CAPACITY ADJUSTMENT IN A SERVICE FACILITY WITH REACTIVE CUSTOMERS AND DELAYS: SIMULATION AND EXPERIMENTAL ANALYSIS.

Carlos Arturo Delgado Alvarez
HEC, School of Business and Economics, University of Lausanne
Dorigny, 1015-Lausanne, Switzerland
+41 21 692 34 67
E-mail: carlos.delgado@unil.ch

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

Santiago Arango
Complexity Center Ceiba, Faculty of Mines, Universidad Nacional de Colombia
Medellin, Colombia
E-mail: saarango@unal.edu.co

ABSTRACT

In this paper, we apply system dynamics to model a queuing system wherein the manager of a service facility adjusts capacity based on his perception of the queue size; while potential and current customers react to the managers' decisions. Current customers update their perception based on their own experience and decide whether to remain patronizing the facility, whereas potential customers estimate their expected waiting time through word of mouth and decide whether to join the facility or not. We simulate the model and analyze the evolution of the backlog of work and the available service capacity. Based on this analysis we propose two alternative decision rules to maximize the manager's cumulative profits. Then, we illustrate how we have developed an experiment to collect information about the way human subjects taking on the role of a manager in a lab environment face a situation in which they must adjust the capacity of a service facility.

KEYWORDS: Queuing system, capacity adjustment management, system dynamics, experimental economics, adaptive expectations

INTRODUCTION

Most typical research in queuing problems has been focused on the optimization of performance measures and the equilibrium analysis of a queuing system. Traditionally, analytical modeling and simulation have been the approaches used to deal with queuing problems. Most simulation models are stochastic and some more recent models are deterministic (van Ackere, Haxholdt, & Larsen, 2010).

The analytical approach describes mathematically the operating characteristics of the system in terms of the performance measures, usually in "steady state" (Albright & Winston, 2009). This method is useful for low-complexity problems whose analytical solution is not difficult to find. For complex problems, a simulation approach is preferable as it enables modeling the problem in a more realistic way, with fewer simplifying assumptions (Albright & Winston, 2009).

We consider those queuing systems in which customers decide whether or not to join a facility for service based on their perception of waiting time, while managers decide to adjust capacity based on their perception of the backlog of work (i.e. the number of customers waiting for service). The analysis of queuing problems could be aimed at either optimizing performance measures to improve the operating characteristics of a system or understanding how the manager and customers interact with the system to achieve their objectives. In the real world, queuing is a dynamic problem whose complexity, intensity and effects on the system change over time. Still, some problems may be modeled using the assumptions of classical queuing theory (Rapoport, Stein, Parco, & Seale, 2004). Considering the complexity of queuing problems, which is due to a set of interactive and dynamic decisions by the agents (i.e. customers and the manager) who take part in the system, we will focus on studying the behavioral aspects of queuing problems.

Haxholdt, Larsen, & van Ackere (2003) and van Ackere, Haxholdt, & Larsen, (2006); van Ackere et al., (2010) have applied deterministic simulation methodologies for studying behavioral aspects of a queuing system. Other authors have included cost allocation as a control for system congestion (queue size) (e.g. Dewan and Mendelson 1990). In this way, customers' decisions on whether or not to join the system are influenced by such costs. Likewise, those decisions can be based on steady-state (e.g. Dewan and Mendelson 1990). or be state-dependent (e.g. van Ackere 1995). The seminal papers on this subject are Naor (1969) and Yechiali (1971). Other authors have included dynamic feedback processes to build perceptions of the behavior of the queue (van Ackere et al., 2006) and/or of demand (van Ackere et al. 2010), which influence the decisions of customers and managers. A more detailed discussion of the state of the art on behavioral aspects in queuing theory can be found in (van Ackere et al., 2010).

We propose two methodological approaches to achieve our goals. Firstly, we use system dynamics to learn about the macro-dynamics of customers and the manager interacting in a service facility. Specifically we analyze how the available service capacity and the queue evolve and how the delay structure affects the manager’s decision. We also want to assess how the manager adjusts capacity based on the evolution of the backlog of work (i.e. the number of customers waiting for service). Haxholdt et al. (2003) and van Ackere et al. (2006 and 2010) applied system dynamics to tackled similar problems. System dynamics is useful for problems, which do not require much detail. That is, those which can be modeled at a high level of abstraction. This kind of problems is usually situated at the macro or strategic level (e.g. marketplace & competition, population dynamics and ecosystem) (Borshchev & Filippov, 2004)

Next we apply experimental economics (Smith, 1982) to capture information about how subjects playing the role of a manager in a lab environment, decide when and by how much to adjust the capacity of a service facility. We use the system dynamics based simulation model as a computational platform to perform the experiment. For more details about how system dynamics models have been used to carry out laboratory experiments, see (Arango, Castaneda, & Olaya, 2011). Experimental economics is a methodology that based on collecting data from human subjects to study their behavior in a controlled economic environment (Friedman & Sunder, 1994).

This paper is organized as follows: Firstly, we discuss the dynamic hypothesis of the problem proposed initially by van Ackere et al. (2010) and explain why we modify the model. Then, we analyze the model behavior of the base case. In the following section, we introduce two alternative strategies to manage the capacity adjustment of the service facility. We determine the optimal parameters for these strategies and analyze the resulting system behavior. We also perform a sensitivity analysis to the parameter values. Finally, we present the experimental laboratory and discuss the collected results.

A SERVICE FACILITY MANAGEMENT MODEL

In this section, we analyze the dynamic hypothesis of the queuing model proposed by van Ackere et al. (2010). This model captures the relationship between customers and manager (referred to as the service provider) as agents who interact in a service system. The causal loop diagram of Figure 1 portrays the feedback structure of these two actors in the system. The model consists of two sectors: the customers' behavior is to the left and that of the manager to the right. Both sectors are connected by the queue, whose evolution determines the dynamics of these actors in the system. Customers decide whether to use the facility based on their estimate of waiting time, while the manager decides to adjust the service capacity based on the queue length. Examples of this kind of system include a garage where customers take their car for maintenance, and workers or students who daily patronize a restaurant to have lunch. In both examples, customers are free to use or not the facility for service and the manager is motivated to encourage customers to use his facility by adjusting its service capacity.

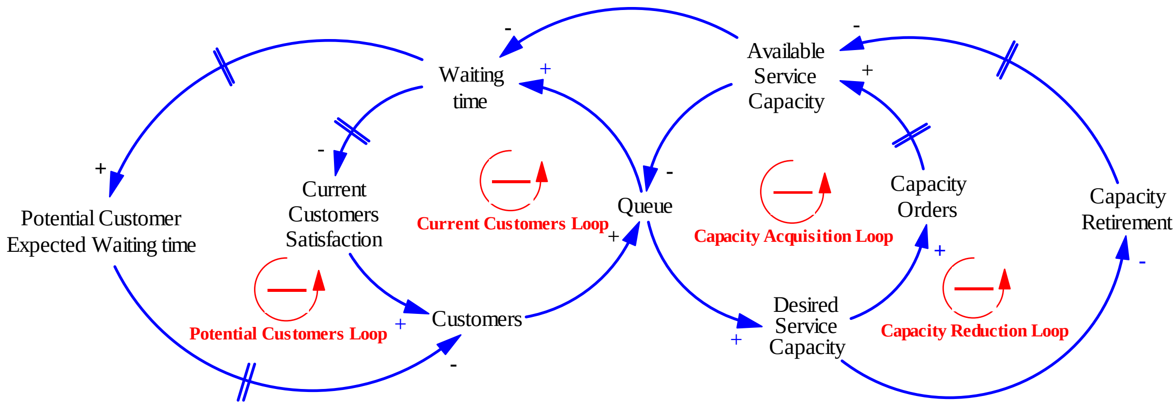


Figure 1. Feedback loop structure for a customers-facility queuing system

Two groups of customers are assumed: current and potential customers. The former make up the customer base of the facility; they periodically patronize it as long as they are satisfied. They consider being satisfied when their expected waiting time is less than the market reference, which they find acceptable. The second group represents those customers who the manager envisages as potentially attractive to the business. They can be either former customers, who left due to dissatisfaction, or new customers who require the service and look for a facility. They decide whether or not to join the facility depending on their expected waiting time, which they also compare to the market reference.

