Inferring Quality from a Queue

Laurens G. Debo*
Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213

Christine A. Parlour†
Haas School of Business
UC Berkeley
Berkeley, CA 94720

Uday Rajan‡
Ross School of Business
University of Michigan
Ann Arbor, MI 48109.

January 16, 2008

Abstract

We study how rational agents infer the quality of a good (a product or a service) by observing the queue that is formed by other rational agents waiting to obtain the good. Agents privately observe the realization of a signal that is imperfectly correlated with the true quality of the good. Based on the queue length and the signal realization, agents decide whether to join the queue and obtain the good or to balk. The time to produce the good is exponentially distributed. Agents arrive according to a Poisson process at a market and are served according to a FIFO discipline. We find that when waiting costs are zero, agents receiving a bad signal join the queue only if it is long enough. When waiting costs are strictly positive, agents do not join the queue if it is too long. Furthermore, there may exist a set of isolated queue lengths ("holes") at which agents with a bad signal do not join. In equilibrium, queues are shorter for low quality goods and an agent is more likely to erroneously join a queue for a low quality good than erroneously balk a queue for a high quality good. If the service rate is decreased, more agents may join queues for high quality goods, even when waiting is costly.

*Tel: (412) 268-8403, E-mail: `laurdebo@andrew.cmu.edu`
†Tel: (412) 268-5806, E-mail: `parlour@haas.berkeley.edu`
‡Tel: (734) 764-2310, E-mail: `urajan@umich.edu`

1 Introduction

Customers frequently have to wait before they can consume a product or service. Lines outside nightclubs, rides at amusement parks and waiting lists for new products are part of everyone's experience. For example, Cabbage Patch Kids in 1983, Beanie Babies in the 1990s, PlayStation 2 in 2000, and the iPod in 2004 had significant waiting times, and queues formed in front of stores when these products first came on the market (*Wall Street Journal*, Dec 2, 2005). In such cases, long queues and attendant waiting times demonstrate that many other agents find it worthwhile to obtain the good. At the same time, longer queues generate higher waiting costs.

We model this consumer tradeoff in a queueing model in which a good can be of either high or low quality. Risk-neutral consumers arrive sequentially at a market. Each consumer receives a binary signal of the quality of the good and must decide whether to purchase it. Purchasing entails joining a "first-come, first-served" queue. There may be a cost associated with waiting. Consumers who have joined the queue leave once they have been serviced. As is standard in queueing models, agents arrive according to a Poisson process and service times are independently and exponentially distributed. Stochastic arrival and departure imply that the length of the queue varies over time. We assume that an arriving consumer does not observe the entire history of the game and so does not know how many people arrived before her. Instead, she sees the number of people currently in the queue. Her decision to join the queue or to balk, of course, is endogenous.

Our model builds upon the extensive literature on queueing and games. Equilibrium joining strategies are examined by Naor (1969) and the subsequent literature on economic aspects of queueing (see Hassin and Haviv, 2003, for an excellent overview). When there are positive waiting costs, agents play a threshold strategy: they join the queue as long as it is not too long. Beyond some threshold, there is a congestion effect: the waiting costs imply that joining the queue is not worthwhile (as in Naor, 1969 and Hassin and Haviv, 2003, chapter 3).

In the absence of waiting costs, the queue exerts a pure informational externality: we find that a long queue leads to an upward revision in the expected value of the good. Similarly, a short queue leads to a downward revision in the expected value of the good. Due to this positive information externality generated by the queue in our model, agents with bad signals play a strategy that includes a different threshold than the classical one: they join the queue *only if it is long enough*. Indeed, we show that a low quality good typically has a shorter queue than a high quality one, so that the queue communicates information about quality to consumers. Further, we find that, due to the one-sided nature of the service process, agents are more likely to erroneously join the queue for a low quality good than erroneously balk from a queue for a high quality good.

If there are positive waiting costs, we demonstrate that the interplay between the congestion and information effects leads to the existence of action "holes": queue lengths for which agents with bad signals will not join the queue, even though they join at queue lengths immediately below and above. Thus, the equilibrium strategy of our model cannot be described by thresholds. To the best of our knowledge, this equilibrium joining structure has not been identified in the queuing literature.

