

## Research Article

# Staffing to Maximize Profit for Call Centers with Impatient and Repeat-Calling Customers

Jun Gong,<sup>1</sup> Miao Yu,<sup>1</sup> Jiafu Tang,<sup>2</sup> and Manru Li<sup>1</sup>

<sup>1</sup>College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

<sup>2</sup>College of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian 116026, China

Correspondence should be addressed to Jun Gong; [gongjun@ise.neu.edu.cn](mailto:gongjun@ise.neu.edu.cn)

Received 1 May 2015; Accepted 6 July 2015

Academic Editor: Tadeusz Kaczorek

Copyright © 2015 Jun Gong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Motivated by call center practice, we study the optimal staffing of many-server queues with impatient and repeat-calling customers. A call center is modeled as an  $M/M/s+M$  queue, which is developed to a behavioral queuing model in which customers come and go based on their satisfaction with waiting time. We explicitly take into account customer repeat behavior, which implies that satisfied customers might return and have an impact on the arrival rate. Optimality is defined as the number of agents that maximize revenues net of staffing costs, and we account for the characteristic that revenues are a direct function of staffing. Finally, we use numerical experiments to make certain comparisons with traditional models that do not consider customer repeat behavior. Furthermore, we indicate how managers might allocate staffing optimally with various customer behavior mechanisms.

## 1. Introduction

Most call center operations can be modeled as queuing systems without feedback from the state of the queue to the arrivals [1]. The majority of research in queuing takes the customer arrival rate as an external factor that is a well-known distribution with no consideration for the potential impact of returns. In fact, arrival with feedback plays an important role in service operations, as a long queue deters some potential customers. For example, if “treated well,” customers might return to form feedback, but dissatisfied customers might never return. However, the potential impact of feedback on the arrival rate has received little attention.

In call center settings, customer satisfaction in terms of waiting time usually affects the feedback arrival rate. The service encounter is unlike face-to-face service encounters at other service sites, such as restaurants, hotels, and banks. In a call center with invisible queues, there is no service environment, so that the best means of providing and controlling customers' satisfaction may be providing products and service efficiently and quickly [2]. Therefore, waiting time

becomes the main issue that impacts customer satisfaction, hurts repeat business, and jeopardizes the company's long-term profit in call centers. In our case, we assume that waiting time is the only factor affecting customer satisfaction; the result is a change of arrival rate as an internal factor.

This paper formulates and solves the staffing problem: determining how many agents should be assigned to achieve the waiting time objective. Optimality of staffing has mostly been viewed as a cost minimization issue [3, 4]. Few papers have focused on call center profit when they make a capacity sizing decision. With this focus, it is difficult to reflect the loss of profit resulting from abandonment and lost repeat behavior. Here, we characterize customer service satisfaction by focusing on waiting time to determine the economically optimal staffing.

Currently, more managers have begun to emphasize the importance of service quality by balancing customer satisfaction and staffing cost. Additionally, some research has also begun to focus on abandonment [5, 6]. We posit that customer repeat behavior is a key controllable determinant based on customer satisfaction with waiting time. This means

that a satisfied customer will have a higher preference to again choose service from the same call center than a dissatisfied customer will. In the field of customer relationship management, Anderson and Sullivan [7] have shown that customer satisfaction is a good predictor for the likelihood of repeated purchases and revenue growth, and prior research has shown that customers react negatively to poor service by abandoning the firm [8]. Furthermore, customer satisfaction can be increased by more staffing in a call center. The key determinant of customer satisfaction is mainly waiting time in the system. A firm can increase its repeat customers by increasing investment in staffing.

The main contributions of this paper can be summarized as follows. (1) We develop an analytical behavioral model with a queuing system based on customer perception of waiting time. This is the first generalization of the queueing model to incorporate customer satisfaction and repeated behavior. (2) We define the staffing optimality of call center as the number of servers maximizing the profit, including repeated revenues and abandonment cost. In sum, our innovative point is to derive a precise relationship between revenue and repeat customers by developing a microlevel queueing model.

The basic model of this paper is based on a queueing model with feedback [9]. Further, we capture the situation in which the arrival rate with repeat customers depends on customer satisfaction with waiting time. The equilibrium arrival rate is calculated by approximating the analytical methods with a Monte Carlo point algorithm. In terms of the queueing model, we extend the Erlang-A expression by formulating customer repeat behavior and develop an approach to calculate performance based on the results of Jouini et al. [10]. We derive a closed-form formula for determining staffing that maximizes a call center's profit. The paper focuses on an inbound call center as an important example of such systems, as in Aksin and Harker [11].

The rest of the paper is organized as follows. In Section 2, we review the literature. In Section 3, we formulate impatient and repeat behavior in a queueing system. In Section 4, we use some extensive numerical experiments to show the role of repeat behavior in maximizing call center profit. In Section 5, we discuss the implications of our study under various circumstances. In Section 6, we generalize the key insight derived from the analysis and conclude with the limitations of our results and future research directions.

## 2. Literature Review

In recent years, call centers have become sophisticated in their use of technology; for a review of the state of call center research, see Gans et al. [12] and Aksin et al. [13]. Most research has focused on models for queueing, staffing, and performance analysis.

The simplest queueing model typically used for a call center is the  $M/M/s$  queue. The Erlang-C system ignores blocking and customer abandonment. The Erlang-B model characterizes the blocking probability for the call center, which is described as  $M/M/s/s$ . The last  $s$  represents  $s$

lines. The simplest model that includes abandonment is the Erlang-A model, which was further developed to incorporate customer impatience and to consider busy signals in the Erlang-C model [14]. Many researchers have made significant progress on various approximation schemes because multiserver queues with abandonment are difficult to analyze exactly.

Another issue in modeling abandonment in call center was emphasized by Garnett et al. [14], Gans et al. [12], and Mandelbaum and Zeltyn [15]. Garnett et al. [14] proposed an asymptotic analysis of a Markovian model with abandonment in the heavy-traffic regime. They mainly characterized the relationships between staffing, the offered load, and system performance measures such as the probability of waiting time and the probability of abandonment. This can be viewed as an extension of Halfin and Whitt [16] by adding abandonment.

Staffing problems are a type of resource allocation problem. This type of optimality has been defined in the economic sense as in Andrews and Parsons [17]. They assume that revenue is a direct function of staffing. Most papers consider only how staffing affects call center operation cost. Borst et al. [18] considered how to determine an asymptotically optimal staffing level to minimize staffing and waiting time costs. Harrison and Zeevi [19] considered how to minimize call center staffing costs and additional abandonment costs. Customer behavior research on staffing is mainly focused on abandonment.

