selection of Most Important Service Attributes

	Call center	In-home repair
Men	—Deliver what they promise in regards to your product issues —Knowledgeable about your product issues —Committed to resolving your product issues —You only have to explain your product issues one time —Take no longer than three minutes to initially answer your call	—Committed to resolving your product issues —Actively listen to you when you are describing your product issues —Deliver what they promise in regards to your product issues
Women	—Able to resolve your product issues —Discuss your issues in a clear and direct manner so that you can easily understand	—Committed to resolving your product issues
Low income	—Able to effectively resolve your issues during the first phone call	—Actively listen to you when you are describing your product issues —Deliver what they promise in regards to your product issues
High income	—Committed to resolving your product issues —You only have to explain your product issues one time	—Clearly explain why the product failed and how they will be fixing it

customer evaluation of specific aspects of the service, both for the call center and in-home repair services. They define the set of available service attributes, P in this application. Stepwise regression was used to determine which attributes were most relevant in explaining the overall quality perception of the company's call center service and in-home repair service quality. The results are summarized in Table 3.

When comparing men and women, for example, the stepwise regression procedure revealed that the two most important service attributes for women were being "able to resolve your product issues" and the ability of the call center representatives to "discuss issues in a clear and direct manner so that you can easily understand." For men, on the other hand, these attributes were not among the most important attributes. The set for men included attributes such as "deliver what they promise in regard to your product issues," "you only have to explain your product issues one time," and "take no longer than three minutes to initially answer your call." Although both segments certainly care about getting their issues resolved when they interact with the company's call center, there are some meaningful differences in the value each group places on particular service attributes, e.g., with women giving higher value to understanding product issues in a clear manner and men placing more value on attributes such as rapid service time.

As an outcome of this two-step process, the set of operating segments $\{(CL_k, P_k)\}_{k:1,\ldots,k'}$ is obtained as reflected in Table 3. Section 6 discusses the link with the operational policies that need to be defined to serve multiple operating segments.

6. Managerial Implications

As noted in Figure 1, defining operating segments leads to defining operational policies for the delivery of differentiated services. More precisely, given the definition of the collection of operating segments $\{(CL_k, P_k)\}_{k:1,\ldots,k'}$, a hierarchy of operational decisions is reflected in specific policies s_k for delivering services for each operating segment (each of the elements in the "How" box in Figure 1). In particular, a firm should make decisions about the service metrics and target levels which are used to serve each operating segment. The resource plan will determine the capacity and capabilities of the resources that are to be deployed to support these service targets. The control policy will define the mechanism to determine how to deliver the differentiated service. Performance evaluation refers to the evaluation of the service delivery performance. Finally, a firm should consider the question of whether to adopt a service differentiation strategy in a way that is consistent with its overall competitive strategy and long-term planning goals. The interrelationship among these operational decisions makes the overall problem very difficult to solve. This section deconstructs this hierarchy through a collection of submodels that can be used to optimize the specific set of operational decisions, i.e., each element of the "How" box in Figure 1 will be discussed separately. For the purpose of illustration, the application to HDTVs and to the call center unit in particular is referenced when useful.

6.1. Service Targets

By definition, having multiple operating segments requires different priorities for various sets of service metrics. Moreover, defining appropriate service targets for each metric is the mechanism that firms can use for imposing these priorities as a part of their service delivery strategy. Some of the key issues to consider for this purpose include the range of targets for each metric, the endogeneity of market demand in reaction to the offered service targets, and the sensitivity of other steps in the decision hierarchy to the specified targets, as discussed next. For the sake of illustrating how to use

Table 4. Differentiated Service Metrics and Targets

Segment	Main operational service metric	Personnel capabilities	Operations decisions and targets
1	First-time resolution rate (FTR)	Empathetic, good communicator—i.e., "able to discuss your issues in a clear and direct manner so that you can easily understand"	FTR > 0.95 Prob(THT < 10 minutes) > 0.70 $N_1 = 10$ Routing rule = FQR
2	Total handling time (THT)	Deep product focus, i.e., "deliver what they promise with regards to product + knowledgeable about product + committed to resolving product issues"	FTR > 0.85 Prob (THT < 10 minutes) > 0.95 $N_2 = 10$ Routing rule = FCFS

existing models in the literature, an existing submodel formulation, which has been adapted to this case of a call center for after-sales services, is introduced.

