OPERATIONS RESEARCH
Vol. 59, No. 6, November–December 2011, pp. 1382–1394
ISSN 0030-364X | EISSN 1526-5463 | 11 | 5906 | 1382

http://dx.doi.org/10.1287/opre.1110.0976

"We Will Be Right with You": Managing Customer Expectations with Vague Promises and Cheap Talk

Gad Allon, Achal Bassamboo, Itai Gurvich
Kellogg School of Management, Northwestern University, Evanston, Illinois 60208
{g-allon@kellogg.northwestern.edu, a-bassamboo@kellogg.northwestern.edu, i-gurvich@kellogg.northwestern.edu}

Delay announcements informing customers about anticipated service delays are prevalent in service-oriented systems. How delay announcements can influence customers in service systems is a complex problem that depends on both the dynamics of the underlying queueing system and on the customers' strategic behavior. We examine this problem of information communication by considering a model in which both the firm and the customers act strategically: the firm in choosing its delay announcement while anticipating customer response, and the customers in interpreting these announcements and in making the decision about when to join the system and when to balk. We characterize the equilibrium language that emerges between the service provider and her customers. The analysis of the emerging equilibria provides new and interesting insights into customer-firm information sharing. We show that even though the information provided to customers is nonverifiable, it improves the profits of the firm and the expected utility of the customers. The robustness of the results is illustrated via various extensions of the model. In particular, studying models with incomplete information on the system parameters allows us also to highlight the role of information provision in managing customer expectations regarding the congestion in the system. Further, the information could be as simple as "high congestion"/"low congestion" announcements, or it could be as detailed as the true state of the system. We also show that firms may choose to shade some of the truth by using *intentional vagueness* to lure customers.

Subject classifications: dynamic programming/optimal control: applications; games/group decisions: noncooperative; probability: stochastic model applications; queues: applications; balking and reneging.
Area of review: Revenue Management.
History: Received April 2009; revisions received October 2009, May 2010, October 2010, December 2010; accepted January 2011.

1. Introduction

Delay announcements that inform customers about anticipated delays are prevalent in service systems. Call centers often use recorded announcements to inform callers of the congestion in the system and encourage them to wait for an available agent. Although some of these announcements do not provide much information—such as the common message, "Due to high volume of calls, we are unable to answer your call immediately," some call centers go as far as providing the customer with an estimate of his waiting time or his place in the queue. In many service systems where the real state of the system is not visible to customers, delay announcements affect customers' behavior and may, in turn, have significant impact on system performance. Consequently, the service provider can use the delay announcements as a tool to induce the appropriate customers' behavior, stirring the system so as to maximize her own profits. How to use delay announcements to manage the service system in an efficient manner is a complex problem, and its answer depends on both the dynamics of the underlying queueing system and the customer behavior.

Previous papers addressing this issue analyzed two categories of information provided to the customer: (i) full-state information—either exact waiting-time information (if such information is available to the service provider) or the state of the system when the customer arrives (or an estimate based on the state), and (ii) no information—where no information is provided, and customers must base their decisions on their expectation regarding the system performance. The main assumption made in these papers is that customers treat the information provided in the delay announcements as a priori verified (i.e., credible), and make their decisions about whether to join or leave the system accordingly. The two main issues with this assumption are: (i) Customers may not be naive in their attitude towards any information provided by interested parties, and thus take such announcements with a "grain of salt." Moreover, under the assumption of "naivety," it makes sense for the firm to deviate from the truth-telling policy. The option that the firm might lie, given that the customer always believes the firm, is never explored in the literature. (ii) Further, prior work implicitly assumes that the announcements have a quantifiable meaning in terms of place-in-queue or average waiting time. However, as stated above, many service providers use verbal messages that need to be further processed in order for customers to make the decision of whether to join or not. For example, without processing, it is not clear what "high volume of calls" means. This problem is clearly a consequence of the first issue, because without processing, only quantifiable announcements are possible. Thus, the customers are able to evaluate their utility only for simple (i.e., no-information or full-information) announcements and only those are discussed. In practice, however, we observe a much richer variety of announcements used by service providers.

This paper addresses these issues by proposing a model in which customers treat information provided by the service provider as unverified and nonbinding (that is, noncontractable). The model treats customers as strategic in the way they process information (as well as in making the decisions whether to join or balk) and the service provider as strategic in the way she provides the information. The customers and the service provider are assumed to be self-interested in making their decisions: the service provider in choosing which announcements to make, and the customers in interpreting these and making their decisions.

This allows us to characterize the equilibrium language that emerges between the service provider and her customers. (We use the term equilibrium language, in accordance with the usage in the cheap talk literature, to refer to the equilibrium emerging in the cheap talk game. The phrase "language" is used to emphasize the signaling rule and its associated actions.) By doing that, not only do we relax the assumption that customers may be naive in their treatment of the announcements, but we also demonstrate that many of the commonly used announcements arise in *equilibrium* in such a model. The spectrum of possible equilibria will range from announcements that are analogous to the nonquantifiable type that describe the volume of arriving customers as high or low, to the more quantifiable ones that provide waiting-time announcements, both of which are common in different systems. Whereas prior work focused on systems with either full-state information or no information, our model shows that a richer language, which usually includes intentional vagueness, arises as an equilibrium outcome in a game played between the service provider and her customers. By intentional vagueness we mean that under any equilibrium the firm provides the same announcement on multiple states of the system. In developing the model and characterizing the emerging equilibrium language, we will account both for the strategic nature of the interested parties—customers and service provider—as well as for the queueing dynamics prevalent to service systems.

This paper proposes what appears to be the first model to deal with the strategic nature of the information transmission in a practical operational setting, where unverifiable,[1] noncommittal, real-time information is provided by a self-interested firm to selfish customers.

We treat the announcements made by the system manager as "cheap talk," i.e., preplay communication that carries no cost. Cheap talk consists of costless,[2] nonbinding, nonverifiable messages that may affect the customer's beliefs. Although providing the information does not *directly* affect the payoffs, it has an indirect implication through the customer's reaction and the equilibrium outcomes. The information has no impact on the payoffs of the different players per se, i.e., the payoffs of both sides depend only on the actions taken by the customers and queueing dynamics. This, in turn, means that if a customer does not follow the recommendation made by the firm, he is not penalized, nor is he rewarded when he follows it. However, as it will be shown, the announcements do have an impact on the service provider's profits and the customers' utility, in equilibrium. This is in agreement with both the cheap talk literature (see Crawford and Sobel 1982) and the queueing literature with strategic customers (see Naor 1969), where the information provided to the customer is in the form of full visibility of the queue and does not alter the customer's utility directly; it allows him to make a knowledgeable decision whether to join or not, and thus affects his utility in an indirect manner. Our model echoes the framework developed by Crawford and Sobel (1982) but the specifics of our model are different from theirs in significant aspects. We examine these differences more closely in §1.2.

This paper focuses on the *strategic* interaction between the customer and the firm in a setting in which their incentives are *misaligned*, when *unverifiable, costless, and nonbinding* information is provided to the customer. In all of the instances described in this paper, the information is always unverifiable and has no contractual bearing (i.e., it is nonbinding). This stands in contrast to service-level *guarantees*, such as those made by Dominos Pizza, Ameritrade, and E∗trade, to name a few, where the commitment is contractually binding; see Allon and Federgruen (2007) for a more detailed discussion of the delay guarantees provided by these firms.

Our main findings are as follows:

(i) We characterize the equilibrium of the game played by the service provider and the customers. The characterization depends both on the queueing dynamics (based on a Markov decision process) and the strategic nature of the players. We show that although the characterization is complex, under pure strategies, an equilibrium can be mapped into a single threshold level. The firm provides two signals to indicate whether the level of congestion is below or above the threshold. The customer makes his decision according to the expected congestion level implied by each signal.

