European Journal of Operational Research 000 (2016) 1–13

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Managing capacity at a service facility: An experimental approach

Carlos A. Delgado-Alvarez[a,*], Ann van Ackere[b], Erik R Larsen[c], Santiago Arango-Aramburo[d]

[a] Faculty of Business, Politecnico Colombiano Jaime Isaza Cadavid, Carrera 48 No. 7-151, Medellin, Colombia
[b] Faculty of Business and Economics (HEC), University of Lausanne, Quartier UNIL-Dorigny, Bâtiment Anthropole, 1015 Lausanne, Switzerland
[c] Institute of Management, Università della Svizzera Italiana, Via Buffi 13, 6900 Lugano, Switzerland
[d] Decision Science Department, Universidad Nacional de Colombia, Carrera 80 N. 65-223, bloque M8-211, Medellín, Colombia

ARTICLE INFO

Article history:
Received 1 September 2015
Accepted 29 September 2016
Available online xxx

Keywords:
System dynamics
Queuing systems
Capacity adjustment management
Experimental economics
Behavioural operations management

ABSTRACT

Most research in queuing has focused on the optimisation of performance and equilibrium analysis, with little attention to the understanding of how managers actually make decisions. In this paper, we use an experimental set-up to investigate the decision-making process in a queuing capacity expansion problem, in the presence of capacity adjustment delays. The experiment represents a queuing system with one facility and virtual customers who decide whether or not to patronise the facility. Subjects play the role of facility managers who adjust the facility's service capacity to maximise profits. We analyse the actions of the manager to provide new insights into how their behaviour affects the evolution and success of the service system. Our results identify three types of managers: incremental, lumpy and reactive investors, and indicate that these groups use different decision rules. The first group achieved the best performance, and the last group the worst. While managers' decisions are influenced by the backlog of work and the available service capacity, they do not correctly account for their past, but not-yet-implemented, decisions.

1. Introduction

Most research in queuing problems has been focused on the optimisation of performance measures and equilibrium analysis, starting with Naor's seminal paper (Naor, 1969). More recent work includes Allon and Federgruen (2008); Bassamboo and Randhawa (2010) and Wang, Debo, Scheller-Wolf, and Smith (2010). Little attention has been given to the understanding of how managers actually make capacity decisions on a day to day basis: capacity is either assumed given, or managers make a one-off capacity decision.

In this paper, we consider 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 present study builds on the previous work of van Ackere, Haxholdt, and Larsen (2013) on behavioural queuing. We adapt their system dynamics (SD) model as an experimental platform to investigate how subjects, playing the role of a manager in a laboratory environment, manage the capacity of a service facility where the customers' satisfaction depends on the waiting time.

We study the capacity management of a service facility, where managers' decisions take time to be implemented and customers take time to update their perceptions. Our focus is thus on the dynamic process of how managers choose to adapt capacity over time, contrary to Allon and Federgruen (2008); Bassamboo and Randhawa (2010) and Wang et al. (2010) where managers make a one-off capacity decision aimed at optimising an equilibrium situation. The problem is framed as the management of a garage where customers must make an appointment to take their cars. The customers' waiting time runs from the time they make their appointment to the time their car is serviced. The task of the subjects is therefore to manage the capacity (i.e., add or remove capacity) in order to satisfy their current customers and attract potential customers. When subjects decide to increase capacity, the capacity orders materialise after a delivery delay. Similarly, the retirement decisions involve a dismantling delay.

Few experiments in economics and management science study queuing problems. Rapoport, Stein, Parco, and Seale (2004) formulated a queuing problem with endogenously determined arrival rates and state-dependent feedback as a non-cooperative n-person game. These experiments focus on studying behaviour from the point of view of the customers. Subjects, playing the role of car

* Corresponding author.
E-mail addresses: cadelgado@elpoli.edu.co, cadelga0@unal.edu.co (C.A. Delgado-Alvarez), ann.vanackere@unil.ch (A. van Ackere), erik.larsen@usi.ch (E.R. Larsen), saarango@unal.edu.co (S. Arango-Aramburo).

http://dx.doi.org/10.1016/j.ejor.2016.09.056
0377-2217/

owners who need to take their car to a garage for the emissions control, should decide each period whether or not to join the queue and when to do so. Their findings show that mixed-strategy equilibrium solutions explain the behavioural arrival patterns of customers very well at an aggregated level, but not at the individual level. Subsequently, Seale, Parco, Stein, and Rapoport (2005) extended the work of Rapoport et al. (2004) to non-cooperative n-person games with complete information (i.e., including the information of the other group-members). They support Rapoport et al. (2004)'s findings, and conclude that these remain valid as long as congestion is unavoidable and information about group behaviour is not provided. Stein, Rapoport, Seale, Zhang, and Zwick (2007) and Rapoport, Stein, Mak, Zwick, and Seale (2010) performed further experiments to study queuing systems with endogenous arrival rates and batch service. They analyse how customers decide whether to join a queue and when to do so. They conclude that players with experience in a constant capacity batch queuing game converge to equilibrium play on an aggregated level when the service capacity is fixed and commonly known, but not when service capacity changes from round to round. Pazgal and Radas (2008) perform an experimental study to compare customer balking and reneging behaviour to that predicted by queuing theory. They conclude that while the subjects' behaviour is consistent with the predictions (little reneging, balking if the queue exceeds a critical value), the level of balking is suboptimal: on average, the subjects' critical value is too large.

