THE IMPACTS OF CUSTOMERS' DELAY-RISK SENSITIVITIES ON A QUEUE WITH BALKING

Pengfei Guo and Paul Zipkin

Probability in the Engineering and Informational Sciences / Volume 23 / Issue 03 / July 2009, pp 409 - 432
DOI: 10.1017/S0269964809000254, Published online: 30 April 2009

Probability in the Engineering and Informational Sciences, **23**, 2009, 409–432. Printed in the U.S.A.
doi:10.1017/S0269964809000254

THE IMPACTS OF CUSTOMERS' DELAY-RISK SENSITIVITIES ON A QUEUE WITH BALKING

PENGFEI GUO
Department of Logistics and Maritime Studies
Hong Kong Polytechnic University
Hung Hom, Hong Kong
Email: lgtpguo@polyu.edu.hk

PAUL ZIPKIN
The Fuqua School of Business
Duke University
Durham, NC 27708, USA
Email: paul.zipkin@duke.edu

Congestion and its uncertainty are big factors affecting customers' decision to join a queue or balk. In a queueing system, congestion itself is resulted from the aggregate joining behavior of other customers. Therefore, the property of the whole group of arriving customers affects the equilibrium behavior of the queue. In this paper, we assume each individual customer has a utility function which includes a basic cost function, common to all customers, and a customer-specific weight measuring sensitivity to delay. We investigate the impacts on the average customer utility and the throughput of the queueing system of different cost functions and weight distributions. Specifically, we compare systems where these parameters are related by various stochastic orders, under different information scenarios. We also explore the relationship between customer characteristics and the value of information.

1. INTRODUCTION

Delay, in most cases, is an unhappy experience for customers. But the degree of unhappiness may differ with customers' types. Facing the same delay, patient customers may join the queue while impatient customers may balk. Thus, customers' sensitivity on delay is an important factor affecting their decision. Worse yet, to join a queue is

 0269-9648/09 $25.00

usually risky, since customers don't know exactly how long they will have to wait. Therefore, customers' attitude towards delay uncertainty also affects their decision to join or to balk. For example, in a transportation system, some travelers need travel for a conference or a meeting and they would prefer a safe route even though it may take a longer time than a short cut; for other travelers with a large time flexibility, they may prefer a risky route with a shorter average delay.

Nowadays there are increasing possibilities, enabled by new technologies, to provide useful delay information to customers, to enhance the values of the services they receive. Customers can thus make better-informed decisions upon arrival. For example, in a transportation network, drivers can learn congestion information through radio and Internet; in a call center, callers are sometimes informed about expected waiting time.

Guo and Zipkin [10] (hereafter G-Z) describe numerous other examples on information service in queues. They study a single-server queue with three levels of delay information, *none*, *partial* (the system occupancy) and *full* (the exact waiting time). Each customer decides whether to stay or leave, based on this information and his own sensitivity to delays. They derive equilibrium behavior for the queues under different information scenarios. They show that information's impact on the whole system is not always positive: It can reduce the server's throughput and even hurt customers' average utility. The most important factor determining these qualitative effects is the shape of the distribution of customers' delay-sensitivity weights. Specifically, they show that more information always improve the server's throughput if customers are heterogeneous enough on the dimension of delay sensitivity.

However, it still remains unclear about the relationship between the equilibrium behavior of the queueing system and customers' characteristics. Neither are we clear about the value of information and customers' characteristics. Here, customers characteristics include the distribution of delay-sensitivity parameter across the whole group and an overall customers' risk-aversion measure. In the former part, we aim to answer questions such as: Is the system more congested with less patient customers? or more-concentrated customers on the dimension of delay sensitivity? or less risk-averse customers? In the latter part, we aim to answer questions such as: Is the value of information greater for less patient customers? Is it larger for more risk-averse customers? Such kind of analysis is called *sensitivity analysis*.

Intuitively, information should be more valuable to a more risk-averse decision maker. This is true in a static sense, but it may not reflect dynamic behavior. See Hilton [15], Freixas and Kihlström [8], Willinger [29] and Nadiminti, et al. [18]. For example, Freixas and Kihlström [8] show that the demand for information may decrease with the level of risk aversion. *Ex post*, information reduces risk, however, *ex ante*, information gathering itself is a risky activity that risk-averse decision makers are less willing to bear. These works consider only a single decision maker facing an exogenous risk. In our system, there is a group of decision makers (customers). Each one's joining decision affects the delays for later-arriving customers, so the system exhibits *negative externality*. The delay risk here is endogenous. The relationship between information and risk aversion is thus even more complex.

The closest work on congestion, customers' risk attitude and the value of information is found in transportation literature. de Palma, et al. [5] consider information service in a traffic system where drivers can choose between a safe route and a risky route. They consider four information scenarios: *No information*, *Free information* which is publicly available at no cost, *Costly information* which is publicly available for a fee, and *Private information* which is available free to a single individual. They found that *Private information* is individually more valuable than either *Free* and *Costly information*. Interestingly, they found that *Free* or *Costly information* can decrease the expected utility of drivers who are sufficiently risk-averse. They assume that drivers are heterogeneous on risk-aversion parameter without consider drivers' sensitivity towards delay while we assume customers are heterogenous on delay-sensitivity parameter and there exists a common risk aversion parameter for all customers.

Different from the conclusion in [5] that information hurts those most risk-averse drivers, we show that if information hurts the server (reduces the throughput), information benefits each *individual* customer. This conclusion strengthens the one in G-Z where they show that if information hurts the server, it benefits the *whole group* of customers. This disparity may come from the different settings in the two works: In our system, customers can choose to balk if the system is too congested; while in [5], drivers have to travel anyhow; information in [5] is an *exogenous* variable indicating travel conditions of the risky route while information in our work is an *endogenous* variable indicating congestion itself.

We show that when customers are more patient towards delay, the system becomes more congested in equilibrium and the average utility for customers is larger under no information and is so only with some special cost functions under partial or full information. We also show that when customers are more concentrated on delay-sensitivity dimension or are more risk-averse, the throughput of the system in equilibrium need not increase. Finally, we show that the value of information is tremendously affected by customers' cost function itself, instead of the shape of cost function. There is no monotone relationship between the value of information and the degree of customers' risk aversion.

The literature on customers influenced by delay information begins with Naor [19], who studies a system like ours with partial information, but with identical customers and linear waiting cost. Subsequent research includes Gavish and Schweitzer [9], Edelson and Hildebrandt [7], and Schroeter [23]. See Stidham [26] and Hassin and Haviv [13] for surveys. More recent work includes Whitt [27], Armony and Maglaras [1,2], Armony, et al. [3], Shimkin and Mandelbaum [25], Guo and Zipkin [11] and Hassin [12].

There has been much research on comparison of systems with different input streams. Ross [22] conjectures that a more regular arrival process leads to better performance. Some counter examples are provided by Heyman [14]. See Rolski [21], Shaked and Shanthikumar [24] and Müller and Stoyan [17] for surveys. Chao and Dai [6] and Dai and Chao [4] show that the conjecture holds for a single-server loss system in random environment. Recently, Whitt [28] analyzes the sensitivity of performance to changes in model parameters in an $M/M/s$ queue with customer abandonment. He shows that performance can be quite sensitive to changes in the arrival and service rates, but relatively insensitive to the abandonment rate.

In our context, the arrival process is formed by a stream of customers who make their own individual decisions, based on their utilities and the available information about the system's state. The overall impact of customers' characteristics is thus far from clear. Understanding this matter is important for the design of systems in different markets with different types of customers.

The remainder of the paper is organized as follows: Section 2 reviews the basic formulation and stochastic orders, and briefly summarizes the three models in G-Z. Section 3–4 develop the sensitivity analysis of system performance with respect to the change of weight distribution and the change of cost function, respectively. Section 5 studies the relationship between the value of information and customers' characteristics. Section 6 concludes. All proofs are included in the Appendix.