Customers form their perception of waiting time ($\dot{W}_t$) each period using adaptive expectations (Nerlove, 1958), as shown in Equation 1:

$$\dot{W}_t = \varphi * W_{t-1} + (1-\varphi) * \dot{W}_{t-1} \quad (1)$$

where φ is called the coefficient of expectations (Nerlove, 1958) and $1/\varphi$ may be considered as the time taken by customers to adapt their expectations. Current customers adjust their expectation based on their last experience (W_t), while potential customers rely on word of mouth. The decision of joining a facility for service based on its reputation often requires more time than when we base this decision on our own experience. Thus, we assume that the time required by potential customers to adapt their expectations is longer than or equal to that of the current customers.

While the current customers' perception determines their loyalty to the facility, the potential customers' perception defines if they will join the customer base. The lower the waiting time perceived by current customers, the more loyal they are, whereas the higher the perceived waiting time, the more customers will leave the customer base. Regarding potential customers, the lower their expected waiting time, the more will become new customers for the facility. The rates at which new customers join the customer base and current customers leave it are modeled using nonlinear functions of the satisfaction level. van Ackere et al. (2010) discuss some alternatives to model these functions.

To summarize the customers' dynamics: longer queues bring about higher waiting times for current customers and increased perceptions of waiting time for potential customers, implying that the level of satisfaction with the facility's service of both customer groups decreases. Consequently, over time this reduction in customers' satisfaction leads current customers to leave the facility and discourages new customers from joining it in the future. Thus, the number of customers waiting for service will decrease until the waiting time tends to acceptable levels compared to the market reference and the customers' perception stabilizes. These dynamics are described by the two balancing loops to the left in Figure 1.

As far as the service provider (the right side of Figure 1) is concerned, van Ackere et al. (2010) model the type of service systems where the capacity adjustment involves an implementation time. For instance, hiring new employees requires new training, laying off staff may imply a notice period, acquiring new IT systems takes time, among others. However, the authors represent this time in the model using an information delay (Sterman, 2000); after the manager estimates the required capacity, any needed adjustment is implemented gradually. This is a simplified view of the delay structure. In a system dynamics context, this kind of delays is better modeled through material delays, which capture the real physical flow of the capacity (Sterman, 2000). Once the adjustment decision has been made, its implementation process does not materialize immediately. We deviate from van Ackere et al. (2010) by incorporating this material delay structure in the model, as the stock and flow diagram of Figure 2 illustrates. In this way, we can model how the manager accounts for his previous decisions, which have not yet taken effect, to make his next decision.

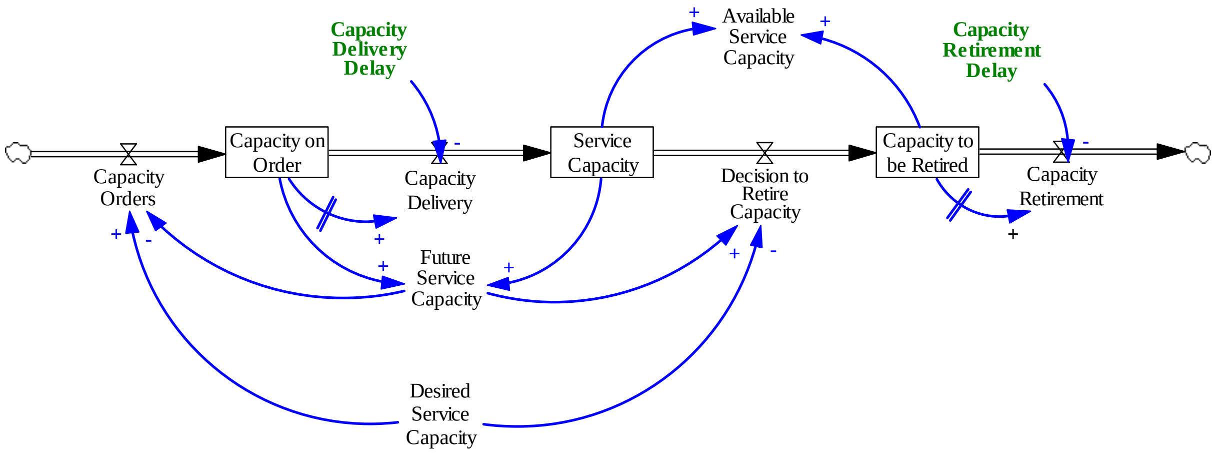


Figure 2. System dynamics representation for the capacity adjustment management of a service facility.

The capacity adjustment process is depicted in Figure 2 by capacity orders and the decision to retire capacity, which determine the available service capacity. Starting from the left, the manager decides how fast and how much to adjust capacity based on his desired service capacity and the future capacity. The latter is explained below and depends on his previous decisions. He estimates the desired service capacity based on his perception of the average queue length and a market reference for the waiting time (τ_{MR}). Like the customers, the manager forms this perception by applying adaptive expectations. He updates his expected average queue length based on the most recent observation of the queue (Q_{t-1}). This expected average queue length (EQ_t) is given by:

$$EQ_t = \beta * Q_{t-1} + (1-\beta) * EQ_{t-1} \tag{2}$$

where β is the coefficient of expectations for the manager and $1/\beta$ may be interpreted as the time required by the manager to adapt his perception. Then, the desired service capacity of the manager is determined as follows:

$$DC_t = \frac{EQ_t}{\tau_{MR}} \tag{3}$$

The longer the queue the greater the desired service capacity and the larger the capacity orders (c.f Figure 1). After the manager decides how much capacity to add (c.f. capacity orders in Figure 2), these orders accumulate as capacity on order (CO) until they are available for delivery (c.f. capacity delivery delay in Figure 2). Some examples of this kind of delayed process in capacity acquisition include construction of new buildings, purchase of new equipment and hiring staff. Once the capacity order is fulfilled, the service capacity (SC) will be increased by the capacity delivery. The greater the service capacity, the higher the service rate and thus fewer customers waiting. In this way, a third balancing loop (c.f. capacity acquisition loop in Figure 1) results from the dynamics between the manager and customers.

The decision of adjusting capacity may also imply removing capacity. When this occurs, the capacity, which the manager decides to withdraw, will be designated as capacity to be retired (CbR). This capacity remains available to the customer during the capacity retirement delay (e.g. end a lease on a building, notice period for staff, etc). Hence, the currently available service capacity at the facility at time *t* is given by,

$$ASC_t = SC_t + CbR_t \tag{4}$$

After the delay involved in the capacity retirement, the available service capacity will decrease due to this retirement, as shown in Figure 1, and the number of customers in the queue will thus increase. This effect yields the fourth balancing loop in the system. This loop describes the behavior caused by the decisions of capacity reduction.

Finally, the capacity that will be available once all the manager's decisions have been implemented, i.e. the future capacity, is given by,

$$FSC_t = CO_t + SC_t \tag{5}$$

Then, Equations (4) implies that FSC_t equals

$$FSC_t = ASC_t + CO_t - CbR_t \tag{6}$$

To summarize the manager's dynamics: longer queues increase his desired service capacity. The higher this desired service capacity, the more capacity the manager orders or the less he removes. Over time, the capacity orders will increase the available service capacity, while the capacity retirement will decrease it. Consequently, the higher (the lower) the available service capacity the lower (the higher) the number of customers queuing. Like the customers' dynamics, the two balancing loops, which describe the manager's behavior, may lead to stabilizing his perception over time. Thus, we are interested in studying how the manager analyzes the customers' behavior in order to adjust capacity and how the multiple delays involved in the system affect his decisions.

