JENA ECONOMIC RESEARCH PAPERS

2014 – 030

An Experimental Investigation into Queueing Behavior

by

Anna Conte
M. Scarsini
Oktay Sürücü

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

An Experimental Investigation into Queueing Behavior

Anna Conte[a,b,*], M. Scarsini[c,†], Oktay Sürücü[d,‡]

[a] *Max Planck Institute of Economics, Kahlaische Str. 10, 07745 Jena, Germany*

[b] *WBS, University of Westminster, EQM Department, 35 Marylebone Road, NW1 5LS London, UK*

[c] *Dipartimento di Economia e Finanza, LUISS, Viale Romania 32, 00107 Roma, Italy*

[d] *Center for Mathematical Economics, Bielefeld University, Germany*

Abstract

We conduct an experiment meant to explore the factors driving customers' decisions in a queueing system. Under different time allowance conditions, the experimental subjects are asked to join one of two queues that differ in their length, server speed, and entry fee. We investigate some aspects of a queue on which subjects base their decision and what is the effect of time pressure on their decision criteria. We find that only a proportion of subjects behave rationally and use the relevant information efficiently. The size of this proportion increases when time limitation is relaxed. The rest of the subjects seem to adopt a rule of thumb that ignores the information on server speed and follows the shorter queue. Consequently, these subjects use less time than rational types when making their decisions. Furthermore, time pressure harms decision performance since the presence of time limitation stresses a great deal of subjects and causes them to use time inefficiently.

JEL Classifications: C91, L00, C33, C35

Keywords: Queues with entry fee, join the shortest queue, laboratory experiments, decision times

*Email: aconte@econ.mpg.de; a.conte@westminster.ac.uk

†Email: marco.scarsini@luiss.it

‡Email: oktay.surucu@uni-bielefeld.de (corresponding author)

I INTRODUCTION

Queues are formed, and customers have to wait, whenever the capacity of a service provider fails to meet the instantaneous demand. There are many instances in everyday life when one encounters queues, for example, when buying museum or concert tickets, conducting a transaction in a bank, calling a hotline, entering in a popular restaurant or club, etc.

Waiting in a queue is costly. Therefore, a customer may decide to balk at the prospect of waiting or to abandon the queue after joining and waiting for a while. Moreover, customers may even be willing to pay extra in order to decrease or eliminate waiting times. Visitors of a Six Flags amusement park, for example, can buy one of three types of pass (regular, Gold and Platinum), in order to eliminate physical wait in queues and reduce the actual waiting time. A driver without any passenger can pay a fee and use high occupancy vehicle (HOV) lanes that are originally designed for carpools of two or more.

What determines customers' behavior is the comparison between the expected benefit of getting the service and the expected cost of waiting. Under the assumption of full rationality, this comparison is made by extracting information about the length, velocity and the entry fee of a queue. However, it is questionable whether people behave rationally and use the information they could extract when making a queueing decision. The present study ought to shed light on this issue by analyzing the characteristics of queues to which people pay attention when making such a decision. More specifically, this study aims to answer the question of whether customers accurately calculate the costs and benefits of joining a queue and make their decision accordingly. If this is not the case, what is the behavioral pattern followed? Which aspects of queues play important roles and affect decisions? What is the effect of time pressure?

Understanding how customers behave in a queueing system helps to determine how to operate a system in the most efficient way. According to Hillier and Lieberman (2001), 37 billion hours per year are spent waiting in queues in the US, which would amount to 20 million person-years of useful work per year, if it were spent productively. This emphasizes the critical importance of designing queueing systems based on customers' behavior from a social welfare perspective.

In this article, we experimentally study and analyze customers' behavior within a simplistic queueing system. In a computerized laboratory setting, we ask subjects to choose between two given queues, each of which is connected to a different server. The servers provide the same service but they differ in entry fee,[1] speed and length of the queue connected. There are 40 such tasks and each of them

[1] The amount one has to pay in order to join the queue of a server.

is repeated three times under different treatment conditions of time allowance: 5 seconds ($5sec$), 10 seconds ($10sec$) and no time limitation ($no-tl$). This is done in order to examine the impact of time pressure on customers' choice.[2]

We analyze the data by means of a finite mixture model. The mixture approach enables us to identify whether there are subjects in our sample who use all the provided information efficiently, that is, base their decisions mainly on the profits they would gain joining each queue, and if there are subjects who adhere, instead, to alternative decision criteria.

As far as the results are concerned, our experimental analysis suggests that only a proportion of the population makes choices exploiting information at best and the size of this group increases when time limitations are relaxed. The remainder of the population appears to consider only part of the provided information when making their decisions concerning the queue to join. Alternatively, we might say that these people ponder information in a less-than-efficient way. In particular, what is interesting here is that they seem to ignore the average waiting time of each queue as a factor in their decisions, even though it is explicitly given, and tend to join the shorter queue. The existence of this type of behavior is also supported by the field experiment conducted by Lu et al. (2013) at a grocery store deli counter.

From the analysis of the actual decision times, we discover that the decision time significantly changes across treatments, and in particular it increases as the time limitation is lengthened. Moreover, a comparative analysis between average decision times of profit maximizers and naïve subjects shows that the former takes longer than the latter in any treatment. Another interesting result we obtain is that, when we introduce a time limitation, decision performance worsens significantly even if the given time is more than what they spend on average without any limitation. When subjects feel time pressure, they use time inefficiently and this harms their performance.

The characterization of customer behavior in a queueing system analyzed in this article provides useful inputs and suggestions for researchers as well as practitioners. The existence of two types of customers raises the question of whether it is possible to design a mechanism that screens and discriminates customers. The effects of this type of discrimination on service provider's profit and on welfare could be further investigated. Moreover, the findings we obtain in this article potentially pave the way for further research on queueing behavior under more complicated settings.