(ii) We show that different types of misalignment may lead to existence or nonexistence of pure-strategy equilibria. Further, we characterize a two-signal equilibrium when it exists, and show that any equilibrium with more than two signals is outcome equivalent to the simpler two-signal one. Our results imply that some commonly used nonquantifiable announcements regarding the congestion level in the system can be reduced to a two-signal equilibrium.

(iii) We show that even if customers are allowed to randomize between multiple actions (i.e., join or balk), any equilibrium with more than two signals will be outcome equivalent to the two-signal equilibrium. Furthermore, under any equilibrium where the customer randomizes, it is not possible for the firm to have a signaling rule that includes both a signal that induces the customers to balk with probability 1 and a signal that induces the customers to join with probability 1.

(iv) We prove that there always exists a *most informative equilibrium*, i.e., an equilibrium with the finest partition of the state space revealed to the customer by the service provider. The firm employs *intentional vagueness* in this equilibrium unless the customers and the firm are perfectly aligned. By perfect alignment, we mean that the self-interested customers agree with the profit-maximizing firm on the preferred action for every state of the system. The role of intentional vagueness is to lure customers to join the system in states in which, under full-state information, the customer would not join, yet the congestion in the system is not too high. The informative equilibria, either with or without intentional vagueness, are shown to be equivalent to the common practices of announcing the location in the queue, an estimate of the average waiting time, or a confidence interval of the expected waiting time.

(v) We show that a babbling equilibrium where the customer disregards any information provided by the firm (or the firm does not provide any information) always exists, either in pure or mixed strategies. We then show that the service provider and the customers always prefer other, more informative, equilibria over a babbling equilibrium.

(vi) We also test the robustness of the above results by considering models in which we relax the information structure and those in which customers adaptively update their beliefs about the meaning of messages adaptively. We find that most of the findings continue to hold with appropriate modifications.

Organization of the remainder of the paper: The rest of this section is divided into two parts: the first reviews the antecedent literature and the second briefly describes the Crawford and Sobel (1982) game. Section 2 provides the detailed description of our model. In §3, we define our notion of equilibrium and state our main results. In §4 we discuss the signaling language that emerges in equilibrium and, in particular, introduce and analyze the notion of most informative equilibria. Section 5 concludes the paper. All the proofs are relegated to the electronic companion, which can be found at http://or.journal.informs.org/. Extensions of the model, as well as a numerical study, are also discussed in the electronic companion.

1.1. Literature Review

Queueing models with strategic customers. The literature on queueing models with strategic customers dates back to Naor (1969), who studied a system in which strategic customers observe the length of the queue prior to making the decision whether to join or balk. There is a (partial) conflict of interest between the self-interested customer and the interests of the social-welfare maximizing service provider. Naor (1969) shows that pricing can be used to achieve the first-best solution. The follow-up literature extends Naor (1969) along multiple dimensions. One such stream studies models where the firm offers different grades of services (see the recent paper by Afeche 2004 and the references therein). Another stream focuses on competition in the presence of congestion-sensitive customers (see Allon and Federgruen 2007 and the references therein). All of the above papers assume that the firms provide long-run averages information rather than real-time information, and that the information is credible and is treated as such by the customers. The above models assume that customers are rational when making the decision of which line to choose or when to join versus balk from a queue. For more-detailed discussion we refer the reader to Hassin and Haviv (2003) and the references therein.

Queueing models with delay announcements. Hassin (1986) studies the problem of a price-setting, revenue-maximizing service provider that has the option to reveal the queue length to arriving customers, but may choose not to disclose this information. The author shows that it may be socially optimal to force the firm to provide full-state information. Armony and Maglaras (2004b) analyzes a service system where arriving customers can decide whether to join, balk, or wait for the provider to call within a guaranteed time. The customers' decisions are based on the equilibrium waiting time. Armony and Maglaras (2004a) extends the above model to allow the service manager to provide the customers with an estimate of the delay, based on the real-time state of the system upon their arrival. The authors show that providing information on the estimated delay improves the system performance. Armony et al. (2009) study the impact of delay announcements on the performance of a many-server queue with customer abandonment.

Dobson and Pinker (2006) develop a stochastic model of a custom production environment with pricing, where customers are heterogeneous in terms of their tolerances for waiting. The authors model intermediate levels of information sharing (with a specific structure) ranging from no sharing to complete state-dependent lead-time information, and compare the resulting performance from the firm's and customer's perspectives. Guo and Zipkin (2007) study a model in which customers are provided with information and make decisions based on their expected waiting times, conditional on the provided information. Three types of information are studied: (i) no information, (ii) queue-length information, and (iii) exact waiting-time information (in systems in which such information is available). The authors provide examples in which accurate delay information improves the system performance, as well as examples in which it does not. Jouini et al. (2011) is the first paper to consider delay announcements in a multiclass setting with priorities. All of these models assume that the information provided to the customer is truthful and that the customer believes the information, even if he will be better off, in terms of his own utility, disregarding it.

1.2. Classical Cheap Talk Game

In this section, we provide an overview of the cheap talk game introduced in Crawford and Sobel (1982) and compare the model to the one studied in this paper. The classical cheap talk game is played between a *sender* who has some private information and a *receiver* who takes the payoffs-relevant actions. The game proceeds as follows: The Sender observes his state/type, which we shall denote by Q. The Sender then sends a signal (or a message) denoted by $m \in \mathcal{M}$, where $\mathcal{M}$ denotes the set of all signals. The receiver, who cannot observe the state/type Q but does know its distribution, processes the signal using Bayes rule and chooses an action y that determines the player's payoff. Both the sender and the receiver obtain utilities that depend on: (a) the action taken by the receiver, y; and (b) the state/type Q. Two distinctive features of their model should be emphasized: first, the state/type, although random, is static: once realized it does not change over time. Also, the distribution of this state/type is exogenous and independent of the actions of the players.

A variety of papers study mixed-motive economic interactions involving private information and the impact of cheap talk on the outcomes. In political context, cheap talk has been studied in multiple papers, including Austen-Smith (1990). Recently, Chen et al. (2008) study questions regarding equilibrium selection in cheap talk games.

Driven by the applications in service operations, our model has two key features: first, the game is played with multiple receivers (customers) whose actions have externalities on other receivers; and second, the stochasticity of the state/type (i.e., the state of the system) is not exogenously given, but is determined endogenously. Indeed, the private information in this model (i.e., the queue length) is driven by the system dynamics, which in turn depend on the equilibrium strategies of both the firm and the customers. In particular, in our model the customers' actions are both payoff relevant and system-dynamic relevant. As we will show, the multiplicity of receivers with externalities as well as the endogenization of the uncertainty impact both the nature of the communication between the firm and the customers and the outcome for the various players. Hence, although the framework used in this paper partially echoes the classical cheap talk model, the abovementioned distinguishing features lead to conceptually different results. The way these features separate our results from those of Crawford and Sobel (1982) will be discussed in §EC.1.

2. Model

We consider a service provider modeled as an $M/M/1$ queue. Customers arrive to the system according to a Poisson process with rate λ. Service times are exponentially distributed with mean $1/\mu$.[3] We assume that $\lambda < \mu$. We assume that all customers are ex ante symmetric: customers obtain a value R if they are served, and incur a waiting cost that is proportional to the time spent in the system, with a unit waiting cost of c.[4] Thus, the utility function of a customer is:

$$U(y, w) = \begin{cases} R - cw & \text{if } y = \text{"join,"} \\ 0 & \text{if } y = \text{"balk,"} \end{cases} \qquad (1)$$

where y is the decision made by the customer and w denotes his sojourn[5] time in the system. Throughout the paper, we shall assume that $R > c/\mu$. This assumption ensures that in the absence of delays, the service is beneficial to the customer, on average. Clearly, if $R < c/\mu$, no customer will join regardless of the system announcements. Throughout the paper we assume that involuntary blocking by the firm is prohibitively expensive.