There exists a large body of literature on customer service quality in queueing systems [20, 21]. A call center is a typical type of invisible queue, which is a type of special service system. de Véricourt and Zhou [22] introduce the call-resolution probability to model call quality. Mehrotra et al. [23] expand it to a routing problem with heterogeneous servers. Generally, call centers use average call waiting time and call waiting probability as measures of service quality, but these measures depend on the actual type of service. Gans et al. [12] posit that the incorporation of human factors is a real challenge for call centers. All of these papers treat customer service quality similarly, as a cost indicator or to satisfy the service quality constraints.

Several authors have begun to study customer behavior as the psychology of the queue, for example, Luo et al. [24], Kumar et al. [25], Carmon et al. [26], and Zhou and Soman [27]. They focus on evaluating how waiting time affects customer satisfaction and how to minimize customer dissatisfaction with the waiting process. However, there are a few papers going one step further. Haxholdt et al. [9] and van Ackere et al. [28] construct a queueing model with feedback and exhibit the phenomenon that the arrival rate depends on waiting time. They exploit some simulation methods with the decision variable of service rate. Recently, van Ackere et al. [29] have started to develop a behavioral model in which customers come and go based on waiting time with a simulation method. Here, we address a similar issue using an analytical method of queueing.

Increasingly, companies are focusing on customer repeat behavior to enhance long-term profitability; much research is beginning to address the issue of the relationship between

customer satisfaction and repeat behavior. Some research concludes that customer satisfaction results in repeat business and increases the firm's profitability [30, 31]. Law et al. [32] focus on the effect of waiting time on repurchasing behavior and customer satisfaction in some service industries. Likewise, Chen [33] posits that the customer satisfaction level is an important factor that may affect the effectiveness of a loyalty program. In addition, customer satisfaction with waiting time affects customer loyalty [34]. In the call center field, Dean [35] notes that some issues of service quality could affect customer loyalty to a call center, and he investigates some consumers of an insurance company and bank using call centers to validate his perspectives. Few studies have addressed the effect of staffing on repeated customer behavior.

Loyalty programs are a structured marketing effort to reward repeat behavior. There is a recent large literature in the area, notably Kim et al. [36] and Singh et al. [37], who treat customer valuation of the product as a random variable. Gandomi and Zolfaghari [38] incorporate the valuation as both a deterministic variable and a stochastic variable for setting product prices, and they make use of customers' surpluses (valuation less price) for driving the probability that customers choose purchase behavior. These papers enable the expression of repeat probability in loyalty programs.

### 3. Model Formulation

We present a model with repeat and impatient customer behaviors. Our model has two sectors, representing the feedbacks of repeat behavior and abandonment. The performance metric of abandonment is the loss of customers based on waiting time; furthermore, the metric of satisfaction with waiting experiences is used to build a link between staffing costs and call center customer revenues. We consider a call center model with a single class of customers and homogeneous and parallel agents.

*3.1. Analysis of Process-Related Metrics of Call Center.* Firstly, the model of the abandonment behavior is developed by the extension of the Erlang-A formula, which can be viewed as an M/M/s+M queuing system with feedback. Let  $T$  denote the random variable measuring patience times with rate  $\gamma$ . The queuing discipline is First-Come-First-Served (FCFS).

The system parameters are

$\lambda_e$ : equilibrium arrival rate following a Poisson distribution;

$\mu$ : service rate ( $1/\mu$  is the average service time) with an exponential distribution;

$s$ : number of servers (decision variable);

$\rho$ : load on system.

Let  $\tau$  denote the acceptable delay, which is the threshold of repeated customer behavior evaluating the service in this time. Customers who can enter service before  $\tau$  are really considered as being satisfied. In practice, such a metric may not be accounted for by managers because of ignoring repeated behavior. A reasonable value of  $\tau$  is approximately

20 seconds. We define  $P_c$  as the service level at which customers react to the waiting time with repeat behavior; this satisfaction index  $P_c$  determines whether the customer will return later or never, that is, the repeat probability. We can write this service level as

$$P_c = \frac{\text{Number answered} \leq \tau}{\text{Number offered}}. \quad (1)$$

Next, we derive the expression of the service level  $P_c$ . The virtual waiting time is defined as the waiting time of customers assuming that they do not renege (abandon). Let  $V$  denote the random variable of the virtual waiting time of a tagged, infinitely patient customer. Note that "answered" means  $V \leq T$  and "reneged" means  $V > T$ . We denote the renege probability by  $P_R$  and the answer probability by  $P_S$ . Let  $W$  denote the waiting time of an arbitrary customer in the queue, the sojourn time of a customer who leaves the queue as a result of renege or start of service:

$$\begin{aligned} P_S &= 1 - P_R, \\ W &= \min \{V, T\}. \end{aligned} \quad (2)$$

$P_c$  can be taken as a function of the random variables  $V$ ,  $T$ , and  $W$ . We can obtain

$$P_c = P(V \leq \tau, V < T). \quad (3)$$

Based on the work of Zeltyn and Mandelbaum [39], we define the following building blocks for performance analysis and refer to Jouini et al. [10] regarding performance. Let patience be generally distributed with cumulative distribution function  $G(x)$ ,  $x \geq 0$ , and  $\bar{G}(x)$  is given by

$$\bar{G}(x) = 1 - G(x). \quad (4)$$

Define  $H(x)$  and  $J(t)$ . Their expressions are then given by

$$\begin{aligned} H(x) &= \int_0^x \bar{G}(u) du, \\ J(t) &= \int_t^\infty e^{\lambda_e H(x) - s\mu x} dx. \end{aligned} \quad (5)$$

We can determine  $\bar{G}(x)$  and  $H(x)$  in closed form using the specific functional form of the exponential patience times; we obtain

$$\begin{aligned} \bar{G}(x) &= e^{-\gamma x}, \\ H(x) &= \frac{1 - e^{-\gamma x}}{\gamma}. \end{aligned} \quad (6)$$

$B(s, \lambda_e/\mu)$  denotes the blocking probability in M/M/s queue, and we define

$$\varepsilon = \frac{\sum_{i=0}^{s-1} (\lambda_e/\mu)^i / i!}{(\lambda_e/\mu)^{s-1} / (s-1)!} = B\left(s-1, \frac{\lambda_e}{\mu}\right)^{-1}. \quad (7)$$

Given these, it is straightforward to derive the probability density function of the virtual waiting time  $V$  for  $x > 0$ :

$$v(x) = \frac{\lambda_e e^{\lambda_e H(x) - s\mu x}}{\varepsilon + \lambda_e J(0)} \quad (8)$$

$$= \frac{\lambda_e e^{\lambda_e H(x) - s\mu x}}{\left(\sum_{i=0}^{s-1} (\lambda_e/\mu)^i / i!\right) / \left((\lambda_e/\mu)^{s-1} / (s-1)!\right) + \lambda_e J(0)}.$$