The paper is organized as follows. Section II provides a historical excursus and discusses the research papers which, to our knowledge, are most closely related to the present study. The experi-

[2] In many real life situations customers should quickly decide which queue to join so that no newly arrived customer joins the system and creates negative externality.

mental design and procedures are discussed in Section III. Section IV outlines the characteristics of the sample. Section V describes the econometric model of the choice data, presents and discusses its results and implications. Section VI concludes.

II LITERATURE REVIEW

The birth of Queueing Theory dates back to 1909, when Agner Krarup Erlang (1878–1929) published his pioneering work on telephone traffic. Since then, his contributions have been widely applied in many different fields. In economics, the first main contribution is due to Naor (1969) and, after this seminal paper, the number of studies in this area has sensibly grown (see Hassin and Haviv (2003) for an excellent survey). In theoretical studies, it is mostly assumed that the arrival and service time distributions are commonly known and well understood. Furthermore, customers are assumed to be fully rational, that is, a customer facing a queue can always accurately and perfectly analyze the given situation and take the optimal action. These restrictions narrow down the real life situations covered by the models.

The experimental studies on Queueing Theory are limited in number. They can be divided into three groups: i) experiments in which the assumption of exogenous arrival times is relaxed; ii) experiments in which the quality of the service is not perfectly known; iii) experiments which question the psychological impact of waiting in a queue.

In the first group fall the contributions by Amnon Rapoport, William Stein, Darryl A. Seale and their colleagues. This group of authors focuses mainly on transient cases by considering queues with non-stationary elements. In particular, they relax the typical assumption of exogenous arrival times and consider systems where arrivals are endogenously determined. This is achieved by letting subjects decide on their arrival times in case they decide to join a (unobservable) queue. Rapoport et al. (2004) study the case in which the serving facility is accessible during a given time period and customers can neither queue before the opening time nor get the service after the closing time. Rapoport et al. (2004) find a strong support for mixed-strategy equilibrium play only at the aggregate but not at the individual level. Seale et al. (2005) extend this study by allowing subjects to arrive before the opening time of the facility. The findings are in complete agreement with those of Rapoport et al. (2004), however the support for mixed-strategy equilibrium play on the aggregate level disappears when congestion is unavoidable and information on the previous round's aggregate behavior is not available. In a follow-up study, Bearden et al. (2005) construct and test a reinforcement learning model based on the experimental results reported in Rapoport et al. (2004) and Seale et al. (2005).

While the model accounts well for the aggregate behavior and generates heterogeneous patterns for the individual decisions similar to those observed in the data, it predicts considerable more switches (changes in the strategy between two consequent rounds) than observed.

Batch queues, where a number of agents in the queue are served at the same time, have also been studied experimentally. Some examples of this type of queue are ferry and bus services, university shuttles, amusement park rides, etc. Stein et al. (2007) conduct a batch queue experiment with endogenously determined arrival times under 4 different conditions: (balking allowed/not allowed)$\times$(private/public information). In the private information condition, subjects are informed about their own performance at the end of each round, whereas in the public information condition on top of their own performance they are also informed about the decisions taken by others (in the form of a cumulative distribution of arrival times). Stein et al. (2007) report that, under each condition, the aggregate but not the individual behavior converges to mixed-strategy equilibrium play. However, such a convergence is faster when balking is not allowed and the information is public. In a follow-up study, Rapoport et al. (2010) extend this experimental study by conducting it in "real time", that is, by making subjects wait for real depending on their decisions, and experience time pressure. Another departure point of this study from Stein et al. (2007) is that the server capacity (the number of people served in a batch) is not always fixed but variable in some treatments. Rapoport et al. (2010) find a strong support for equilibrium play only at the aggregate level, when the server capacity is fixed. When it is variable, the aggregate behavior diverges and results in a pareto superior outcome. Another study that provides evidence of convergence to the mixed-strategy equilibrium at an aggregate level when arrival times are endogenous is Daniel et al. (2009).

The second branch of experimental studies on Queueing Theory considers situations in which the quality of the service is not perfectly known and investigates weather the length of the queue could be perceived as a signal of quality. Giebelhausen et al. (2011) find strong evidence that wait is a positive predictor of quality perception, satisfaction and purchase intentions when quality is uncertain. Koo and Fischbach (2010) arrive to a similar conclusion: one's perception of quality increases with the number of others behind him/her in the queue. Kremer and Debo (2012) study herding behavior in an asymmetric information structure by introducing informed agents. They find support for the hypothesis that long queues are excessively associated with high quality and therefore, purchasing frequency may increase in waiting time.

The third group of experiments associated to the Queueing Theory literature studies the psychological impact of waiting in queues. Leclerc et al. (1995) examine whether agents, in making decision, treat time as they treat money. The results suggest that the way time is treated depends on the context, integration of time losses is preferred over segmentation and agents are risk-averse in the domain of losses. Based on this last result, Kumar and Krishnamurthy (2008) argue that on the one hand it is in the service providers' interest to reduce the uncertainty about service times since agents are risk averse. However, on the other hand, such a reduction in uncertainty increases congestion, which, on its turn, results in a decreased demand due to congestion averseness. Kumar and Krishnamurthy (2008) report that congestion aversion is more dominant than risk aversion. That is, when possible, people tend to avoid congestion, when this is not possible (or if no congestion is anticipated) they avoid to take risk in waiting times. Another psychological impact of waiting is studied by Oxoby et al. (2005), who investigate the effect of the manner in which a waiting situation occurs on the inference of time costs. Their results suggest that, after being exposed to unoccupied waiting time, there is a decrease in inequality aversion and an increase in negative reciprocity.