SD models have been used extensively to study different aspects of capacity management problems using an experimental approach. One of the most-studied aspects is supply chain management, with a particular focus on the well-known beer game example with its bullwhip effect, see for instance Sterman (1989a); Villa, Gonçalves, and Arango (2015), and Choi and Messinger (2016). Other studies have focussed on the industry level. For instance, Paich and Sterman (1993) study the "boom and bust" cycle, where sales of a new product first grow exponentially, before falling back to the replacement level. At a national level, Sterman (1989b) studies the economic longwave using a capital investment accelerator model. Arango and Moxnes (2012) show that in the presence of capacity vintages and long delays, price cycles can emerge endogenously. A review of SD models used to perform laboratory experiments can be found in Arango, Castañeda, and Olaya (2012), while Wang and Disney (2016) review more specifically work on the bullwhip effect.

Our work differs from the above in several respects. First, the context we study is different: we consider a service facility with two customer classes (current and potential), and subjects play the role of the manager, not of a customer. Second, we consider a significantly more complicated delay structure than previous work: the two customer types have different reaction times, and the manager faces asymmetric capacity addition and dismantling delays; in the underlying SD model, the manager also had a perception and a decision delay. Third, previous work implicitly assumes that the different model behaviours can be generated by a single decision rule. Based on our results, we argue that generating the behaviours of the different types of investors actually requires using different decision rules.

Our experimental results indicate that subjects can be classified into three types of managers according to the way in which they adjust the service capacity: those who make incremental investments in service capacity; those who make lumpy investments; and those who overreact to changes in the backlog by repeatedly adding and removing capacity without any logic. Managers' decisions are influenced by the current backlog of work and the available service capacity, but they do not correctly account for their past decisions, which have not yet been implemented, when making future decisions.

The remaining sections of this paper are organised as follows: first, we briefly describe the SD model used for the experiment, focusing on the adaptations made to the original model. Next, we present the experimental design and discuss the results: we provide a descriptive analysis of the way subjects take their decisions and test the experimental hypothesis. The final section contains our conclusions and suggestions for further work.

2. A service facility management model

The system dynamics queuing model used in this paper was originally proposed by van Ackere et al. (2013) to study the feedback and delay structure involved in the relationship between customers and the manager of a service facility. Customers decide whether or not to use the facility for service, while the manager adjusts the service capacity in order to attract more customers, while at the same time trying to minimise spare capacity. Most queuing models assume that customers only use the service once; they thus do not create expectations about waiting times. However, we consider a situation where customers repeatedly require the same service, i.e., depending on the expected service time they will return to the facility or turn to a competitor. We therefore need to be able to model both feedback processes and expectation formation processes. Given its focus on feedback and lags, System Dynamics (SD) is the most appropriate method to achieve our goals (Morecroft, 2015; Sterman, 2000).

As for any modelling approach, its validity and credibility depends on how well the model captures the main feedback structures (Barlas, 1996). This can be an issue when modelling socio-economic systems, where controversies can arise due to different perspectives and focus of the various stakeholders; a well-known example concerns the World model, see Myrtveit (2005) for details. This issue does not arise in our model, given its stylised nature. Another limitation of the SD approach is its continuous and aggregate nature (perfect mixing hypothesis, Sterman, 2000). For instance, we do not distinguish between individual customers, but instead model flows of customers; all expected customers have the same expected waiting time and so do all the potential customers. While in the underlying SD model investments and divestments were also modelled as continuous processes, in the experiments they are replaced by discrete decisions taken by the subjects.

The causal loop diagram in Fig. 1 shows the feedback structure of the actors of the system. The model consists of two sectors, customers (left) and the manager (right), connected by the queue. Customers decide whether to use the facility based on their expectation of the waiting time, while the manager decides to adjust the service capacity based on the queue length. Examples of this kind of system include car maintenance facilities, and workers or students who daily patronise a restaurant for lunch. In both examples, customers have no obligation to use the facility and the manager is motivated to encourage customers to use his facility by adjusting its service capacity. The capacity ordering process is described in detail in Fig. 2. We next describe the main equations of the model. A notation table is given in Appendix A and the full equation listing can be found in Appendix B.

The number of customers queuing is modelled as the backlog of work which the facility has to serve over the following periods. Given the continuous nature of System Dynamic models, the estimate of the waiting time is considered to be the time between the moment customers make an appointment for service and the moment their service is completed. Once the customers make their appointment they become part of the backlog of customers waiting for service. Thus, the waiting time of these customers depends on the service rate (μ_t) of the facility. We compute the waiting time (W_t) each period as the ratio between the backlog of customers