

BEHAVIOURAL QUEUING WITH INTERACTING CUSTOMERS AND SERVICE PROVIDERS: A SIMULATION BASED APPROACH

Carlos Arturo Delgado-Alvarez
Department of Industrial
Engineering
ICESI University
Calle 18 No. 122-135, Cali,
Colombia
E-mail: cadelgado@icesi.edu.co

Ann van Ackere
HEC, School of Business and
Economics
University of Lausanne
Dorigny, 1015, Lausanne,
Switzerland
E-mail: ann.vanackere@unil.ch

Erik R. Larsen
Institute of Management
University of Lugano
6904, Lugano, Switzerland
E-mail: erik.larsen@usi.ch

KEYWORDS

Queuing problems, endogenous service and arrival rates, cellular automata (CA), adaptive expectations.

ABSTRACT

We address a service facility problem with captive interacting customers and service providers. This problem is modelled as a deterministic queuing system. Customers must routinely decide which facility to join for service, whereas service providers must decide how much to adjust the service capacity of their facilities. Both service providers and customers base their decisions on their perceptions about the system. Customers use their previous experience and that of their neighbours to update their perceptions about the average sojourn time, while service providers form their perceptions based on the queue length. We use cellular automata (CA) to model the interaction between customers and service providers. We perform a simulation to assess the way the customers' and service providers' decisions evolve and affect the system behaviour. Our results show that the more conservative the service providers, the larger the market share they achieve and the lower probability that their facility closes down.

INTRODUCTION

Most queuing problems are modelled assuming static conditions, and exogenous arrival and service rates. They are analysed in steady-state, despite the fact that they are dynamic and that agents' decisions depend on the state of the system. The analysis of queuing problems could be aimed either at optimising performance measures to improve the operating characteristics of the system without accounting for customer behaviour or at understanding the agents' behaviour through the analysis of their decisions.

Over the last decades, some researchers have attempted to move away from these predominant assumptions of traditional queuing theory towards a more dynamic context in which agents' decisions are increasingly considered. The present paper goes in this direction. We will focus on studying the behavioural aspects of queuing problems by using deterministic simulation.

A very broad range of studies has addressed the behaviour of customers and service providers in queuing problems. Nevertheless, this literature is scattered and not well-organised. The literature related to customer behaviour has been broadly discussed by Delgado (2012). The research on customer behaviour in queuing systems has been mainly tackled by marketing researchers, who study the relationships among waiting times, customer satisfaction and service quality in service facilities (Davis and Heineke 1998; Hui and Tse 1996; Taylor 1994). These studies attempt to understand the influence of waiting time on customer satisfaction, customer loyalty and service quality (Bielen and Demoulin 2007; Law, et al. 2004). Their aim is to endow service providers with information about customers' attitudes to enable them to redesign their service facility accordingly. For a review of the literature, see in Bielen and Demoulin (2007) and Gallay (2010).

As service providers (throughout the remaining of the paper we will use the term "managers" instead of service providers) value their customers because they increase the value of the firm, customers value their time (Delgado et al. 2011a). "Time is money", such as the adage says. Whatever the service customers require, waiting for service represents a waste of time for them which affects their utility. This impact is even stronger when customers repeatedly patronise a facility for service. When customers perceive that their utility is being affected, they look for another manager who maximises their utility. Some examples of this kind of systems include car owners who annually or biannually (depending on the country) must choose a garage for the emission, students or workers who daily has to choose an hour and/or a restaurant for lunch, a person who goes monthly to the bank to pay her bills, and a person who goes weekly to the supermarket, and so forth. In all these examples, customers take into account their previous experience to decide which facility to use for service. This experience enables them to choose the time and/or place that they consider less crowded. There are also situations in which customers are not necessarily humans. For instance, customers could be a stack of files waiting at an office to be dealt with, jobs at a factory waiting to be performed or dispatched, and vehicles waiting at a garage to be repaired, among

others. Still, note that behind each of these objects is a human being waiting (e.g. the final customer of the garage is the car owner).

The complexity of the relationship between managers and customers increases in real life when many service facilities compete to render a same service. In this case, the managers' actions will affect their future decisions, those of the customers as well as those of the rival managers.

Concerning the managers' decisions, the literature have focused on analysing policies of optimal pricing and capacity decisions to control problems associated with congestion in service facility systems. P. Naor (1969) is the seminal paper on this subject. He formalised the insight originally formulated by W. A. Leeman (1964) and then discussed by T. L. Saaty (1965) and W. A. Leeman (1965). These authors suggest using pricing to help reduce queues in many service systems. Naor's model was subsequently generalised by Yechiali (1971), Edelson (1971), Edelson and Hilderbrand (1975), Stidham Jr. (1985, 1992), Mendelson and Whang (1990), Dewan and Mendelson (1990), among others. More recently, Sinha, et al. (2010) applied an optimal pricing scheme of surplus capacity to control the joint problem of existing and potential customers who are differentiated according to a pre-specified service quality level.