MODEL BEHAVIOR

Before trying out some alternative policies or strategies to model the manager's decisions and discussing descriptively some experimental results, we analyze the typical behavior of the system occurring when one of the equilibrium conditions is modified. The model is initially set under the equilibrium conditions, which are described in Table 1. Then we illustrate the impact on the system behavior of increasing the size of the initial customer base from 175 to 200. The other initial values remain as shown in Table 1. We simulate the model for 100 time units using a simulation step of 0.0625 time units.

State Variables	Equilibrium Value	Unit
Customer base	175	People
Queue	50	People
Average queue	50	People
Capacity on order	0	People / Time
Service capacity	25	People / Time
Capacity to be retired	0	People / Time
Perceived waiting time of current customer	2	Time unit
Perceived waiting time of potential customers	2	Time unit
Exogenous Variables	**Value**	**Unit**
Visit per time unit	0.15	1 / Time unit
Market reference waiting time (τ_{MR})	2	Time unit
Delays	**Value**	**Unit**
Time to perceive queue length ($1 / \beta$)	4	Time unit
Capacity delivery delay	4	Time unit
Capacity retirement delay	2	Time unit
Perception time of current customers ($1 / \varphi_c$)	2	Time unit
Perception time of potential customers ($1 / \varphi_p$)	4	Time unit

Table 1. Initial conditions of equilibrium

Figure 3 illustrates the evolution of the available service capacity and the number of customers waiting for service. We can observe that the manager adjusts the service capacity by imitating the evolution of the queue (i.e. the backlog of work). In this sense, he is trying to keep the average waiting time close to the market reference and while keeping the utilization rate close to 1, as shown in Figure 4. The lags involved in the manager and customer dynamics in addition to the manager's reaction result in the oscillating phenomenon and a certain decreasing tendency, as shown in Figure 3. Next, we go into more detail of the causes of this pattern.

An increase in the customer base will raise the arrival rate. Considering that the service capacity remains constant due to the lags involved in the capacity adjustment process and the formation of perceptions by the manager, more customers will wait for service. As the queue increases, the manager adjusts gradually his desired service capacity. According to Figure 1, the higher the desired service capacity, the larger the capacity orders. However, the capacity is delivered after 4 periods. The average waiting time therefore increases initially as plotted in Figure 4, affecting the perception of current customers and the expected waiting time of potential customers. When the perception of waiting time exceeds the market reference (2 time units), the customer base starts to decrease because more current customers are dissatisfied and fewer potential customers wish to join the facility. Hence, when the manager's decisions to add capacity start to materialize, the backlog of work (i.e. the queue) is falling. Consequently, the available service capacity reaches its peak at about the time the queue is reaching its nadir. Moreover, the manager reacts again to this behavior of the customers, but on this occasion by reducing his available service capacity to avoid having idle capacity. Neither manager nor customers consider the delays inherent in the reaction of each other. Hence, the backlog soars because of the manager's decision. Thus, despite the manager trying to adjust the service capacity by imitating the evolution of the queue, the multiple delays in the system bring about a fluctuating pattern as illustrated in figure 3.

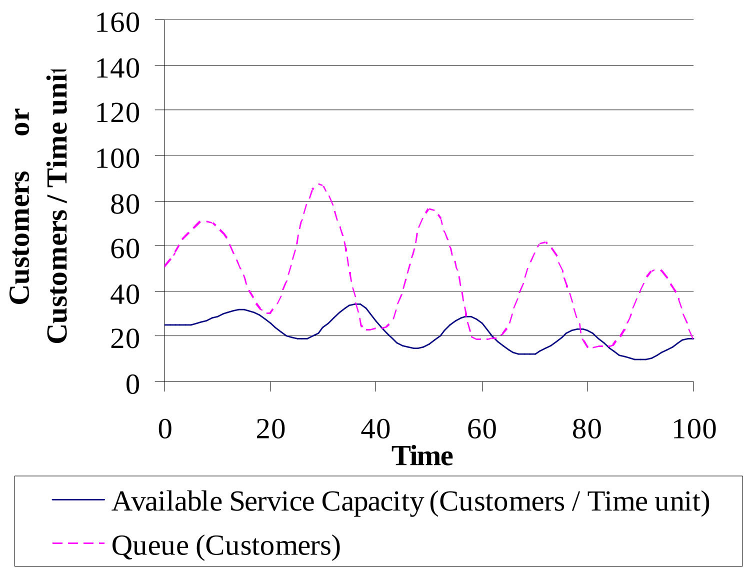


Figure 3. Illustrative behavior of the available service capacity and queue length

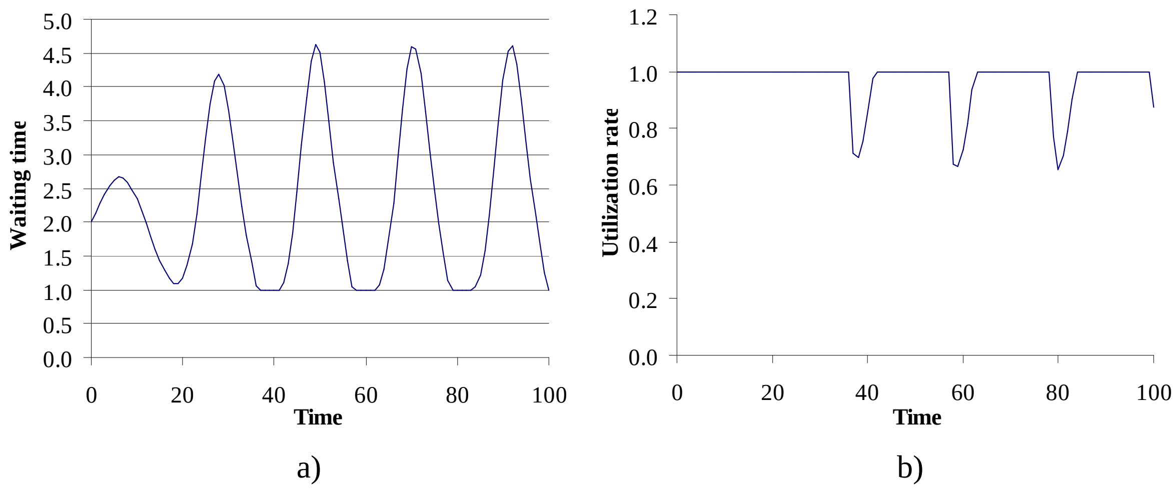


Figure 4. Illustrative behavior of (a) the average waiting time and (b) the utilization rate.

We have explained the model and illustrated a typical case where the manager reacts to customers' dynamics. In the next section, we propose other alternative decision rules to enable the manager to adjust capacity more effectively. These rules are based on the manager's perception of the backlog of work. Two alternative ways to form this perception based on the evolution of the queue are introduced. The decision rules consider both the required capacity adjustment and the speed at which this adjustment is carried out.