2. FORMULATION AND PRELIMINARIES

2.1. Notation and Utility

As in G-Z, we assume a single-server queue with exponential service times. Potential customers arrive in a Poisson process. We suppose that a customer's utility equals a reward for receiving service minus a waiting cost. This waiting cost depends on a customer-specific delay-sensitivity parameter and the expectation of a (common) function of the waiting time. Denote

- $\lambda =$ arrival rate of potential customers
- $\mu =$ service rate
- $W =$ waiting time in queue
- $\theta =$ customer sensitivity towards delay, $\theta \in [0, 1]$
- $H =$ cumulative distribution function of θ, assumed continuous on $[0, 1]$, with density h
- $c(w) =$ basic cost to wait time w, a positive, increasing, unbounded, continuous function
- $r =$ reward to the customer for receiving service, $r > 0$
- $u(w, \theta) =$ utility for a customer with weight θ to wait time w
- $u(\theta|I) =$ expected utility for a customer with θ, given delay information I
- $u =$ average utility for a whole group of customers

A customer with delay-sensitivity θ has a utility function for receiving service but waiting time w, which is expressed as

$$u(w, \theta) = r - \theta c(w).$$

The service reward is the same for all customers (this assumption is convenient but not essential). Customers differ in the importance of time. This difference is expressed by the customer delay-sensitivity parameter θ. Each customer's weight is independent of all other events; it follows the common distribution function $H(\theta)$. The shape of $H(\cdot)$ characterizes the heterogeneity of customers in delay sensitivity. A very concentrated distribution indicates nearly identical customers and a dispersive distribution means that customers are different: Some are patient while others are impatient.

The waiting time w can be a random variable. The customer assesses the distribution of this random variable, based on the available information. Let I denote the information variable. The expected waiting cost $c_I = E[c(W)|I]$ is a function of that information. The expected utility for the customer to remain in the system is thus $u(\theta|I) = r - \theta c_I$. The customer remains in the system if $u(\theta|I)$ is non-negative and otherwise balks. We assume that there is no reneging.

We normalize $r = 1$ and assume $c(0) = 1$. Under this assumption, a customer seeing an empty system will always join. Define $\theta_I = 1/c_I$. The effective arrival rate given delay I is hence $\lambda H(\theta_I)$. Define the function

$$J(\theta) = \frac{\int_0^\theta H(\phi)d\phi}{\theta}.$$

G-Z show that the average utility for all customers in the system (including those balking customers), u, equals $E_I[J(\theta_I)]$. Also they show that $J(1/x)$ is decreasing and convex in x.

2.2. Brief Summary of G-Z's Three Information Models

G-Z consider three levels of delay information. With *no information*, customers still estimate their waiting times, but these estimates are based only on long-term (equilibrium) experience, not real-time information. The occupancy provides *partial information*; the remaining uncertainty comprises the actual service times of the waiting customers. The exact waiting time gives the customer *full information*. We briefly summarize the models and solutions and then give our general stochastic comparison results.

Let '-' denote no information. Also, denote N as the system occupancy and V as the workload at the moment of arrival.

Under no information, the resulting system is an $M/M/1$ queue with the equilibrium arrival rate λ_-, which solves

$$\lambda_- = \lambda H\left(\frac{1}{E[c(W|-)]}\right). \tag{1}$$

Here, $E[c(W|-)]$ indicates the expected cost given λ_-. Hence the equilibrium system is still an $M/M/1$ queue with effective arrival rate λ_-.

Under partial information, N can be modeled as a birth-death process. The birth rate in state n is $\lambda_n = \lambda H(\theta_n)$, where $\theta_n = 1/c_n = 1/E[c(W)|N = n]$, and the death rate is μ. The equilibrium distribution of N can be expressed as

$$p_n = \left(\prod_{m=0}^{n-1} \lambda_m/\mu\right) p_0 = \Theta_n(\lambda/\mu)^n p_0,$$

where

$$\Theta_n = \prod_{m=0}^{n-1} H(\theta_m),\ n > 0.$$

Let

$$\Theta = \sum_{n>0} \Theta_n(\lambda/\mu)^n.$$

Then,

$$p_0 = \frac{1}{1+\Theta}.$$

Define the cumulative effective arrival rate $\Lambda(v) = \int_0^v \lambda(v)dv$. Under full information, the pdf for the equilibrium workload V, $f(v), v > 0$, solves the integral equation

$$f(v) = \lambda p_0 e^{-\mu v} + \int_0^v \lambda H[1/c(w)]e^{-\mu(v-w)} f(w)dw \tag{2}$$

with the normalization condition

$$p_0 + \int_0^\infty f(v)dv = 1. \tag{3}$$

The solution is

$$f(v) = \lambda p_0 e^{\Lambda(v)-\mu v}, \tag{4}$$

where

$$p_0 = \frac{1}{1+\lambda\int_0^\infty e^{\Lambda(v)-\mu v}dv}. \tag{5}$$

Some auxiliary comparison results with different effective arrival rates are included in the Appendix A.

3. IMPACT OF CUSTOMERS' DELAY SENSITIVITY

Consider two systems, identical except for H and c. Use the superscript $k = 1, 2$ to index the systems. In this section, we assume $c^1 = c^2 = c$ and consider specific conditions on the H^k. We consider two situations. In one situation, customers in one group are stochastically more patient than those in another group. In the other situation, customers in one group are more concentrated on delay-sensitivity dimension than the other group. We are interested in the performance measures in equilibrium with different groups of customers.

3.1. Impact of Average Delay Sensitivity

We first give the definition of two stochastic orders indicating the concept of *stochastically smaller*. If $H^1(x) \geq H^2(x)$, $x \in [0,1]$, θ^1 is *stochastically smaller* than θ^2 (denoted $\theta^1 \preceq_{st} \theta^2$), or H^2 is said to dominate H^1 according to *first-order stochastic dominance*. This means that system 1's customers are stochastically more patient than system 2's. A stronger condition is that the ratio h^1/h^2 is monotonically decreasing, a condition called *monotone likelihood ratio*. Then θ^1 is said to to be smaller than θ^2 in the *likelihood-ratio order* (denoted $\theta^1 \preceq_{lr} \theta^2$). Detail discussions on these concepts can be found in Shaked and Shanthikumar [24] and Müller and Stoyan [17].

We have the following conclusion about the system occupancy and the workload.

Proposition 1: *If $\theta^1 \preceq_{st} \theta^2$, then $N^1 \succeq_{lr} N^2$ under no or partial information, and $V^1 \succeq_{lr} V^2$ under full information.*

The above proposition implies that when system 1's customers are stochastically more patient than system 2's, system 1 becomes more congested than system 2.

Next, we compare the average utilities for the two systems.

Proposition 2: *If $\theta^1 \preceq_{st} \theta^2$, then $u^1 \geq u^2$ under no information.*

Therefore, for no information, when customers become more patient, their average utility becomes larger in equilibrium. For partial and full information, first consider the special case of the power distribution, $H(\theta) = \theta^\alpha$ for constant $\alpha > 0$. Note that, for two distributions of this form, $\alpha^1 \leq \alpha^2$ implies $H^1 \geq H^2$.

Proposition 3: *If each H^k is a power distributions with $\alpha^1 \leq \alpha^2$, then $u^1 \geq u^2$ under partial or full information.*

Therefore, for a special case with power distribution of delay-sensitivity parameter, when customers become more patient, their average utility is larger, under partial and full information. Beyond this special case, the relation between u^1 and u^2 is not clear. A stronger condition is needed to conclude that customers on average are better off in one system than the other.

Condition $\theta^1 \preceq_{lr} \theta^2$ means that $h^1(x)/h^2(x)$ is decreasing in x. This condition is stronger than $\theta^1 \preceq_{st} \theta^2$. We now consider the relation between u^1 and u^2 under partial and full information.

PROPOSITION 4: *If $\theta^1 \preceq_{lr} \theta^2$, then*

$$\frac{u^1}{u^2} \geq \frac{J^1(1)p_0^1}{J^2(1)p_0^2},$$

under partial or full information.

Since $p_0^1 \leq p_0^2$ and $J^1(1) \geq J^2(1)$, it is not necessarily true that $[J^1(1)p_0^1]/[J^2(1)p_0^2] \geq 1$. But at least we obtain a lower bound on u^1/u^2.

3.2. Impact of Dispersion of Delay Sensitivity

We first provide some concepts on stochastic orders indicating the dispersion of the distribution function H. If $\int_v^1 \bar{H}^1(x)dx \leq \int_v^1 \bar{H}^2(x)dx$, for all $v \in [0,1]$, θ^1 is smaller than θ^2 in the *increasing convex order* (denoted $\theta^1 \preceq_{icx} \theta^2$). If $E[\theta^1] = E[\theta^2]$ and $\int_v^1 \bar{H}^1(x)dx \leq \int_v^1 \bar{H}^2(x)dx$, for all $v \in [0,1]$, θ^1 is smaller than θ^2 in the *convex order* (denoted $\theta^1 \preceq_{cx} \theta^2$), or H^1 is said to dominate H^2 according to *second-order stochastic dominance*. This condition implies that $var[\theta^1] \leq var[\theta^2]$. Intuitively, it means that system 1's customers are less heterogeneous than system 2's.