Although some managers' strategies effectively consider either the demand or the supply perspective when adjusting their service capacity, optimal strategies should incorporate the perspective of the two conflicting parts of the system (Pullman and Thompson 2002). Our research is motivated by the logic behind this assertion and the complexity of the interaction between the decisions of customers and managers in service facility systems. Consequently, our modelling approach considers a service facility system where competing facilities render a service which customers require routinely. Each facility has its own queue and manager. Queues are assumed to be invisible to the customers. We assume that customer interact with their neighbours and share information about their most recent experience. They use their experience and that of their best performing neighbour to update their expectations of their previously chosen queue and the one used by their quickest neighbour. Then, based on their expectations, customers choose a facility for the next time. Managers take their decision to adjust service capacity on the basis of their desired service capacity which they determine based on their perception of the queue length at their facility and a market reference sojourn time. This market reference is a benchmark whereby managers compete with each other to attract more customers to their facilities. In other words how managers perform compared to this benchmark is a competitiveness index of the facilities.

In order to study this complex problem we propose an idealised queuing model with reactive and adaptive customers and managers in which the decisions of both types of agents are interdependent. This model is built using a cellular automata-based framework. The interaction between customers is portrayed in a one-dimensional cell lattice. The main structure of the cellular automata (CA) model is similar to that of the model proposed by Delgado et al. (2011a and 2011b). However, the factors which determine the average sojourn time customer experience at the facilities are different. While in those papers this experience depended only on customers' decisions because the service capacity remained constant (i.e. exogenous service rates), now this experience is also influenced by the managers' decisions (i.e. endogenous service capacity).

Our results show that the more conservative a manager is, the larger his market share. Additionally, his facility is less likely to close down. Similarly, the facilities of reactive managers are less likely to remain in operation in the long term.

This paper is organised as follows. The next section describes the one-dimensional CA model we use to study the agents' behaviour in a multichannel service facility system. We deal in turn with the managers' and the customers' decision rules. Then, we describe the managers' and customers' profile depending on the model parameters. The next section presents the simulation results. We perform an experiment in which we analyse the influence which the different managers' parameters have on the performance of the facilities. Finally, we present the conclusion and contributions of the paper.

MODEL DESCRIPTION

Consider the queuing system and the CA model explained in Delgado et al. (2011a and 2011b). This model represents a fixed population of N reactive and adaptive customers arriving periodically at a service facility system. Each period, they must choose one between m facilities for service. Delgado et al. (2011a and 2011b) consider endogenous arrival rates (λ_{ji}) and exogenous service rates (μ). We deviate from these papers by assuming both arrival and service rates as endogenously determined. We endow managers with the ability to adjust the facilities' service capacity. In this sense, we model a system in which customers are free to choose a facility for service and the managers adjust their capacity depending on the customers' behaviour. The managers' actions can either encourage or discourage customers to use a certain facility (Delgado 2012). Consequently, the average sojourn time (i.e. customer's experience) depends on both the customers' and managers' decisions. In other words, the ability of a facility (state) to be more attractive for customers (cells) than the others depends on the behaviour of all agents in the system. For the remaining

of the paper we will use the term “agents” when referring to both the managers and the customers.

Customers’ Decisions

We model the service facility’s customers as a social network of colleagues, friends, or neighbours who interact in a one-dimensional K -neighbourhood (K is the number of neighbours each customer interacts with on each side). When customers patronise a facility for service, they experience an average sojourn time (W_{jt}) which depends on the arrival rate (λ_{jt}) and the service rate (μ_{jt}) of that facility. This average sojourn time represents a congestion measure of the facilities and is given by Equation (1). This equation is proposed and explained by Delgado et al. (2011b) and considers that arriving customers can temporarily surpass the service rate in a transient period, but it also satisfies the behavioural characteristics of steady-state.

$$W_{jt} = \frac{\lambda_{jt}}{\mu_{jt}^2} + \frac{1}{\mu_{jt}} \quad (1)$$

Customers use their most recent experience to form expectations of the average sojourn time for their most recently chosen facility. Customers also share their experiences with their neighbours and use this information to update their expectations regarding the average sojourn time for the previously facility chosen by their quickest neighbour. Then, we endow customers with computational memory which enables them to update their expectations for their previously patronised facility and that chosen by their best performing neighbour.