Then, for the origin value,

$$P(V=0) = \frac{\varepsilon}{\varepsilon + \lambda_e J(0)}. \quad (9)$$

And the probability to renege of  $P_R$  is

$$P_R = P(V > T) = \frac{1 + (\lambda_e - s\mu) J(0)}{\varepsilon + \lambda_e J(0)}. \quad (10)$$

Hence, the service level  $P_c$  can be obtained from (3) and the building blocks in the following equation:

$$P_c = P(V=0) + \int_0^\tau \bar{G}(x) v(x) dx \quad (11)$$

$$= \frac{\varepsilon}{\varepsilon + \lambda_e J(0)} + \int_0^\tau \bar{G}(x) \frac{\lambda_e e^{\lambda_e H(x) - s\mu x}}{\varepsilon + \lambda_e J(0)} dx.$$

Finally, based on the work of Jouini et al. [10], we can obtain

$$P_c = \frac{\varepsilon + s\mu (J(0) + J(\tau)) + e^{\lambda_e H(\tau) - s\mu\tau}}{\varepsilon + \lambda_e J(0)}. \quad (12)$$

**3.2. Dimensioning a Call Center with Repeat Customer Behavior.** We consider the background that a firm sells a good or service through a call center in which customers purchase service repeatedly, so this kind of call center is different from what offers after-sales and technical support services. Based on this assumption, customer calls can be taken as purchases and their rates of occurrence depend on the satisfaction of the most recent experiences. Customers use call center to get service at a higher rate when they are satisfied than when they are dissatisfied. Then, we model customer satisfaction referring to Ho et al. [40], and a closed-form formula is derived for dimensioning the call center. This closed-form formula reveals that approximating the mixture arrival processes by a single aggregate Poisson process can systematically estimate the revenue of call center with repeated behavior.

In the queue model-handling method, we refer to more details on analysis of call centers with retrials by Aguir et al. [41] and Pustova [42]. The proportion is modeled as the parameter  $\theta$ .  $\theta$  is the probability that a customer becomes a potential repeat customer in the next period.  $(1 - \theta)$  is the proportion of customers that fail to proceed to the call center. The potential probability of repeat behavior  $\theta$  depends greatly on the specific industry of the call center. For example,  $\theta$  is expected to be relatively higher for a call center providing

a ticket booking service because the service frequency is sufficiently high that people would prefer a service with a higher service quality. We consider different levels of repeat behavior,  $\theta \in [0, 1]$ . We can directly derive the equilibrium arrival rate  $\lambda_e$  and take  $P_c$  as the repeat fraction joining the queue in the following analysis. Let  $\lambda_{\text{new}}$  be the original arrival rate of customers. We assume that  $\lambda_{\text{new}}$  depends greatly on the size of the call center. Customers who have reneged have experienced the process with relatively lower satisfaction. For ease of calculation, assume that they will no longer join the queue, which means that the repeated probability of reneging customers is zero. Hence, we obtain

$$\lambda_e = \lambda_{\text{new}} + \theta P_c \lambda_e. \quad (13)$$

Then, this equilibrium model falls into the class of product form networks from Baskett et al. [43], so the stationary behavior of the queueing model will not depend on the distribution of repeated behavior delays. Therefore, they can be ignored. Furthermore, we should discuss how to compute  $\lambda_e$ . Specifically, (13) can be denoted in a continuous function  $f$  in  $\lambda_e$ :

$$f(\lambda_e) = \frac{\lambda_{\text{new}}}{1 - \theta P_c} = \lambda_e. \quad (14)$$

Then, the solution of the equilibrium arrival rate becomes the root of (14), where  $P_c$  is functions of  $\lambda_e$ , and  $\lambda_{\text{new}}$  is a constant. Therefore, the equilibrium rate  $\lambda_e$  can be obtained with the Monte Carlo point algorithm (see, e.g., [44]).

In what follows, we want to numerically compute the optimal staffing level with repeat behavior. Based on the firm's profit from selling a service or product with repeat customer behavior, we find that the firm's expected profit can be stated as

$$\max_s \text{TP} = c_1 \lambda_e \cdot P_S - c_2 s. \quad (15)$$

The first term is the customer expected service income, which can be expressed as  $c_1 \cdot \lambda_e \cdot P_S$ , where  $c_1$  is the customer service income parameter; given an average revenue per customer served, there is customer loss from some reneged customers, such that system revenues net of loss is  $\lambda_e \cdot P_S$ . The second term is the human resource cost  $c_2 \cdot s$ , which is in the form of salaries. With a fixed cost per server, the staffing cost increases monotonically with the number of servers.

Based on the model in (15), we can find and evaluate the optimal values of  $s$ . The staffing problem formulated above is to be solved with the renege probability and equilibrium rate embedded in the revenue expression. After obtaining the equilibrium arrival rate  $\lambda_e$ , we make use of the enumeration method to obtain the desired economically optimal staffing solution.

## 4. Numerical Experiments

We make use of the enumeration method to obtain the optimal staffing level  $s$  and validate the model. To illustrate the model with different scenarios, we consider the regular model without repeat customers ( $\theta = 0$ ), three models of

partial customer repeat behavior  $\theta = 0.2, \theta = 0.5, \theta = 0.8$ , and a model with full repeat customers  $\theta = 1$ . We then compare their profit and staffing through experiments with these three scenarios and draw some conclusions, in contrast to traditional research without repeat behavior. The common parameters are  $\lambda_{\text{new}} = 5, \tau = 1/3, c_1 = 5, c_2 = 2, \mu = 1$ , and  $\gamma = 0.1$ .

**4.1. Comparison between the Models with and without Repeat Behavior.** If the potential probability of repeat behavior  $\theta$  is equal to 0, the equilibrium arrival rate is just the original value, and this situation is the traditional research setting without repeat customers. Figure 1 shows that there is no doubt that the staffing cost curve presents a linear increasing trend with increasing staffing, and the revenue always rises with increasing arrival rate. This trend results in the linearly decreasing profit.

Call centers in different service industries have different sensitivities to waiting time. The models of partial customer repeat behavior can demonstrate these different situations. These three models have almost similar trends, except for the different rate of reaching the optimal staffing. Figure 2 shows that the revenue rises and then levels off, which drives the profit trend. As staffing increases and begins to reach the optimal point, TP increases as a result of more staffing, leading to more customer arrival with repeat behavior. TP will then decrease because of the unchanged revenue. The system reaches its optimal staffing at the highest point of revenue growth; the maximum point of profit appears once the rate of revenue growth is lower than the rate of staffing growth. On the other side, it conforms to the law of diminishing returns (see, e.g., [45]), which states that the marginal increase in service level declines in the staffing level.