In this subsection, we discuss the relationship between the conditions $\theta^1 \preceq_{icx} \theta^2$ and $\Lambda^1(v) \geq \Lambda^2(v)$ and $\Lambda_n^1 \geq \Lambda_n^2$. First, the condition $\theta^1 \preceq_{icx} \theta^2$ is not a sufficient condition for $\Lambda^1(v) \geq \Lambda^2(v)$. The condition $\theta^1 \preceq_{icx} \theta^2$ means that $\int_x^1 H^1(y)dy \geq \int_x^1 H^2(y)dy$, $x \in [0,1]$. Let $y = 1/c(t)$, $dy/dt = -c'(t)/(c(t))^2$. Then this condition becomes

$$\int_v^0 H^1(1/c(t))[-c'(t)/(c(t))^2]dt \geq \int_v^0 H^2(1/c(t))[-c'(t)/(c(t))^2]dt.$$

This condition is very different from

$$\int_0^v H^1(1/c(t))dt \geq \int_0^v H^2(1/c(t))dt.$$

Hence, the condition $\theta^1 \preceq_{icx} \theta^2$ doesn't imply $\Lambda^1(v) \geq \Lambda^2(v)$. Similarly, $\theta^1 \preceq_{icx} \theta^2$ is not a sufficient condition for $\Lambda_n^1 \geq \Lambda_n^2$. The former depends on the integrals of the H^k on the whole interval $[0, v]$, while the latter depends on the products of the discrete values of the H^k on $\{\theta_0, \theta_1, \theta_2, \ldots\}$.

Hence, more concentrated customers need not imply a larger cumulative effective arrival rate and therefore, the impact on throughput is unclear. Table 1 shows the busy probability with two H functions. In this example, H^1 is a beta distribution with $\alpha = 2$ and $\beta = 2$ and H^2 is a beta distribution with $\alpha = 0.5$ and $\beta = 0.5$. For such two distributions, one can show that $H^1 \preceq_{cx} H^2$. We assume a linear cost function and fix $\mu = 2$ but change λ over $\{0.5, 1, 2, 3, 4, 5, 6, 7\}$. We can see that within each information scenario, the busy probability with H^1 is larger than the one with H^2 when utilization is small; smaller when utilization is large. Therefore, when customers are

TABLE 1. Busy Probability with Different H Functions $\mu = 2$

	$\alpha = 2, \beta = 2$			$\alpha = 0.5, \beta = 0.5$		
λ	No	Partial	Full	No	Partial	Full
0.5	0.2376	0.2311	0.2371	0.1965	0.2263	0.2320
1	0.4151	0.4187	0.4395	0.3482	0.4088	0.4249
2	0.6026	0.6716	0.7305	0.5669	0.6658	0.7062
3	0.6900	0.8105	0.8879	0.7062	0.8173	0.8722
4	0.7402	0.8864	0.9587	0.7943	0.9024	0.9550
5	0.7731	0.9292	0.9861	0.8508	0.9487	0.9878
6	0.7966	0.9542	0.9957	0.8882	0.9733	0.9976
7	0.8144	0.9694	0.9987	0.9137	0.9862	0.9997

TABLE 2. Average Utility with Different H Functions $\mu = 2$

	$\alpha = 2, \beta = 2$			$\alpha = 0.5, \beta = 0.5$		
λ	No	Partial	Full	No	Partial	Full
0.5	0.4247	0.4485	0.4564	0.4562	0.4701	0.4741
1	0.3437	0.3995	0.4108	0.4197	0.4421	0.4482
2	0.2315	0.3194	0.3217	0.3555	0.3926	0.3960
3	0.1710	0.2635	0.2466	0.3015	0.3519	0.3444
4	0.1351	0.2252	0.1905	0.2575	0.3192	0.2954
5	0.1115	0.1983	0.1513	0.2224	0.2931	0.2519
6	0.0949	0.1787	0.1241	0.1944	0.2721	0.2158
7	0.0827	0.1640	0.1048	0.1720	0.2549	0.1871

more concentrated, the throughput of the system is larger in light traffic but smaller in heavy traffic. We also compute the average utility for customers (see Table 2) and observe that the average utility with H^1 is always smaller than the one with H^2. Therefore, more concentrated customers obtain a smaller average utility. We tried quadratic and square-root cost functions and obtained the same conclusions.

4. IMPACT OF CUSTOMERS' COST FUNCTION AND RISK ATTITUDE

In this section, we fix $H^1 = H^2 = H$ but consider different conditions on the c^k.

4.1. Inequality

Condition $c^1 \leq c^2$ means that customers in system 1 care less about waiting than those in system 2. We have the following conclusion about the system occupancy and the workload.

PROPOSITION 5: *If $c^1 \leq c^2$, then $N^1 \succeq_{lr} N^2$ under no or partial information, and $V^1 \succeq_{lr} V^2$ under full information.*

The above proposition implies that when customers' cost function is smaller, the system becomes more congested.

About average utilities, we have the following conclusions.

PROPOSITION 6: *If* $c^1 \leq c^2$, *then* $u^1 \geq u^2$ *under no information.*

The above conclusion means that when customers have smaller cost function, their average utility is larger. However, this conclusion need not hold for partial and full information. On one hand, since $c^1 \leq c^2$, $\theta_i^1 \geq \theta_i^2$ and thus $J(\theta_i^1) \geq J(\theta_i^2)$ since $J(\theta)$ is increasing in θ. On the other hand, since $J(\theta_i)$ is decreasing in i, $I^1 \succeq_{st} I^2$ implies $E[J(\theta_{I^1}^k)] \leq E[J(\theta_{I^2}^k)]$ for $k = 1, 2$. Thus, it is unclear whether $E[J(\theta_{I^1}^1)]$ or $E[J(\theta_{I^2}^2)]$ is larger.

Intuitively, *given* information I, system 2 has larger expected utility, but its larger arrival rate pushes the system to a more congested state, which decreases the utilities. The overall effect is unclear. Tables 3 and 4 show that either effect can dominate the other. All the cases considered have linear costs: $c^k(w) = 1 + \gamma^k w$. In Table 3 the average utility decreases with c^k, while in Table 4 it increases for partial and full information.

In a system with no balking, where all customers stay, the distribution of W is just that of the standard $M/M/1$ system with arrival rate λ. The average waiting cost for system k is $E[\theta^k]E[c^k(W)]$, which is larger for system 1. Thus, the customers in system 1 get lower average utility. In the system allowing balking but with no information, this conclusion is still true, by Propositions 2 and 6. However, in the system with balking and either partial or full information, the average utility in system 1 can be larger. Here, customers make their own decisions to maximize their expected

TABLE 3. Compare Linear Cost Functions with Beta H; $\alpha = \beta = 2$, $\lambda = 2$, $\mu = 2$

	Busy Probability			Average Utility		
γ	No	Partial	Full	No	Partial	Full
0.1	0.8601	0.8922	0.8959	0.3613	0.3956	0.4001
0.5	0.6972	0.7668	0.7848	0.2750	0.3324	0.3460
1.0	0.6026	0.6954	0.7259	0.2315	0.3038	0.3222

TABLE 4. Compare Linear Cost Functions with Beta H; $\alpha = \beta = 2$, $\lambda = 8$, $\mu = 2$

	Busy Probability			Average Utility		
γ	No	Partial	Full	No	Partial	Full
0.1	0.9768	1.0000	1.0000	0.0870	0.0893	0.0894
0.5	0.9002	0.9998	1.0000	0.0798	0.0899	0.0898
1.0	0.8284	0.9940	0.9996	0.0732	0.0908	0.0907

utilities, and this leads to less congestion in system 1. Less congestion in turn increases the average utility for those served customers.

4.2. Risk Aversion

In this subsection, we consider a different condition on c^k which indicates the degree of customers' risk aversion. For a utility function of one variable $u(x)$ which is increasing in x, the Arrow-Pratt measure of absolute risk aversion is $A(x) = -u''(x)/u'(x)$. Here, customers have an increasing *disutility* function $c(\cdot)$, hence we use $A(w) = c''(w)/c'(w)$ to measure risk aversion. Let $A^k(w)$ denote the risk aversion of the cost function c^k. We suppose customers in system 1 are more risk averse than those of system 2. As we shall see, this condition is related to others we have seen above.

PROPOSITION 7: *If* $A^1(w) \geq A^2(w)$ *for all* w *and* $c^{1'}(0) \geq c^{2'}(0)$, *then* $c^1(w) \geq c^2(w)$*; if* $A^1(w) \geq A^2(w)$ *for all* w *and* $c^{1'}(0) < c^{2'}(0)$, *then* $c^1(w)$ *crosses* $c^2(w)$ *once from below on* $(0, \infty)$.