We follow Delgado et al. (2011a) and assume that customers can apply different weights to update their memory depending on the source of information. In this sense, we denote by α the weight that customers give to their own information and by β the weight for the information provided by their best performing neighbour. We apply adaptive expectations (Nerlove 1958) to model the updating process of the customer’s memory (M_{ijt+1}). The adaptive expectations concept, also known as exponential smoothing (Theil and Wage, 1964), is based on the weighted average of two sources of evidence: the latest evidence (the most recent observation, W_{ijt}), and the value computed one period before (M_{ijt}). (Theil and Wage 1964). Then M_{ijt+1} is given by:

$$M_{ijt+1} = \theta * M_{ijt} + (1 - \theta) * W_{ijt} \quad (2)$$

$$s.t. \quad \theta = \begin{cases} \alpha & \text{when using own information} \\ \beta & \text{when using neighbours' information} \end{cases}$$

where θ denotes the coefficient of expectations and takes two different values depending on the source of information, as explained above. The logic behind this

coefficient is explained in Delgado et al. (2011a) and W_{ijt} is computed using Equation (1). We label “conservative” those customers who give more weight to their memory than to the new information, i.e. α and β greater than 0.5. In contrast, when they give more weight to new information, we call them “reactive”.

Once customers update their memories, they will decide to patronise the facility with the lowest expectation of sojourn time. Longer (shorter) queues bring about higher (lower) sojourn times and increase (decrease) customers’ perceptions. When customers’ perception about a certain facility exceeds the expectations they have regarding some other, they decide to switch facility. Otherwise they remain at the same facility

Managers’ Decisions

We endow managers with similar abilities as the customers. In this sense, we assume that managers have a memory and react to customer behaviour by adjusting the service capacity of their facility. Although customers cannot observe the queues before choosing a facility, managers have information about the number of customers arriving at their facilities. They thus use this information to form their perceptions about the future arrival rate, $\hat{\lambda}_{jt}$. The managers’ memory also enables them to update their perceptions each period using an adaptive expectation model, as follows:

$$\hat{\lambda}_{jt+1} = \delta * \hat{\lambda}_{jt} + (1 - \delta) * \lambda_{jt} \quad (3)$$

where hats indicate the expected queue length, and δ the coefficient of expectations (Nerlove, 1958). δ can be interpreted as the speed at which managers adjust their perceptions. This parameter follows the same logic as explained above for the customers’ parameters (α and β).

Managers use their estimate about the future demand (i.e. arrival rate) to determine the service capacity required to meet their customers’ expectations of sojourn time. Managers do not have accurate information regarding these expectations, but they know a reference average sojourn time, τ_{MR} , which is considered by the market to be acceptable to the customers. This market reference can be interpreted as a benchmark the managers use to evaluate the competitiveness of their firms. This benchmark is assumed to be exogenous and fixed. Given $\hat{\lambda}_{jt}$ and τ_{MR} , managers can determine their desired service capacity, $\hat{\mu}_{jt}$, by using Equation 1. Rewriting this equation in terms of these three variables we have:

$$\tau_{MR} = \frac{\hat{\lambda}_{jt}}{\hat{\mu}_{jt}^2} + \frac{1}{\hat{\mu}_{jt}} \quad (4)$$

Note that this is a second order equation for which there are two possible solutions. Nevertheless, due to the

nature of the problem it is impossible to have negative arrival and service rates. Hence, we only consider the positive solution of Equation 4. This solution is given by:

$$\dot{\mu}_{jt} = \begin{cases} 0 & \text{if } \hat{\lambda}_{jt} = 0 \\ 1 + \frac{\sqrt{1 + 4 * \tau_{MR} * \hat{\lambda}_{jt}}}{2 * \tau_{MR}} & \text{if } \hat{\lambda}_{jt} > 0 \end{cases} \quad (5)$$

We assume $\dot{\mu}_{jt}$ to be the service capacity which managers consider as sufficient to satisfy the customers' needs regarding expected sojourn times. Then, the aim of managers is to adapt their available service capacity (μ_{jt}) to their desired service capacity. They must therefore decide when and how much capacity to add or remove. Nonetheless, once the adjustment decision has been made, its implementation process does not materialise immediately. In fact, when managers decide how much capacity they wish either to add or remove, there is usually a lag between the moment they take their decision and it is implemented. Examples of this kind of delays include the delivery delay entailed when purchasing new machines; the time required to build new infrastructure; the period for training new employees; and the legal notice period to lay off staff.