Our result is related to, but, significantly different from Hassin and Haviv (1997), who consider equilibrium queue joining behavior when an individual agent's tendency to act in a certain way increases when more individuals act in this way. They term this "Follow The Crowd" (FTC) behavior, and show that it arises in a situation in which where agents can pay to join a queue with higher priority. In Hassin and Haviv's model, the equilibrium joining strategies are threshold strategies[1] in which agents do not join when the queue is too long. In our model, the information externality generated by a queue generates an FTC motive. The FTC characteristic of our model results in a fundamentally different equilibrium strategy structure than in Hassin and Haviv (with a threshold *below* which agents do not join and with holes), due to our information structure that causes the value and cost to increase as a function of the queue length.

Furthermore, we find that for some parameter values, the arrival rate at a queue for a high quality good and the total consumer surplus generated both *decrease* as the service rate increases, even when the waiting costs are strictly positive. That is, the benefits of congestion, in terms of the additional information about quality provided to consumers, outweigh the effects of the extra waiting costs borne by consumers. In the queueing literature, increases in service rates typically lead to an increase in total welfare. Thus, relaxing the information assumptions of Naor's (1969) classical model leads to a significant reinterpretation of the effects of changes in service rates.

Our model also builds on the extensive literature on herd behavior. The canonical herding models are due to Bikhchandani, Hirshleifer and Welch (1992) and Banerjee (1992). In these models, agents observe the entire history of the game, in terms of actions taken by previously arriving agents. Each agent has a private signal, and infers the quality of a good after viewing the actions of all prior agents. With positive probability, after some point in time, an information cascade forms, i.e. all agents take the same action irrespective of the realization of their private signal. In a cascade, the true quality of the good may never be learned by agents even in the long run. This is striking because if agents would have shared their private information with each other, in the long run, they would have learned the true quality of the good. A comprehensive summary of the herding literature appears in Chamley (2004).

The assumption that agents observe the entire history of play is typically relaxed in the literature on "word of mouth communication" and social learning, where it is assumed that agents observe only samples of the history. Banerjee (1993) presents a model in which agents optimally choose whether to invest in a project, and information is communicated to subsequent investors via a rumor process. He finds that not all investors with good signals will invest. Ellison and Fudenberg (1993, 1995) find that with boundedly rational agents, word of mouth communication can lead all agents to a superior outcome, even if each individual receives little social information. If each agent observes only her predecessor's action, beliefs and actions can cycle forever, with longer periods of uniform behavior and rare switches (Celen and Kariv, 2004). Smith and Sorensen (1997) consider a sequential action model in which agents receive a random sample of the history. They find that the true quality of the good will never be learned, even on the long run.

In our model, agents observe the queue when they arrive, but have no other public information, not even how many agents have arrived previously. It is natural to consider that an agent arriving at the market has no information about how many agents arrived previously and decided not to consume the good, or indeed consumed the good and left the market. Instead, an agent observes only those agents currently waiting for the good. The queueing model thus provides a way to restrict the histories that an agent sees. Further, as a result of the stochastic service process, the queue length changes over time even when no one else joins. Therefore, agents arriving at different times observe different histories, and may come to different conclusions about the quality of the good, even if they had similar private information about quality.[2] When agents with good and bad signals play the same strategies at a given queue length, we say they display herd behavior. As in our model an empty queue is bad news for agents with a bad signal, there is no herding at the empty queue—an agent with a bad signal balks, while an agent with a good signal joins the queue. Starting with a long queue, if a lot of agents leave the queue because they were serviced, the queue may dwindle before the next agent comes along. In other words, empty queues appear with positive probability, and would do so even if all agents chose to join the queue. At any point of time, the random arrival and departure rates ensure that queue lengths change for newly arriving agents. Hence, there cannot be an information cascade, or a point in time beyond which all agents take the same action.

We describe our model in Section 2, outline some properties of equilibria in Section 3, discuss social learning in Section 4, and consider the effect of changing the service rate in Section 5. We conclude in Section 6.