Thus, greater risk aversion implies either no crossing or single crossing of the two cost functions, depending on their derivatives at 0. However, single crossing of the two cost functions implies nothing about the relation between the throughputs. Hence, a system with more risk-averse customers need not have a smaller throughput.

For example, consider c^1 to be a quadratic cost function, i.e., $c^1(w) = 1 + w^2$, which represents risk-averse customers, and c^2 to be a quare-root cost function, i.e., $c^2(w) = 1 + \sqrt{w}$ which represents risk-seeking customers. It can be easily shown that $A^1(w) > A^2(w)$ and $c^{1'}(0) < c^{2'}(0)$. $c^1(w)$ crosses $c^2(w)$ once from below. We assume $\lambda = 0.5$ and $\mu = 1.5$ and H to be a beta distribution with parameters (α, β). We fix $\beta = 4$ and change α over $\{0.5, 1, 2, 4, 6, 8\}$. When α increases, the pdf of θ skews to the right side, that is, there is a larger proportion of customers to be impatient. Table 5 shows the busy probability with the two cost functions. Table We can see that under no and partial information, the busy probability with square-root cost is always larger than the one with quadratic cost. However, under full information, it is larger when α is small, but smaller when α is larger than a certain level. Therefore, under full information, when most of customers are patient, risk-seeking behavior brings a larger throughput for the server while when most of customers are impatient, risk-seeking behavior brings a smaller throughput for the server. Table 6 shows the average utility with the two cost functions. We observe similar comparison result as the one with busy probability. We tried other settings with different α, β, λ and μ values and observe the same phenomena.

5. THE VALUE OF INFORMATION AND DELAY-RISK SENSITIVITIES

In this section, we discuss the relationship between the value of information and customer characteristics, namely, the weight on delay and the degree of risk aversion. We

TABLE 5. Busy Probability with Different-Shape Cost Functions; $\lambda = 0.5$, $\mu = 1.5$, $\beta = 4$

	Quadratic Cost			Square-Root Cost		
α	No	Partial	Full	No	Partial	Full
0.5	0.3306	0.3264	0.3274	0.3331	0.3314	0.3316
1	0.3257	0.3193	0.3222	0.3324	0.3280	0.3279
2	0.3116	0.3062	0.3144	0.3298	0.3183	0.3183
4	0.2800	0.2847	0.3046	0.3184	0.2961	0.2988
6	0.2529	0.2693	0.2985	0.3016	0.2780	0.2841
8	0.2308	0.2596	0.2943	0.2833	0.2658	0.2742
10	0.2128	0.2544	0.2911	0.2657	0.2584	0.2676
12	0.1978	0.2519	0.2885	0.2495	0.2542	0.2632
14	0.1850	0.2508	0.2865	0.2349	0.2520	0.2602
16	0.1741	0.2503	0.2847	0.2219	0.2509	0.2581

TABLE 6. Average Utility with Different-Shape Cost Functions; $\lambda = 0.5$, $\mu = 1.5$, $\beta = 4$

	Quadratic Cost			Square-Root Cost		
α	No	Partial	Full	No	Partial	Full
0.5	0.8171	0.8389	0.8527	0.8562	0.8576	0.8583
1	0.6756	0.7228	0.7475	0.7413	0.7460	0.7478
2	0.4800	0.5663	0.6029	0.5701	0.5860	0.5903
4	0.2781	0.3965	0.4369	0.3632	0.4058	0.4129
6	0.1832	0.3073	0.3430	0.2491	0.3116	0.3184
8	0.1310	0.2529	0.2827	0.1810	0.2548	0.2604
10	0.0989	0.2168	0.2414	0.1375	0.2177	0.2218
12	0.0776	0.1922	0.2127	0.1079	0.1926	0.1956
14	0.0626	0.1732	0.1905	0.0867	0.1734	0.1755
16	0.0516	0.1570	0.1716	0.0711	0.1570	0.1586

first give the expression of the value of information, then give two general conclusions and at last we give numerical computation results.

A simple and direct measure of the value of information is the difference of the average utilities under more and less information. We use superscripts *no*, *part* and *full* to indicate the parameters and performance measures of the system under no, partial and full information, respectively. Define $VI^{fn}(\theta) = u^{full}(\theta) - u^{no}(\theta)$. Then $VI^{fn}(\theta)$ measures the value of full over no information for a customer with weight θ. Similarly, define $VI^{pn}(\theta) = u^{part}(\theta) - u^{no}(\theta)$ and $VI^{fp}(\theta) = u^{full}(\theta) - u^{part}(\theta)$.

5.1. Value of Information and Delay Sensitivity

We have two propositions about the value of information.

PROPOSITION 8: *If* $p_0^{full} \geq p_0^{no}$*, then for all* θ*,* $VI^{fn}(\theta) \geq 0$*. Similarly, if* $p_0^{part} \geq p_0^{no}$*, then for all* θ*,* $VI^{pn}(\theta) \geq 0$*.*

That is, information benefits every *individual* customer, if it hurts the server. This conclusion is stronger than a similar conclusion in G-Z where they show that information benefits the *whole group* of customers if it hurts the server.

PROPOSITION 9: *When* $\theta > \theta_-$*,* $VI^{fn}(\theta)$ *is decreasing in* θ*. That is, the value of information is smaller for less patient customers. When* $\theta \leq \theta_-$*, if* $VI^{fn}(\theta) > 0$*, then* $VI^{fn}(\theta)$ *is increasing with* θ*. Otherwise, if* $VI^{fn}(\theta) < 0$*,* $VI^{fn}(\theta)$ *is decreasing with* θ*. That is, the value of information is larger for less patient customers, if it benefits those customers. Similar conclusions hold for* $VI^{pn}(\theta)$*.*

Thus, the value of information is not necessarily larger for less patient customers.

5.2. Value of Information and Degree of Risk Aversion

In this subsection, we study the relationship between the value of information and the degree of risk aversion. We assume H is a uniform distribution and consider the cost function

$$c(w) = e^{\gamma w}$$

for $\gamma > 0$. This function is increasing and convex. To guarantee the stability of the no-information model, we restrict $\gamma < \mu$.

Here, $A(w) = c''(w)/c'(w) = \gamma$. Thus, a larger γ means a higher level of risk-aversion. This risk aversion measure is independent of w, a property called *constant absolute risk aversion* (CARA). Also, the cost function is increasing with γ.

For the numerical computation, we consider $\theta = 0.1, 0.5, 0.9$, which represent very patient, moderately patient and impatient customers, respectively. We fix $\mu = 2$ and change λ over $\{0.5, 1, 2, 3\}$. Figures 1, 2 and 3 show the value of full versus no information with $\theta = 0.1, 0.5, 0.9$, respectively. Figures 4, 5 and 6 show the value of partial versus no information, and Figures 7, 8 and 9 show the value of full versus partial information.

First, we observe that the relationship between the value of information and the level of customers' risk aversion is not monotone. When $\theta = 0.1$, the value of information roughly increases with the level of risk aversion. However, for $\theta = 0.5$ and 0.9, it doesn't. For very patient customers, they join the system under no and full information in most of time. What they care about is congestion. The cost function $c(w) = e^{\gamma w}$ is increasing in γ. So, when γ increases, the system becomes less congested with and without information. However, it is likely that the congestion in the system with information is lessened more than the one without information. Hence, for patient customers, the value of information tends to increase with γ, since they usually join the system, and less congestion with information benefits them most. However, customers with moderate to severe impatience often leave the system with and without information, and thus the value of information is less promising to them.

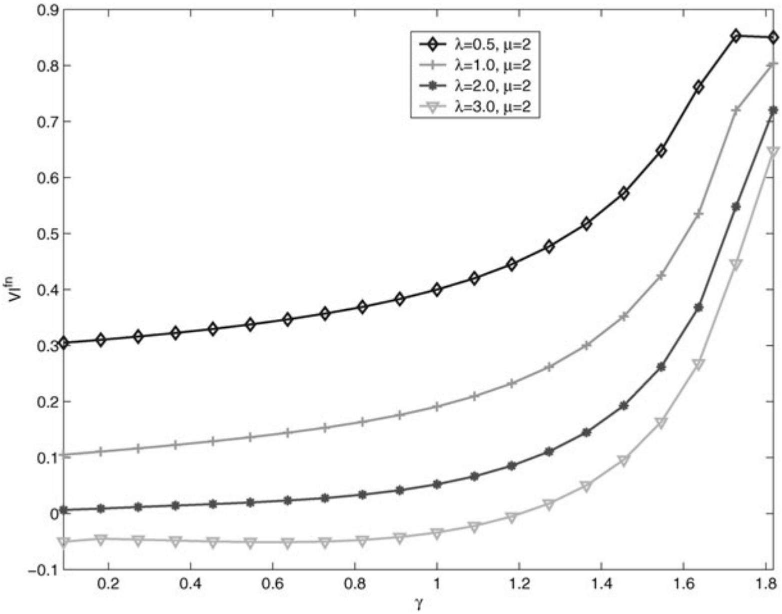

FIGURE 1. Full/no information with $\theta = 0.1$.