ALTERNATIVE DECISION RULES

The aim of the manager is to maintain sufficient available service capacity (ASC_t) in his facility in order to satisfy the customers. He thus decides whether to adjust the service capacity and at what time to do so. We propose a heuristic to determine the required capacity adjustment (RCA_t) by incorporating the speed at which the manager decides to adjust it. Let α be the service provider's speed to adjust capacity, i.e. how fast he decides to either add or reduce capacity. We defined above that the capacity adjustment decisions depend on the future service capacity (FSC_t), and the desired capacity (DC_t). Thus, including α in this definition, we may state RCA_t as follows:

$$RCA_t = \alpha * (DC_t - FSC_t), \qquad (7)$$

where α must be nonnegative and less than 1. This adjustment involves either an increase in capacity (when DC_t - FSC_t >0), a decrease in capacity (when DC_t - FSC_t < 0), or leaving capacity unchanged (when DC_t - FSC_t = 0). Taking into account that the capacity delivery delay may be different from the capacity retirement delay (c.f. Figure 2), we assume that the speed to either add or remove capacity can also be different. In this sense, the parameter α is determined as follows:

$$\alpha = \begin{cases} \alpha_1 & \text{if } DC_t - FSC_t < 0 \\ \alpha_2 & \text{if } DC_t - FSC_t >= 0 \end{cases} \qquad (8)$$

where DC_t and FSC_t are as defined in Equation 3 and 6. Consider now that the manager does not necessarily keep in mind all his previous decisions, some of which are still in the process of execution. Thus, the future service capacity (FSC_t), which the manager perceives, would be modeled as:

$$\mathrm{FSC_t = ASC_t} + \gamma * (\mathrm{CO_t - CbR_t}) \qquad (9)$$

where γ represents the proportion of the capacity adjustment that has not yet been implemented, which the manager takes into account. Replacing α, DC_t and FSC_t using Equations 8, 3 and 9, respectively, in Equation 7, the decision of how much to adjust capacity each period is determined by

$$RCA_t = \alpha * \left(\frac{\beta * Q_{t-1} + (1-\beta) EQ_{t-1}}{\tau_{MR}} - ASC_t - \gamma * (CO_t - CbR_t) \right) \qquad (10)$$

s.t.

$$\alpha = \begin{cases} \alpha_1 & \text{if} \quad \dfrac{\beta * Q_{t-1} + (1-\beta) EQ_{t-1}}{\tau_{MR}} - ASC_t - \gamma * (CO_t - CbR_t) < 0 \\ \alpha_2 & \text{if} \quad \dfrac{\beta * Q_{t-1} + (1-\beta) EQ_{t-1}}{\tau_{MR}} - ASC_t - \gamma * (CO_t - CbR_t) >= 0 \end{cases} \tag{11}$$

We propose a second manner to estimate DC_t. Instead of using adaptive expectations, the manager may simply consider the most recent backlog, i.e. customers waiting for service (Q_t), to estimate demand. That is, he looks at his current order book to decide how much capacity is required. Such an attitude is meaningful in situations where capacity can be adjusted fairly cheaply and quickly, e.g. by using temporary staff. In this case Equations 10 and 11 become:

$$RCA_t = \alpha * \left(\frac{Q_t}{\tau_{MR}} - ASC_t - \gamma * (CO_t - CbR_t) \right) \tag{12}$$

s.t.

$$\alpha = \begin{cases} \alpha_1 & \text{if} \quad \dfrac{Q_t}{\tau_{MR}} - ASC_t - \gamma * (CO_t - CbR_t) < 0 \\ \alpha_2 & \text{if} \quad \dfrac{Q_t}{\tau_{MR}} - ASC_t - \gamma * (CO_t - CbR_t) >= 0 \end{cases} \tag{13}$$

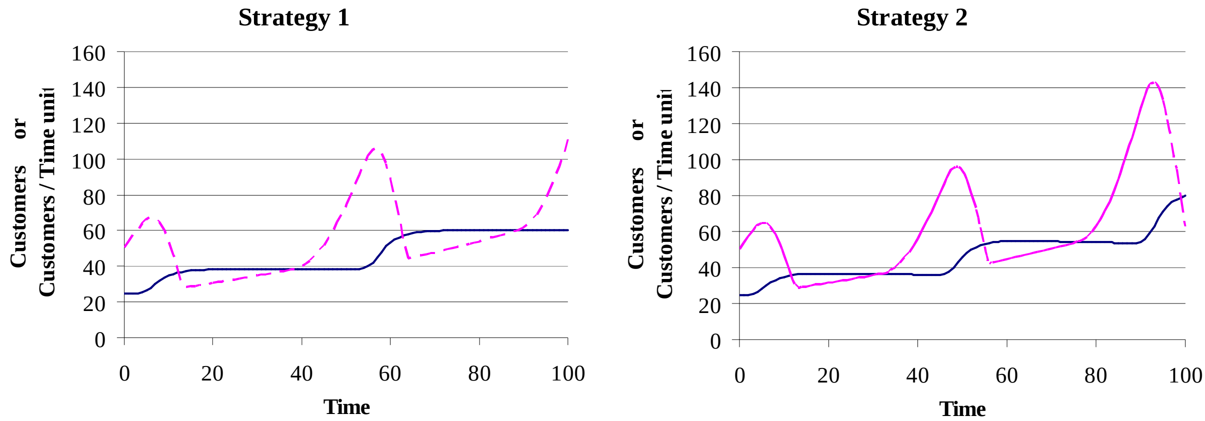
Optimal Strategies

Our objective is to find optimal values for the parameters $\alpha_1, \alpha_2, \beta$ and γ, which determine the above two strategies, to maximize the manager's cumulative profits over 100 time units. In order to calculate this profit we introduce a fixed cost and revenue resulting from providing the service. The equations 10 to 13 are nonlinear and thus complicated to optimize analytically. Thus, we apply simulation optimization (Keloharju & Wolstenholme, 1989; Moxnes, 2005) in order to find the optimal parameter values.

We use the optimizer toolkit of Vensim where the cumulative profits are set as the payoff function. The optimal parameter values we obtain are given in Table 2. According to this table, the second strategy, i.e. when the manager forms his perception based on the most recent value of the backlog, reaches the best payoff (2'151 compared to 2'059 for strategy 1). This occurs because when using strategy 2 the manager makes decisions a bit more aggressively than when using strategy 1, as shown figure 5. Hence, the manager reaches higher profits when he relies on the most recent information about the customers' behavior, i.e. Q_t. The optimal value of β (i.e. the coefficient of expectations), which equals 1 (see table 2), for strategy 1 strengthens the above remark. A coefficient of expectation equal to 1 means that the manager updates his expectation by using only the most recent information regarding the backlog. That is, the manager does not account for the past. In that case, Q_{t-1} is the latest information about the backlog the manager has to update his perception, EQ_t, at time unit *t*.

Strategy	Alpha 1	Alpha 2	Beta	Gamma	Maximum Payoff Value
Adaptive expectations	1.00	0.00	1.00	0.40	1'950
Most recent value of the backlog	1.00	0.00	N.A	0.37	2'071

Table 2. Optimal values of the parameters which define each strategy



(a) (b)

Figure 5. Evolution of the queue (i.e. backlog) and the available service capacity for the two capacity adjustment strategies with the optimal parameter values.

Figure 5 shows the behavior of the two parts of the system (customers and the manager) for both strategies. Their optimal behaviors are similar. Like in the base case, when the manager applies either of these two strategies, the backlog grows at the beginning of the simulation and the manager reacts by increasing capacity. However, as he bases his decisions on the most recent information about the backlog, he notices quickly that the backlog goes down. Thus, his decision to increase capacity becomes less aggressive resulting in the utilization rate gradually increasing back to 1 (see Figure 6). Consequently, the manager's decisions encourage current customers to remain loyal which in turn encourages the manager to keep the available service capacity constant. The manager's behavior brings about current customers being satisfied and thus inducing potential customers to patronize the facility through word of mouth. New customers joining the customer base imply that the arrival rate steeply increases. The manager responds by slowly increasing the available service capacity, which quickly reduces the queue. From this point onwards, an oscillating phenomenon starts to emerge. This oscillating pattern differs from that of the base case in that it grows exponentially over time.