2 Model

Consider an experience good (a physical product or a service) which can have high (h) or low (ℓ) quality. The utility an agent obtains from purchasing and consuming a good of quality j is v_j, with $v_\ell < v_h$. This utility is net of the good's price, which is not explicitly modeled.

It takes time to service (i.e., provide the good to) each consumer. The service time for each consumer is exponentially distributed with parameter μ, and is independent across consumers. Consumers may suffer a disutility from waiting to obtain the good, with a waiting cost of $c \geq 0$ per unit time. Consumers are risk-neutral, and arrive sequentially at the market according to a Poisson process with parameter λ. If agents arrive faster than they are serviced, they queue. The queue is served on a first-come first-served basis.

The quality of the good is not directly observed by consumers. Agents' prior belief that the good is high quality is p. In addition, each agent receives a private signal $s \in \mathcal{S} = \{g, b\}$ about the quality of the good, where $\Pr(s = g \mid j = h) = \Pr(s = b \mid j = \ell) = q \in (\frac{1}{2}, 1)$. We assume symmetry in the conditional probabilities of signals for analytic convenience. Agents know neither the order in which they arrive at the market, nor the history of actions taken by previously arriving agents. However, each agent does observe the length of the queue when he arrives at the market.

The agent takes an action $a \in \{\text{join,balk}\}$, where $a = \text{join}$ indicates a decision to acquire the good, or to join the queue. Once he joins the queue, he may not renege; i.e., he cannot leave until he has been served. Joining the queue is therefore synonymous with consuming the good. If he chooses to balk (i.e., not acquire the good), he obtains a reservation utility of zero.

The state space is defined by $\mathcal{S} \times \mathbb{Z}_+$ where $\mathbb{Z}_+$ is the set of non-negative integers. The information set of an agent is completely defined by (s, n), the private signal and queue length observed upon arrival. Since an agent has no other distinguishing feature, the equilibrium we consider is perforce symmetric. A mixed strategy for an agent is a mapping $\alpha : \mathcal{S} \times \mathbb{Z}_+ \to [0,1]^{\mathcal{S} \times \mathbb{Z}_+}$, where $\alpha(s, n)$ denotes the probability that an agent with signal s who sees queue length n joins the queue.

Consider a randomly arriving agent. Suppose all other agents are playing the strategy α. Then, given signal, s, and queue length, n, the randomly arriving agent's posterior belief that the good is of high quality is denoted as $\theta_s(n; \alpha)$. Due to the memoryless property of the exponential distribution, the expected time to serve each agent in the queue (including the one currently being served) is $\frac{1}{\mu}$. Thus, for the randomly arrived agent, the expected total waiting time before he leaves the system is $(n + 1)\frac{1}{\mu}$. Given his signal and observed queue length (s,n), the agent's expected net utility from joining the queue is

$$u(s,n,\alpha) \quad = \quad v_\ell + \theta_s(n;\alpha)(v_h - v_\ell) - \frac{n+1}{\mu}c. \tag{1}$$

We make the following assumptions about the parameters. Let $p_s = \text{Prob}(j = h \mid s)$. Then, from Bayes' rule, $p_g = \frac{p\,q}{p\,q+(1-p)(1-q)}$ and $p_b = \frac{p(1-q)}{p(1-q)+(1-p)q}$.

Assumption 1
(i) Either $c > 0$ *or* $\frac{\lambda}{\mu} < 1$.
(ii) $p_g v_h + (1-p_g)v_\ell > \frac{c}{\mu} > p_b v_h + (1-p_b)v_\ell$

Part (i) ensures that the system is stationary, i.e. the expected queue length remains finite. If there are congestion costs (so that $c > 0$), even if all agents believe the good to be of high quality there is a maximum queue length. If there are no congestion costs, then the arrival rate of agents to the market is assumed to be smaller than the service rate μ. Part (ii) states that an agent who acts only on the basis of her own signal, and ignores any information in the observed queue length, joins an empty queue if and only if her signal is good, i.e. when the updated valuation upon receiving a good (bad) signal is higher (lower) than the cost of waiting until the service is done. This is similar to the usual assumption in the cascades literature (see, for example, Bikhchandani, Hirshleifer, and Welch, 1992) that an agent who ignores the information provided in other agents' actions will acquire the good if she has a good signal, but not if she has a bad one.