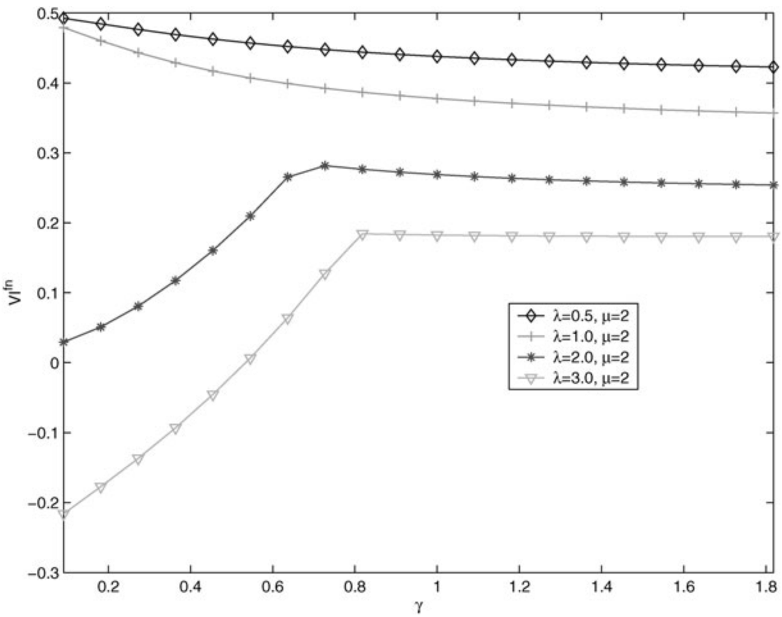

FIGURE 2. Full/no information with $\theta = 0.5$.

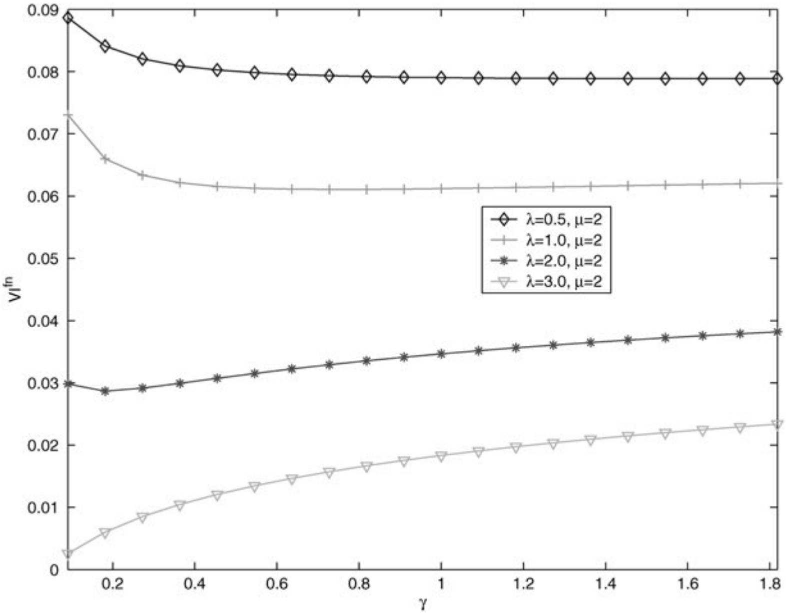

FIGURE 3. Full/no information with $\theta = 0.9$.

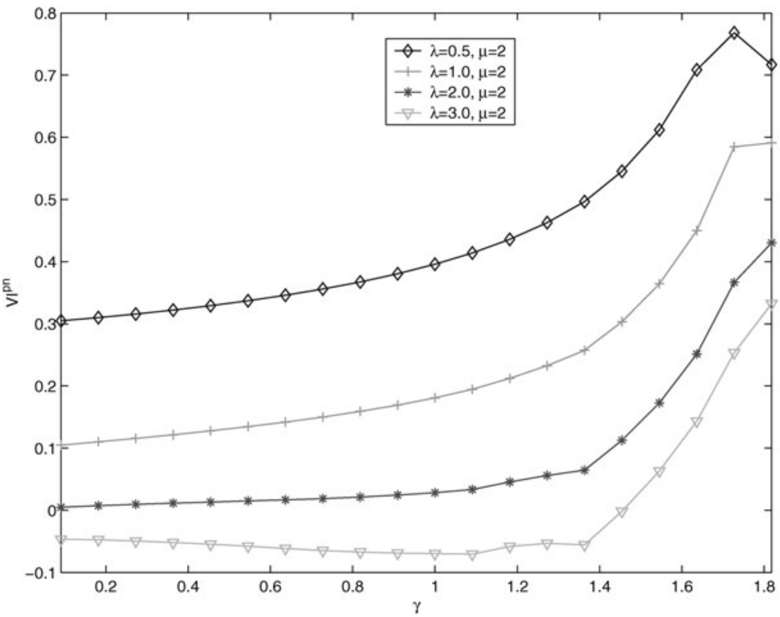

FIGURE 4. Partial/no information with $\theta = 0.1$.

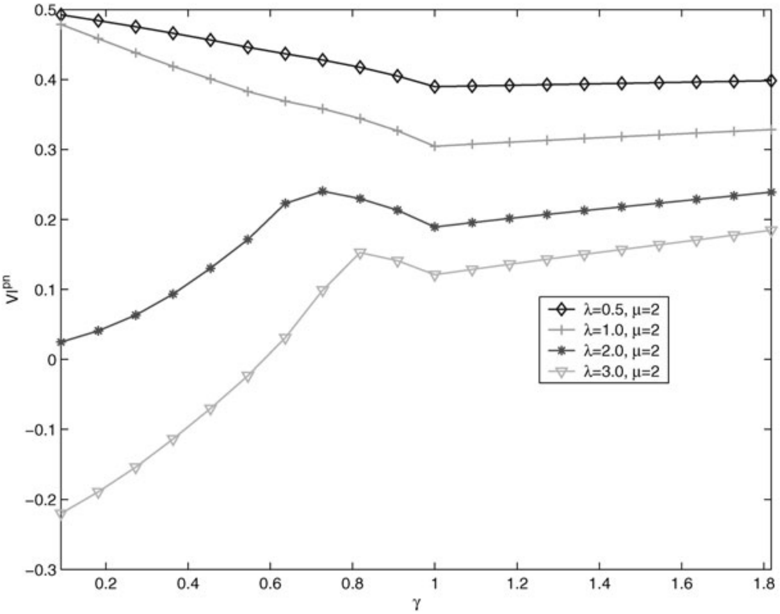

FIGURE 5. Partial/no information with $\theta = 0.5$.

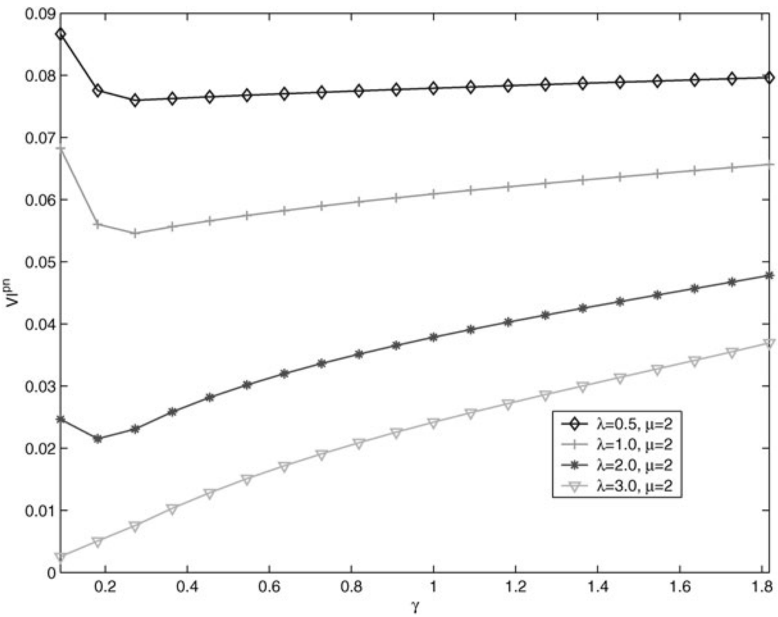

FIGURE 6. Partial/no information with $\theta = 0.9$.