Motivated by the results in Table 3, measurable segment-specific service metrics for the different groups can be defined. For illustration, consider the first-time resolution rate (FTR) and total handling time (THT) as the operational metrics in a two-segment case. Table 4 considers a potential realization for this case.

The relative values for the target levels for each segment-metric combination (Table 4) could be based on industry standards and an understanding of the strategic importance of each segment to the business. Alternatively, these targets and decisions could be generated as solutions to particular submodel optimization problems. Several papers have specifically modeled the decision of setting appropriate service targets for different customer segments. This paper notes that the single-metric case (i.e., differentiation based on resource availability levels exclusively) is possibly the approach most commonly encountered in the operations literature. For example, Gurvich et al. (2008) adopts different versions of this idea, considering a service-level constraint in which the probability that the waiting time W_k exceeds a certain threshold T_k for customer class k is bounded by class-specific service levels α_k—i.e., ($P\{W_k > T_k\} \leq \alpha_k$). Relatedly, Cohen et al. (2006b) report on Cisco's after-sales service support strategy, which offers its customers service-level agreements with different service standards for response time for support (i.e., ranging from several hours to several business days).

To further illustrate how marketing and operations factors interact through the market response to the service targets, an adaptation of the model formulated in Ho and Zheng (2004) to the HDTV case can be considered. Ho and Zheng (2004) use the gap model of service quality from marketing to select service performance targets in a service queueing setting where the demand rate is endogenously determined and capacity is fixed. The objective function is to maximize the demand rate λ, which is obtained based on an equilibrium condition as a function of the total market demand Λ, the firm's market share S, and consumer's utility U, as $\lambda = \Lambda \cdot S(U)$. Because a consumer's utility depends on the service attributes, the market share of a particular firm i is a function of the values of its service targets relative to the values for service offered by all of its (m) competitors: $S_i = e^{U_i}/(\sum_{j=1}^{m} e^{U_i})$. Consumer's utility U depends on the FTR rate and the THT; i.e., $U(\text{FTR}, \text{LTH}) = \beta_0 + \beta_1\,\text{FTR} + \beta_2\,\text{THT}$. The stepwise regressions described in Section 5 could inform the relative values for the coefficients β_i. The decisions are to set target levels for FTR and LTH for each segment for the customer call center. FTR is a property of the classes of servers that could be assigned to each segment; if T is the customer wait time and $1/\mu$ is the expected service time for the service call, then $\text{THT} = T + 1/\mu$. Thus, THT is endogenously determined through the selection of a target level, which in turn impacts the demand rate. Let

$$\varnothing[\text{FTR}, \text{THT}, \lambda] = \Lambda S[U(\text{FTR}, \text{THT}), F(\text{FTR}, \text{THT}, \lambda)], \tag{2}$$

which can be interpreted as "tomorrow's demand rate" given "today's demand rate." Market equilibrium is reached when tomorrow's demand rate is the same as today's rate, and the decision problem can thus be stated as maximizing λ subject to $\lambda = \varnothing[\text{FTR}, \text{THT}, \lambda]$. Ho and Zheng (2004) discuss conditions required for a unique equilibrium and extend the model to a case where there is competition among firms. One key attribute of this formulation is the ability to illustrate explicitly how the market responds to service targets (i.e., market shares and the level of demand for the service). This model, however, assumes that capacity is fixed. These capacity decisions are dealt within a different submodel in the current framework, discussed next.

6.2. Process Resource Plan

The definition of a resource plan involves several operational decisions. Perhaps the most relevant is that of service capacity management. Thus, Table 4 indicates the number of servers assigned to each group (i.e., N_1 and N_2). This step also defines the capability of the resources (e.g., agents) involved in the provision of the required services. Some servers may have flexible capabilities, i.e., the ability to serve more than one operational segment. Other servers may only be able to serve a particular segment. As noted previously, servers could also be characterized by their capability to deliver a level of FTR, which could be based on their past performance or training. Server capability, therefore, is a direct consequence of staff management decisions, e.g., training and hiring.