The firm's profit from a customer served is $v > 0$. The firm incurs a holding cost $h(w)$ per customer who waits w in the system. This holding cost is incurred by the firm due to the customer waiting in the system. This cost constitutes, among other factors, the cost associated with loss of goodwill, the actual cost of holding the customer, and in some settings, the opportunity cost associated with the customer not being able to generate revenues. Disney's theme parks, for example, incur two costs due to waiting: the opportunity cost of having a customer standing in line without the ability to spend money, and the wages of the entertainment staff that is in charge of alleviating the pain of waiting. Thus, the firm's profit from a customer is $v - h(w)$ for a customer who received service from the firm and spent w in the system and is 0 for the customer who decides not to join. Here, h is a nonnegative and convex-increasing function. We assume that $v - \mathbb{E}h(X) > 0$, where X is an exponential random variable with mean $1/\mu$ to avoid cases in which even customers that arrive without any waiting are not profitable. We assume that customers who join the system are served in a first-come-first-served manner.

Information Provision and Sequence of Events. The static structure described above is assumed to be common knowledge. The real-time state of the system, corresponding to the number of customers waiting in queue, is the firm's private information. The evolution of this state will depend on the equilibrium strategies of both the provider and the customers. The interaction between the customers and the firm is described as follows: upon a customer's arrival to the system, the firm, which can observe the queue length, makes an announcement to this customer. The customer reacts to that signal by deciding whether to join or not join (balk). In making the decision, the customer uses the information, denoted by I, that he can infer from the firm's announcement regarding the current state of the system and chooses an action that maximizes his expected utility. Therefore, the customer will join if and only if $R \geqslant c\mathbb{E}(w \mid I)$. The type of signals that are used by the firm, and the way the customers interpret these in making their decisions, are described precisely in the next section.

Note that the customers' and the service provider's incentives are not completely misaligned: both prefer short waiting times, which result in higher utility for the customer and higher profits for the service provider. At the same time, we observe that these incentives are not perfectly aligned, and this would lead to the equilibria described in the next section. We refer the reader to Farrell and Rabin (1996) for a discussion of settings in which incentives are perfectly misaligned. We formally define the concept of misalignment for our setting in §3.

3. Main Results

3.1. Problem Formulation

In practice, one observes different types of messages that are conveyed to the customers. These messages could provide tangible information such as expected wait in the queue or the position of the customer in the queue. However, in some cases the message may only have intangible information such as the congestion level or volume of calls being high or low. We propose a framework that allows us to study the provision of unverifiable information using a unified approach that does not assume a specific type of the announcement. The framework accounts for the following key features: (a) the state of the world changes dynamically; (b) the customer cannot verify the information provided by the firm; and (c) the customers would process any information provided to them by the firm and base their action on it. The structure that we lay out next covers the suggested framework that allows us not to restrict attention to any specific type of announcements. In particular, it allows us to account for announcements of the variety mentioned above.

In this section we formally define the game between the service provider and the customers. The equilibrium concept we employ is one of Markov-perfect Bayesian Nash equilibrium (MPBNE), which in this case is simply a Nash equilibrium in the decision rules that relate agents' actions to their information and to the situation in which they find themselves, while allowing for actions to depend only on payoff-relevant histories.

To formally define the game, let $\mathcal{M} = \{m_0, m_1, m_2, \ldots\}$ represent the discrete set of feasible signals that the firm can provide to the customer. For example, the set of feasible messages may contain nonnegative integers, rational numbers, concrete messages (those with some tangible information), or nonquantifiable messages (those with only intangible information), among others. We represent the signaling rule by a function $g\colon \mathbb{Z} \mapsto \mathcal{M}$, where $g(q) = m$ if the firm uses the signal m when the queue length is q.

Recall that customers are indistinguishable and their strategies are ex ante symmetric, both in their interpretation of the signals and in their actions. Let $y\colon \mathcal{M} \mapsto \{0, 1\}$ denote the strategy of the customer, where $y(m)$ is the probability that a customer joins when the firm signals m. Consequently, we interpret $y(m) = 1$ as a "join" decision and $y(m) = 0$ as a "balk" decision. We initially restrict ourselves to pure strategies. These provide valuable insights into the structure of information transmission, while keeping the model simple. We will relax this restriction in §EC.6.

Under a signaling rule g and decision rule y, the queue evolves as a birth and death process on the positive integers where, for any state q, the birth rate is $\lambda y(g(q))$ and the death rate is $\mu \mathbb{1}\{q > 0\}$. Here, we use the notation $\mathbb{1}\{\cdot\}$ to denote the indicator function. As we restrict attention to $\lambda < \mu$, this birth-death process has a unique steady-state distribution. Let $p_q(y, g)$ be the steady-state probability of having q customers in the system. We let $Q_{y,g}$ be a random variable with the steady-state distribution of the number of customers in the system.

The requirements of a Markov-perfect Bayesian Nash equilibrium in our context are explained as follows: Given a signaling rule for the system, customers with an action rule that dictates joining the system when the signal is m will not deviate from this rule if their expected conditional utility from joining the system, given by $\mathbb{E}[R - c(q+1)/\mu \mid g(q) = m]$, is positive. Given the customer's action rule y, the firm will not deviate from its signaling rule g if it maximizes its steady-state profit, i.e., if g solves an appropriate Markov decision process (MDP, see below) with respect to the action rule y. This discussion is formalized in the following definition.

Definition 3.1. (Markov-perfect Bayesian Nash equilibrium MPBNE). We say that the signaling rule $g(\cdot)$ and the action rule $y(\cdot)$ constitute a Markov-perfect Bayesian Nash Equilibrium (MPBNE),[6] if they satisfy the following conditions:

1. For each $m \in \mathcal{M}$, we have

$$y(m) = \begin{cases} 1 & \dfrac{\sum_{\{q:\, g(q)=m\}} [R - c((q+1)/\mu)] p_q(y, g)}{\sum_{\{q:\, g(q)=m\}} p_q(y, g)} \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

2. With $f(j) = v - \mathbb{E}[h(W(j+1))]$, where $W(j+1)$ is the time spent by a customer who joins the system with j customers, there exists constants $J_0, J_1, \ldots,$ and γ that solve the following set of equations:

$$\begin{aligned} J_0 &= \max_{m\in\mathcal{M}} \left\{ \frac{f(0)y(m) - \gamma}{\lambda} + J_0(1 - y(m)) + J_1 y(m) \right\} \\ &= \frac{f(0)y(g(0)) - \gamma}{\lambda} + J_0(1 - y(g(0))) + J_1 y(g(0)) \\ J_q &= \max_{m\in\mathcal{M}} \left\{ \frac{f(q)y(m) - \gamma}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} J_{q-1} \right. \\ &\qquad \left. + \frac{\lambda}{\lambda + \mu} (J_q(1 - y(m)) + J_{q+1} y(m)) \right\} \end{aligned}$$

$$= \left\{ \frac{f(q)y(g(q)) - \gamma}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} J_{q-1} + \frac{\lambda}{\lambda + \mu} (J_q(1 - y(g(q))) + J_{q+1} y(g(q))) \right\}. \quad (2)$$

In the above definition of MPBNE, the first condition uses the Bayesian rule for the customer based on the signaling function g to determine whether to join or balk. The second condition states that the composite function $y \circ g$ solves the firm's MDP, which is closely related to the *admission control problem* in the MDP literature. In the optimality equations (EC.8), the constant γ represents the long-run average profit (referred to as the gain in MDP literature) made by the firm under optimal policy, and constants $J_0, J_1, \ldots$ represent the *relative cost or the bias* for states $0, 1, \ldots$. Thus, we require that y, g form an equilibrium only if g is bias optimal. Our setting is very close to Haviv and Puterman (1998), which characterizes the bias-optimal policy for an $M/M/1$ queue.