Our study differs from the above mentioned ones in that it analyzes the behavior of a customer who finds herself in a basic queueing environment and focuses on the aspects she considers when making a queueing decision. In our setting, the quality of the service is perfectly known and therefore, the length of a queue does not serve as a signaling device. The most closely related study to ours is Lu et al. (2013). It is an empirical study that analyzes customers' purchasing behavior in a queueing environment through a field experiment conducted at a deli counter of a grocery store. One of the key findings of this study is that queueing decisions are made mainly based on the length of a queue rather than its speed. This particular result chimes nicely with our experimental finding that there exists a type of customer who considers the length and ignores all the other characteristics of a queue.

III EXPERIMENTAL DESIGN

In our experimental design, subjects were presented with two servers, that were providing the same service. One of the servers was always faster than the other but this premium service was not free of charge, whereas the slower server did not require any entry fee. Subjects were informed about the speed[3] as well as the length[4] of each queue, and asked to choose between joining the faster queue and paying its fee or joining the slower one for free. We use the notation NEF for a queue *without* an entry fee and EF for a queue *with* an entry fee.

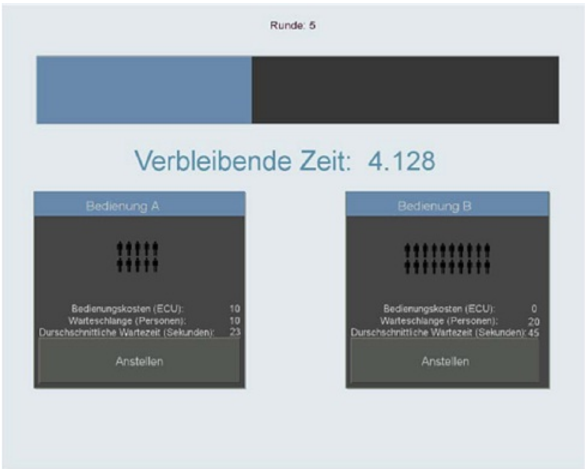
The experiment included 40 different tasks and each of them was repeated three times under different treatment conditions of time allowance: 5 seconds ($5sec$), 10 seconds ($10sec$) and no time

[3] The information on the speed of a server was given in terms of the average waiting time per person.

[4] The information on the length of a queue was not only given numerically but also visually using figures. See the snapshot in Figure1.

limitation ($no - tl$). That is, each subject was asked to make 120 decisions in total. The order the subjects were presented these tasks was randomized. The time restriction for each task was displayed both by a visual and a digital count-down timer. Fig. 1 displays a snapshot from the experiment.

Figure 1
A Snapshot from the Experiment



The score of a subject in a round was calculated by subtracting the entry fee and total waiting cost related to the queue to which s/he decided to join from the initial endowment. Each round's initial endowment was 100 ECUs (Experimental Currency Units) and the waiting cost per minute was 3 ECUs. Thus, the score of a round was $(100 - \text{entry fee} - 3 * \text{wait})$. Experimental Currency Units were converted into euros at the rate of €0.30 and subjects were paid according to their score in a randomly chosen round.

III.I <u>*Procedures*</u>

The experiment was programmed in C++ using a Z-tree interface (Fischbacher, 2007) and conducted in the experimental laboratory of the Max Planck Institute of Economics in Jena (Germany), directed by Prof. Werner Güth.

Participants were undergraduate students from the University of Jena, recruited by the ORSEE (Greiner, 2004) software. Upon entering the laboratory, participants were randomly assigned to visually isolated computer terminals. The instructions were distributed and then read aloud to establish public knowledge.

Overall, we collected 11,640 observations from 97 subjects across five sessions and on average each session lasted about 75 minutes including the time being used up for reading the instructions and paying the participants. Average earnings per subject were €17 (inclusive of a €2.50 show-up fee).

IV DESCRIPTIVE DATA ANALYSIS

This section presents our findings obtained from basic analysis of our data. We initially consider the success rates across treatments. As seen from Table 1, of all the decisions made in $5sec$ treatment 65% is optimal and this rate increases as time limitation is relaxed. A considerable proportion of decisions are suboptimal and having more time to decide improves the performance.

Table 1
Success Rates across Treatments

	$5sec$	$10sec$	$no-tl$
success rate	65%	68%	73 %

In order to gain more insight into the subjects' behavior, we deepen our analysis by examining the success rate when it is optimal to join each queue separately. The first row of Table 2 presents the success rates for the tasks in which the profit of joining the slower with no entry fee queue ($\pi(NEF)$) is higher than the profit of joining the faster with an entry fee queue ($\pi(EF)$). The second row gives the complementary rates concerning the tasks where it is optimal to join the faster queue. The success rate increases when the time limitations is relaxed even when we consider tasks separately depending on the identity of the optimal queue. However, the comparison of the rows of Table 2 shows that in each treatment, the success rate is higher when the optimal decision is to join the queue with no entry fee. This observation suggests that there may be a tendency towards the no entry fee queue.

Table 2
Success Rates across Treatments

	$5sec$	$10sec$	$no-tl$
$\pi(NEF) > \pi(EF)$	70%	72%	77%
$\pi(NEF) < \pi(EF)$	58%	62%	68%

As a final analysis, we further examine success rates by introducing another criteria on top of the identity of the optimal queue. The new criteria is the identity of the shorter queue. Now we have four categories[5] of tasks and the success rate for each category is presented in Table 3. Due to their similarity, we present the rates not for each treatment separately but in aggregate terms.