In a symmetric equilibrium, the agent's expected payoff in state (s,n) from playing the strategy α is $\alpha(s,n)u(s,n,\alpha)$. Hence, α defines a perfect Bayesian equilibrium if it maximizes this expected payoff in each state (s,n).

Definition 1 A strategy α is a stationary Markov perfect Bayesian equilibrium if, for each $s \in \{g,b\}$ and each $n \in \mathbb{Z}_+$,

$$\alpha(s,n) \;\in\; \arg\max_{x\in[0,1]} \; x\,u(s,n,\alpha),$$

where $u(s,n,\alpha)$ is defined by (1), and $\theta_s(n;\alpha)$ is defined by Bayes' rule for any n that is reached on the equilibrium path with a positive probability.

2.1 Posterior beliefs

Recall that p_s is the posterior probability that the good has high quality when an agent observes signal s but does not see the queue length. Let $\pi_j(n;\alpha)$, for $j = h,\ell$, be the probability that a randomly arriving agent observes n people in the queue, given that the true quality of the good is j and all agents play the strategy α. Then, the agent's posterior belief that the good has high quality, after having observed signal s and queue length n is

$$\theta_s(n;\alpha) \quad = \quad \frac{p_s\pi_h(n;\alpha)}{p_s\pi_h(n;\alpha)+(1-p_s)\pi_\ell(n;\alpha)} \qquad \text{if } \pi_j(n;\alpha)>0 \text{ for some } j\in\{h,\ell\}. \quad (2)$$

As may be noted from Definition 1, perfect Bayesian equilibrium places no restrictions on belief $\theta_s(n;\alpha)$ if $\pi_j(n;\alpha)=0$ for both $j=h$ and $j=\ell$.

To characterize equilibrium, we exploit a result from the queuing literature: the PASTA (Poisson Arrivals See Time Averages) property (see Wolff, 1982). Let $\{X_{t,j}(\alpha), t\geq 0\}$ be the stochastic process indicating the number of agents in the system at time t when the state of the world (i.e., the quality of the good) is j and the action profile of the agents is α. $X_{t,j}(\alpha)$ takes values on $\mathbb{Z}_+$. Let the system be empty at time $t=0$, and let $\pi_{t,j}(n;\alpha)=\Pr(X_{t,j}(\alpha)=n|X_{0,j}(\alpha)=0)$ be the probability that at time t there are n agents in the system, conditional on the state of the world and the action profile α.

Lemma 1 *[PASTA, Wolff (1982)] For any strategy α and good quality j, suppose the limiting probability, $\pi_j(n;\alpha)=\lim_{t\to\infty}\pi_{t,j}(n;\alpha)$ exists. Then, the probability that a randomly arriving agent observes queue length n is $\pi_j(n;\alpha)$.*

Assumption 1 ensures that for each $j=\ell,h$ and each strategy α, the number of customers in the system in our model is described by an ergodic Markov process. Therefore, the limiting probabilities $\pi_j(\cdot)$ exist for each n. The PASTA property then implies that the distribution over queue length faced by a random arriving agent equals the long run queue length distribution.

We derive the stationary distribution for our process as follows. Let $r_j(n;\alpha)$ be the ex ante probability (i.e., given a queue length n, but before the signal s is observed) that an agent who plays strategy α joins a queue that already has n agents, given that the good's quality is j. Then,

$$r_j(n;\alpha) \quad = \quad \begin{cases} q\alpha(g,n)+(1-q)\alpha(b,n) & \text{if} \quad j=h \\ (1-q)\alpha(g,n)+q\alpha(b,n) & \text{if} \quad j=\ell \end{cases}. \qquad (3)$$

Now, if there are n agents in the queue, the rate at which a new agent joins the queue is $\lambda r_j(n;\alpha)$. This "joining rate" depends both on the true quality of the good, $j\in\{\ell,h\}$, and on the strategy being played by the agents, α. As the service rate is constant and equal to μ, for any (j,α) pair, the induced queuing system is a birth and death process.