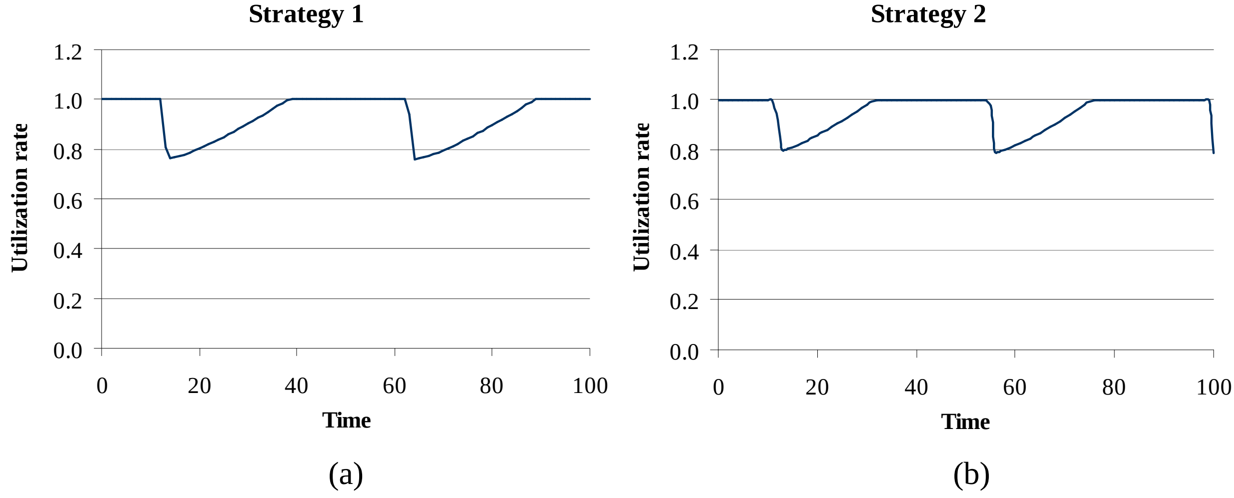


Figure 6. Evolution of the utilization rate for the two capacity adjustment strategies set up with the optimal parameter values.

Sensitivity analysis

We perform a sensitivity analysis to understand the impact of the different parameters, which define the alternative strategies, on the model behavior. In particular, we analyze the effect of a change in the values of these parameters on the manager's cumulative profits and the evolution of the queue.

First we illustrate the case in which we change α_2 (i.e. the speed at which the manager removes capacity). We select this parameter because it has the strongest impact. Figure 7 illustrates how changing the value of α_2 in both strategies affects the evolution of the queue and the manager's cumulative profits. We can observe that changes in these two variables emerge after about 27 time units, particularly, when α_2 is large (e.g. 0.5 or 1.0), i.e. when the manager quickly removes capacity. For instance, using both strategies with α_2 equal to 1.0 the cumulative profits decrease about 70% compared to the optimal value, while the backlog decreases by about 98% for strategy 1 and 94% for strategy 2. Likewise, the higher the parameter, the more the backlog oscillates.

Changes in the other parameters have small impacts on the evolution of the cumulative profits and the queue. As far as α_1 (i.e. the speed at which the manager add capacity) is concerned, for very small values (e.g. 0.0 and 0.1) the manager's cumulative profits and the queue are slightly reduced using both strategies. Regarding the speed at which the manager updates his perception in Strategy 1, i.e. β, varying this parameter results in similar effects as changing α_1. Finally, by trying different values of γ we found that they do not have any significant impact.

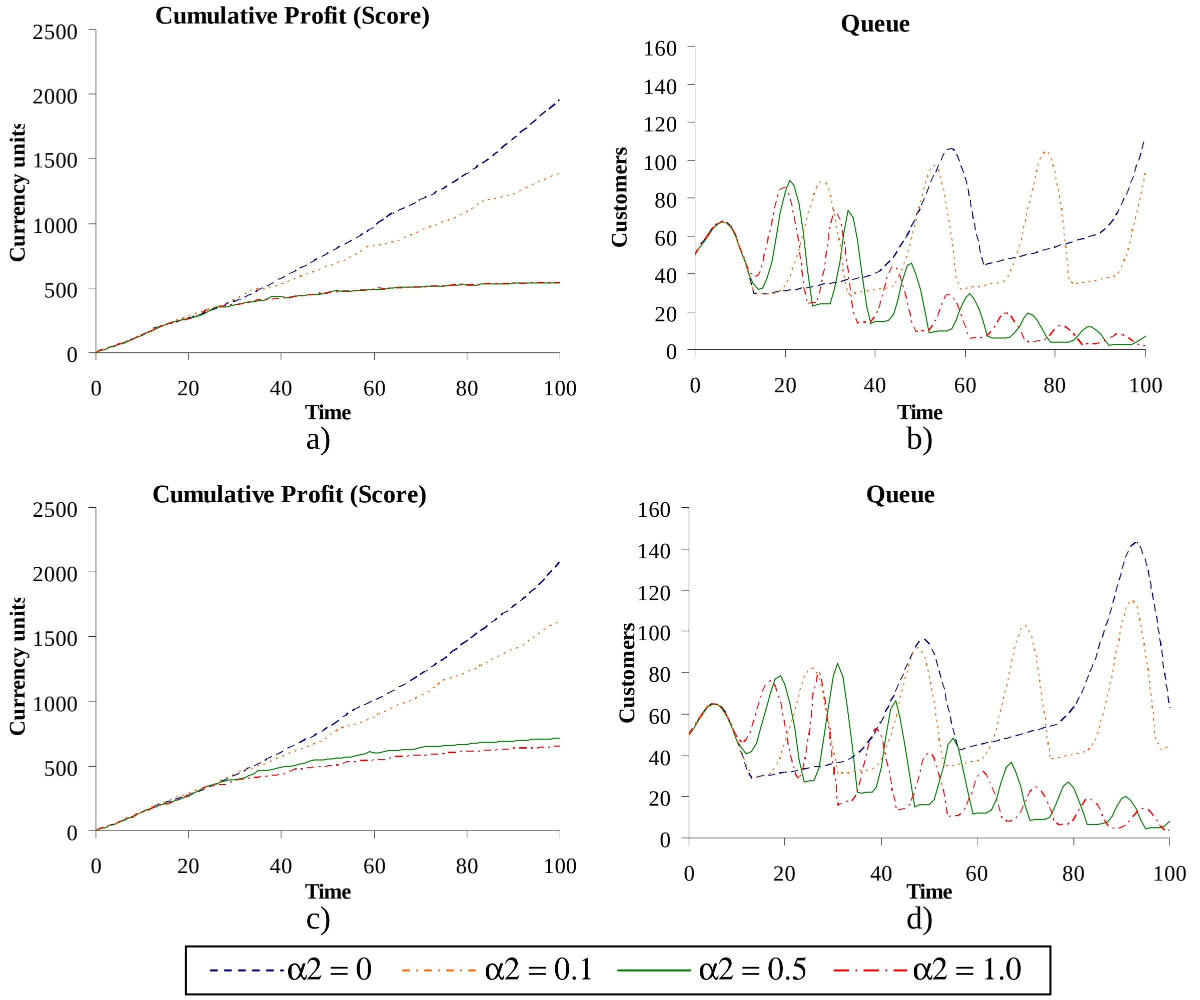


Figure 7. The cumulative profits and queue length when strategies 1 (Figs a and b) and 2 (Figs c and d) are simulated for selected values of α_2, keeping values of α_1, β, and γ constant as shown in Table 2.

A SERVICE FACILITY MANAGEMENT EXPERIMENT

We use the model described above as a computational platform to implement a laboratory experiment (c.f. Smith, 1982). The objective behind this experiment is to collect experimental information to assess how human subjects taking on the role of a manager face a situation in which they must adjust the capacity of a service facility. We also want to analyze how they use the available information to make capacity adjustment decisions. The subjects have information about the behavior of both the facility and the customers. Regarding the facility, they know the past and current available service capacity and utilization rate. As for customers, subjects know the past and current backlog (i.e. the number of customers waiting for service).