The best performed task category is the one presented in the first cell of Table 3, where the optimal

Table 3
Success Rates across Task Types

	$\|EF\|>\|NEF\|$	$\|NEF\|>\|EF\|$
$\pi(NEF)>\pi(EF)$	79%	61%
$\pi(NEF)<\pi(EF)$	53%	70%

decision is to join the no entry fee (NEF) queue, which is also shorter. The second best performed task type is the second cell on diagonal. This time, the faster queue with entry fee (EF) is more profitable and shorter. When we consider the off-diagonal cells the success rate drops considerably. What one might conclude from this observation is that there is a tendency towards shorter queue.

In the following section, we introduce the mixture model that we use to test the hypothesis that agents use the information they could extract efficiently and make the optimal queueing decision. The mixture model confirms our conjectures that relaxing time constraint improves the performance, and that there are tendencies towards the no entry fee and shorter queue.

V THE MIXTURE MODEL

Let us assume that there are G different types of decision maker in the population, denoted by the subscript g. Let i indicate the subject and $\tau \in \{5sec, 10sec, no{-}tl\}$ denote the experimental treatment. In each round, subject i is faced with the choice between two queues whose servers provide the same service: a queue with no entry fee (NEF) and a queue with entry fee (EF).

Subject i's decision is based on the following equation:

$$\begin{aligned} d_{igt}^{\tau *} &= \gamma_g^\tau + X_{it}'\beta_g^\tau + \delta_{ig}^\tau + \epsilon_{igt}^\tau \qquad t_i^\tau = 1, \ldots, T_i^\tau \\ \delta_{ig}^\tau &\sim N\left(0, \sigma_g^{\tau 2}\right) \\ \epsilon_{ig}^\tau &\sim N(0,1) \end{aligned} \tag{1}$$

Here, $d_{it}^{\tau *}$ is the latent dependent variable representing subject i's propensity to choose queue NEF in treatment τ; γ_g^τ is a type-specific intercept; X_{it} is a vector of explanatory variable describing the

[5] We exclude the cases where both queues have the same length or are equally profitable.

characteristics of the two queues and β_g^τ is a vector of coefficients on such variables; δ_{ig}^τ is a subject-specific time-invariant intercept, which follows a Normal distribution with mean 0 and variance $\sigma_g^{\tau 2}$; finally, ϵ_{ig}^τ is an idiosyncratic error term distributed Standard Normal.

We do not observe $d_{igt}^{\tau *}$ directly, but a $\{-1, 1\}$ indicator, which is linked to $d_{igt}^{\tau *}$ by the following observational rule:

$$d_{igt}^\tau = \begin{cases} 1 & \text{if } d_{igt}^{\tau *} \geq 0, \\ -1 & \text{else.} \end{cases}$$

This is the well-known random-effects probit model, whose assumptions lead to subject i's likelihood contribution, given that he/she is of type $g \in 1, \ldots, G$, being

$$l_{ig}^\tau = \int_{-\infty}^{\infty} \prod_{t=1}^{40} \Phi \left[d_{igt}^\tau \times \left(\gamma_g^\tau + X_{it}' \beta_g^\tau + \delta_{ig}^\tau \right) \right] \varphi \left(\delta_g^\tau ; 0, \sigma_g^{\tau 2} \right) \mathrm{d}\delta_g^\tau . \tag{2}$$

Here, $\Phi[.]$ is the Standard Normal Cumulative Distribution Function and $\varphi\left(\delta_g^\tau; 0, \sigma_g^{\tau 2}\right)$ is the Normal density function with mean 0 and variance $\sigma_g^{\tau 2}$, evaluated at δ_g^τ.

Types differ in the decision rules adopted, that is, in the variables that they consult when choosing their preferred queue. We want to isolate groups of subjects who adopt similar decision rules when choosing between the NEF and the EF queue. For this purpose, we adopt a finite mixture model approach and assume that there are two types of subjects ($G = 2$): (i)"profit maximizer" who uses all the relevant information and decides rationally, as suggested by the theory; (ii) "naïve" who uses the provided information in a less-than-efficient way. We assume that each subject is either profit maximizer or naïve, and cannot change type within a treatment. As data from each treatment is analyzed separately, the mixture model hypothesis made here *does* allow subjects to change type (decision rules) but only across treatments. Verifying whether subjects change type and understanding the evolution of decision rules across treatments are indeed among the main scopes of our analysis.

The likelihood contribution of subject i in treatment τ is then

$$L_i^\tau = \sum_g \pi_g \times l_{ig}^\tau . \tag{3}$$

Here, π_g, termed "mixing proportion", represents the fraction of the total population who are type $g \in \{\textit{profit-maximizer}, \textit{naïve}\}$, so that $\sum_g \pi_g = 1$. The mixing proportions are estimated along with the other parameters of the model by maximizing the full sample log-likelihood,

$$LogL^\tau = \sum_{i=1}^{n} \ln[L_i^\tau]. \tag{4}$$

The mixture model is estimated using the method of Maximum Simulated Likelihood for each treatment $\tau \in \{5sec, 10sec, no-tl\}$ separately. In each component g of the mixture, integration over δ_g^τ is performed by simulation using 100 draws for each contestant based on Halton sequences (Train, 2003).

V.I *Estimation Results*

The results from the maximization of Eq. (4) are reported in Table 4. For each treatment, there are two columns displaying the parameter estimates of the mixture models for each type. The type "profit maximizer" is characterized by using the difference in profits of the two queues (Δ(profit)=$\pi(NEF) - \pi(EF)$) as explanatory variable. On the other hand, sub-optimal behavior of a "naïve" decision maker is modeled by means of the average waiting times and the lengths of the two queues, and we also control for the entry fee level of the EF queue.