Discussion of the Customers' Ability and Beliefs. The above definition takes the approach that the customer is rational, has full-state information about the system parameters, and is capable of forming beliefs about the system performance given a set of strategies. The assumption that the customer has full-state information about the arrival rate is relaxed in §EC.2. In that section, we study a model in which the firm has private information on the arrival rate, and show that all the results of the above model extend to that setting as well. We also relax the assumption that customers are capable of forming beliefs on the system performance given the strategies of all players in the game, i.e., the firm and the customers. This is done in §EC.3.1 by conducting a numerical study in which customers form their action rules through repeated interactions with the firm. We numerically show that such an iterative mechanism would lead the system to the equilibria characterized in this section. Another implicit assumption made in this definition is that customers make a decision only regarding joining versus balking, and do not leave the system after joining, but before receiving service. This assumption is relaxed in Appendix EC.4. There, we show that as long as the customer experiences linear waiting costs, even if customers are allowed to update their belief about the system and renege the queue, the equilibria characterization remains unchanged.

All of these relaxations allow us to test the robustness of the assumptions made in the model above, and show that the resulting characterizations continue to hold when the assumptions are relaxed.

One of the goals of this paper is to identify the conditions under which a firm can credibly communicate unverifiable information. Our litmus test for such credibility will be the existence, or lack thereof, of an equilibrium with influential cheap talk. When such an equilibrium exists, it means that the firm can induce, by virtue of using at least two distinct messages, two distinct actions. An equilibrium with influential cheap talk is formally defined as follows:

Definition 3.2. We say that an equilibrium (y, g) has influential cheap talk if there exist two distinct signals m_i, m_j, that are used with positive probability in equilibrium, i.e., $\sum_{\{q:\, g(q)=m_i\}} p_q(y, g) > 0$ and $\sum_{\{q:\, g(q)=m_j\}} p_q(y, g) > 0$ such that $y(m_i) \neq y(m_j)$.

As in most cheap talk games, the equilibrium, even if it exists, is never unique. This stems from the fact that one can relabel the signals or introduce new signals and still have the same outcomes and payoffs for the players.

Definition 3.3 (DOE). We say that two MPBNE (y_1, g_1) and (y_2, g_2) are dynamics and outcome equivalent (DOE), if $y_1(g_1(q)) = y_2(g_2(q))$ for all $q = 0, 1, \ldots$.

3.2. Characterizing Key Equilibria

Although Definition 3.1 of the pure-strategy MPBNE in the previous section is complete, it is not directly amenable for further analysis. Thus, our first step toward characterizing the equilibria is to show that any pure-strategy MPBNE can be described by a threshold level.

Proposition 3.1. *For any pure-strategy equilibrium (y, g) with influential cheap talk, there exists $N := \inf\{q\colon y(g(q)) = 0\}$, which is finite such that $y(g(q)) = 0$ for all $q \geqslant N$ and $y(g(q)) = 1$ for all $q < N$.*

The next proposition shows that one can always reduce a pure-strategy influential equilibrium to an equilibrium where the firm uses exactly two signals.

Proposition 3.2. *Let the pair (y, g) be a pure-strategy MPBNE with influential cheap talk and let $N := \inf\{q\colon y(g(q)) = 0\}$. Then the pair $(\tilde{g}(\cdot), \tilde{y}(\cdot))$ given by*

$$\tilde{g}(q) = \begin{cases} m_1 & q \leqslant N, \\ m_0 & \textit{otherwise}, \end{cases} \qquad \tilde{y}(m) = \begin{cases} 1 & m = m_1, \\ 0 & \textit{otherwise}, \end{cases} \quad (3)$$

forms a MPBNE with the same firm profit and customer utility. Further, (y, g) and $(\tilde{y}, \tilde{g})$ are DOE.

Proposition 3.1 implies that instead of studying the customers' actions and the firm's announcement in each state of the system (i.e., queue length), it suffices to focus on the threshold queue length, below which the customer's action is "join," and above which it is "balk." Note that every equilibrium characterized using Proposition 3.2 requires that the equilibrium has influential cheap talk and thus that the constant $N = \inf\{q\colon y(g(q)) = 0\}$ is finite. There may exist equilibria where the constant N is infinite. We shall discuss these in §3.3.

Definition 3.4. We say that the threshold q induces a pure-strategy MPBNE if the pair $(\tilde{g}(\cdot), \tilde{y}(\cdot))$ given by (3) forms an MPBNE.

Before delving into the analysis of the model and the characterization of the equilibrium, we take a step back and develop intuition into the possible regimes and outcomes. In order to do that, and knowing that we can restrict attention to threshold-based equilibria, we introduce two important threshold levels: the first, q^*, denotes a threshold value above which a customer *will not* join if he has *full-state information* of the state of the system, and below which he *will join*. The second threshold level, $\hat{q}$, is motivated by the service provider's point of view. It denotes the threshold level below which the service provider would like the customers to join, and above which she would like them to balk, if she had *full control* of their actions.

Full state information. We define q^* to be the threshold above which customers do not obtain positive utility, in expectation, given full queue-length information, that is,

$$q^* = \left[\frac{R\mu}{c}\right], \tag{4}$$

where $[\,\cdot\,]$ is the bracket function; i.e., q^* is the largest integer not exceeding $R\mu/c$. Note that this threshold pertains to the marginal customer who decides to balk. We will refer to this as the first-best from the customer's perspective, because this maximizes the utility for the individual (self-interested) customer. As shown in Naor (1969), this threshold, which is based on self-optimization (to use the terminology of Naor 1969), does not maximize the overall expected utility of the customer population. For each $q \geqslant 0$ we let $y_{\mathrm{FI}}(q)$ be the probability of a customer joining the system when there are q customers in the system and customers have full-state information.

Full control. From the service provider's point of view, deciding on a threshold level amounts to fixing the waiting space in a $M/M/1/k$ queueing system. For each value of k, the expected number of customers joining the queue per unit of time equals $\lambda(1-\rho^k)/(1-\rho^{k+1})$ where $\rho = \lambda/\mu$. Let $\hat{q}$ denote the optimal waiting space. Thus, $\hat{q}$ solves the following *full-control optimization problem*:

$$\hat{q} = \arg\max_{k\in\mathbb{Z}_+} \lambda \frac{1-\rho^k}{1-\rho^{k+1}}[v - \mathbb{E}[h(W_k)]], \tag{5}$$

where W_k is the steady-state sojourn time of the customers who join the $M/M/1/k$ queue. The following proposition, which relies on Knudsen (1972), shows that such a maximizer exists, and characterizes properties of the objective function in (5).

Proposition 3.3. *The function* $\Pi(\cdot)$ *defined by*

$$\Pi(k) := \lambda \frac{1-\rho^k}{1-\rho^{k+1}}[v - \mathbb{E}[h(W_k)]], \quad k = 1, 2, \ldots$$

is unimodal in k, *i.e., there exists* $k^* \in \{1, 2, \ldots, \infty\}$ *such that the function* $\Pi(k)$ *is strictly increasing for* $k < k^*$ *and strictly decreasing for* $k \geqslant k^*$.

Based on the optimal threshold $\hat{q}$ we define $y_{\mathrm{FC}}(q)$ for each $q \geqslant 0$ to be the probability of a customer being admitted to the system when there are q customers in the system and the firm has full control.

Our equilibrium analysis will be based on the level of misalignment between the firm and the customers. To quantify this intuitive concept of misalignment, we make the following definition:

Definition 3.5. We define ϕ as the *misalignment* between the firm and the customers as follows:

$$\phi = \sum_{j=0}^{\infty}(y_{\mathrm{FI}}(j) - y_{\mathrm{FC}}(j)).$$

For our game, the misalignment ϕ equals the difference in thresholds q^* and $\hat{q}$, i.e., $\phi = q^* - \hat{q}$. This misalignment is related to the bias in Crawford and Sobel (1982). Using this misalignment measure ϕ, which is based on unilateral optimization under full-state information to the customers and the full control of the service provider, respectively, we can identify three regions. Each of these regions results in a different conflict of interest, and thus in different equilibria and outcomes for the customers and the firm. We will initially outline the key equilibrium in each of the three regions and explain the intuition behind them. We provide a formal statement in Proposition 3.4. The three cases are as follows.