When managers decide how much capacity to add (x_t), these orders accumulate as capacity on order (μ_{jt}^+) until they are available for delivery. Once the delivery time (d^+) expires the ordered capacity is available for service. That is, order x_t is fulfilled in period $t + d^+$. Assuming that there is no capacity on order at the beginning of the simulation, i.e. $\mu_{jt}^+ = 0$ for $t \leq 1$ and $x_t = 0$ for $t < 1$, the cumulative orders are given by:

$$\mu_{jt}^+ = \sum_{k=t-d^++1}^{t-1} x_k \quad t > 1 \quad (6)$$

Similarly, when the capacity adjustment implies removing capacity, the capacity managers decide to withdraw (y_t) is designated as capacity to be retired (μ_{jt}^-). This capacity remains available for customers until the dismantling time (d^-) expires, i.e. y_t is removed from the service capacity in period $t + d^-$. Assuming that there are no previous retirement decisions at the start of the simulation, i.e. $y_t = 0$ for $t < 1$ and hence $\mu_{jt}^- = 0$ for $t \leq 1$, the capacity retirements accumulate as follows for $t > 1$:

$$\mu_{jt}^- = \sum_{k=t-d^-+1}^{t-1} y_k \quad (7)$$

The capacity decisions that have not yet been implemented (μ_{jt}^\pm) are given by:

$$\mu_{jt}^\pm = (\mu_{jt}^+ - \mu_{jt}^-) \quad (8)$$

We propose a heuristic which enables managers to know how much capacity either to add (x_t) or remove (y_t) and when to do so. The required capacity adjustment depends on the gap managers observe between their desired capacity ($\dot{\mu}_{jt}$) and the service capacity, which they perceive to have currently. When this gap is positive, new capacity orders will be placed, whereas new capacity retirements will be carried out when the gap is negative.

The current available service capacity (μ_{jt}) and the managers' previous decisions, which are still in the process of implementation (μ_{jt}^\pm), make up the capacity that managers are expecting to have in service if no further changes are decided. Nevertheless, managers do not necessarily keep in mind all their previous decisions, which have not been yet implemented. Denoting by ψ the proportion of the not yet implemented capacity adjustment, which managers remember, we obtain that the service capacity they perceive to have at time t , is given by:

$$\bar{\mu}_{jt} = \mu_{jt} + \psi * \mu_{jt}^\pm \quad (9)$$

where ψ is nonnegative and less than or equal to 1. We call ψ the "coherence factor" of managers. If managers are rational when making capacity decisions they should take into account their previous decisions, which are still in process of execution. In this case, $\psi = 1$. Otherwise, if they only account for part of these past decisions, $\psi < 1$.

By computing the difference between the service capacity, which managers consider they currently have, and their desired capacity, we obtain the required capacity adjustment ($\Delta\mu_{jt}$):

$$\Delta\mu_{jt} = \dot{\mu}_{jt} - \bar{\mu}_{jt} \quad (10)$$

Managers may make this adjustment as fast or as slow as they wish. That is, we assume that managers can be prudent when taking their decisions. The second element of the heuristic tackles this issue. Let ζ be the speed at which managers decide to adjust capacity, i.e. how fast they decide to either add or remove capacity. Then, when the desired capacity exceeds the current capacity, which managers perceive to have, they decide to add capacity and the ordered capacity (x_{jt}) will be:

$$x_{jt} = \zeta * (\dot{\mu}_{jt} - \bar{\mu}_{jt}) \quad \text{if } \dot{\mu}_{jt} > \bar{\mu}_{jt} \quad \forall \quad 0 \leq \zeta \leq 1 \quad (11)$$

while if this decision implies to withdraw capacity, the capacity to be retired is:

$$y_{jt} = \zeta * (\bar{\mu}_{jt} - \dot{\mu}_{jt}) \quad \text{if } \dot{\mu}_{jt} < \bar{\mu}_{jt} \quad \forall \quad 0 \leq \zeta \leq 1 \quad (12)$$

ζ is small when managers take their decisions slowly, i.e. they are prudent decision makers. On the contrary, a high ζ implies that managers take actions quickly, i.e. they are aggressive decision makers.

To summarise the managers' dynamics: The more customers patronise a facility, the higher its manager's perception of the arrival rate is. High (low) manager's expectations increase (decrease) his desired service capacity. The higher (lower) the desired service capacity the more capacity the managers order (remove). With a delay, the capacity orders will increase the service capacity, while the capacity retirements will decrease it. This will affect the number of customer arriving at that facility.

CUSTOMERS' AND MANAGERS' PROFILE

Both customers and managers can be characterised according to their attitude towards new information when updating their perceptions (α , β , and δ). Managers are additionally described according to their coherence when taking decisions (ψ) and the speed at which they implement these decisions (ζ). Considering their attitude towards new information, customers and managers can be defined as conservative, hesitant or reactive. We say they are conservative or reluctant regarding new information when they have more confidence in their memory than in recent experiences (i.e. α or β high for customers and δ high for managers). When the contrary occurs, we call customers reactive (i.e. α or β low for customers and δ low for managers). When a roughly equal weight is given to memory and to the new information, we call hesitant customers (i.e. α or β intermediate).