Table 4
Maximum likelihood estimates of the mixture model's parameters

τ	$5sec$		$10sec$		$no-tl$	
g	*prof.max.*	*naïve*	*prof.max.*	*naïve*	*prof.max.*	*naïve*
regressors						
Δ(profit)	0.0316***		0.0432***		0.0517***	
	(0.0032)		(0.0033)		(0.0031)	
average waiting time (NEF)		-0.0097*		-0.0082		-0.0014
		(0.0052)		(0.0054)		(0.0069)
average waiting time (EF)		-0.0022		0.0077		0.0095
		(0.0046)		(0.0047)		(0.0064)
length (NEF)		-0.0368***		-0.0278***		-0.0259***
		(0.0051)		(0.0051)		(0.0062)
length (EF)		0.0360***		0.0344***		0.0481***
		(0.0034)		(0.0036)		(0.0048)
entry fee EF server		0.0178***		0.0164***		0.0220***
		(0.0033)		(0.0034)		(0.0043)
γ_g	0.1651***	0.1418	0.0987*	-0.2924	0.1384***	-1.3057**
	(0.0467)	(0.4134)	(0.0559)	(0.4260)	(0.0424)	(0.5207)
σ_g^τ	0.1448***	0.3655***	0.2020***	0.3550***	0.2185***	0.3012***
	(0.0574)	(0.0520)	(0.0669)	(0.0513)	(0.0428)	(0.0582)
π_g	0.4322***	0.5678***	0.5036***	0.4964***	0.6478***	0.3522***
	(0.0802)	(0.0802)	(0.0764)	(0.0764)	(0.0667)	(0.0667)
LogLikelihood	-2371.93		-2335.28		-2182.46	
number of observations	3860		3873		3880	
number of subjects	97		97		97	

The results show that the coefficient of the variable of interest for the profit maximizer type, i.e. Δ(profit), is of the expected sign and statistically significant in each treatment[6]. It amounts to say that this type uses the difference in profits as the decision criteria and joins the queue that provides

[6] ***, ** and * denote p-values < 0.01, < 0.05 and < 0.10, respectively.

higher profit. The naïve type, on the other hand, seems to ignore the information on average waiting times and considers the length of the two queues and the entry fee as the decision criteria. In $5sec$ treatment, however, there is a small exception that in addition to the previously named variables, the average waiting time of NEF queue is also significant, but only slightly so. Furthermore, all the significant variables are of the expected sign.

The mixing proportions π_g indicate that as time constraints are relaxed decision performance improves and the fraction of profit maximizers rises. In $5sec$ treatment, there are mildly more naïve types than profit maximizers and when the time allowance is doubled (in $10sec$ treatment), the fraction of rational type increases and constitutes half of the population. Finally, when no time limitation is imposed the fraction of profit maximizers jumps to 65%.

The mixture model assigns subjects to a type probabilistically and therefore, it does not specify which subject is assigned to which type. Furthermore, the model is run separately for each treatment. This means that the model does not help us in determining whether a subject assigned to a specific type in one treatment is also assigned to the same type in other treatments. The next subsection investigates this issue and examines how types evolve across treatments.

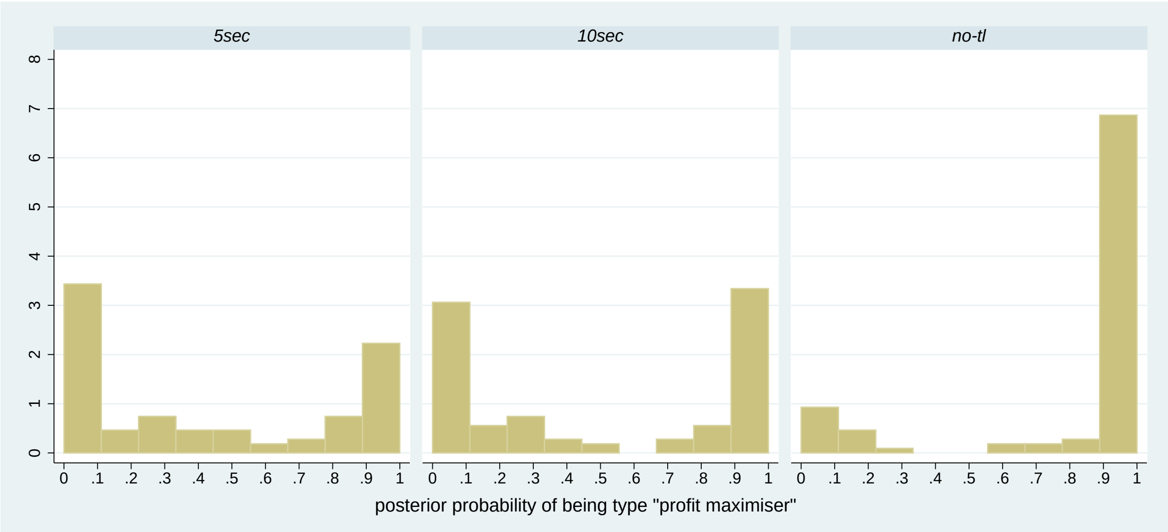
V.II *The Evolution of Decision Rules*

We start investigating the evolution of types by calculating the posterior probability of being a specific type for each subject in each treatment $\tau \in \{5sec, 10sec, no-tl\}$. We derive these probabilities using the Bayes' rule and the estimation results from our mixture model (see Table 4). Subject i's posterior probability of being type $g \in \{\textit{profit maximizer}, \textit{naïve}\}$ is the given by