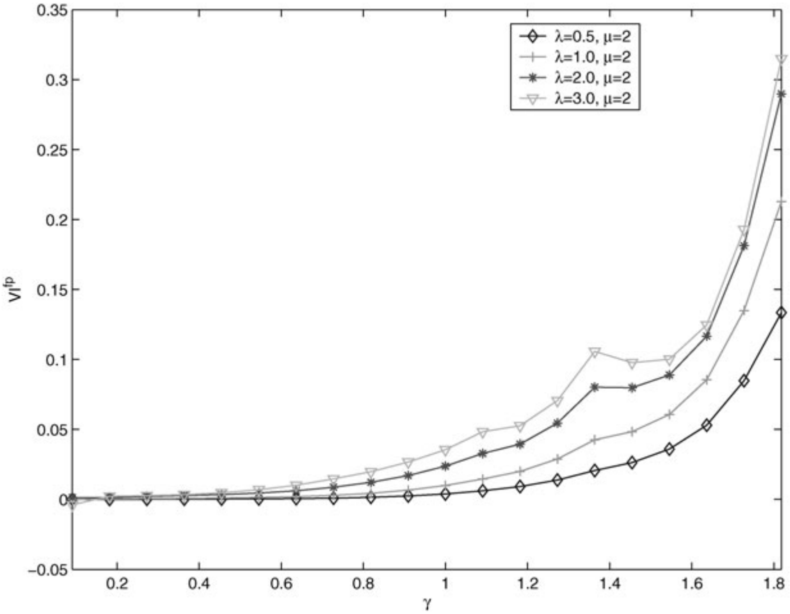

FIGURE 7. Full/partial information with $\theta = 0.1$.

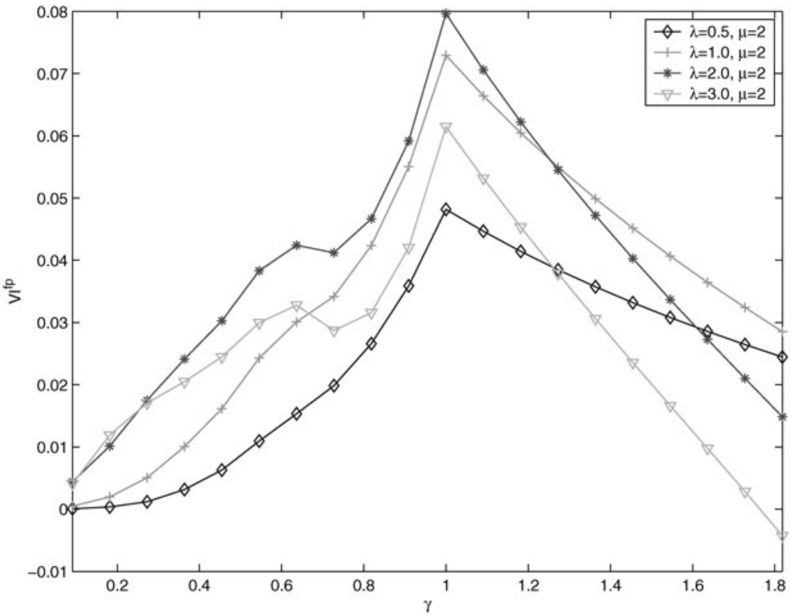

FIGURE 8. Full/partial information with $\theta = 0.5$.

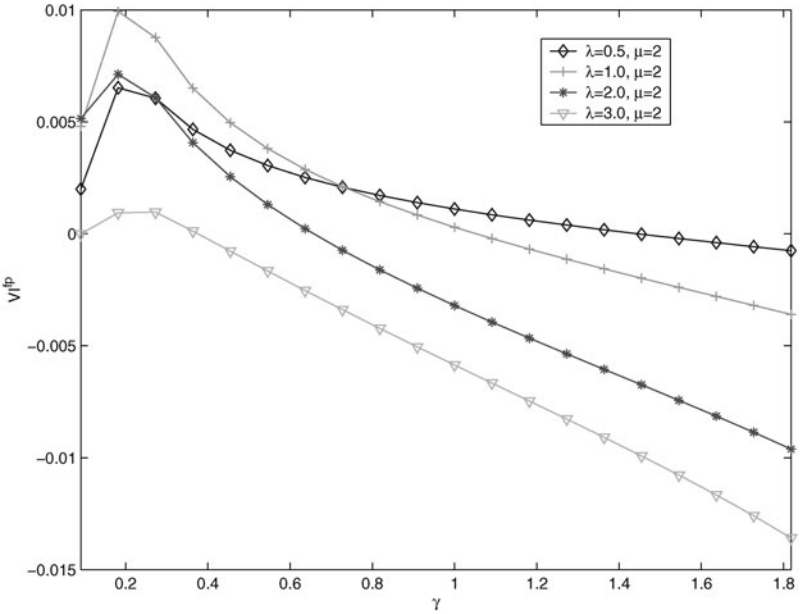

FIGURE 9. Full/partial information with $\theta = 0.9$.

In summary, information here not only helps one customer to make his decision, but also affects the congestion of the system itself. Hence, the risk here is endogenous. In this situation, the value of customers' cost function, instead of the shape of it, is an important factor deciding customers' delay risk. And there is no monotone relationship between the value of information and the degree of risk aversion.

We also observe that the value of full versus no information, and that of partial versus no information, decrease with the system's utilization. However, the value of full versus partial information need not behave in this way.

6. CONCLUSIONS

In this paper, we explore the impacts of customers' delay sensitivities and risk attitudes on balking queues.

We first consider different stochastic orders on the distribution of delay-sensitivity parameters of two systems. We show that when one system's customers are stochastically more patient than those of the other system, the system becomes more congested than the other. With no information, the average utility for customers in that system is larger than the one in the other system. But this conclusion only holds for some special cost function under partial or full information. We also numerically show that when customers are more concentrated in one system than the other system, the throughput of that system is larger in light traffic but smaller in heavy traffic than the other's.

We then discuss the relationship on customers' cost functions of the two systems and discuss the relationship between them and the effective arrival rates. When customers' cost function becomes smaller, the system becomes more congested. When customers' cost function is smaller, customers' average utility is larger under no information. However, this conclusion need not hold for partial and full information. We also explore risk-aversion relations between cost functions. Our numerical result shows that under no and full information, risk-seeking customers brings a larger throughput for the server; however, under full information, customers' risk-seeking behavior brings a larger throughput for the server when most of customers in the system are patient but it reduces throughput for the server when most of customers are impatient.

Finally, we examine the relationship between the value of information and customers' characteristics. We show that, when information hurts the server, it always benefits each *individual* customer. However, the value of information need not increase with the weight on delay cost. We also explore numerically the value of information for different risk-aversion customers. We show that there is no simple relationship between the value of information and the degree of customers' risk aversion.

In this paper, we consider the information service for *all* customers. Everyone can react to the information and the aggregate behavior makes the value of information not always positive. It is especially true in a traffic system: If everyone is informed that one road is not crowed and everyone rushes to that road, soon that road will become crowed and nobody benefits from the information. Therefore, one may consider another case where information is only provided to some minority while majority people are informed nothing. A further research on this direction could be interesting.

Acknowledgment

We would like to thank HK PolyU CRG(Competitive Research Grant), "StrategicCustomers" Grant A.44.PB0W for their financial support.

References

1. Armony, M. & Maglaras, C. (2004). On customer contact centers with a call-back option: Customer decisions, routing rules, and system design. *Operations Research* 52: 271–292.
2. Armony, M. & Maglaras, C. (2004). Contact center with a call-back option and real-time delay information. *Operations Research* 52: 527–545.
3. Armony, M., Shimkin, N. & Whitt, W. (2005). The impacts of delay announcements in many-server queues with abandonment. Working paper, New York University.
4. Dai, L. & Chao, X. (1996). Comparing single-server loss systems in random environment. *IEEE Transaction on Automatic Control* 41: 1079–1083.
5. de Palma, A., Lindsey, R. & Picard, N. (2007). Congestion, risk aversion and the value of information. Working paper. University of Alberta, Canada.
6. Chao, X. & Dai, L. (1995). A monotonicity result for a single-server loss system. *Journal of Applied Probability* 32: 1112–1117.
7. Edelson, M. & Hildebrand, K. (1975). Congestion tolls for Poisson queueing processes. *Econometrica* 43: 81–92.