Experimental Protocol

We design this experiment based on the protocol for experimental economics (e.g. Smith, 1982; Friedman and Sunder 1994). We recruited undergraduate and master students in Finance, Management and Economics from the University of Lausanne. They were invited to participate in an experiment designed to study decision making in a service industry, through which they could earn up to 80 Swiss Francs. We received about 400 replies and selected 187 subjects following the principle of "first come, first served" in order to perform six experimental treatments. Each treatment had at least 30 participants. Subjects were allocated across eleven experimental sessions; each involved around 16 subjects and lasted, on average 90 minutes. Two facilitators supervised each session. The task of the subjects was to use a computer based interface, which portrayed the service capacity adjustment problem of a garage, to decide each period how much capacity to add or remove. They had to perform this task for 100 experimental periods.

This experiment was conducted in the informatics laboratories of the School of Business and Economics. Upon arrival at the laboratory, the subjects were allocated to a PC and separated from their neighbor by another PC. Communication between the subjects was forbidden. Once they were seated, we gave them written instructions and a consent form, which they had to sign before starting the experiment. Then, a short introduction to the experiment was presented to them. The instructions were quite simple and provided subjects with a short explanation of the system that they had to manage in the experiment and all the information, which they had available to carry out their task. We present the instructions and the interface used to run the experiment in the appendix of this paper.

We gave the subjects the payoff scale through which they earned their reward depending on their performance in the experiment. Performance was measured based on the cumulative profits that subjects had at the end of the experiment, i.e. at the period 100 or when the available service capacity reached 0, If that happened before than the period 100.

Experimental Treatments

In addition to the base case, we have designed other five experimental treatments to understand how the manager adjusts the capacity of an industry service. These five treatments are divided in two groups to study the effect of different factors. The first group is composed of four treatments and its objective is to analyze how the delay structure, inherent to the system, affects how the manager decides to adjust capacity. This delay structure includes the delays the manager knows (i.e. the implicit lags in capacity adjustment), and those which are unknown to him (i.e. the time required by potential and current customers to update their perceptions). The last group has a single treatment, which includes a cost to add or remove capacity. Table 3 summarizes the conditions of each treatment.

Treatment	**Current customers Delay**	**Potential customers Delay**	**Time to increase capacity**	**Time to decrease capacity**	**Cost per unit change in capacity**
Base Case	4	2	4	2	-
A	10	2	4	2	-
B	6	4	4	2	-
C	4	2	8	4	-
D	4	2	2	1	-
E	4	2	4	2	1

Table 3. Treatment conditions.

EXPERIMENTAL RESULTS

All subjects overreact to the initial increase of the backlog. This sudden rise is independent of subjects' decisions since it depends on the initial conditions. Thus, we can interpret this first reaction of the subjects as a learning process in which they are trying to adapt to the system behavior. In other words, we can call this initial period a transition period. Recall that we observed a similar pattern of the backlog in the simulation results.

From this point onwards, we identify three groups of subjects, whose decisions result in similar behavioral patterns. Figure 8 illustrates the evolution of the backlog and the available service capacity of two typical subjects of each group. The first group is composed of those subjects who overreact strongly to the initial overshoot of the backlog and then they make many small decisions to gradually adjust capacity over time (e.g. Subjects 5 and 11). Most of these decisions concern capacity addition. Consequently, the garage's available service capacity for this kind of managers presents an exponential increase over time. After the initial transition, the available service capacity and the queue behave in the same way. Thus, we can consider that these subjects quickly learn to manage the system to achieve sustainable growth. The subjects in this group achieved the higher scores of the experiment.

The second group (e.g. Subjects 12 and 18) represents those subjects who, after their slight overreaction to the initial backlog, make fewer but more aggressive capacity adjustment decisions than the subjects of the first group. Moreover, they continue to overreact to the evolution of the backlog over time. This behavior results in an oscillating pattern for both the backlog and the available service capacity: they increase exponentially, but more slowly than for the first group. These two groups, despite achieving quite different behavioral patterns compared to the two optimal strategies discussed before, attain similar total profits.

The last group includes subjects who, even after the transition period, continue to overreact significantly to the evolution of the backlog (e.g. Subjects 3 and 30). Although in some cases the backlog evolves as when simulating the optimal strategies (see Figure 5), the subjects did not capture the customers' behavior. We can consider that these subjects were unable to handle the delay structure inherent to the system. They performed poorly, achieving the lower payoffs, and occasionally finding themselves with zero service capacity before the end of the experiment.

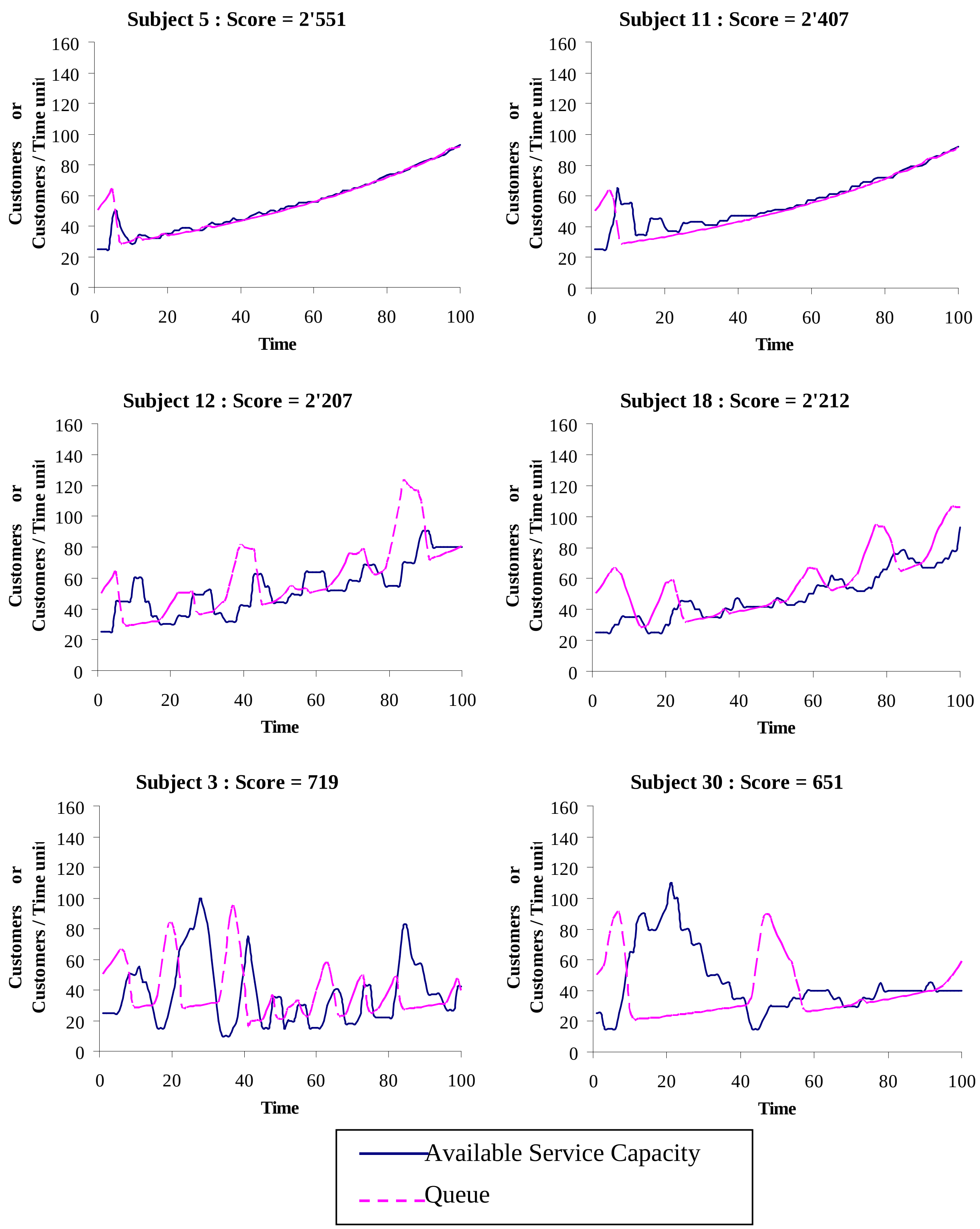


Figure 8. Experimental results for six typical subjects.