$$\begin{aligned}\mathrm{pp}_{i,g}^{\tau}(\mathrm{obs}_i^{\tau}) = \Pr\left[i = \text{type } g \mid \mathrm{obs}_i^{\tau}\right] &= \frac{\Pr\left[i = \text{type } g\right] \times \Pr\left[\mathrm{obs}_i^{\tau} \mid i = \text{type } g\right]}{\Pr\left[\mathrm{obs}_i^{\tau}\right]} \\ &= \frac{\pi_g^{\tau} \times l_{ig}^{\tau}}{L_i^{\tau}},\end{aligned} \tag{5}$$

where obs_i^{τ} represents the observations collected from subject i in treatment τ. In practice, π_g^{τ}, l_{ig}^{τ} and L_i^{τ} are replaced by their estimated counterparts, obtained by maximizing Eq. (4) from treatment τ data, $\forall g$. Obviously, subject i's posterior probability of being naïve type is obtained by $\mathrm{pp}_{i,\text{naïve}}^{\tau} = 1 - \mathrm{pp}_{i,\text{prof.max.}}^{\tau}$, $\forall \tau$.

Figure 2
Histograms of Posterior Probabilities of Being Profit Maximizer Type



The histograms of the posterior probabilities of being profit maximizer type are displayed in Fig. 2. The resulting posterior probabilities are consistent with the mixing proportions estimated by the mixture model. In $5sec$ treatment, naïve type is mildly preponderant. When the time allowance is 10 seconds, the posterior probabilities of being one of the two types are almost equal. With no time limitation, naïve type is decidedly recessive. Most of the subjects are concentrated at the extremes of the distributions. This finding testifies that our mixtures are rather powerful at segregating subjects, except for a small number of them for whom there is some uncertainty. We assign subjects to types according to the maximum posterior probability. Figure 3 shows the cumulative percentage of subjects assigned to a type with maximum posterior probability less than the probability indicated on the horizontal axis, for each treatment. The figure confirms that the power of our mixture model at segregating subjects is quite impressive: 60%, 70% and 85% of them are assigned to type with posterior probability larger than 0.90 in treatment $5sec$, $10sec$ and $no-tl$, respectively. Overall, the assignment to a type is remarkably good in the treatment with no time limitation, only marginally less in the other two cases.

Having assigned each subject to a type in each treatment, now we are ready to consider subjects' profiles throughout the experiment. We consider two possible types for each of the three treatments, therefore, there are eight possible profiles. Table 5 reports frequencies and proportions for all these eight profiles.

The most popular profile is the one where subjects are assigned to naïve type in $5sec$ and $10sec$

Figure 3
Cumulative proportion of subjects assigned to a type with maximum posterior probability $<$ posterior probability indicated on the horizontal axis, by treatment.

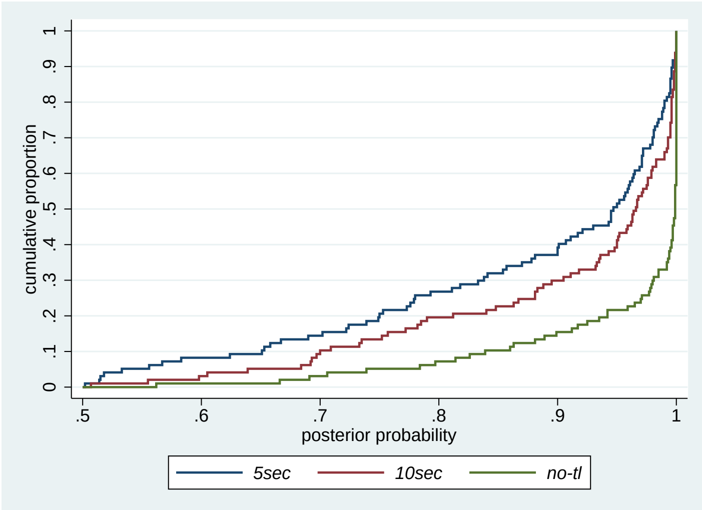


Table 5
Profile Frequencies and Proportions

	τ			
$5sec$	$10sec$	$no-tl$	Frequency	Proportion
naïve	naïve	naïve	11	11%
naïve	naïve	prof.max.	29	30%
naïve	prof.max.	naïve	1	1%
naïve	prof.max.	prof.max.	17	18%
prof.max.	naïve	naïve	3	3%
prof.max.	naïve	prof.max.	7	7%
prof.max.	prof.max.	naïve	1	1%
prof.max.	prof.max.	prof.max.	28	29%

treatments, but to profit maximizer type in $no-tl$ treatment (the second row of the table). When there were time limitations, thirty percent of our sample failed to make the optimal decision but managed to do so when no limitation was imposed. A possible explanation for this behavior could be that the given time allowances were not long enough to think throughly and choose the more profitable queue, but when limitations were removed subjects could take the time necessary to make the optimal decision. Another possibility is that it was not the tightness of time allowances (especially for $10sec$ treatment) that caused these subjects to perform badly in $5sec$ and $10sec$ treatments, but the presence of a limitation. Under time pressure, they might have panicked and used their time inefficiently, and therefore failed to make the optimal decision. We postpone investigating this issue to the next subsection where we consider the decision times.

Table 5 reveals that the second most popular profile, that is adopt by 29% of our subject pool, is being profit maximizer type in each treatment. These subjects showed consistent behavior throughout the experiment and different time limitations did not affect their being rational. The other profile that is consistent throughout the experiment is the one which is assigned as naïve type in each treatment and is adopted by 11% of our population. In total 40% of the subjects behaved consistently and did not change type. This amounts to say that time pressure did not have any effect on these subjects.