SIMULATION RESULTS AND DISCUSSION

Due to the number of parameters the model has, we limit our simulation analysis to evaluating the dynamics of a system which is configured with 3 facilities and 120 customers (i.e. a one dimensional lattice of 120 cells, where each cell can take exactly one of three states). Each facility is initially provided with a service capacity of 5 customers per time unit, a manager and its own queue. Agents are endowed with an initial memory (i.e. expected average sojourn times for customers and expected arrival rates for managers). This initial memory is randomly allocated to the agents using a uniform distribution, whose maximum and minimum values are respectively 10% above and below the sojourn time of the Nash equilibrium. Given that all facilities have the same service capacity at the beginning, the Nash equilibrium occurs when customers are split equally among the three facilities, i.e. 40 customers patronising each. This distribution yields an average sojourn time of 1.8 time units.

We assume the implementation and dismantling delays involved in the managers' decisions to be fixed and equal to 4 and 2 periods, respectively. That is, once

managers decide to increase capacity, this order will be delivered 4 periods later. Similarly, when they decide to reduce capacity, the capacity to be retired will still be available for service during the next 2 periods.

We develop and simulate the model using the numerical computing environment MATLAB and use STATA to test statistical hypotheses related to the performance of the facilities.

In this paper we focus on analysing the impact of the managers' profile on the system behaviour. To do this, we simulate 1,000 iterations (i.e. 1,000 different random seeds) of the model for a number of different combinations of the managers' parameters ($\delta 1$, $\delta 2$, $\delta 3$, ψ , ζ) for the case where $\alpha = 0.3$ and $\beta = 0.7$. In order to validate if 1,000 iterations are enough to draw conclusions about the different scenarios the system exhibits in steady state, we have run 10,000 iterations of the model and extended the simulated time to 10,000 time periods for several parameter combinations. The steady-state period was computed for the last 100 time period. We found that there were no significant differences in the number of facilities closing compared to 1,000 iterations over 500 periods. We therefore assume that 1,000 simulations of the model over 500 time periods are appropriate for our analysis. The results are discussed in terms of the possible scenarios (regarding the number of facilities remaining open at the end of the simulation period) which we can obtain when simulating the model. Table 1 contains the eight possible scenarios according to the number of facilities considered in the model.

Table 1: Possible scenarios generated by simulating the model.

Numerical Code	Scenario
0	All facilities close
1	Facility 1 is the only one open
2	Facility 2 is the only one open
3	Facility 3 is the only one open
12	Facilities 1 and 2 remain open, while facility 3 closes
13	Facilities 1 and 3 remain open, while facility 2 closes
23	Facilities 2 and 3 remain open, while facility 1 closes
123	All facilities remain open

Figure 1 shows the relative frequency of each scenario described in Table 1 for the nine possible combinations of parameters (ψ , ζ) using the three values $\{0.2, 0.5, 0.8\}$. The coefficients of expectations of managers and customers are: $\{\delta 1 = 0.2, \delta 2 = 0.5, \delta 3 = 0.8, \alpha = 0.3$ and $\beta = 0.7\}$. The nine combinations of parameters (ψ , ζ) are

on the horizontal axis. The first three combinations (i.e. I, II, and III) illustrate the cases where the managers are slow decision makers ($\zeta = 0.2$) and have different degrees of rationality when accounting for their not-yet implemented decisions (ψ). The next three combinations (i.e. IV, V, VI) represent those cases where managers are moderate decision makers and the last three (i.e. VII, VIII, IX) those where managers are fast decision makers.

Figure 1 indicates that the scenario where the three facilities remaining open (see the line labelled as “Facilities 123”) is the most likely when managers take their decisions slowly (see cases I, II and III). This probability decreases as the decision making process is faster and it is close to zero when managers are faster decision makers ($\zeta = 0.8$) and almost rational when accounting for their not-yet implemented decisions ($\psi = 0.8$) (case IX). When one facility closes, this is mostly facility 1 (see line “Facilities 23”), whose manager is the most reactive ($\delta 1 = 0.2$). This scenario is the most likely in all the cases in which managers are moderate decision makers (cases IV, V and VI) and in those cases where they are fast decision makers and either slightly irrational or almost rational when considering their not-yet implemented decisions (cases VIII and IX).

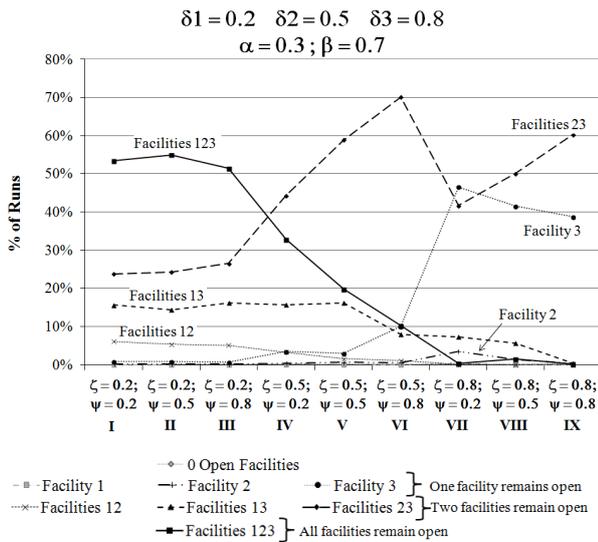


Figure 1: Percentage of runs which yield each possible scenario depending on the consistency factor (ψ) and the speed factor (ζ).