1. *Complete alignment*: $\phi = q^* - \hat{q} = 0$ (*zero misalignment*). In this region, the interests of the two parties are completely aligned, and thus a pure-strategy MPBNE is as follows: The firm gives two signals: (i) the first for low congestion, which can be denoted as "Low." This signal is announced if the queue length is below q^*. (ii) A second signal denoted by "High," which indicates high congestion, and is given when the queue length exceeds q^*. Thus, we have $g(q) =$ "Low" if $q < q^*$ and $g(q) =$ "High" if $q \geqslant q^*$; the customer joins the queue when he/she receives the signal "Low" and balks otherwise, i.e., y("Low") = "join", y("High") = "balk".

As stated before, this is the key equilibrium in this region; however, this need not be the unique pure-strategy MPBNE. As we will show in §4, there are multiple equilibria in this model. There, we will, however, be able to show that even as the equilibria become progressively more informative, it is outcome and dynamics equivalent to the one described above.

2. *Overly patient customers*: $\phi = q^* - \hat{q} > 0$ (*strictly positive misalignment*). In this region, if customers are endowed with full-state information, they would like to join the system even when the service provider would like them to balk (if she had full control). Thus, we use the term "overly patient" to emphasize the fact that, in this case, customers are willing to join a more congested system than what the firm would like. Specifically, when the queue length is between $\hat{q}$ and q^*, the customers would like to join, whereas the firm would like them to balk.

We will show that there is no threshold that is immune to defection by both the customers and the firm, and consequently that there is no pure-strategy MPBNE with influential cheap talk. Indeed, for pure-strategy MPBNE to exist, the firm should be able to signal "High" and customers who receive "High" should balk. The only threshold immune to profitable deviation by the firm is $\hat{q}$. Given that under any pure-strategy MPBNE the customers respond to "High" by balking, a profitable deviation for the firm from any other candidate threshold is to announce "High" when the state of the system is equal to and greater than $\hat{q}$. The customers, however, know that $\hat{q} < q^*$ so that $\hat{q}$ cannot induce an equilibrium: an arriving customer that receives the signal that instructs him to "balk," can deviate from the prescribed equilibrium strategy by joining; the customer will then earn positive utility (because the *only* state in which he can receive such a signal is on the threshold itself, which is, by assumption, below q^*), and thus *detect* that such a deviation is profitable—hence ruling out the possibility of a pure-strategy MPBNE with influential cheap talk.

3. *Impatient customers*: $\phi = q^* - \hat{q} < 0$ (*strictly negative misalignment*). In this region, the service provider would like the customers to join a more congested system than the one they wish to join. Specifically, when the queue length is between q^* and $\hat{q}$, the firm would like the customers to join, whereas the customers would like to balk. In order to study this region, we define $F(q)$ to be the customer's expected utility if he finds q customers in the system upon arrival and decides to join the queue; i.e., $F(q) := R - c(q+1)/\mu$. We define for $\ell < k$,

$$G(\ell, k) = \sum_{q=\ell}^{k-1} p_q^k F(q), \tag{6}$$

where $p_q^k := (\rho^q(1-\rho))/(1-\rho^{k+1})$ is the steady-state probability of q customers in an $M/M/1/k$ queue. Here, $G(l, k)$ is interpreted as the average utility of a customer joining the $M/M/1$ queue given that the number of customers in the system is between l and k. Then, we have two subcases to consider:

(a) $G(0, \hat{q}) \geqslant 0$: if the firm announces "Low" when the queue length is below $\hat{q}$ and "High" otherwise, the customer would like to join when they get the "Low" signal, because their expected utility is positive (because $G(0, \hat{q}) > 0$). Further, because in equilibrium "High" would be announced only when the queue exactly equals $\hat{q}$, the customer would balk because they know that $q^* < \hat{q}$. Alternatively, the customer can experience the negative utility, and thus will not follow the prescription of the firm. This is optimal for the firm and also describes our pure-strategy MPBNE for this setting. Thus, the firm is capable of achieving its first-best profits and operates as if it has full control over the customers' decisions.

(b) $G(0, \hat{q}) < 0$: In this case there is no pure-strategy MPBNE with influential cheap talk. For pure-strategies equilibrium with influential cheap talk to exist, the firm should be able to signal "Low" and customers who receive "Low" should join. As in case 2, the only threshold immune to profitable deviation of the firm is $\hat{q}$. However, the customers know that $\hat{q} > q^*$; thus, threshold $\hat{q}$ cannot constitute an equilibrium; an arriving customer that receives a signal that instructs him to "join" would receive negative expected utility, and thus can deviate from the prescribed equilibrium strategy by balking and obtaining zero utility. This rules out the possibility of a threshold-induced pure-strategy MPBNE.

The intuition is simple: if the expected utility of the customers under an $M/M/1/\hat{q}$ system, as given by $G(0, \hat{q})$, is positive, they will have no incentive to deviate. Any deviation here will lead to zero utility for the customers. If, on the other hand, their utility is negative, they would be better off by not joining at all. Consequently, the threshold $\hat{q}$ cannot induce a pure-strategy MPBNE. Further, no other threshold is immune to profitable deviation on the firm's part. Thus, in case 3(b) there does not exist a pure-strategy MPBNE with influential cheap talk. We emphasize, however, that in case 3(a) the customers can be lured to join the system even in states in which they obtain negative expected utility as long as their utility averaged over all states in which they join is positive.

We turn now to the formal statement of the equilibria we have discussed thus far. To this end, we let Π_{FI} and Π_{FC} be the firm's profit under full-state information and full control, respectively. Let U_{FI} and U_{FC} denote the expected utility of the customers under full-state information and full control, respectively. As discussed before, Π_{FC} is the first-best profit for the firm and U_{FI} is the first-best utility for the customer. The next proposition summarizes the above observations and also compares the firm's profit and expected customer utility under the different equilibria.

Proposition 3.4. 1. *If $q^* = \hat{q}$, then q^* induces a pure-strategy MPBNE. Under this equilibrium the firm's profit equals Π_{FC} and the expected utility of the customers is U_{FI}.*

2. *If $q^* > \hat{q}$, there is no finite q that induces a pure-strategy MPBNE.*

3. *If $q^* < \hat{q}$, then there are two cases*:

(a) *If $G(0, \hat{q}) \geqslant 0$, $\hat{q}$ induces a pure-strategy MPBNE. Under this equilibrium, the firm's profit equals Π_{FC} and the expected utility of the customers is U_{FC}.*

(b) *If $G(0, \hat{q}) < 0$, there is no finite q that induces a pure-strategy MPBNE.*

First, it is important to note that all pure-strategy equilibria that have influential cheap talk are neologism proof. Neologism proof merely means that given an equilibrium, the firm (i.e., the sender of the information) does not have incentives to use words that are not used in equilibrium but may have a "focal meaning" for the customer (i.e., the receiver) and are profit enhancing. Having said that, it is important to recall that there are multiple pure-strategy equilibria in the basic game that can be generated by relabeling the messages the firm provides. However, when pure-strategy equilibria exist, all equilibria result in the same profit for the firm, the same average utility for all customers, and the same system dynamics. That is, all equilibria are DOE (dynamics and outcome equivalent). The main implication of showing that all equilibria are DOE is that there is a unique outcome for the firm and the customers.