The resource plan allows the firm to operationalize its service delivery processes. Some key issues for this step include the trade-offs and constraints associated with meeting multiple service targets for multiple segments, the integration of resource planning with more detailed control decisions concerning customer prioritization, and the workforce scheduling of resources in a manner that meets the service targets for each segment. A wide range of models in the literature focus on optimizing capacity levels when facing multiple classes of customers. Such models take the demand rate as an input and generate capacity decisions as an output. Most of these studies use a single service metric with segment-specific service levels and jointly determine staffing capacity along with the real-time control rule, which allocates servers to arriving customers (Gurvich et al. 2008, 2010; Gurvich and Whitt 2010, Mehrotra et al. 2012). Such control rules define the mechanisms for matching each arriving customer to a particular queue staffed by those servers capable of meeting the requirements of the customer's particular segment. For example, Armony and Ward (2010) formulate an optimization problem for a call center with heterogeneous agent pools, in which each agent has a different speed of service. The optimization problem is formulated with the goal of minimizing customer wait time.

As noted earlier, the firm could decide to forego flexibility, dedicate capacity to each segment, and use these servers exclusively for customers within each group. This is analogous to what airlines do when they provide different call center contact numbers to different priority classes of customers (gold versus silver). An alternative model would be the use of one class of servers for all service interactions and ensuring that the service providers have capabilities sufficient to meet the needs of all segments, (i.e., a flexible capacity option). This approach would leverage scale economies through risk pooling.

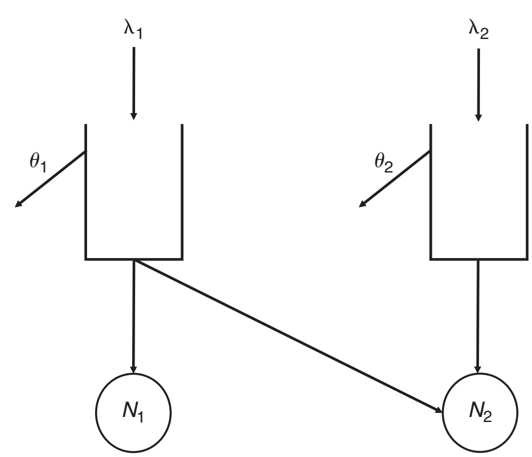
The definition of service targets, along with the capacity decisions, leads to definition of the capabilities and training requirements for the call center personnel who would be used to interact with each customer segment.

6.3. Process Control Policy

The control problem involves defining service priorities and using control mechanisms to determine the real-time routing of arriving customers through the service delivery process. There are multiple aspects to solving this problem, such as the use of service priority rules based on attributes of the servers and/or the segment, as well as rationing mechanisms based on the realized state of the system. There is extensive literature concerned with the definition and selection of such control policies for stochastic service systems. Such routing rules define the pool of servers to which an arriving customer is assigned, and the admission to service from a queue to a pool of servers following the completion of a service. Issues that have been considered in this literature include consideration of attributes of the routing rule. A common assumption is to restrict attention to fixed control policies whose rules do not depend on the state of the system and do not evolve dynamically. Rationing introduces control policies that are state dependent, i.e., where priority-based allocations may only be invoked, for example, when the system is heavily loaded and FCFS is used otherwise. Alternative priority rules exist; e.g., Gurvich and Whitt (2010) use the fixed-queue-ratio (FQR) rule, which assigns a customer waiting in the queue to an available server based on the realized wait times of the customers, according to the longest observed idle time of the servers. (Table 4 indicates how routing rules are assigned to each group in the current example.)

A recent stream of literature in this area has developed solutions to jointly solve the staffing and control problems. Gurvich et al. (2010) consider a call center model with multiple customer classes and agent types operating under quality–service constraints and demand rate uncertainty. Figure 2 represents a simplified visualization of their model, adapted to illustrate how resource and control decisions can be jointly considered. For the sake of illustration, it is assumed that the values for λ_1 and λ_2 have been determined (as a result of solving the model for setting service targets for each segment) and that there are two classes of

Figure 2. A Two-Class, Two-Pool Call Center Network



call center servers: one with the capabilities denoted in Table 4 for both segments and one with capabilities limited to the requirements for only one segment. The decisions are to determine the number of servers of each type to be assigned to the call center along with a routing rule that specifies (i) which server should be assigned to an arriving customer, if there are multiple agents available and capable of serving that customer, and (ii) which customer should be assigned to a newly available server, given that there are customers waiting in the queues that this agent can serve. Thetas represent abandonment rates of the two operating segments.