Thirdly, full repeat customers means waiting time has absolute influence on customer repeat behavior; that is,  $\theta = 1$  in Figure 3. In practice, there is no situation with complete repeat behavior,  $\theta = 1$ ; that is, “the satisfaction trap” will always exist and some customers will not choose repeat behavior however high satisfaction is. Therefore, the parameter is always set to  $\theta < 1$ . However, for comparing this case to the models with zero and partial repeat behaviors, the trend of profit with full repeat can be seen in a small system as a result of the linearly increasing revenue. In theory, this is the effect of full repeat behavior, so that optimal staffing can be reached only in the relatively larger system. Specific rules are drawn in the latter.

**4.2. Comparison of the Metrics.** Let us consider the impact of the metrics  $P_c, \lambda_e,$  and  $P_s$  on the profit of a call center and then analyze the variation of profits. These three metrics can themselves almost predict the profit of a call center. First, note that  $P_c$  plays a key role for the equilibrium rate according to the profit trend in Figure 4, especially if the repeat probability of the model is zero. As  $P_c$  increases to its maximum, the equilibrium arrival rate also reaches its maximum because of the effect of repeat behavior. Figure 5 illustrates the trend of the equilibrium arrival rate according to  $P_c$  and shows that the larger the potential repeat probability  $\theta$  is, the larger the

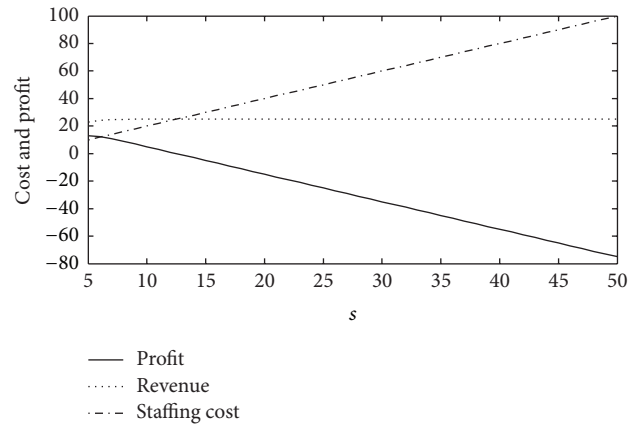


FIGURE 1: Staffing cost and revenue without repeat behavior  $\theta = 0$ .

effects of repeat behavior and the equilibrium arrival rate are. In particular, the effect of repeat behavior is so large that the equilibrium arrival rate cannot reach a maximum even at the staffing level of 50 in the model with full repeat behavior, and the repeat probability also cannot reach value 1.

Furthermore, the probability for entering service can demonstrate the trend of the reneging probability of the system. Figure 6 shows that abandonment always decreases with increasing staffing level. It also shows that higher staffing levels will result in more customers obtaining service up to the maximum profit, whatever the equilibrium arrival rate is. In terms of revenue, the equilibrium arrival rate and the probability of obtaining service both play roles in increasing revenue with increasing staffing level.

## 5. Implications for Different Customer Behavior Mechanisms

In this section, the optimal staffing and profit will be analyzed under different assumptions of customer behavior mechanisms, including potential repeat probability, first arrival rate, and sensitivity to waiting time. The common parameters are  $c_1 = 5, c_2 = 2, \mu = 1$ , and  $\gamma = 0.1$ .

**5.1. Implications for Potential Repeat Probability.** We conducted the analysis involving the optimal staffing and profit with different value of  $\theta$ . The feasible staffing level of a call center was set between 5 and 100. We set the parameters  $\tau = 1/3$  and  $\lambda_{\text{new}} = 5$ . Figure 7 shows that the larger the parameter  $\theta$  is, the more optimal the staffing  $s$  is. This means that a larger  $\theta$  will result in more demand for service because of the repeat behavior effect, so managers need to deploy more agents for more repeat customers.

In terms of growth, both the repeat probability  $P_c$  and staffing  $s$  have roles in the equilibrium arrival rate as a result of  $\lambda_e = \lambda_{\text{new}} / (1 - \theta P_c)$ ; there is a change in the rate of growth at  $\theta = 0.7$ . Figure 8 shows that it is a reflection of the cumulative effect of repeat behavior; any increase of  $\theta$  means more demand from customers, so that the system has higher repeat probability  $P_c$ . More staffing is needed to provide satisfaction.