The faster the decision making process, the lower the probability of the most conservative manager (i.e. manager 3, $\delta 3 = 0.8$) being the only one who closes his facility (Facilities 12). The probability of the hesitant manager (i.e. manager 2, $\delta 2 = 0.5$) being the only one to close his facility (Facilities 13) is higher when the decision making process is slow (cases I, II and III) and lower when such a process is fast (cases VII, VIII and IX). When the most conservative manager (i.e. manager 3) decides quickly he is more likely to achieve a monopoly position (See line labelled as Facility 3 in cases VII, VIII and IX). In the case where managers are

almost rational and take decisions fast, the probability of the most conservative manager closing his facility is negligible (less than 0.1%) (case IX). Finally, it is worth mentioning that the scenario in which the three facilities shut down is very unlikely (less than 0.2%) for this case.

Concerning the market share of the facilities in each scenario, we hypothesise that the most conservative managers’ facilities capture a larger market share than the others. We use the Mann-Whitney-Wilcoxon test (MWW) (Newbold 1988), a non-parametric statistical test also called the Mann-Whitney U-test, to assess the null hypothesis that the median of the average arrival rates of two facilities during steady-state are the same. The alternative hypothesis assumes that the median of the average number of customers arriving at the most conservative manager’s facility in steady-state is greater than that at the other facility. We have applied a non-parametric test because we do not know the distribution of the data and for some scenarios we have very little data. Parametric tests are not appropriate in these cases. Instead, the MWW-test is appropriate because the distributions have enough symmetry to assume that the median and the average are similar.

Table 2 provides the test-statistics of the MWW-test to assess the difference between the medians of the distributions of average arrival rates at each facility in steady-state for the scenarios where at least two facilities remain open, as monopolistic situations are irrelevant in this context. This table contains the median of the number of customers arriving at each facility for the same nine cases of Figure 1.

Table 2: The values of the median for the number of customers arriving at each facility during steady-state for 1,000 simulations of the model as a function of ψ and ζ

Case	2 FACILITIES OPEN						3 FACILITIES OPEN					
	Facilities 12		Facilities 13		Facilities 23		Facilities 123			F1- F2	F1- F3	F2- F3
	F1	F2	F1	F3	F2	F3	F1	F2	F3			
I	48	72 a	48	72 a	60	60	30	40	50	a	a	a
II	49	71 a	46	74 a	57	63 a	26	40	52	a	a	a
III	51	69 a	44	76 a	54	66 a	27	39	53	a	a	a
IV	35	86 a	39	81 a	50	70 a	27	45	48	a	a	a
V	37	83 a	27	93 a	47	73 a	25	39	56	a	a	a
VI	38	82 a	34	86 a	38	82 a	20	43	52	a	a	a
VII	66	54 a	60	60	66	54 c	43	30	20		b	
VIII	-	-	6	114 a	50	70 a	49	33	40			a
IX	15	106 d	6	114 a	36	85 a	14	61	45	d	d	d

The letters (i.e. “a”, “b”, “c”, and “d”) next to the median value of the second facility of each scenario indicate the results of the MWW-test for the null

hypothesis that the median arrival rate is the same for the assessed facilities. The letters “a” and “b” indicate that this hypothesis is rejected at a significance level of 0.01 and 0.1, respectively, against the alternative hypothesis that the median arrival rate of the most conservative manager’s facility is significantly greater than that of the other facility. The letter “c” indicates that the null-hypothesis is rejected at a significance level of 0.01, against the alternative hypothesis that the median arrival rate of the most conservative manager’s facility is significantly lower than that of the other facility. The letter “d” indicates that the null hypothesis cannot be tested because of lack of data. For instance, the letter “a” in case I when facility 1 ($\delta 1 = 0.2$) and 3 ($\delta 3 = 0.8$) remain open (medium green scenario) indicates that according to the MWW-test the median number of customers using facility 3 (i.e. with the most conservative manager) is significantly greater than that using facility 1 at a significance level of 0.01. In the scenario where the three facilities remain open, this test is assessed by pairs of facilities and the significance of the test is indicated in the last three columns for each pair of facilities. For instance, in case I, the MWW-test indicates that at a significance level of 0.01, the median of facility 3 ($\delta 3 = 0.8$) is significantly greater than that of facility 2 ($\delta 2 = 0.5$) and facility 1 ($\delta 1 = 0.2$) and that the median of facility 2 ($\delta 2 = 0.5$) is significantly greater than that of facility 1 ($\delta 1 = 0.2$).