Treatment Results

The outcomes of the treatments were compared using the *Wilcoxon Rank-Sum test or Mann-Whitney U test*. Table 4 shows the corresponding p-values. Using a 0.05 significance level, these p-values enable us to interpret that the cumulative profits achieved in treatments C (i.e., slow adjustment) and D (i.e., fast adjustment) are, on average, significantly different compared to the cumulative profits achieved in the other treatments. By looking at the box plots in Figure 9 we can get an idea of such a difference as the mean cumulative profits of treatments C and D are either above or below the mean cumulative profits of the other treatments, supporting the remark inferred from the Wilcoxon Rank-Sum tests. We can also observe that the variability in treatment D is less compared to that of the other treatments. In addition, the distributions of treatments A, C, D and E are reasonably more symmetric than those of treatment B and the base case.

Col Mean - Row Mean P-Values	Basecase	Treatment A	Treatment B	Treatment C	Treatment D
Treatment A	0.2805				
Treatment B	0.9035	0.1772			
Treatment C	0.0008	0.0029	0.0000		
Treatment D	0.0002	0.0000	0.0003	0.0000	
Treatment E	0.2871	0.7562	0.2310	0.0003	0.0000

Table 4. P-values of the Wilcoxon Rank-Sum test for the cumulative profits

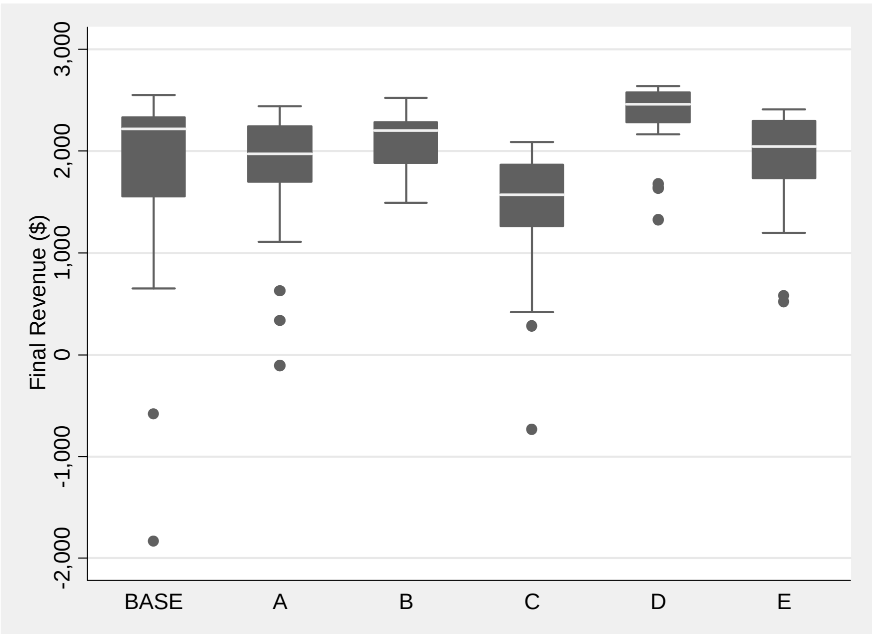


Figure 9. Box plots for the cumulative profits by treatment

CONCLUSIONS AND FURTHER WORK

In this paper, we have applied a system dynamics model to study how the manager of a service facility adjusts capacity based on his perception of the queue length, whereas potential and current customers react to the managers' decisions. While current customers update their perception based on their own experience and decide whether to stay in the customer base, potential customers update their perception through word of mouth and decide whether to join the customer base. We have simulated the model and analyzed the evolution of the backlog of work and the available service capacity. Based on this analysis we have proposed two alternative decision rules to maximize the manager's cumulative profits. Then, we have illustrated how we developed an experiment to collect information about how human subjects taking on the role of a manager in a lab environment face a situation in which they must adjust the capacity of a service facility.

Simulating this queuing model showed that when the manager tries to adjust the service capacity by imitating the evolution of the queue (i.e. the backlog of work), the multiple delays in the system bring about an oscillatory phenomenon. Optimizing the parameters, which set the alternative strategies, we found that the manager reaches higher profits when he relies on the most recent information about the customers' behavior, i.e. the most recent backlog. The sensitivity analysis enables to conclude that changes in the speed at which the manager removes capacity have a strong impact on the evolution of the available service capacity and the backlog. Varying the other parameters results in small impacts on the evolution of these two variables.

As far as the experiment is concerned, we identify three groups of subjects, whose decisions bring about similar behavioral patterns. The first group included the subjects who overreact strongly to the initial sudden increase of the backlog and make many small decisions to gradually adjust capacity over time. The second group represented the subjects who, after overreact to the initial backlog slightly, they make fewer but more aggressive capacity adjustment decisions than the subject of the first group. The last group included subjects who even, after the transition period, overreact significantly to the backlog. The two first groups, despite quite different behavioral patterns compared to the two optimal strategies discussed, achieved similar total profits.

The next step will be estimate a decision rule which adjusts to collecting data from Subjects. Extensions include incorporating prices to manager' decisions, i.e. a unit cost for each unit of capacity which the manager decides to add or remove. An interesting approach would be to conduct another experiment wherein another group of human subjects will assume the role of customers.

REFERENCES

Albright, S. C., & Winston, W. L. (2009). *Management Science Modeling* (Revised 3r., p. 992). South Western Cengage Learning.

Arango, S., Castaneda, J., & Olaya, Y. (2011). Laboratory Experiments and System Dynamics. Medellin.

Borshchev, A., & Filippov, A. (2004). From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools. *The 22nd International Conference of the System Dynamics Society*. Oxford.

Dewan, S., & Mendelson, H. (1990). User Delay Costs and Internal Pricing for a Service Facility. *Management Science*, *36*(12), 1502-1517.

Friedman, D., & Sunder, S. (1994). *Experimental methods: a primer for economists* (1st ed.). Cambridge: Cambridge University Press.

Haxholdt, C., Larsen, E. R., & van Ackere, A. (2003). Mode Locking and Chaos in a Deterministic Queueing Model with Feedback. *Management Science*, *49*(6), 816-830.

Keloharju, R., & Wolstenholme, E. F. (1989). A Case Study in System Dynamics Optimization. *Journal of the Operational Research Society*, *40*(3), 221-230. doi:10.1057/palgrave.jors.0400303

Moxnes, E. (2005). Policy sensitivity analysis: simple versus complex fishery models. *System Dynamics Review*, *21*(2), 123-145. doi:10.1002/sdr.311

Naor, P. (1969). The Regulation of Queue Size by Levying Tolls. *Econometrica*, *37*(1), 15-24.

Nerlove, M. (1958). Expectations and Cobweb Phenomena. *The Quarterly Journal of Economics*, *72*(2), 227-240.

Rapoport, A., Stein, W. E., Parco, J. E., & Seale, D. A. (2004). Equilibrium play in single-server queues with endogenously determined arrival times. *Journal of Economic Behavior & Organization*, *55*(1), 67-91. doi:10.1016/j.jebo.2003.07.003

Smith, V. L. (1982). Microeconomic Systems as an Experimental Science. *The American Economic Review*, *72*(5), 923-955.

Sterman, J. D. (2000). *Business Dynamics: Systems thinking and modeling for a complex world* (p. 982). Chicago, IL: Irwin-McGraw Hill.

van Ackere, A. (1995). Capacity management: Pricing strategy, performance and the role of information. *International Journal of Production Economics*, *40*(1), 89-100.