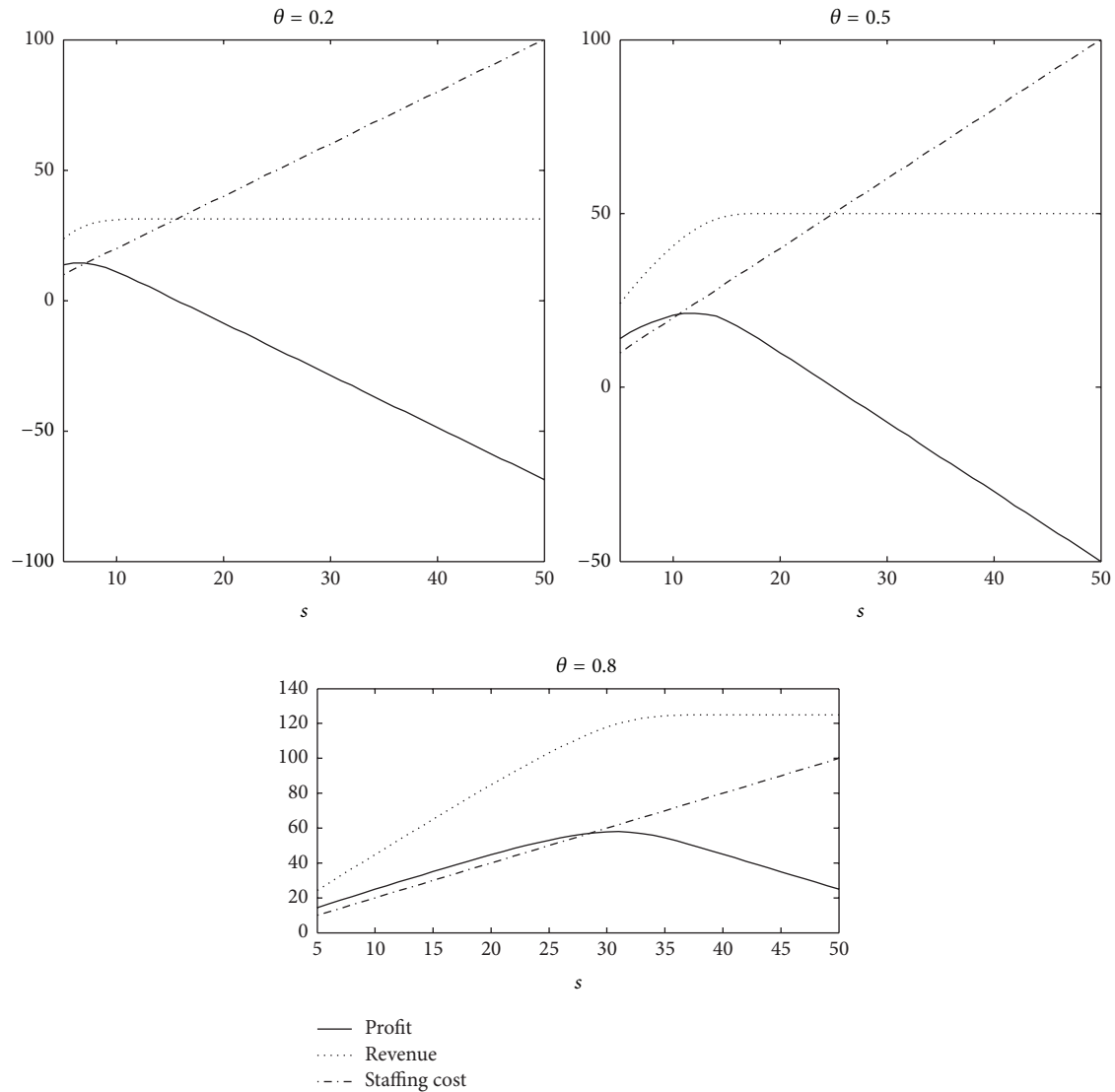


FIGURE 2: Staffing cost and revenue with partial repeat behavior.

In parallel,  $P_S$  must reach its top value within a small range. This cumulative effect leads to an amplified equilibrium rate, and it appears to be the beginning of a cumulative effect when  $\theta = 0.7$ . The rapidly increasing equilibrium arrival rate can bring more profits.

**5.2. Implications for the First Arrival Rate.** Call centers in different service industries have different performance measures of the first time arrival rate, so the proportions of  $\lambda_{\text{new}}$  are also different. The effect of the first time arrival rate  $\lambda_{\text{new}}$  is presented in Table 1, which shows the performance with optimal staffing and profit. We set the potential probability parameter  $\theta$  to 0.5, and  $\tau = 1/3$ . Table 1 shows that the larger the parameter  $\lambda_{\text{new}}$  is, the higher the optimal profit is.

Any increase of  $\lambda_{\text{new}}$  means more demand from customers, so that the system has a higher repeat probability  $P_c$  and needs more agents. The repeat probability  $P_c$  is balanced nearly to 1, so that  $\lambda_e$  would increase to the maximum

TABLE 1: The impact of first arrival customers on call center.

	$\lambda_{\text{new}}$						
	5	7	10	15	20	30	50
$s$	12	17	23	34	45	65	106
TP	21.35	31.82	48.14	76.03	104.4	162.09	279.06
$\lambda_e$	9.14	13.23	18.93	28.93	39.02	58.64	98.55
$P_S$	0.992	0.995	0.995	0.996	0.996	0.996	0.997
$P_c$	0.906	0.942	0.943	0.963	0.975	0.977	0.985
$\rho$	0.762	0.778	0.823	0.851	0.867	0.902	0.93

value. The increasing rate needed for staffing is less than the equilibrium rate because of the interaction, which can result in increasing load until the repeat probability  $P_c$  reaches nearly 1.

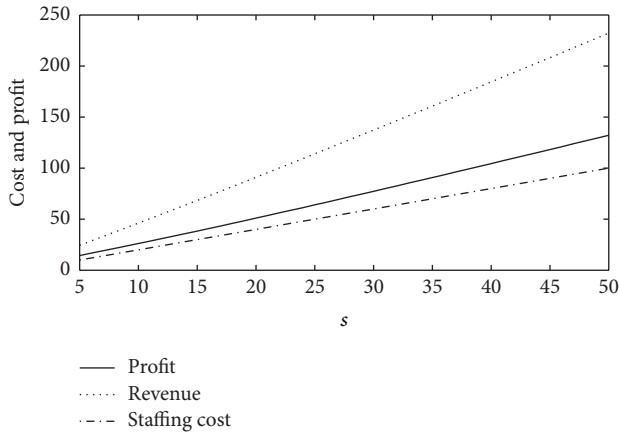


FIGURE 3: Staffing cost and revenue with full repeat behavior  $\theta = 1$ .

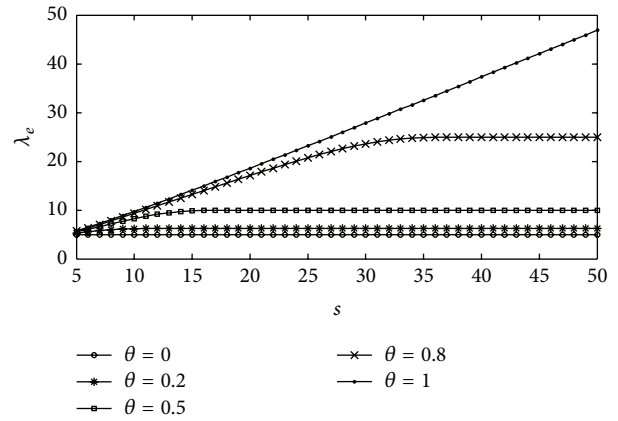


FIGURE 5: The equilibrium arrival rate at different values of  $s$ .

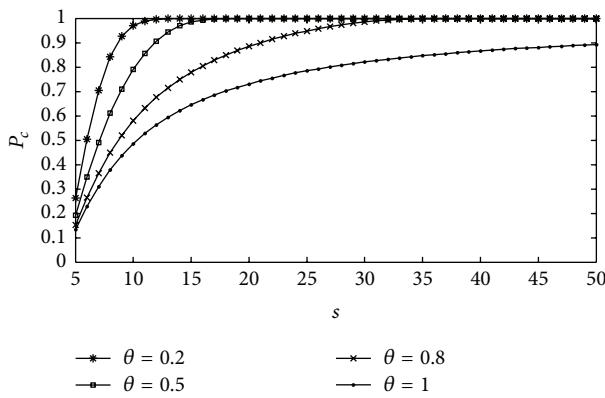


FIGURE 4: The repeat probability at different values of  $s$ .