8. Freixas, X. & Kihlstrom, R. (1984). Risk aversion and information demand. In M. Boyer and R. Kihlstrom, eds., *Bayesian Models in Economic Theory.* Amsterdam: North Holland.
9. Gavish, B. & Schweitzer, P. (1973). Queue regulation policies using full information. Working paper, Israel Scientific Center.
10. Guo, P. & Zipkin, P. (2007). Analysis and comparison of queues with different levels of delay information. *Management Science* 53: 962–970.
11. Guo, P. & Zipkin, P. (2006). The effects of information on a queue with balking and phase-type service times. Forthcoming in *Naval Research Logistics*.
12. Hassin, R. (2007). Information and uncertainty in a queueing system. *Probability in the Engineering and Informational Sciences* 21: 361–380.
13. Hassin, R. & Haviv, M. (2003). *To Queue or Not to Queue: Equilibrium Behavior in Queuing Systems*. Boston/Dordrecht/London: Kluwer Academic Publishers.
14. Heyman, D. (1982). On Ross's conjectures about queues with non-stationary Poisson arrivals. *Journal of Applied Probability* 19: 245–249.
15. Hilton, R. (1981). The determinants of information value: Synthesizing some general results. *Management Science* 27: 57–64.
16. Karlin, S. (1968). *Total Positivity Vol. 1*. Stanford, California: Stanford University Press.
17. Müller, A. & Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons.
18. Nadiminti, R., Mukhopadhyay, T. & Kriebel, C. (1996). Risk version and the value of information. *Decision Support Systems* 16: 241–254.
19. Naor, P. (1969). The regulation of queue size by levying tolls. *Econometrica* 37: 15–24.
20. Pratt, J. (1964). Risk aversion in the small and in the large. *Econometrica* 32: 122–136.
21. Rolski, T. (1990). Queues with nonstationary inputs. *Queueing Systems* 5: 113–130.
22. Ross, S. (1978). Average delay in queues with nonstationary arrivals. *Journal of Applied Probability* 15: 602–609.
23. Schroeter, R. (1982). The costs of concealing the customer queue. Working paper. Department of Economics. Iowa State University.
24. Shaked, M. & Shanthikumar, J. (1994). *Stochastic Orders and Their Applications.* Boston, San Diego, New York, London, Sydney, Tokyo, Toronto: Academic Press.
25. Shimkin, N. & Mandelbaum, A. (2004). Rational abandonment from tele-queues: nonlinear waiting cost with heterogeneous preferences, *Queueing Systems* 47: 117–146.
26. Stidham, S. (1985). Optimal control of admission to a queuing system. *IEEE Transaction on Automatic Control* 30: 705–13.
27. Whitt, W. (1999). Improving service by informing customers about anticipated delay. *Management Science* 45: 192–207.
28. Whitt, W. (2006). Sensitivity of performance in the Erlang-A queueing model to changes in the model parameters. *Operations Research* 54: 247–260.
29. Willinger, M. (1989). Risk aversion and the value of information. *Journal of Risk and Insurance* 56: 320–328.

APPENDIX A

Auxiliary Comparison Results of Three Information Models

PROPOSITION 10: *Under no information, if* $\lambda^1_- \geq \lambda^2_-$, *then* $N^1 \succeq_{lr} N^2$.

PROOF: Whitt [27] (Theorem 4.1) shows that, for any pair of birth-death processes, $N^1 \succeq_{lr} N^2$, provided

$$\frac{\lambda_n^1}{\mu_{n+1}^1} \geq \frac{\lambda_n^2}{\mu_{n+1}^2}, n \geq 0.$$

Here, each system k is an $M/M/1$ queue with arrival rate λ_-^k and service rate μ. Hence the above condition is satisfied. ■

Define the cumulative effective arrival rate $\Lambda_n = \lambda \prod_{m=0}^{n} H(1/c_m)$.

PROPOSITION 11: *Under partial information,*

1. *if* $\Lambda_n^1 \geq \Lambda_n^2$, *then* $p_0^1 \leq p_0^2$;
2. *if* $\lambda_n^1 \geq \lambda_n^2$ *for all* $n = 0, 1, 2, ...,$ *then* $N^1 \succeq_{lr} N^2$;
3. *if* λ_n^1 *crosses* λ_n^2 *once from above (i.e., there exists* $\hat{n} > 0$ *such that* $\lambda_n^1 \geq \lambda_n^2$, $n \leq \hat{n}$, *and the inequality is reversed for* $n > \hat{n}$*), then* $N^1 \preceq_{uv} N^2$;
4. *if the ratio* λ_n^1/λ_n^2 *is decreasing in n, then* $N^1 \preceq_{lc} N^2$.

PROOF: For part 1, if $\Lambda_n^1 \geq \Lambda_n^2$ for all n, then $\Theta_n^1 \geq \Theta_n^2$ for all n, and hence $p_0^1 \leq p_0^2$.

Part 2 follows by the same argument as in Proposition 10.

For part 3, if λ_n^1 crosses λ_n^2 once from above, consider the ratio

$$\frac{p_{n+1}^1/p_{n+1}^2}{p_n^1/p_n^2} = \frac{\lambda_n^1}{\lambda_n^2}.$$

For $n \leq \hat{n}$, this fraction is greater than 1, thus p_n^1/p_n^2 is increasing; for $n > \hat{n}$, similarly, p_n^1/p_n^2 is decreasing. Thus p_n^1/p_n^2 is unimodal. Hence, $N^1 \preceq_{uv} N^2$.

Finally, for part 4, if λ_n^1/λ_n^2 is decreasing in n,

$$\frac{p_{n+1}^1/p_{n+1}^2}{p_n^1/p_n^2} = \frac{\lambda_n^1}{\lambda_n^2}.$$

is decreasing in n. So p_n^1/p_n^2 is log-concave, and $N^1 \preceq_{lc} N^2$. ■

PROPOSITION 12: *Under full information,*

1. *if* $\Lambda^1(v) \geq \Lambda^2(v)$, *then* $p_0^1 \leq p_0^2$;
2. *if* $\lambda^1(v) \geq \lambda^2(v)$ *for all* $v \geq 0$, *then* $V^1 \succeq_{lr} V^2$;
3. *if* $\lambda^1(v)$ *crosses* $\lambda^2(v)$ *once from above, then* $V^1 \preceq_{uv} V^2$;
4. *if the ratio* $\lambda^1(v)/\lambda^2(v)$ *is decreasing in v, then*

$$\frac{\ln(f^1(v))'}{\ln(f^2(v))'}$$

is decreasing in v.

PROOF: Part 1 follows immediately from (5).

For part 2, if $\lambda^1(v) \geq \lambda^2(v)$ for all $v \geq 0$, then $\Lambda^1(v) - \Lambda^2(v)$ is positive and increasing. So, $p_0^1 \leq p_0^2$, and

$$\frac{f^1(v)}{f^2(v)} = \frac{p_0^1}{p_0^2} \exp\{[\Lambda^1(v) - \Lambda^2(v)]\}$$

is increasing. Hence, $V^1 \succeq_{lr} V^2$.

For part 3, if $\lambda^1(v)$ crosses $\lambda^2(v)$ once from above, $\Lambda^1(v) - \Lambda^2(v)$ is positive and increasing for $v \leq \hat{v}$, but decreasing for $v \geq \hat{v}$. That is, $\Lambda^1(v) - \Lambda^2(v)$ is unimodal, and therefore so is $f^1(v)/f^2(v)$.

Finally, for part 4, if the ratio $\lambda^1(v)/\lambda^2(v)$ is decreasing in v, then

$$\frac{\ln(f^1(v))'}{\ln(f^2(v))'} = \frac{\Lambda^{1'}(v)}{\Lambda^{2'}(v)} = \frac{\lambda^1(v)}{\lambda^2(v)}$$

is decreasing in v. ■

APPENDIX B

Proofs of Statements

PROOF FOR PROPOSITION 1: The condition $\theta^1 \preceq_{st} \theta^2$ means $H^1(\theta) \geq H^2(\theta)$, for all θ in [0, 1]. Hence, $H^1(1/c(v)) \geq H^2(1/c(v))$ for all $v \geq 0$ and $H^1(1/c_n) \geq H^2(1/c_n)$ for all $n \geq 0$. So, $\lambda^1(v) \geq \lambda^2(v)$, and $\lambda^1_n \geq \lambda^2_n$. Also, $\lambda^1_- \geq \lambda^2_-$ from (1). From Propositions 10, 11 and 12, the conclusion follows. ■

PROOF FOR PROPOSITION 2: By the definition of J^k, we have $J^1 \geq J^2$. By Proposition 1, $\theta^1_- \geq \theta^2_-$. Thus, $u^1 = J^1(\theta^1_-) \geq J^1(\theta^2_-) \geq J^2(\theta^2_-) = u^2$. ■

PROOF FOR PROPOSITION 3: In this case, suppressing k for the moment,

$$J(\theta) = \frac{1}{\alpha+1}\theta^{\alpha} = \frac{1}{\alpha+1}H(\theta).$$

so the average utility $u = E[J(\theta_I)] = \frac{1}{\alpha+1}E[H(\theta_I)] = \frac{1}{\alpha+1}(1-p_0)\mu/\lambda$. Under either partial or full information, $1 - p^1_0 \geq 1 - p^2_0$, by Proposition 1. Thus, $u^1 \geq u^2$. ■

PROOF FOR PROPOSITION 4: We demonstrate the result for partial information. The proof for full information is similar.