In most cases, the p-values computed by MWW-test suggest rejecting the null hypotheses at the 1% level of significance. This enables us to conclude that the facility of the most conservative manager (i.e. manager 3, $\delta 3 = 0.8$) usually attracts more customers than the other facilities. Most of the exceptions are due to lack of data. For instance, the yellow scenario (i.e. the three facilities remain open) in case IX ($\psi = 0.8$ and $\zeta = 0.8$), which is unlikely (0.1% of iterations barely visible in Figure 1).

In case VII, where managers are moderately irrational ($\psi = 0.2$) and fast decision makers ($\zeta = 0.8$), the scenario in which the most reactive manager (i.e. manager 1, $\delta 1 = 0.2$), shuts down his facility (i.e. light green scenario) contrasts with the expected behaviour. That is, the less conservative manager of the two still active managers, who is manager 2 (i.e. $\delta 2 = 0.5$), usually captures the largest market share.

We repeated the experiment with the same combinations of the managers’ coefficient of expectations for another combination of the customers’ parameters. We have tested all the extreme and intermediate cases regarding the customers’ attitudes toward the new information. Again, the main insights, about that the most conservative managers’ facility captures a larger market share than the others, remain valid. One significant observation is that the scenario in which the three facilities are still open at time 500 is much more frequent when customers are either hesitant

or very conservative regarding new information, no matter its provenance. Conversely, the scenario where the conservative managers achieve a monopoly position is more likely when customers are more reactive with respect to their own information and more conservative regarding that of their neighbours.

CONCLUSIONS

In this paper we have extended the CA model proposed by Delgado et al. (2011a) by incorporating the service rate as an endogenous variable. We have endowed the managers with the ability to adjust the service capacity of their facility. Managers are provided with a memory which enables them to update their expectations regarding the number of customers arriving at their facilities each period. Other attributes, which characterise the managers’ profile, are the extent to which they account for their previous decisions when deciding by how much to adjust capacity and the speed at which they take decisions. The former indicates the level of irrationality of managers when the decision making involves delays (i.e. coherence factor).

We have performed some experiments to analyse the sensitivity of the model to the managers’ profiles. We conclude that the facility of the most conservative manager usually achieves the highest market share. Additionally this facility is the most likely to remain open until the end of the simulation period (i.e. it is less likely that this facility shuts down).

This work can be extended by analysing the sensitivity of the model to the customers’ parameters and the delays involved in the implementation of the managers’ decisions. Further work in this field includes adding uncertainty parameters to the customers and managers’ decision rules (as was done in the previous paper for the customers) and assessing the influence of other service factors in the customers and managers’ decisions, such as price and quality.

REFERENCES

- Bielen, F., and N. Demoulin. 2007. “Waiting Time Influence on the Satisfaction-Loyalty Relationship in Services”. *Managing Service Quality* 17, No. 2, 174-193.
- Davis, M. M. and J. Heineke. 1998. “How Disconfirmation, Perception and Actual Waiting Times Impact Customer Satisfaction”. *International Journal of Service Industry Management* 9, No. 1, 64-73.
- Delgado, C. A.; A. van Ackere; E. R. Larsen; and K. Sankaranarayanan. 2011a. “Collective Behavioral Patterns in a Multichannel Service Facilities System: a Cellular Automata Approach”. In *Operations Research, Computing, and Homeland Defense: Proceedings of the 12th INFORMS Computing Society Conference* (Monterey CA, 9-11). INFORMS, Hanover, MD, 16-27.
- Delgado, C. A.; A. van Ackere; E. R. Larsen; and K. Sankaranarayanan. 2011b. “Modelling Decisions under Uncertainty in a Behavioural Queuing System”. In *Proceedings of the 25th European Conference on*