When we ignore the tightest time limitation, $5sec$ treatment, we can define three categories of profile pattern depending subjects' assigned types in treatments $10sec$ and $no-tl$. We do this in order to study the effects of the presence of a time limitation. The first profile category is *consistent pattern* where subjects are assigned to either profit maximizer or naïve type in both treatments. Time pressure does not have any impact for those who follow this pattern, who constitute 61% of our population.The second category is *improving pattern* where subjects change their type from naïve to profit maximizer when time limitation is removed. To this category belongs 37% of our subjects. It is striking that when time pressure is cut out more than one-third of the population's decision performance is improved. As mentioned earlier, we will discuss whether this improvement is due to the fact that 10 seconds were not enough to make the optimal decision in the following subsection. The last category is *worsening pattern* where subjects switch from profit maximizer to naïve type when passing from $10sec$ to $no-tl$ treatment. This type of unexpected behavior is quite rare in our population; in fact, it is adopted only by 2%.

Finally, we would like to note that if instead of $10sec$ treatment we compare $5sec$ with $no-tl$ treatment, the fractions of the population who shows consistent, improving and worsening pattern are 48%, 48% and 4%, respectively. This time we see that while almost half of the subjects does not change their types between the two treatments, a great deal of them improve their decision performance. Furthermore, the unexpected behavior is again rarely observed.

V.III *Decision Times*

In this subsection, we analyze the decision time of the subjects in our sample under the three treatments. The descriptive statistics, given in Table 6, reveals that decision times are different across treatment.[7] Paired t-tests confirm that the differences are statistically significant.[8] Furthermore, we

[7]In treatment $5sec$ and $10sec$, 20 and 7 decision times are missing, respectively, because subjects did not make a decision within the given time limit. In these few cases, for the tables and the tests reported in this section, we have replaced the missing decision times with the upper time limit. Even if we neglect these missing observations, the reported tests' results do not alter.

[8]The t-statistic takes values -11.604, -12.621 and -11.597, for comparisons $5sec$ vs. $10sec$, $5sec$ vs. $no-tl$ and $10sec$ vs. $no-tl$, respectively. These tests' p-values under the null hypothesis that the means from the

see that subjects use more time to make a decision as time limitations are relaxed. This implies that a possible explanation why decision performance is improved across treatments could be that subjects take more time to analyze the parameters of the environment and hence make a better decision.

Table 6
Summary Statistics of Decision Times.

τ	$5sec$	$10sec$	$no-tl$
Mean	2.413	3.338	8.180
Std. Dev.	0.773	1.429	4.952
Min	0.595	0.670	1.392
Max	3.901	6.244	22.564
Number of subjects	97	97	97

An interesting observation revealed by Table 6 is that the average decision time in $no{-}tl$ treatment, under which subjects performed best, is around 8.2 seconds. However, the average decision time drops to 3.3 seconds in $10sec$ treatment. This implies that the presence of time limitation puts a great deal of subjects under pressure and hinders them from using the given time efficiently, and this, in turn, harms the decision performance.

We deepen our analysis by examining decision times by types. The descriptive statistics, given in Table 7, verify our previous observation that subjects take more time to decide when the limitations are relaxed holds regardless of types. Furthermore, profit maximizers spend more time than naïve types no matter what the time limitation is. In fact, a two-sample t-test with unequal variances confirms that regardless of the treatment we can reject the hypothesis that both types spend the same amount of time, on average, to make a decision. The test statistics and p-values are reported in the third and fourth rows of the table.

In the previous subsection we have seen that one-third of the population exhibits an improving profile pattern and change from naïve to profit maximizer type when $10sec$ and $no-tl$ treatments are compared. These subjects spend, on average, 2.7 seconds in $10sec$ treatment and 8.9 seconds in $no{-}tl$ treatment. This observation shows that even though the subjects had enough time to make better decisions in $10sec$, they failed to use it efficiently. Time pressure harmed the decision performance of these subjects.

A final remark from Table 7 is that decision time cannot be the only explanation for naïve behavior. Subjects classified as naïve type in $no-tl$ treatment use more time than profit maximizer type of

two treatments are equal against the alternative that they are not are always < 0.001.

Table 7
Summary Statistics of Decision Times by Types

τ	$5sec$		$10sec$		$no-tl$	
g	*prof.max.*	*naïve*	*prof.max.*	*naïve*	*prof.max.*	*naïve*
Mean	2.746	2.189	4.036	2.683	8.864	4.716
Std. Dev.	0.661	0.767	1.396	1.125	5.050	2.365
t-statistic	-3.817		-5.235		-5.089	
p-value	0.000		0.000		0.000	
Number of subjects	39	58	47	50	81	16

$10sec$ treatment, but perform relatively worse. They spend some time to analyze the parameters of the queueing system, but this does not improve their decisions either because their cognitive capacity fall short or they intentionally follow a rule of thumb.

VI DISCUSSION

This study contributes to the understanding of customer queueing behavior by experimentally examining the situation in which subjects need to make decision between two queues under different treatment conditions of time allowance. Our econometric analysis suggests that a considerable proportion of the population behaves rationally and base their decisions mainly on the profits they would gain joining each queue. The size of this proportion increases as time limitations are relaxed. The rest of the population, however, does not use the provided information in a normative way. They pay no attention to server speed, which is given as the average waiting time per person in the queue. They seem to use rule of thumbs that exhibit a tendency towards the shorter queue.

An important contribution of this paper is the finding that a great deal of subjects exhibit consistent behavior throughout the experiment. The change of time allowance across treatments does not alter the types they are assigned to. Hence, time pressure does not have any effect on their decision performance. However, on the other hand, there are also those who improve their performance when no time limitation is imposed. The underlying reason of this improvement is not because the limitation in time is too tight to make a good decision but because the presence of such a limitation puts subjects under pressure and harms their performance. That is to say, no trifling portion of subjects systematically follows the shorter queue when there is time pressure, but acts almost rationally when the pressure vanishes.