Social Optimization. By setting $h(w) = cw$ and $v = R$ (where c is the disutility experienced by customers due to waiting and R is the value obtained by the customer from service), the system manager's problem amounts to maximizing the social welfare in the system. It follows from Naor (1969) that $\hat{q} \leqslant q^*$. In this case, the solution of the first-best and full-control problem lies in Region I or in Region II. Thus, either the customers are fully aligned and the social planner can announce the true state of the system, or there is no pure-strategy equilibrium with influential cheap talk. The firms' myopic optimization (based on each customer separately) may seem to be aligned with the customer's utility-maximizing problem. However, due to the externalities and the fact that $\hat{q} \leqslant q^*$, influential communication between a social planner and customers is impossible in pure strategies.

To summarize our findings so far: we have identified three regions, each with a different equilibrium behavior. We observed that a pure-strategy MPBNE with influential cheap talk exists only if the firm's and the customers' incentives are perfectly aligned or if the customers are mildly impatient. Further, we prove that any MPBNE with influential cheap talk is DOE with these equilibria that have only a two-signal language.

Next, we would like to understand if an MPBNE exists with noninfluential cheap talk. We shall study the following three types of equilibria: First, in §3.3 we show the existence of a *babbling* equilibria, where the customers disregard any information that the firm provides and hence render the cheap talk noninfluential. Next, in §EC.6, we extend the definition of MPBNE to allow customers to randomize their actions. We characterize the *mixed*-strategy MPBNE with *noninfluential* cheap talk as well as the one with *influential* cheap talk. Lastly, we examine whether or not firms can provide more detailed information regarding the state of the system. As we have already seen in this section, such revelation cannot change the outcome or the dynamics in equilibrium (if such revelation leads to an MPBNE with influential cheap talk). To this end, in §4 we will define a notion of *most informative* equilibria which characterizes an equilibria where the information provided to the customer cannot be "refined" by the firm in a credible manner.

3.3. Babbling Equilibrium

The equilibria constructed above are based on a signaling rule with two signals. In practice, however, there are many service providers that share no information whatsoever with the customer, whether it is direct information or one that is implicit in the type of recorded music heard while waiting. Are these systems, where no meaningful information is transmitted, in equilibrium? Furthermore, is it possible to have an equilibrium in which the firm does provide information, but due to the lack of its credibility, customers do not follow the firm's recommendation? It turns out that such an equilibrium may indeed exist in our setting. When it does exist, it is referred to as a "babbling" equilibrium (see Farrell and Rabin 1996) to denote that the information transmitted is uncorrelated with the state of the system, and any information provided is treated by the customers as meaningless. We first provide a formal definition of babbling equilibrium. In the following, let $Q_{y,g}$ be a random variable with the steady-state distribution of queue length under an MPBNE (y, g).

Definition 3.6. We say that a pure-strategy MPBNE equilibrium (y, g) is a babbling equilibrium if the random variables $g(Q_{y,g})$ and $Q_{y,g}$ are independent and $y(m_i) = y(m_j)$ for all $m_i, m_j \in \mathcal{M}$.

Definition 3.6 states that the firm provides signals that are uncorrelated with the state, and that the customers are disregarding any information that is provided by the firm, either because the messages provide no information (i.e., they are identical) or because the firm lacks credibility. In the setting of Crawford and Sobel (1982), such an equilibrium always exists, and it is sometimes the only one. In contrast, in our model, such an equilibrium exists in pure strategies only under the conditions that are stated in Proposition 3.5 below.[7] Using the definition of pure-strategy MPBNE, it is clear that for an equilibrium to be a pure-strategy babbling equilibrium it must be the case that for each $m_j \in \mathcal{M}$, $y(m_j)$ must either be 0 or 1. However, it is easy to see that $y(m_j) = 0$, for all $m_j \in \mathcal{M}$ cannot form an equilibrium. Thus, in our model, a "babbling equilibrium" exists in pure strategies if, in the absence of information, all customers join. If indeed all customers join, the resulting queueing system is an $M/M/1$ queue (i.e., with infinite waiting space), in which case the average steady-state waiting time is $E[W] = 1/(\mu - \lambda)$ and customers join if and only if $R \geqslant cE[W]$, i.e., if $R \geqslant c/(\mu - \lambda)$. To provide rigorous characterization, we have the following result.

Proposition 3.5. *There exists a pure-strategy babbling equilibrium if and only if* $R \geqslant c/(\mu - \lambda)$. *Further, if* $q^* < \hat{q}$ *and* $G(0, \hat{q}) < 0$, *i.e., Proposition* 3.4, *Case* 3(b), *there does not exist a pure-strategy babbling equilibrium.*

Whenever a babbling equilibrium exists, the firm's profit under such an equilibrium is denoted by Π_{NI}, and the customer's utility under such an equilibrium is denoted by U_{NI}. (NI stands for noninfluential.) The following proposition shows that even though a babbling equilibrium may exist, the firm's profit under this equilibrium is dominated by the firm's profit under the two-signal equilibria described in §3.2. Further, the customer's expected utility is lower under the babbling equilibrium as compared to his utility under the two-signal equilibrium.

PROPOSITION 3.6. *Assume that both a pure-strategy babbling equilibrium and an equilibrium with influential cheap talk exist, then,* $U_{\text{NI}} < U_I$ *and* $\Pi_{\text{NI}} \leqslant \Pi_I$.

Proposition 3.6 underscores the value of communication. Even though an equilibrium with noninfluential cheap talk does exist, both the service provider and the customers are better-off when moving from a babbling equilibrium to an equilibrium with influential cheap talk (if such equilibria exist). This communication does not necessarily maximize the customer's overall expected utility, but it does improve it. The rationale behind Proposition 3.6 is as follows: Naor (1969) shows that when customers are self-interested and can observe the length of the queue prior to joining, their optimal threshold, q^*, is higher than what the social optimum prescribes, but *it is finite.* In our setting, we observe that the threshold queue length for the two-signal equilibrium is at least as high as q^*. Further, for the babbling equilibrium, when it exists, the threshold is infinite. Thus, using information improves the customer's overall expected utility when compared to settings where the service provider is giving no information. Note that this improvement in utility and profit is achieved even though the information is unverifiable.

At this point, we remind the reader that in the region where the customers are overly impatient (region III(b)), there is no pure-strategy MPBNE, either influential nor babbling. Without expanding the strategy set for the customers or the firm, it is unclear how the system would behave in this parameter regime. To alleviate this unresolved matter, we next explore the possibility of a MPBNE in which the customers are allowed to randomize.

3.4. Does Randomization Enable Communication?

In this section we summarize the results of Appendix EC.6, where we relax the assumption that the customer, given an announcement, either joins with probability 1 or balks with probability 1, i.e., we allow the customers to randomize. We discuss the impact of allowing the firm to randomize in Appendix EC.5.

We show that in this setting it is never optimal for the firm to use the same signal for two states and use some other signal for a state in-between. Thus, a firm will never provide the same signal on two disjoint intervals and provide a different signal in-between.

We have shown in §3.3 that babbling equilibria need not exist in pure strategies. When customers are allowed to randomize, however, such equilibria always exist. We characterize the details of such an equilibria in Appendix EC.6.1.

In §EC.6.2 we show that in addition to the babbling equilibria, there may exist more *informative* MPBNE in mixed strategies. We start by showing that it suffices to consider equilibria with certain characteristics. Specifically, in the following proposition we show that if customers randomize, they will do so only within an interval. Furthermore, we show that balking with probability 1 in some states and joining with probability 1 in other states cannot coexist. Thus, there are only two possible types of two-signal mixed-strategy MPBNE in which randomization is used. The two types can be described as follows: The firm announces "High" and "Low" based on the threshold q_{mix}; the customers then react as follows: (a) in the first type of MPBNE, which we shall refer to as *Join or Randomize* equilibria, the customers who receive "Low" join the system and the customers who receive "High" join the system with probability $\theta \in (0, 1)$; (b) in the second type of MPBNE, which we shall refer to as *Randomize or Balk* equilibria, the customers who receive "Low" join the system with probability θ, and the customers who receive "High" balk. Both of these equilibria are completely defined by two parameters: the threshold q_{mix} used by the firm for signaling and the randomization parameter θ. We provide a characterization of the pair (q_{mix}, θ) which induces the mixed-strategy MPBNE in Appendix EC.6.3.