Consider a queue length of $n>0$. There are only two ways in which this could have occurred: (i) there was a queue of $n+1$ consumers, and the agent at the front of the queue was served, or (ii) the queue was of length $n-1$ and with a new agent joined. Thus, when $n > 0$, the expected number of entrances into state n per unit of time is $\pi_j(n+1;\alpha)\mu + \pi_j(n-1;\alpha)\,\lambda\, r_j(n;\alpha)$. Conversely, if the queue is at n, it can either increase due to an arriving agent or decrease due to a service completion. The expected number of departures from state n per unit of time is $\pi_j(n;\alpha)(\lambda\, r_j(n;\alpha) + \mu)$. In steady state, the expected number of entrances into any state n is equal to the expected number of departures from the state, generating a flow balance equation for each n. The stationary probabilities are the solution to all such flow balance equations. With (i) of Assumption 1, the solution is unique for a given quality j and strategy α. The following Lemma characterizes two properties of the solution (all proofs are in the Appendix).

Lemma 2 *Suppose all agents follow the strategy profile α. Then, for $j = h, \ell$, the stationary probability of observing a queue of length n satisfies*

$$\pi_j(n;\alpha) = \pi_j(0;\alpha)\left(\frac{\lambda}{\mu}\right)^n \prod_{k=0}^{n-1} r_j(k;\alpha), \tag{4}$$

$$\text{with} \qquad \pi_j(0;\alpha) = \frac{1}{1+\sum_{n=1}^{\infty}\left(\frac{\lambda}{\mu}\right)^n \prod_{k=0}^{n-1} r_j(k;\alpha)}. \tag{5}$$

Notice that $\pi_j(n+1;\alpha) = \pi_j(n;\alpha)\frac{\lambda}{\mu}\, r_j(n;\alpha)$. Therefore, the stationary probability of observing $(n+1)$ agents in the queue is zero if and only if $r_j(n;\alpha) = 0$; i.e., if no signal will induce an agent to join a queue that already has n agents. Further, if $\pi_j(n;\alpha) = 0$ for any n, then $\pi_j(k;\alpha) = 0$ for all $k \geq n$, and the queue cannot grow beyond length n.

Suppose all other agents are playing the strategy α, and consider the optimal action of an agent who, upon arrival, observes signal s and is faced with a queue length n. Recall that $u(s,n,\alpha) = v_\ell + \theta_s(n;\alpha)(v_h - v_\ell) - \frac{n+1}{\mu}c$. A consumer's best response is then $\alpha^*(s,n) = 1$ if $u(s,n,\alpha) > 1$ and $\alpha^*(s,n) = 0$ if $u(s,n,\alpha)$, with all $\alpha \in [0,1]$ providing equal payoffs if $u(s,n,\alpha) = 0$. That is, the agent is willing to join the queue only if his expected payoff $u(s,n,\alpha)$ is non-negative, or, in other words, if the posterior belief $\theta_s(n;\alpha)$ that the good is of high quality is sufficiently high.

We now show that a perfect Bayesian equilibrium exists. In the proof of Proposition 1 we distinguish between the cases when the maximal queue length is finite and when it is not. When the waiting cost is strictly positive, there is a maximum queue length beyond which no agent will join the queue, even if the good is known to be of high quality. When $c = 0$, by contrast, the queue can grow arbitrarily large.

Proposition 1 *For all parameter values satisfying Assumption 1, there exists a stationary Markov perfect Bayesian equilibrium.*

The information content in some queues may induce more agents to join the queue, while congestion costs encourage agents to balk. As the information content depends on the equilibrium actions of other agents, multiple equilibria can arise. In Example 1, we will show an example with multiple equilibria.

3 Properties of Equilibrium

We now characterize some properties that are common to all equilibria, even when there may be multiple equilibria. Given the structure of the posterior belief in equation (2), it is immediate that, in any state in which an agent with a bad signal joins the queue, so will an agent with a good signal. Hence, if an agent with a good signal does not join the queue at some length n, the queue can never grow beyond n, since an agent with a bad signal will not join either.