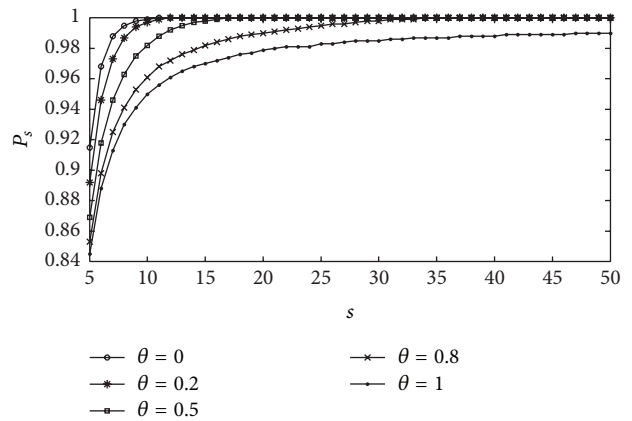


FIGURE 6: The probability to obtain service at different values of  $s$ .

5.3. *Implications for Sensitivity to Waiting Time.* As the independent CDF variable of waiting time, the sensitivity to  $\tau$  controls the probability of repeat behavior. We set the parameter  $\theta = 0.5$ , and  $\lambda_{new} = 5$ . To some extent, we can take the parameter  $\tau$  as the sensitivity to waiting time for all customers. Next, we explore the effect of  $\tau$  on the profit of the call center. Figure 9 shows the impact of  $\tau$  on the call center profit with the staffing increasing. As expected, the optimal profit is always greater when  $\tau$  is greater because of customers' relative insensitivity to waiting time, so that customers would have much higher satisfaction.

Figure 9 shows that the larger the parameter  $\tau$  is, the more the optimal profit of call center is because of the need for fewer agents and costs. In a specific industry, customers can be divided into different types based on their sensitivity to waiting time. For example, if  $\tau$  is rescaled to the range  $[0, 1]$ , we can take it as the range of high sensitivity to time in which most customers usually have less patience; therefore, managers need to assign more staff to satisfy demand. Then, the profit will be relatively low. As  $\tau$  increases to more than 3, we note that the optimal profit is almost constant when the marginal sensitivity to waiting time gradually decreases

to zero. That is, for callers insensitive to waiting time, when  $\tau$  is long, the system deals with all calls within some time less than  $\tau$ ; the call center can obtain an optimal profit with the top equilibrium arrival rate.

## 6. Conclusion

We present a deterministic queuing system with repeat and impatient behavior, assuming that satisfied customers will return and disappointed customers will not. Based on a queuing model, we formulate and analyze the staffing that maximizes the profit. This model of customer behavior agrees better with business reality and preserves the Markovian property.

Numerical analysis illustrates that repeat behavior has an important impact on the optimal staffing. Ignoring this behavior could lead to significant understaffing. Our analysis of the interrelationship between repeat behavior and service satisfaction is expressed by the performance measure of waiting time, which can be derived from research in fully deterministic queuing analysis.

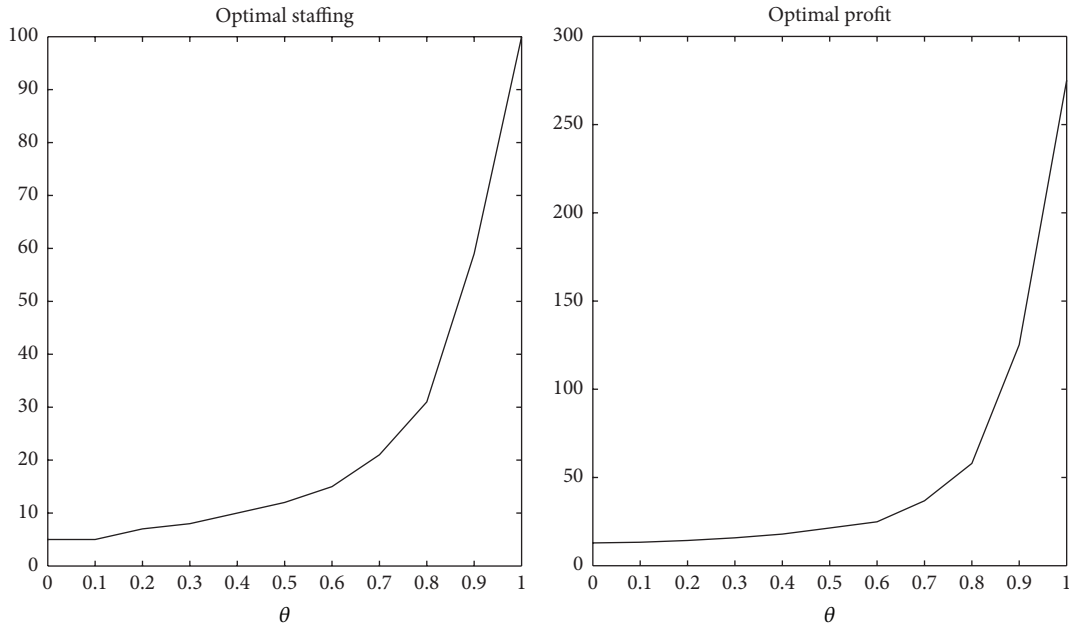


FIGURE 7: Relative optimal staffing and profit with different  $\theta$ .