Finally, our analysis reveals that the average time subjects take to make decision increases as we relax time limitation. An investigation of decision times shows that profit maximizers take significantly more time than their naïve peers under any time limitation condition. The importance of this result is twofold. Firstly, it testifies that the decision criterion used by profit maximizers is cognitively more demanding than the decision process of naïve subjects. Secondarily, it indirectly emphasizes our mixture model's success in segregating subjects into types and justifies our reasons for choosing this analytical approach.

Although the queueing environment studied in this paper is very simplistic, the findings potentially pave the way for further research by providing useful inputs. One possibility is to deepen the investigation of non standard behavior. In this study, we classified any behavior that is not rational as naïve. However, using a large enough number of subjects, the non-rational behavior could be further categorized by isolating different decision rules used by subjects. The characterization of different types would bring about the consideration of a mechanism design that screens and discriminates customers based on their types, and the welfare effect of this discrimination.

There are some limitations to this experiment. We provided subjects with explicit information on the server speed in order to see how they react to it. In a real life situation, this piece of information is not explicitly available but could be extracted by observing the queue for a while. However the fact that subjects pay no attention to this explicitly given information suggests that they would not even try to extract it when facing a similar situation in everyday life.

A more serious limitation to our experiment is that the subjects did not experience the irritation and annoyance of waiting in a queue. Each round finished and a new one began immediately after one made his/her decision to which queue to join, without waiting for real. Designing an experiment that involves real waiting is problematic because the cost of waiting is subjective and not observable. That is, each subject's annoyance due to waiting may be different, and moreover, measuring or deducing it may not even be possible. Finally, due to the accumulation effect of this cost, a robust analysis would require a huge number subjects since no more than a few observation could be obtained from a subject.

Acknowledgments This work was partially supported by PRIN 20103S5RN3 and MOE2013-T2-1-158. This work was carried out while Marco Scarsini was a member of GNAMPA-INdAM.

REFERENCES

Bearden, J. N., A. Rapoport and D. A. Seale. “Entrée times in queues with endogenous arrivals: Dynamics of play on the individual and aggregate levels”, *Experimental Business Research* 2, 2005, 201–221.

Conte, A. and J.D. Hey. “Assessing Multiple Prior Models of Behaviour under Ambiguity”, *Journal of Risk and Uncertainty* 46(2), 2013, 113–132.

Choi, S., R. Fisman, D. Gale and S. Kariv. “Consistency and Heterogeneity of individual behavior under uncertainty”, *American Economic Review* 97 (5), 2007, 1921–1938.

Daniel, T.E., E. Gisches and A. Rapoport. “Departure time in Y-shaped traffic networks with multiple bottlenecks”, *American Economic Review* 99 (5), 2009, 2149–2176.

Fischbacher, U. “Zurich toolbox for readymade economic experiments”, *Experimental Economics* 10(2), 2007, 171–178.

Giebelhausen, M.D., S.G. Robinson and J.J. Cronin. “Worth waiting for: increasing satisfaction by making consumers wait”, *Journal of the Academy of Marketing Science* 39(6), 2011, 889–905.

Greiner, B. “An online recruitment system for economic experiments”, in: Kremer, K., Macho, V. (Eds.), *Forschung und wissenschaftliches Rechnen* 2003. Ges. für Wiss. Datenverarbeitung, Göttingen, 2004, 79–93.

Hassin, R. and M. Haviv. *To queue or not to queue: Equilibrium behavior in queueing systems*, Kluwer Academic Publishers, 2003.

Koo, M. and A. Fischbach. “A silver lining of standing in line: Queuing increases value of products”, *Journal of Marketing Research*, 47 (4), 2010, 713–724.

Kremer, M. and L. Debo. “Herding in a queue: A laboratory experiment”, Working paper, Smeal College of Business, Penn State University, University Park, PA, 2012.

Kumar, P. and P. Krishnamurthy. “The impact of service-time uncertainty and anticipated congestion on customers’ waiting-time decisions”, *Journal of Service Research*, 10(3), 2008, 282–292.

Leclerc, F., B.H. Schmitt and L. Dubè. “Waiting time and decision making: Is time like money?”, *Journal of Consumer Research* 22(1), 1995, 110–119.

Lu, Y., A. Musalem, M. Olivare and A. Schilkrut. “Measuring the effect of queues on customer purchases”, *Management Science* 59(8), 2013, 1743–1763.

Naor, P. "The regulation of queue size by levying tolls", *Econometrica* 37, 1969, 15–24.

Oxoby, R.J. and D. Bischak. "Passing the time: Other-regarding behavior and the sunk cost of time", Discussion paper, University of Calgary, Department of Economics, 2005.

Rapoport, A., W. E. Stein, V. Mak, R. Zwick and D. A. Seale. "Endogenous arrivals in batch queues with constant or variable capacity", *Transportation Research Part B* 44(10), 2010, 1166–1185.

Rapoport, A., W.E. Stein, J.E. Parco and D.A. Seale. "Equilibrium play in single-server queues with endogenously determined arrival times", *Journal of Economic Behavior and Organization* 55 (1), 2004, 67–91.

Seale, D.A., J.E. Parco, W.E. Stein and A. Rapoport. "Joining a queue or staying out: effects of information structure and service time on arrival and staying out decisions", *Experimental Economics* 8 (2), 2005, 117–144.

Stein, W.E., A. Rapoport, D.A. Seale, H. Zhang and R. Zwick. "Batch queues with choice of arrivals: equilibrium analysis and experimental study", *Games and Economic Behavior* 59 (2), 2007, 345–363.

Train, K. *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, 2003.