In Gurvich et al. (2010), a general version of this problem is solved with the objective of minimizing the total cost of employing servers required to satisfy a quality of service constraint. In particular, the model selects the segment-specific upper bound on THT, determined in the solution to the submodel for setting service targets, as well as the routing policy from a class of admissible routing rules. The decisions are N_1 and N_2 (number of servers) and the choice of a routing rule. As noted earlier, this satisfies the second performance metric, FTR, by selection of an appropriate class of agents who could be used for each segment.

The introduction of multiple classes of agents and customers provides an opportunity for the firm to adopt a rationing mechanism for allocating customers to servers on a real-time basis. In this case, the allocation mechanism would be a function of the state of the system, i.e., restricting particular agents to specific segments for which they are uniquely qualified when the system is heavily loaded, and otherwise all agents will be available to all customers on an FCFS basis. A related example is Deshpande et al. (2003), who introduce a threshold inventory rationing model for delivering differentiated service to multiple customer classes. This model was developed for the allocation of inventory to support maintenance and repair. As noted in that model, it is possible to maintain priorities by using separate inventory stockpiles dedicated to each customer segment (analogous to no flexibility for the servers in the current example). Doing so introduces a cost, because of the loss of the benefit of pooling. An alternative is to use a common stockpile for all customer classes, which maintains the pooling advantage but leads to inappropriate priorities, e.g., if one sets the service level by rounding up the target to the level associated with the highest-priority customer class. The trade-offs introduced in this model further illustrate the operational control decisions for service delivery.

6.4. Performance Evaluation

As noted in the extensive SERVQUAL literature (Zeithaml et al. 1996), to evaluate service performance, customers' perceptions can be contrasted to expectations, and the gap between both can be used to monitor service quality. For example, for a given service metric, one can calculate the ratio of the perceptions score to the expectations score to evaluate whether the company is exceeding customers' expectations.

A similar approach was used in the application with the HDTV OEM, measuring the ratio of perceptions to expectations to monitor service KPIs such as the ones in Table 3. Indeed, based on this study, the company was motivated to use customer-centric service KPIs, in addition to the efficiency KPIs traditionally used in this industry. To do so, a survey similar to the one in the online appendix, but focusing on expectations regarding what an excellent company would do, was implemented to get a score for the expectations of customers regarding each service metric. The scores from the expectations survey can be combined with the scores from the perceptions survey (e.g., using averages) to obtain the KPI ratios. Figure 3 displays the results obtained for the subset of metrics that were found to be most relevant for each segment in Table 3, for the case of the company's call center.