The theory of total positivity (see, Karlin [16]) shows that monotone ratios are preserved under integration. Thus, decreasing h^1/h^2 implies decreasing H^1/H^2. Decreasing H^1/H^2, in turn, implies decreasing J^1/J^2. Hence $J^1(\theta_n)/J^2(\theta_n)$ is increasing in n. From proposition 1, we know that $N^1 \succeq_{lr} N^2$ under partial information, that is, p^1_n/p^2_n is increasing in n. Hence,

$$\begin{aligned}\frac{u^1}{u^2} &= \frac{\sum_{n\geq 0} J^1(\theta_n)p^1_n}{\sum_{n\geq 0} J^2(\theta_n)p^2_n} \\ &\geq \frac{J^1(\theta_0)p^1_0}{J^2(\theta_0)p^2_0} = \frac{J^1(1)p^1_0}{J^2(1)p^2_0}.\end{aligned}$$ ■

PROOF FOR PROPOSITION 5: From the condition $c^1 \leq c^2$, we get $H(1/c^1(v)) \geq H(1/c^2(v))$ for all $v \geq 0$ and $H(1/c^1_n) \geq H(1/c^2_n)$ for all $n = 0, 1, 2....$ Hence, $\lambda^1(v) \geq \lambda^2(v)$ and $\lambda^1_n \geq \lambda^2_n$. Also we can derive that $\lambda^1_- \geq \lambda^2_-$ from (1). From Proposition 10, 11 and 12, we obtain the conclusion. ■

PROOF FOR PROPOSITION 6: By Proposition 5, $\lambda^1_- \geq \lambda^2_-$, so $\theta^1_- \geq \theta^2_-$. Also, $J^k = J$. Thus, since J is increasing, $u^1 = J(\theta^1_-) \geq J(\theta^2_-) = u^2$. ■

PROOF FOR PROPOSITION 7: According to Pratt [20], the condition $A^1(w) \geq A^2(w)$ is equivalent to the condition that function $\tau = c^2 \circ (c^1)^{-1}$ be increasing and concave; that is, c^2 is an increasing and concave transform of c^1. In out context, $c^2(0) = c^1(0) = 1$. Hence, $\tau(1) = 1$. Consider the derivative of τ,

$$\frac{d\tau}{dx} = \frac{c^{2'} \circ (c^1)^{-1}(x)}{c^{1'} \circ (c^1)^{-1}(x)}.$$

Condition $c^{2'}(0) \leq c^{1'}(0)$ is equivalent to $c^{2'} \circ (c^1)^{-1}(1) \leq c^{1'} \circ (c^1)^{-1}(1)$ or $\frac{d\tau}{dx}(1) \leq 1$. Since $\tau(1) = 1$ and τ is increasing and concave, it follows that $\tau(x) \leq x$. Insert $x = c^1(w)$, and we get $c^2(w) \leq c^1(w)$, $w \geq 0$.

Similarly, the condition $c^{2'}(0) > c^{1'}(0)$ is equivalent to $c^{2'} \circ (c^1)^{-1}(1) > c^{1'} \circ (c^1)^{-1}(1)$ or $\frac{d\tau}{dx}(1) > 1$. Since $\tau(1) = 1$ and τ is increasing and concave, it follows that the graph of $\tau(x)$ starts from $(1, 1)$ and crosses the diagonal line $y = x$ once from above. Insert $x = c^1(w)$, and we get $c^2(w)$ crosses $c^1(w)$ once from above. ■

PROOF FOR PROPOSITION 8: We prove the result for full versus no information. The proof for partial versus no information is similar.

Under no information, if $\theta > \theta_-$, the customer will balk and gets utility 0. Hence

$$u^{no}(\theta) = 0.$$

If $\theta \leq \theta_-$, the customer will join and obtain a nonnegative expected utility $1 - \theta E[c(W^{no})]$ and hence

$$u^{no}(\theta) = 1 - \theta E[c(W^{no})].$$

Under partial information, the customer will join if $1 - \theta c_n \geq 0$ or $n \leq c^{-1}(1/\theta)$. Denote n^* to be n satisfying $c_{n^*} = 1/\theta$ (n^* may not be an integer). The expected utility for the customer u^{part} is

$$u^{part}(\theta) = \sum_{n \leq n^*} (1 - \theta c_n) p_n^{part}.$$

Now define a new variable $\tilde{N}^{part}$ such that $\tilde{p}_n^{part} = p_n^{part}$ for all $n < n^*$ and $\tilde{p}_{n^*}^{part} = 1 - \sum_{n<n^*} p_n$. Then we can write

$$u^{part}(\theta) = 1 - \theta E[c_{\tilde{N}^{part}}].$$

Under full information, if $v \leq c^{-1}(1/\theta)$, customer will join and obtain the utility $1 - \theta c(v)$; otherwise, he will leave with utility 0. Hence, the customer's expected utility is

$$u^{full}(\theta) = \int_0^{c^{-1}(1/\theta)} (1 - \theta c(v)) dF^{full}(v).$$

Define a new distribution $\tilde{F}^{full}$ such that $\tilde{F}^{full}(v) = F^{full}(v)$, for $v < c^{-1}(1/\theta)$ and $\tilde{F}^{full}(v) = 1$, for $v \geq c^{-1}(1/\theta)$. Denote a random variable with this distribution by $\tilde{V}^{full}$. One can easily verify that

$$\tilde{V}^{full} \preceq_{st} V^{full}. \tag{6}$$

Then we can write

$$u^{full}(\theta) = 1 - \theta E[c(\tilde{V}^{full})]$$

G-Z show that $V^{full} \preceq_{st} W^{no}$ if $p_0^{full} \geq p_0^{no}$. From (6), we derive that $\tilde{V}^{full} \preceq_{st} W^{no}$. Hence $E[c(W^{no})] \geq E[c(\tilde{V}^{full})]$, and thus $u^{no}(\theta) \leq u^{full}(\theta)$ for $\theta \leq \theta_-$. For $\theta > \theta_-$, the customer has utility 0 under no information and nonnegative expected utility under full information, hence, $u^{no}(\theta) \leq u^{full}(\theta)$. ■

Proof for Proposition 9: We present the proof for full versus no information. The case of partial versus no information is similar.

When $\theta > \theta_-$, let's first write

$$VI^{fn} = u^{full}(\theta) = \int_0^{c^{-1}(1/\theta)} (1 - \theta c(v)) f^{full}(v) dv + (1 - \theta) p_0^{full}.$$

We have

$$\frac{dVI^{fn}}{d\theta} = -\int_0^{c^{-1}(1/\theta)} c(v) f^{full}(v) dv - p_0^{full} < 0.$$

When $\theta \leq \theta_-$,

$$\begin{aligned} VI^{fn} &= u^{full}(\theta) - u^{no}(\theta) \\ &= \int_0^{c^{-1}(1/\theta)} (1 - \theta c(v)) f^{full}(v) dv + (1 - \theta) p_0^{full} - (1 - \theta E[c(W^{no})]). \end{aligned}$$

We have

$$\begin{aligned} \frac{dVI^{fn}}{d\theta} &= -\int_0^{c^{-1}(1/\theta)} c(v) f^{full}(v) dv - p_0^{full} + E[c(W^{no})] \\ &= -E[c(\tilde{V}^{full})] + E[c(W^{no})]. \end{aligned}$$

When $u^{full}(\theta) > u^{no}(\theta)$, $E[c(\tilde{V}^{full})] < E[c(W^{no})]$ and hence $dVI^{fn}/d\theta > 0$; when $u^{full}(\theta) \leq u^{no}(\theta)$ then $dVI^{fn}/d\theta \leq 0$. ■