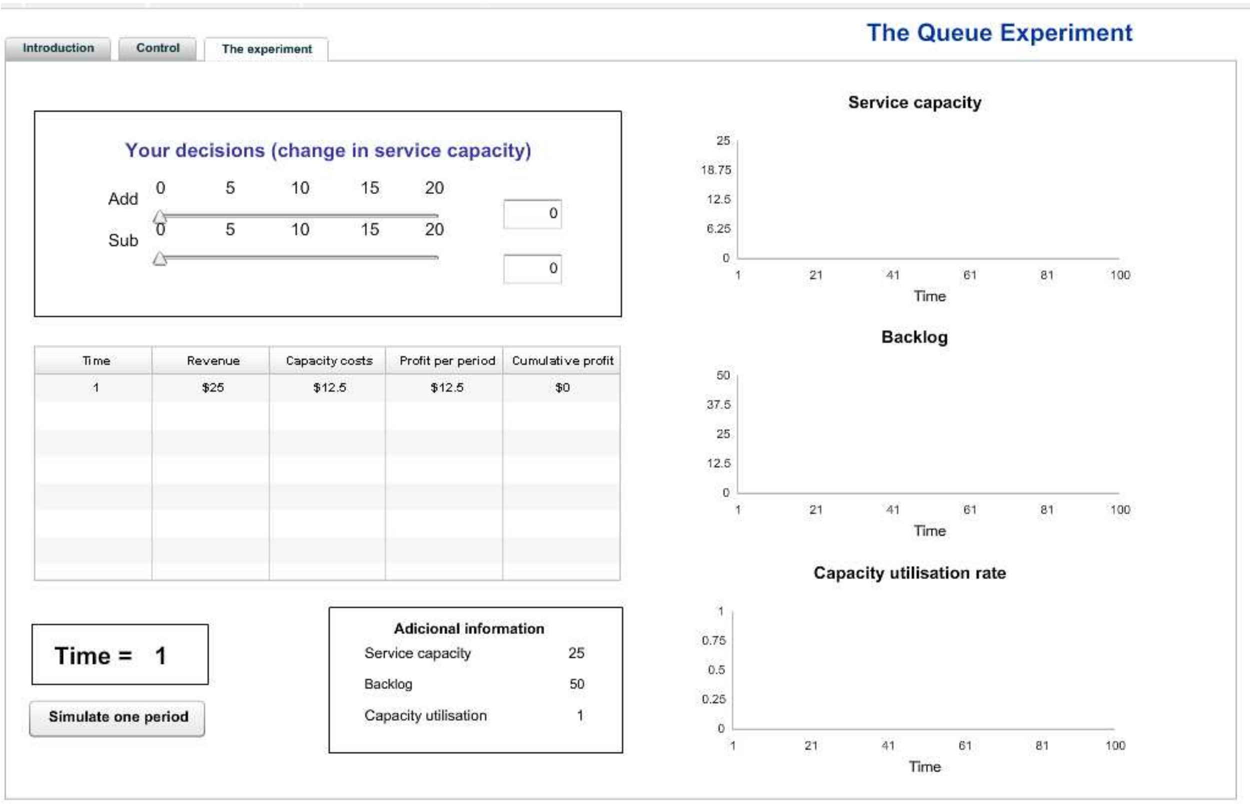
van Ackere, A., Haxholdt, C., & Larsen, E. R. (2006). Long-term and short-term customer reaction: a two-stage queueing approach. *System Dynamics Review*, *22*(4), 349-369.

van Ackere, A., Haxholdt, C., & Larsen, E. R. (2010). Dynamic Capacity Adjustments with Reactive Customers. *Management*. Lausanne.

Yechiali, U. (1971). On optimal balking rules and toll in the GI/M/1 queuing process. *Operations Research*, *19*(2), 349-370.

APPENDIX

A. Computer Interface



B. Subjects' Instructions (Base case)

Instructions for the participants

NOTE: PLEASE DO NOT TOUCH THE COMPUTER BEFORE BEING ASKED TO DO SO

Welcome to the experiment on decision making in a service industry. The instructions for this experiment are quite simple. If you follow them carefully and make good decisions, you may earn a certain amount of money. The money will be paid to you, in cash, at the end of the experiment. You are free to halt the experiment at any time without notice. If you do not pursue the experiment until the end, you will not receive any payment. The University of Lausanne has provided funds to support this experiment. If you have any questions before or during the experiment, please raise your hand and someone will come to assist you.

We assure you that the data we collect during the course of this experiment will be held in strict confidence. Anonymity is guaranteed; information will not be reported in any manner or form that allows associating names with individual players.

Description of Experiment

This experiment has been designed to study how managers adjust service capacity in a service facility. Below is a short explanation of the system that you will have to manage in the experiment. It is a relative simple system and you only have to make two decisions each time period (increasing capacity and/or decreasing capacity).

The situation

You are the manager of a large garage, which repairs and maintains cars. You have an existing customer base as well as many potential customers who currently are not using your services, but might consider doing so in the future. Both groups are sensitive to the waiting time.

Waiting time: is the average time between the moment a customer calls your garage to make an appointment and the time the car has been serviced. This depends on two factors, how many other customers have made reservations previously (i.e. how long is the queue) and the service capacity of the garage (i.e. how many cars can on average be serviced per time period). Due to planning constraints, this waiting time cannot be less than one month.

Customers: These customers use your garage on average every twice a year. They evaluate the expected waiting time (which is based on (an average of) the last few times they have used your garage) and compare this expected waiting time to the time they consider acceptable (the average for the industry, which is 2 months: the elapsed time between the moment a customer calls, and the moment he can pick up his car after servicing averages 2 months). If they are satisfied (i.e. the expected waiting time is comparable to or better than the average for the industry) they will remain your customer and return again to use your garage. If they consider that the waiting time is too long compared to the industry average they will switch to another garage.

Potential customers: These are people who might become customers if they consider that your waiting time is attractive (i.e. less than the industry average). However, given that they are currently not among your customers, they only hear about the waiting time at your place through word of mouth. Consequently, their estimate of the waiting time at your place is based on less recent information than the estimate of your current customers. Note: the number of potential customers is unlimited.

Service Capacity: This is the number of cars the garage can service on average in one month. You, as the manager, control the service capacity of the garage, i.e. you have the possibility to increase and/or decrease capacity. However, this cannot be done instantaneously: it takes 4 months to increase capacity (e.g. ordering more tools, hiring people, acquiring more buildings etc) and 2 months to decrease capacity (end a lease on a building, lay off people, etc). Note: If at some point your decisions result in a service capacity equal to zero (0), the garage will be closed and the experiment is ended.

Your Task

As the manager, you make decisions regarding any change in capacity for the garage each month. To help you make these decisions you have information about the number of customers currently waiting for service or whose car is currently being serviced (referred to as the queue), profit, the current capacity of the garage, and the capacity utilization rate. You goal is to maximize the total profit over 100 months.

Cost and revenue information:

Profits [E$/month] = Revenue – Cost

Revenue [E$/month]

= number of customers served [cars/month]*Average Price per Customer [E$/car]

Average Price per Customer = 1 $/car

Cost [E$/month]

= Service capacity [units]* Unit cost of service capacity [E$/unit/month]

Unit cost of service capacity = 0.5 $/unit/month

Interface

In front of the computer, you will have the interface where all interactions will take place. The information is the same as what we have provided in these instructions. Please ask the facilitator to have a trial run to test out the software.

Payment

At the end of the experiment, you will receive a cash reward. This will consist of a guaranteed participation fee of 20CHF, plus a bonus which will depend on the total profit you have achieved. This bonus will vary between 0 and 60CHF. ***If you do not pursue the experiment until the end, you will not receive any payment.***

You will be asked to complete and sign a receipt with your name, email address, and student ID number. Thereafter, you can collect your payment. We will be happy to answer any questions you may have concerning this experiment.

If you want to participate in this experiment, please sign the consent form on your desk. ***This form must be signed before the start of the experiment***

If you have no further questions, please ask the experiment facilitator to begin. Good luck and enjoy the experiment.