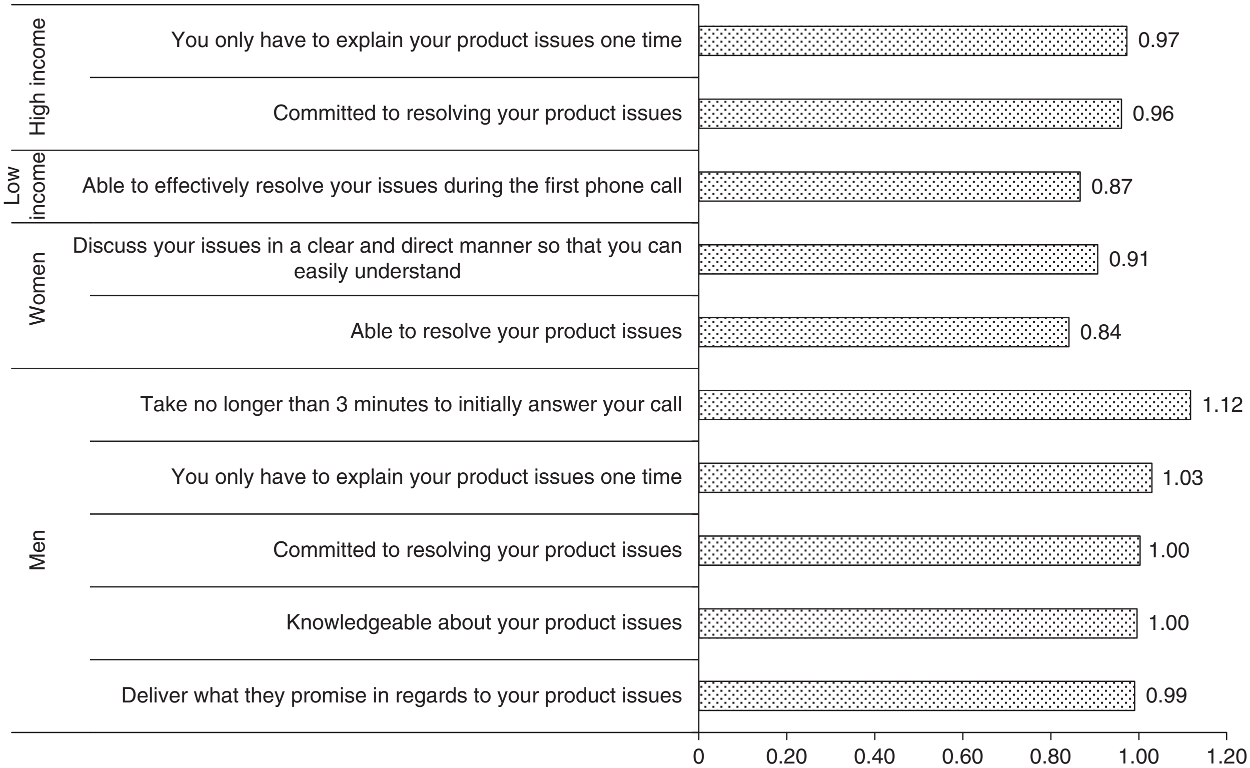
As shown, the company was exceeding expectations in some cases (e.g., response times for the men's segment) but was also below expectations for some metrics (e.g., FTR rate for the low-income segment). More importantly, this example illustrates how the proposed KPIs can be used to track service performance for the metrics of relevance for each operating segment specifically, as opposed to an overall (undifferentiated) performance evaluation that would usually be the norm when a company does not use service differentiation. Tracking service KPIs at the segment level allows companies to more easily identify improvement opportunities that may be specific for each segment, and it provides yet another illustration of how the use of operating segments can change operational decisions and metrics under differentiation.

6.5. Service Differentiation Decision

All of the steps introduced by this framework answer the question of *how* a firm should differentiate its services. From this input, a firm should ultimately evaluate the costs and benefits brought by service differentiation, then contrast them with the costs and benefits of providing a single service. In other words, a firm should ultimately consider the question of *whether* to offer differentiated services. This decision could be based on an evaluation of the predicted costs and service performance under each scenario, i.e., differentiated or not, possibly based on the resource planning models discussed earlier or through a simulation study.

As noted earlier, the OM literature has recently considered this question in different ways, such as, e.g., in the context of queueing models (Sainathan 2015). As detailed in Section 2, there is a growing OM literature

Figure 3. Service KPIs = Perceptions/Expectations



contrasting pooling and dedicated services. As extrapolated from this paper, an overall answer for the question involves a number of complexities. For example, arriving at an "optimal" solution for the differentiated services scheme is in itself a difficult task because of the nature of the interactions between the decisions in different steps. This discussion of the various submodels illustrates some of the most important trade-offs.

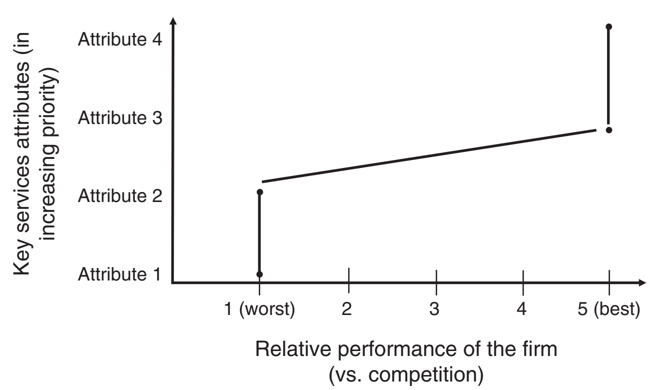
7. Limitations and Extensions

It is important to note that the HDTV application in this paper ignores the consideration of the impact of competition by firms providing similar services to the market. The attribute map introduced by Frei and Morriss (2012) addresses this issue by combining an operating segment with an assessment of the relative performance for each attribute, by the firm, when compared to the performance in those same attributes by competing firms. The attribute map is illustrated in Figure 4, where the vertical axis can correspond to the definition of an operating segment for a particular group of customers.