- Modelling and Simulation* (Krakow, Poland, Jun. 7-10), ECMS, Dudweiler, Germany, 34-40.
- Delgado, C.A. 2012. "Behavioural Queueing: Cellular Automata and a Laboratory Experiment". *PhD Diss.*, Université de Lausanne.
- Dewan, S. and H. Mendelson. 1990. "User Delay Costs and Internal Pricing for a Service Facility". *Management Science* 36, No. 12 (Dec), 1502-1517.
- Edelson, N. M. 1971. "Congestion Tolls under Monopoly". *The American Economic Review* 61, No. 5 (Dec), 873-882.
- Edelson, N. M. and Hildebrand, D. K. 1975. "Congestion Tolls for Poisson Queuing Processes". *Econometrica* 43, No. 1 (Jan), 81-92.
- Gallay, O. "Agent-Based Routing in Queueing Systems". *PhD Diss.*, Ecole Polytechnique Fédérale de Lausanne, 2010
- Hui, M. K. and D. K. Tse. 1996. "What to Tell Consumers in Waits of Different Lengths: an Integrative Model of Service Evaluation". *The Journal of Marketing* 60, No. 2 (Apr), 81-90.
- Law, A. K. Y.; Y. V. Hui; and X. Zhao. 2004. "Modeling Repurchase Frequency and Customer Satisfaction for Fast Food Outlets". *International Journal of Quality & Reliability Management* 21, No. 5, 545-563
- Leeman, W. A. 1964. "The Reduction of Queues Through the Use of Price". *Operations Research* 12, No. 5 (Sep-Oct): 783-785.
- Leeman, W. A. 1965. "Comments' on Saaty's 'The Burdens of Queuing Charges'". *Operations Research* 13, No. 4 (Jul-Aug), 680-681.
- Mendelson, H. and S. Whang. 1990. "Optimal Incentive-Compatible Priority Pricing for the M/M/1 Queue". *Operations Research* 38, No. 5 (Sep-Oct), 870-883.
- Naor, P. 1969. "The Regulation of Queue Size by Levying Tolls". *Econometrica* 37, No. 1 (Jan), 15-24.
- Nerlove, M. 1958. "Adaptive Expectations and cobweb phenomena". *The Quarterly Journal of Economics* 72, No. 2 (May), 227-240.
- Newbold, P. 1998. *Statistics for business and economics*. (2nd ed.). Prentice Hall, Englewood Cliffs, NJ.
- Pullman, M. E. and G. M. Thompson. 2002. "Evaluating capacity- and demand- management decisions at sky resort". *Cornell Hotel and Restaurant Administration Quarterly* 43, No. 6 (Dec), 25-36.
- Saaty, T. L. 1965. "The Burdens of Queuing Charges-Comments on a Letter by Leeman." *Operations Research* 13, No. 4 (Jul-Aug), 679-680.
- Sinha, S. K.; N. Rangaraj; and N. Hemachandra. 2010. "Pricing surplus server capacity for mean waiting time sensitive customers". *European Journal of Operational Research* 205, No. 1(Aug), 159-171
- Stidham, S. Jr. 1985. "Optimal Control of Admission to a Queueing System". *IEEE Transaction on Automatic Control* 30, No. 8 (Aug), 705-713.
- Stidham, S. Jr. 1992. "Pricing and Capacity Decisions for a Service Facility: Stability and Multiple Local Optima". *Management Science* 38, No. 8 (Aug): 1121-1139.
- Taylor S. 1994. "Waiting for Service: the Relationship Between Delays and Evaluations Of Service". *The Journal of Marketing* 58, No. 2 (Apr), 56-69.
- Theil, H. and S. Wage. 1964. "Some observations on adaptive forecasting". *Management Science* 10, No.2 (Jan), 198-206.
- van Ackere, A. and E. R Larsen. 2004. "Self-Organising Behaviour in the Presence of Negative Externalities: a Conceptual Model of Commuter Choice". *European Journal of Operational Research* 157, No. 2 (Sep), 501-513.
- van Ackere, A.; C. Haxholdt; and E. R Larsen. 2013. "Dynamic Capacity Adjustments with Reactive Customers". *Omega* 41, No. 4 (Aug), 689-705.
- Yechiali, U. 1971. "On Optimal Balking Rules and Toll Charges in the Gi/M/1 Queueing Process". *Operations Research* 19, No. 2 (Mar-Apr), 349-370.

AUTHOR BIOGRAPHIES

CARLOS A. DELGADO-ALVAREZ is Professor of Operations Research and Modeling and Simulation at the Department of Industrial Engineering of the ICESI University, Cali, Colombia since 2012. He obtained his PhD in Information Systems from the School of Business and Economics, University of Lausanne, Switzerland. He received his M.Sc. in Systems Engineering and a B.Eng. in Industrial Engineering from the Faculty of Mines, National University of Colombia, where he also worked for three years as a research engineer and teaching assistant for bachelor courses in systems simulation and Algorithms and Programming. He is a member of INFORMS society. His research interests include operations management, system simulations, social complex systems and experimental economics.

ANN VAN ACKERE is Professor of Decision Sciences at HEC Lausanne, the School of Business and Economics of the University of Lausanne, Switzerland since 1999. She obtained her PhD from the Stanford Graduate School of Business and joined the faculty of London Business School, UK, upon graduation.

ERIK R. LARSEN is Professor at the Institute of Management and vice dean of the Faculty of Economics at the University of Lugano, Switzerland. Previously he held appointments at Cass Business School (London), London Business School, and Copenhagen Business School. During the period 1996-1998, he was an EU Marie Curie Fellow at the University of Bologna, Italy. He obtained his PhD from the Institute of Economics, Copenhagen Business School and his M.Sc. from the Technical University of Denmark.