So far, we have characterized and proved the existence (or nonexistence) of two-signal communication between the firm and its customers that form an MPBNE. We next turn to the question regarding the existence of additional equilibria.

4. Most Informative Equilibria

In practice, firms frequently use announcements regarding the congestion in the system or the volume of calls that are equivalent to the two-signal equilibrium. In this section we show that other types of equilibria, in which more information is provided to the customers, are also possible. Although these equilibria may be different in terms of the firm's announcements, they are equivalent in terms of the customers' actions, their utility, and the firm's profits.

To formally examine this question, we introduce the notion of *most informative equilibria*.

DEFINITION 4.1. Consider an ordered set $\mathbb{S} = \{a_0 = 0, a_1, a_2, a_3, \ldots, a_J = \infty\}$, where $a_i \in \mathbb{Z}_+$.[8] Consider the signaling rule defined as follows: $g(q) = m_j$ for all $q \in [a_{j-1}, a_j)$, for all $1 \leqslant j \leqslant J$. We say that the set $\mathbb{S}$ induces a mixed-strategy MPBNE if there exists an action rule $y(n) \in [0, 1]$ for all $n \in \mathbb{N}$, such that $(y(\cdot), g(\cdot))$ forms a mixed-strategy MPBNE. Further, if there does not exist an action rule $y(\cdot)$, such that $(y(\cdot), g(\cdot))$ form an MPBNE, we say that the ordered set does not induce a mixed-strategy MPBNE.

DEFINITION 4.2. We say that an ordered set $\mathbb{S}$ induces the *most informative* pure-strategy MPBNE if the following holds: (a) $\mathbb{S}$ induces a mixed-strategy MPBNE in the sense of Definition 4.1, and (b) every ordered set $\mathbb{S}' \supsetneq \mathbb{S}$ does not induce a mixed-strategy MPBNE.

For the two types of mixed equilibria defined in the §EC6.2, we put:

$q_{\text{mix}}^{\text{JR}} = \max\{q$: there exists θ such that (q, θ) induces a Join/Randomize-type mixed-strategy MPBNE.

$q_{\text{mix}}^{\text{RB}} = \min\{q$: there exists θ such that (q, θ) induces a Randomize/Balk-type mixed-strategy MPBNE.

The following result describes the *most informative* pure-strategy MPBNE.

Proposition 4.1. 1. *If $q^* = \hat{q}$, the set $\mathbb{Z}_+$ induces the most informative MPBNE. In this equilibrium the firm's profit equals Π_{FC}, and, the expected utility of the customers is U_{FI}.*

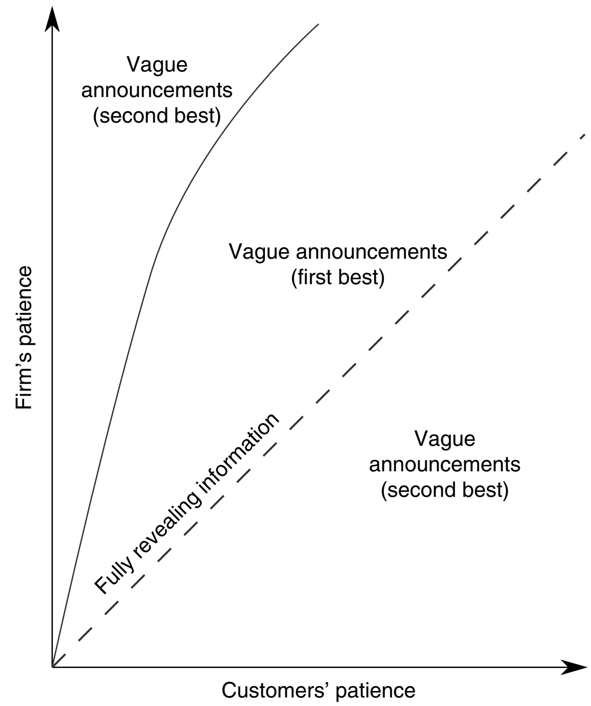
2. *If $q^* > \hat{q}$, we have the following cases*: (a) *If there exists a mixed-strategy MPBNE of type Join/Randomize with parameters (q_{mix}, θ), then the set $\{0, 1, \ldots, q_{\text{mix}}^{\text{JR}}, \infty\}$ induces the most informative equilibrium*; *and* (b) *if there exists a mixed-strategy MPBNE of type Randomize/Balk with parameters (q_{mix}, θ), then the set $\{0, q_{\text{mix}}^{\text{RB}}, q_{\text{mix}}^{\text{RB}} + 1, q_{\text{mix}}^{\text{RB}} + 2 \ldots, \infty\}$ induces the most informative equilibria.* (c) *If there does not exist any informative mixed-strategy MPBNE, then the most informative equilibria is the babbling equilibria.*

3. *If $q^* < \hat{q}$, either*: (a) $G(0, \hat{q}) > 0$, *in which case the set $\{0, 1, 2, \ldots, \bar{q}, \hat{q}, \hat{q}+1, \ldots, \infty\}$ induces the most informative MPBNE where $\bar{q} = \arg\max\{q: G(q, \hat{q}) > 0\}$. Under this equilibrium the firm's profit equals Π_{FC}. Further, the expected utility of the customers is U_{FC}.* (b) *If $G(0, \hat{q}) \leqslant 0$, and the mixed-strategy MPBNE of type Randomize/Balk has parameters (q_{mix}, θ), then the set $\{0, q_{\text{mix}}^{\text{RB}}, q_{\text{mix}}^{\text{RB}} + 1, q_{\text{mix}}^{\text{RB}} + 2, \ldots, \infty\}$ induces the most informative equilibria.*

One consequence of Proposition 4.1 is that if the firm is not perfectly aligned with the customers, it will always maintain some level of *intentional vagueness*. Recall that by intentional vagueness we mean that under the most informative equilibrium, the firm provides the same signal on multiple states of the system. This occurs, for example, in case 3(a) where the firm is intentionally vague when the state of the system is between $\bar{q}$ and $\hat{q}$ so that customers are lured to the system. If the firm provides any further information in these states, the firm would forgo some of its profits, and this cannot be sustained in equilibrium. It is important to note that even though the most informative equilibria could have countably infinite signals (under Randomize/Balk and full alignment), only a finite subset of these are transmitted with strictly positive probability.

In discussing the opportunities for the service provider to lure a customer to take an action against his best interest, we need to distinguish between two types of strategies: (i) misrepresentation of the state of the system, and (ii) intentional vagueness. The latter was shown in Proposition 4.1 to be not only feasible, but an actual equilibrium that may arise under certain conditions. The former, however, is not possible. There are several reasons for this impossibility. These are based on the fact that customers are strategic and thus can detect such a lie, in the same way that they can detect a profitable deviation from a strategy profile.

Figure 1. The three regions as defined in Proposition 3.4, based on the customers' patience and the firm's patience.



Summary of the Results and Implications. The following figure summarizes our findings thus far. Note that if the customer and firm have identical impatience the firm can reveal complete information regarding the system state; this corresponds to the 45-degree line shown in Figure 1. Further, if the customers are more patient than the firm, then any equilibrium that exists must entail randomization by the customers, and thus the firm cannot achieve its first-best profits. However, if the customers are less patient than the firm, then the firm achieves its first-best profit by using vague announcements unless the difference between the patience is not too large. This is depicted in the region above the 45-degree line in the figure. Thus, there always exists intentional vagueness in the announcement unless the firm and the customers are perfectly aligned.