Attribute maps suggest that when delivering services, firms will be more competitive for some attributes than for others. If the map exhibits the shape in Figure 4 (i.e., increasing), the firm is most successful in those attributes to which it attaches a higher priority. A key observation made by Frei and Morriss is that an effective service strategy requires a firm to accept that it is not strategic to be good at everything. Data on the relative performance of a firm with respect to its competitors would be needed to draw the attribute map, which can be used as an evaluation tool for a suggested service strategy.

The definition of operating segments does not need to be constrained to the use of customer demographics data. Indeed, the case study illustrated the notion of operating segments using demographic data (e.g., men versus women, low-income versus high-income customers), given that this information was all that was available to characterize a customer. This is a

Figure 4. The Attribute Map



limitation of the analysis and in no way the specific results obtained in the application to HDTVs should be taken as general guidance for defining operating segments in other settings; rather, the application serves the purpose of illustrating how the general methodology can be applied. Relatedly, although the use of demographic variables have some appealing features (they are easy to collect and can make the operationalization of the policies simple), enriching the characterization of a customer (e.g., by measuring actual behavior, as opposed to preferences responses to survey instruments) would make possible refined definitions of operating segments, possibly leading to a better characterization of service needs in contrast to a definition based solely on customer demographics. More generally, and considering that the main distinction behind operating segments is the heterogeneous service priorities and operational requirements inherent in optimally serving different groups of consumers, much more refined definitions of operating segments than the one used in this application are possible.

Although the methodology proposed in this paper is in principle very general, additional considerations can be useful when dealing with high-dimensional data. For example, if the space of possible service attributes becomes very large, additional tools from machine learning can be incorporated in the definition of operating segments. Indeed, machine learning research has been concerned with the problem of feature selection in high-dimensional spaces for a long time (see Blum and Langley 1997), and a wide range of tools including cross-validation, methods based on regularization parameters, and dimensionality reduction methods has been developed and can be easily incorporated into the current proposed empirical framework when the dimensionality of the data grows (see Hastie et al. 2009). For example, LASSO (Tibshirani 1996) could be used to perform variable selection using principles from regularization.

Finally, there are a variety of implications of this framework for research in the service operations field. These implications include addressing model formulations and solution algorithms to better capture the interactions between the various decisions denoted in the framework. Empirical research could be conducted to test and evaluate hypotheses derived from the framework in different industry settings, e.g., to explore the differences between firms that do adopt a differentiation strategy and those that do not. The availability of richer data on the benefits achieved by service differentiation (beyond the likelihood to recommend the brand, which was the only performance metric available in the current application) and the costs (unavailable in the case of the HDTV application) should also facilitate further empirical examination of the involved trade-offs.

8. Conclusions

This paper proposed a conceptual framework for service differentiation, highlighting in particular the role of operating segments in defining an overall service differentiation strategy. It presented a general methodology to identify operating segments and showed an application in the context of a consumer electronics OEM to illustrate how the methodology can be used in practice. It also discussed the managerial implications associated with implementing operational policies required to support a service differentiation strategy. This discussion also expounded upon how existing models and methods can be applied to support the adoption of the framework.

Regarding the application of the framework in the context of after-sales services for product–service bundles, the proposed approach is contrary to current practice at the OEM, where, for example, only one type of call center employee is hired and trained, consistent with their current nondifferentiated service approach. Their practice, in particular, was to use KPIs that are not differentiated by segments and to deliver service by common resources on an FCFS basis. It is quite common for companies to offer such nondifferentiated services; thus, there is a significant opportunity to improve practice based on the segment differentiation concepts suggested by the results.

The discussion in Section 6 included some of the main drivers of the costs and benefits of a service differentiation strategy. In practice, those costs include issues such as fairness, brand weakening, and increased complexity. Although ultimately these issues will be context specific, the proposed framework and methodology for identifying operating segments provide valuable conceptual guidance for firms interested in implementing such a differentiation strategy. Finally, it is the hope of the authors that this paper could be a step toward focusing the operations community on the important problem of managing differentiation in service settings, a domain in which differentiation strategies have not been sufficiently studied.