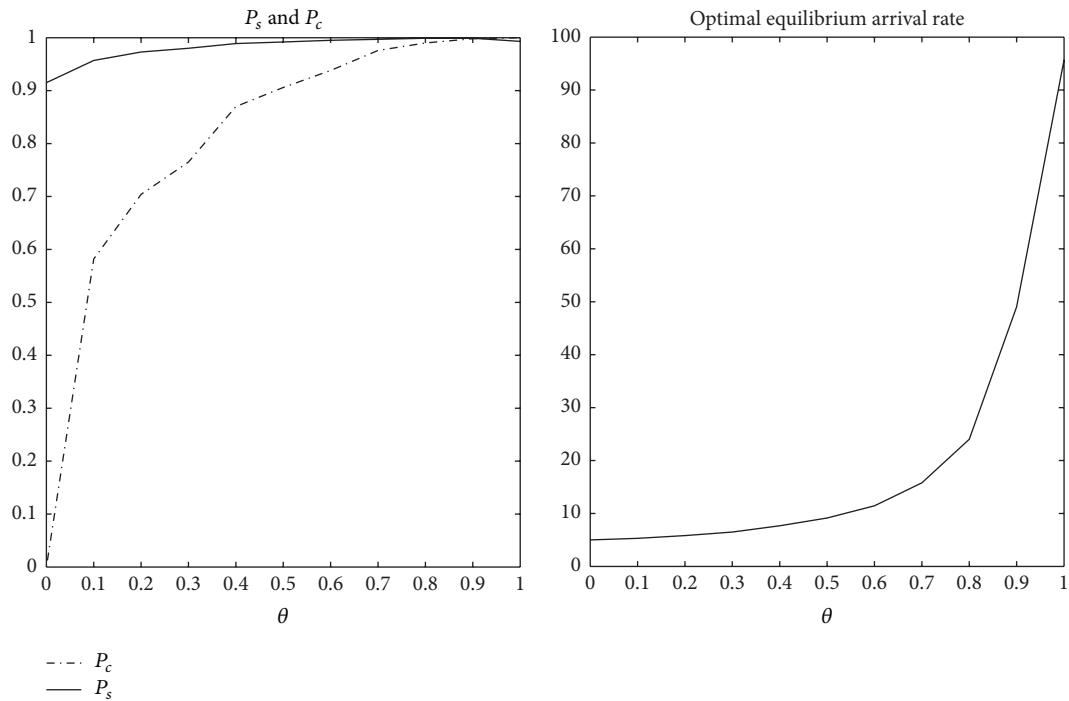
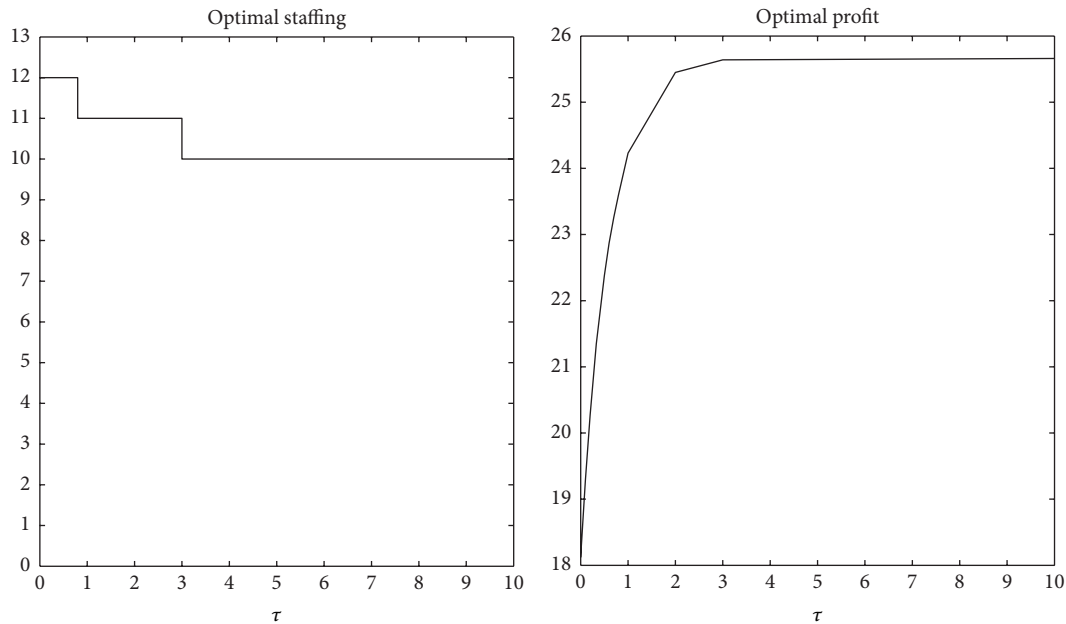


FIGURE 8: The impact of  $P_c$ ,  $P_s$ , and  $\lambda_e$  on the optimal staffing level.

We have additionally constructed some efficiency analyses intended to help managers deploy staffs for their particular call centers; the analyses allow different sets of input parameters, such as potential probability, first arrival rate, and sensitivity to waiting time, which play important roles in optimal staffing and profit. We focus exclusively on customer

behavior, including both the economic considerations of staffing in a simple call center and the impact of delays on customer satisfaction. This latter effect is an obvious first direction for extending this research. Finally, with real call center data and customer choices, this work may be especially useful for managers making staffing decisions.



FIGURE 9: Impact of the sensitivity to waiting time  $\tau$ .

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This project is supported by the National Nature Science Foundation under Grant 71271052 and Fundamental Research Funds for the Central Universities (N120804003).

## References

- [1] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, Wiley, New York, NY, USA, 3rd edition, 1999.
- [2] E. K. W. Lau and W.-S. Chan, "An integrative framework capturing customer satisfaction and service quality in call centers," *Review of Business Research*, vol. 12, no. 3, pp. 76–82, 2012.
- [3] A. Mandelbaum, W. A. Massey, M. I. Reiman, A. Stolyar, and B. Rider, "Queue lengths and waiting times for multiserver queues with abandonment and retrials," *Telecommunication Systems*, vol. 21, no. 2–4, pp. 149–171, 2002.
- [4] O. Baron and J. Milner, "Staffing to maximize profit for call centers with alternate service-level agreements," *Operations Research*, vol. 57, no. 3, pp. 685–700, 2009.
- [5] A. I. Pazgal and S. Radas, "Comparison of customer balking and renegeing behavior to queueing theory predictions: an experimental study," *Computers & Operations Research*, vol. 35, no. 8, pp. 2537–2548, 2008.
- [6] O. Jouini, Z. Akşin, and Y. Dallery, "Call centers with delay information: models and insights," *Manufacturing and Service Operations Management*, vol. 13, no. 4, pp. 534–548, 2011.
- [7] E. W. Anderson and M. W. Sullivan, "The antecedents and consequences of customer satisfaction for firms," *Marketing Science*, vol. 12, no. 2, pp. 125–143, 1993.
- [8] G. J. Fitzsimons, "Consumer response to stockouts," *Journal of Consumer Research*, vol. 27, no. 2, pp. 249–266, 2000.
- [9] C. Haxholdt, E. R. Larsen, and A. van Ackere, "Mode locking and chaos in a deterministic queueing model with feedback," *Management Science*, vol. 49, no. 6, pp. 816–830, 2003.
- [10] O. Jouini, G. Koole, and A. Roubos, "Performance indicators for call centers with impatient customers," *IIE Transactions*, vol. 45, no. 3, pp. 341–354, 2013.
- [11] O. Z. Aksin and P. T. Harker, "Capacity sizing in the presence of a common shared resource: dimensioning an inbound call center," *European Journal of Operational Research*, vol. 147, no. 3, pp. 464–483, 2003.
- [12] N. Gans, G. Koole, and A. Mandelbaum, "Telephone call centers: tutorial, review, and research prospects," *Manufacturing & Service Operations Management*, vol. 5, no. 2, pp. 79–141, 2003.
- [13] O. Z. Aksin, M. Armony, and V. Mehrotra, "The modern call center: a multi-disciplinary perspective on operations management research," *Production and Operations Management*, vol. 16, no. 6, pp. 665–688, 2007.
- [14] O. Garnett, A. Mandelbaum, and M. Reiman, "Designing a call center with impatient customers," *Manufacturing and Service Operations Management*, vol. 4, no. 3, pp. 208–227, 2002.
- [15] A. Mandelbaum and S. Zeltyn, "Staffing many-server queues with impatient customers: constraint satisfaction in call centers," *Operations Research*, vol. 57, no. 5, pp. 1189–1205, 2009.
- [16] S. Halfin and W. Whitt, "Heavy-traffic limits for queues with many exponential servers," *Operations Research*, vol. 29, no. 3, pp. 567–588, 1981.
- [17] B. Andrews and H. Parsons, "Establishing telephone-agent staffing levels through economic optimization," *Interfaces*, vol. 23, no. 2, pp. 14–20, 1993.