Announcements and Managing Customer Expectations. The announcements in our model are used to build customer expectations about the congestion in the system. In this sense, the announcement plays a role of managing customer expectations regarding the system congestion. It is interesting to note that the message provided to the customer might not exhibit this information literally, but is captured by the Bayesian rule employed by the customer (Condition 1 in Definition 3.1 of MPBNE). In this section, we explore the extent to which the firm can reveal the information, i.e., how precisely can we tell the customers "what to expect" while sustaining the equilibrium. The above results indicate that in our setting, because the incentives of the parties are not perfectly aligned, the firm almost always needs to use intentional vagueness when managing expectations regarding the system congestion. In pure strategies, the firm uses intentional vagueness to lure customers into the system, in order to make sure the firm itself does not have incentive to deviate from this announcement. When allowing for mixed strategies, the firm is using intentional vagueness, even though the firm's profit is below its first best, in order to sustain an equilibrium—this time making sure that the incentives for the customers are aligned.

5. Conclusions

We study a model of a service system where customers are not only strategic in their actions, but also in the way they interpret information. The service provider on her end is strategic in the way she provides information. We show that even though the information is costless and unverifiable, it can improve the outcome for all players. Through a stylized model, we are able to provide a theoretical basis for intentional vagueness on the part of the service provider.[9]

The framework used in the paper echoes the cheap talk model proposed in Crawford and Sobel (1982). Driven by the application to services, we proposed a model that has two key features: First, the model considers multiple receivers (customers) whose actions have externalities on other receivers. Second, the stochasticity of the type (i.e., state of the system) is not exogenously given, but is determined by the equilibrium strategies of the players. We show that these have nontrivial impacts on the structure of equilibria. The framework developed in this paper can be applied to other operations management settings where the customers cannot credibly verify the information provided to them. Allon and Bassamboo (2011a) study a setting where a retailer signals strategic information to the potential buyers.

Throughout this paper, we assumed that the queue is Markovian and that there is a single server. The results in the paper are extendable to the case of multiple servers. Another assumption that can be relaxed is the assumption that $\lambda < \mu$ (i.e., demand is less than capacity). The structure provided in the paper also holds for $\lambda > \mu$. The model in the paper can be suitably extended to allow for multiple customer classes. The MPBNE that arises from this model would have a richer equilibrium language. However, it can be shown that under mild conditions, the solution would exhibit a similar threshold-type structure. (See Allon and Bassamboo 2009 for analysis of such models.) In this paper, we assumed that announcements are provided immediately upon the customer's arrival. Allon and Bassamboo (2011b) shows that delaying the information providing may assist the firms in improving its credibility for some parameters, yet it may hurt its credibility as well.

Future models should also account for the fact that customers may not be expected-utility maximizers. Under such modeling assumptions, customers may prefer receiving information over no information, and in particular, customers may prefer more accurate announcements. Further, in many settings one may expect the value obtained from the service and the cost of waiting to be affected by the announcements made by the firm.

6. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Endnotes

1. In our context we use the term unverifiable to describe situations in which the customer cannot verify statements regarding the state of the system.

2. We assume that the cost associated with conveying the message is negligible. In most practical service organizations, although the provider needs to incur fixed costs, for example, by investing in a more sophisticated IT infrastructure to identify the state of the system, the marginal cost of providing the information to the customer is insignificant. There is a voluminous literature that deals with models in which signaling is not costless, and the mere fact that players are willing to incur a cost provides a signal.

3. We discuss a model where the firm selects the service rate μ optimally in Appendix EC.3.2.

4. If the customer's cost of waiting is nonlinear, then the main results of the paper continue to hold. However, some of the results in the appendices hinge on this linearity. We will comment more elaborately on those in the appendices.

5. For a model with waiting time in the queue, see Allon and Bassamboo (2011b).

6. Note that $p_q(y, g)$ can be thought of as the beliefs of the agents on the state of the system. These beliefs must be consistent with the strategy of the other players.

7. Seidmann (1992) provides an example of a setting in which a babbling equilibrium does not exist.

8. Note that J could possibly be infinite.

9. It is important to note that different firms may have different reasons for using intentional vagueness. For example, real service systems are more complex than simple $M/M/1$ queues (both operationally and in terms of information availability to both parties), which implies that the service provider may be unsure about the mapping between the number of customers in the system and the resulting waiting that a customer would experience. This uncertainty may be a driver for the intentional vagueness observed in practice.

Acknowledgments

The authors are thankful for comments by the participants of the conference on customer-oriented operations models at the Olin Business School. Also, they are grateful to David Austen-Smith, Sandeep Baliga, Martin Lariviere, Jim Dana, Peter Eso, Gabriel Weintraub, Robert Swinney, Ramandeep Randhawa, Gerard Cachon, Robert Shumsky, Moshe Haviv, Avi Mandelbaum, Noah Gans, Zeynep Aksin, Fikri Karaesmen, and Jan Van Mieghem for long discussions on the paper.

References

Afeche, P. 2004. Incentive-compatible revenue management in queueing systems: Optimal strategic idleness and other delaying tactics. Working paper, Kellogg School of Management, Northwestern University, Evanston, IL.

Allon, G., A. Bassamboo. 2009. Managing customer expectations and priorities in service systems. Working paper, Kellogg School of Management, Northwestern University, Evanston, IL.

Allon, G., A. Bassamboo. 2011a. Buying from the babbling retailer? The impact of availability information on customer behavior. *Management Sci.* **57**(4) 713–726.

Allon, G., A. Bassamboo. 2011b. The impact of delaying the delay announcements. *Oper. Res.* **59**(5) 1198–1210.

Allon, G., A. Federgruen. 2007. Competition in service industries. *Oper. Res.* **55**(1) 37–55.

Armony, M., C. Maglaras. 2004a. Contact centers with a call-back option and real-time delay information. *Oper. Res.* **52**(4) 527–545.

Armony, M., C. Maglaras. 2004b. On customer contact centers with a call-back option: Customer decisions, routing rules, and system design. *Oper. Res.* **52**(2) 271–292.

Armony, M., N. Shimkin, W. Whitt. 2009. The impact of delay announcements in many-server queues with abandonment. *Oper. Res* **57**(1) 66–81.

Austen-Smith, D. 1990. Information transmission in debate. *Amer. J. Political Sci.* **34**(1) 124–152.

Chen, Y., N. Kartik, J. Sobel. 2008. Selecting cheap-talk equilibria. *Econometrica* **76**(1) 117–136.

Crawford, V. P., J. Sobel. 1982. Strategic information transmission. *Econometrica* **50**(6) 1431–1451.

Dobson, G., J. Pinker. 2006. The value of sharing lead time information. *IIE Trans.* **38**(3) 171–183.

Farrell, J., M. Rabin. 1996. Cheap talk. *J. Econom. Perspect.* **10**(3) 103–118.

Guo, P., P. Zipkin. 2007. Analysis and comparison of queues with different levels of delay information. *Management Sci.* **53**(6) 962–970.

Hassin, R. 1986. Consumer information in markets with random product quality: The case of queues and balking. *Econometrica* **54**(5) 1185–1195.

Hassin, R., M. Haviv. 2003. *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*. Kluwer Academic Publishers, Boston.

Haviv, M., M. L. Puterman. 1998. Bias optimality in controlled queueing systems. *J. Appl. Probab.* **35**(1) 136–150.

Jouini, O., Z. Aksin, Y. Dallery. 2011. Call centers with delay information: Models and insights. *Manufacturing Service Oper. Management* **13**(4) 534–548.

Knudsen, N. C. 1972. Individual and social optimization in a multiserver queue with general cost-benefit structure. *Econometrica* **40**(3) 515–528.

Naor, P. 1969. The regulation of queue size by levying tolls. *Econometrica* **37**(1) 15–24.

Seidmann, D. 1992. Cheap talk games may have unique, informative equilibrium outcomes. *Games Econom. Behav.* **4**(3) 422–425.