References

Abhishek V, Guajardo JA, Zhang Z (2016) Business models in the sharing economy: Manufacturing durable goods in the presence of peer-to-peer rental markets. Working paper, Carnegie Mellon University, Pittsburgh.

Allon G, Federgruen A (2009) Competition in service industries with segmented markets. *Management Sci.* 55(4):619–634.

Allon G, Federgruen A, Pierson M (2011) How much is a reduction of your customers' wait worth? An empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing Service Oper. Management* 13(4):489–507.

Armony M, Ward AR (2010) Fair dynamic routing in large-scale heterogeneous-server systems. *Oper. Res.* 58(3):624–637.

Bassamboo A, Zeevi A (2009) On a data-driven method for staffing large call centers. *Oper. Res.* 57(3):714–726.

Blum A, Langley PP (1997) Selection of relevant features and examples in machine learning. *Artificial Intelligence* 97(1–2):245–271.

Buell R, Campbell D, Frei FX (2016a) How do customers respond to increased service quality competition? *Manufacturing Service Oper. Management* 18(4):585–607.

Buell R, Campbell D, Frei FX (2016b) The customer may not always be right: Customer compatibility and service performance. Harvard Business School Working Paper 16-091, Harvard University, Cambridge, MA.

Cachon PC, Harker PT (2002) Competition and outsourcing with scale economies. *Management Sci.* 48(10):1314–1333.

Campbell D, Frei F (2011) Market heterogeneity and local capacity decisions in services. *Manufacturing Service Oper. Management* 13(1):2–19.

Chen T, Kalra A, Sun B (2009) Why do consumers buy extended service contracts? *J. Consumer Res.* 36(12):611–623.

Cohen MA, Whang S (1997) Competing in product and service: A product life-cycle model. *Management Sci.* 43(4):535–545.

Cohen I, Cohen MA (2017) On sourcing and stocking policies in a two-echelon, multiple location, repairable parts supply chain. *J. Oper. Res. Soc.* 68(6):617–629.

Cohen MA, Agrawal N, Agrawal V (2006a) Winning in the aftermarket. *Harvard Bus. Rev.* 84(5):129–138.

Cohen MA, Agrawal N, Agrawal V (2006b) Achieving breakthrough service delivery through dynamic asset deployment strategies. *Interfaces* 36(3):259–271.

Deshpande V, Cohen MA, Donohue K (2003) A threshold inventory rationing policy for service differentiated demand classes. *Management Sci.* 49(6):683–703.

Ernst R, Cohen MA (1990) Operations related groups (ORGs): A clustering procedure for production/inventory systems. *J. Oper. Management* 9(4):574–598.

Eppen GD (1979) Effects of centralization on expected costs in a multi-location newsboy problem. *Management Sci.* 25(5): 498–501.

Fader PS, Lattin JM (1993) Accounting for heterogeneity and nonstationarity in a cross-sectional model of consumer purchase behavior. *Marketing Sci.* 12(3):304–317.

Frei F, Morriss A (2012) *Uncommon Service: How to Win by Putting Customers at the Core of Your Business* (Harvard Business Review Press, Cambridge, MA).

Gilbert SM, Weng ZK (1998) Incentive effects favor nonconsolidating queues in a service system: The principal-agent perspective. *Management Sci.* 44(12):1662–1669.

Guajardo JA, Cohen MA, Netessine S (2016) Service competition and product quality in the U.S. automobile industry. *Management Sci.* 62(7):1860–1877.

Guajardo JA, Weber R, Miranda J (2010) A model updating strategy for predicting time series with seasonal patterns. *Appl. Soft Comput.* 10(1):276–283.

Gurvich I, Whitt W (2010) Service-level differentiation in many-server service systems via queue-ratio routing. *Oper. Res.* 58(2):316–328.

Gurvich I, Armony M, Mandelbaum A (2008) Service level differentiation in call centers with fully flexible servers. *Management Sci.* 54(2):279–294.

Gurvich I, Luedtke J, Tezcan T (2010) Staffing call centers with uncertain demand forecasts: A chance constrained optimization approach. *Management Sci.* 56(7):1093–1115.

Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J. Machine Learn. Res.* 3(March):1157–1182.

Halek M, Eisenhauer J (2001) Demography of risk aversion. *J. Risk Insurance* 68(1):1–24.

Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, New York).

Ho TH, Zheng Y-S (2004) Setting customer expectation in service delivery: An integrated marketing-operations perspective. *Management Sci.* 50(4):479–488.

Iyengar R, Jedidi K, Kohli R (2008) A conjoint approach to multipart pricing. *J. Marketing Res.* 45(2):195–210.

Kohavi R, John G (1997) Wrappers for feature selection. *Artificial Intelligence* 97(1–2):273–324.

Koller D, Sahami M (1996) Toward optimal feature selection. *Proc. 13th Internat. Conf. Machine Learn.* (Morgan Kaufmann, San Diego), 284–292.

Lal K, Chapelle O, Weston J, Elisseeff A (2006) Embedded methods. *Feature Extraction, Foundations Appl., Stud. Fuzziness Soft Comput.* 207(5):167–185.

Lindsey C, Sheather S (2010) Variable selection in linear regression. *Stata J.* 10(4):650–669.

Liu Y, Ram S, Lusch R, Brusco M (2010) Multi-criterion market segmentation: A new model, implementation and evaluation. *Marketing Sci.* 29(5):880–894.

Lu Y, Olivares M, Musalem A, Schilkrut A (2013) Measuring the effect of queues on customer purchases. *Management Sci.* 59(8): 1743–1763.

Mandelbaum A, Reiman MI (1998) On pooling in queueing networks. *Management Sci.* 44(7):971–981.

Mehrotra V, Ross K, Ryder G, Zhou Y-P (2012) Routing to manage resolution and waiting time in call centers with heterogeneous servers. *Manufacturing Service Oper. Management* 14(1):66–81.

Mittal V, Kamakura W (2001) Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics. *J. Marketing Res.* 38(2):131–42.

Moorthy K (1984) Market segmentation, self-selection and product line design. *Marketing Sci.* 3(4):288–307.

Moorthy K, Png I (1992) Market segmentation, cannibalization and the timing of product introductions. *Management Sci.* 38(3): 345–359.

Png I, Reitman D (1994) Service time competition. *RAND J. Econom.* 25(4):619–634.

Propper C (1995) The disutility of time spent on the United Kingdom's national health service waiting lists. *J. Human Resources* 30(4):677–700.

Ratchford B (2001) The economics of consumer knowledge. *J. Consumer Res.* 27(4):397–411.

Sainathan A (2015) Prioritization and price-plus-delay competition with self-selecting, heterogeneous, time-sensitive customers. Working paper, Nanyang Business School, Nanyang Technological University, Singapore.

Shunko M, Niederhoff J, Rosokha Y (2018) Humans are not machines: The behavioral impact of queueing design on service time. *Management Sci.* 64(1):453–473.

Song H, Tucker A, Murrell K (2015) The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Sci.* 61(12):3032–3053.

Tibshirani R (1996) Regression shrinkage and selection via the Lasso. *J. Roy. Statist. Soc. Ser. B* (*Methodological*) 58(1):267–288.

van Dijk NM, van der Sluis E (2008) To pool or not to pool in call centers. *Production Oper. Management* 17(3):296–305.

van Dijk NM, van der Sluis E (2009) Pooling is not the answer. *Eur. J. Oper. Res.* 197(1):415–421.

Wang Y, Cohen MA, Zheng Y-S (2002) Differentiating customer service on the basis of delivery lead-times. *IIE Trans.* 34(11):979–989.

Wind Y (1978) Issues and advances in segmentation. *J. Marketing Res.* 15(3):317–337.

Wind Y, Bell D (2008) Market segmentation. Baker M, Hart S, eds. *The Marketing Book*, 6th ed. (Elsevier, Oxford, UK), 222–244.

Wind J, Green P, Shifflet D, Scarbrough M (1989) Courtyard by Marriott: Designing a hotel facility with consumer-based marketing models. *Interfaces* 19(1):25–47.

Zeithaml V, Berry L, Parasuraman A (1996) The behavioral consequences of service quality. *J. Marketing* 60(4):31–46.