- [18] S. Borst, A. Mandelbaum, and M. I. Reiman, "Dimensioning large call centers," *Operations Research*, vol. 52, no. 1, pp. 17–34, 2004.
- [19] J. M. Harrison and A. Zeevi, "Dynamic scheduling of a multi-class queue in the Halfin-Whitt heavy traffic regime," *Operations Research*, vol. 52, no. 2, pp. 243–257, 2004.
- [20] Z. J. Ren and Y.-P. Zhou, "Call center outsourcing: coordinating staffing level and service quality," *Management Science*, vol. 54, no. 2, pp. 369–383, 2008.
- [21] K. S. Anand, M. F. Paç, and S. Veeraraghavan, "Quality-speed conundrum: trade-offs in customer-intensive services," *Management Science*, vol. 57, no. 1, pp. 40–56, 2011.
- [22] F. de Véricourt and Y.-P. Zhou, "Managing response time in a call-routing problem with service failure," *Operations Research*, vol. 53, no. 6, pp. 968–981, 2005.
- [23] V. Mehrotra, K. Ross, G. Ryder, and Y.-P. Zhou, "Routing to manage resolution and waiting time in call centers with heterogeneous servers," *Manufacturing and Service Operations Management*, vol. 14, no. 1, pp. 66–81, 2012.
- [24] W. Luo, M. J. Liberatore, R. L. Nydick, Q. B. Chung, and E. Sloane, "Impact of process change on customer perception of waiting time: a field study," *Omega*, vol. 32, no. 1, pp. 77–83, 2004.
- [25] P. Kumar, M. U. Kalwani, and M. Dada, "The impact of waiting time guarantees on customers' waiting experiences," *Marketing Science*, vol. 16, no. 4, pp. 295–314, 1997.
- [26] Z. Carmon, J. G. Shanthikumar, and T. F. Carmon, "A psychological perspective on service segmentation models: the significance of accounting for consumers' perceptions of waiting and service," *Management Science*, vol. 41, no. 11, pp. 1806–1815, 1995.
- [27] R. Zhou and D. Soman, "Looking back: exploring the psychology of queuing and the effect of the number of people behind," *Journal of Consumer Research*, vol. 29, no. 4, pp. 517–530, 2002.
- [28] A. van Ackere, C. Haxholdt, and E. R. Larsen, "Long-term and short-term customer reaction: a two-stage queueing approach," *System Dynamics Review*, vol. 22, no. 4, pp. 349–369, 2006.
- [29] A. van Ackere, C. Haxholdt, and E. R. Larsen, "Dynamic capacity adjustments with reactive customers," *Omega*, vol. 41, no. 4, pp. 689–705, 2013.
- [30] N. Au, E. W. T. Ngai, and T. C. E. Cheng, "A critical review of end-user information system satisfaction research and a new research framework," *Omega*, vol. 30, no. 6, pp. 451–478, 2002.
- [31] W. Nie, "Waiting: integrating social and psychological perspectives in operations management," *Omega*, vol. 28, no. 6, pp. 611–629, 2000.
- [32] A. K. Y. Law, Y. V. Hui, and X. Zhao, "Modeling repurchase frequency and customer satisfaction for fast food outlets," *International Journal of Quality and Reliability Management*, vol. 21, no. 5, pp. 545–563, 2004.
- [33] L.-F. Chen, "A novel approach to regression analysis for the classification of quality attributes in the Kano model: an empirical test in the food and beverage industry," *Omega*, vol. 40, no. 5, pp. 651–659, 2012.
- [34] F. Bielen and N. Demoulin, "Waiting time influence on the satisfaction-loyalty relationship in services," *Managing Service Quality*, vol. 17, no. 2, pp. 174–193, 2007.
- [35] A. M. Dean, "Service quality in call centers: implications for customer loyalty," *Managing Service Quality*, vol. 12, no. 6, pp. 414–423, 2002.
- [36] B.-D. Kim, M. Shi, and K. Srinivasan, "Reward programs and tacit collusion," *Marketing Science*, vol. 20, no. 2, pp. 99–120, 2001.
- [37] S. S. Singh, D. C. Jain, and T. V. Krishnan, "Customer loyalty programs: are they profitable?" *Management Science*, vol. 54, no. 6, pp. 1205–1211, 2008.
- [38] A. Gandomi and S. Zolfaghari, "Profitability of loyalty reward programs: an analytical investigation," *Omega*, vol. 41, no. 4, pp. 797–807, 2013.
- [39] S. Zeltyn and A. Mandelbaum, "Call centers with impatient customers: many-server asymptotics of the  $M/M/n+G$  queue," *Queueing Systems*, vol. 51, no. 3-4, pp. 361–402, 2005.
- [40] T.-H. Ho, Y.-H. Park, and Y.-P. Zhou, "Incorporating satisfaction into customer value analysis: optimal investment in lifetime value," *Marketing Science*, vol. 25, no. 3, pp. 260–277, 2006.
- [41] M. S. Aguir, F. Karaesmen, O. Z. Aksin, and F. Chauvet, "The impact of retrials on call center performance," *OR Spectrum*, vol. 26, no. 3, pp. 353–376, 2004.
- [42] S. V. Pustova, "Investigation of call centers as retrieval queuing systems," *Cybernetics and Systems Analysis*, vol. 46, no. 3, pp. 494–499, 2010.
- [43] F. Baskett, K. M. Chandy, R. R. Muntz, and F. G. Palacios, "Open, closed, and mixed networks of queues with different classes of customers," *Journal of the Association for Computing Machinery*, vol. 22, pp. 248–260, 1975.
- [44] R. Y. Rubinstein, *Simulation and the Monte Carlo Method*, John Wiley & Sons, New York, NY, USA, 1981.
- [45] G. Koole and A. Pot, "A note on profit maximization and monotonicity for inbound call centers," *Operations Research*, vol. 59, no. 5, pp. 1304–1308, 2011.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