Lemma 3 *Let α^* denote an equilibrium strategy. Then,*
(i) the probability an agent joins the queue is higher when the agent has a good signal and when the good is of high quality; i.e., for all $n \in \mathbb{Z}_+$, $\alpha^*(g,n) \geq \alpha^*(b,n)$ and $r_h(n;\alpha^*) \geq r_\ell(n;\alpha^*)$.
(ii) for any queue length n, if $\alpha^*(g,n) = 0$, then $\pi_\ell(\tilde{n};\alpha^*) = \pi_h(\tilde{n};\alpha^*) = 0$ for all $\tilde{n} \geq n+1$.

Given equation (4), the stationary probabilities of a high and a low quality good at a queue length of zero, and thus the equilibrium strategy at zero ($\alpha^*(s,0)$) are an important characteristic of the equilibrium. The first part of Proposition 2 establishes that a low quality good is more likely to have an empty queue than a high quality good. Further, an agent with a bad signal does not join an empty queue. By contrast, an agent with a good signal joins an empty queue with positive probability.

Proposition 2 *In equilibrium,*
(i) a queue length of zero is more likely if the good is of low quality, or $\pi_\ell(0;\alpha^*) > \pi_h(0;\alpha^*)$.
(ii) an agent with a bad signal never joins the empty queue, whereas an agent with a good signal joins the empty queue with positive probability; i.e., $\alpha^*(b,0) = 0$ and $\alpha^*(g,0) > 0$.

It follows immediately from Proposition 2 that from a queue length of zero, a length of one can only be reached if an agent with a good signal joins. In general, observing a longer queue length is "good news." By contrast, a queue length of zero is "bad news" for an agent: the likelihood that the good is of low quality is higher. With no information about queue length, an agent believes with probability p_s that the good is of high quality. Once she sees a queue length of zero, she revises this belief downward as $\frac{\pi_\ell(0;\alpha^*)}{\pi_h(0;\alpha^*)} < 1$.

Further, Proposition 2 and Lemma 3 imply that agents with good signals play "threshold" strategies: if the waiting cost is strictly positive, there is a threshold $\bar{n}(c)$ such that an agent with signal g joins the queue with positive probability until the queue reaches $\bar{n}(c)$. At this point, the agent does not join the queue. From Proposition 2, neither does an agent with signal b. Hence, $\bar{n}(c)$ is the maximal queue length observed in equilibrium.[3] Let $\lfloor x \rfloor$ denote the largest integer less than or equal to x. If the queue length is $\lfloor \frac{v_h\,\mu}{c} \rfloor$, an agent who knows the good has high quality finds the waiting costs too high to join the queue. Thus, when $c > 0$, it is immediate that $\bar{n}(c) \leq \lfloor \frac{v_h\,\mu}{c} \rfloor$. When $c = 0$, we define $\bar{n}(0) = \infty$.

Let $W_j(\alpha^*)$, for $j = h, \ell$, be the expected waiting time for a randomly arriving agent at a good with quality v_j. From the PASTA property, randomly sampled agents face the stationary distribution. Thus,

$$W_j(\alpha^*; c) \quad = \quad \sum_{n=0}^{\bar{n}(c)} \pi_j(n;\alpha^*)\, r_j(n;\alpha^*)\, \frac{n+1}{\mu}, \text{ for } j = h, \ell.$$

The queue length distribution induced by a high quality good first order stochastically dominates the distribution induced by a low quality good, so that on average a new consumer observes a longer queue at the high quality good. Given that the average queue length is longer for the high quality good, it follows that the equilibrium expected waiting time is higher also for the high quality good.

Proposition 3 *In equilibrium,*
(i) the expected queue length is higher for a high quality good, and
(ii) the expected waiting time for a high quality good is greater than the expected waiting time for a low quality good, or $W_h(\alpha^*; c) > W_\ell(\alpha^*; c)$.

Queues, therefore, are physical manifestations of good quality. While high quality goods may sometimes find themselves with an empty queue due to the random service rate, on average, a random arrival faces a high probability of a long wait.

3.1 Pure informational externality: $c = 0$.

To isolate the informational role of queues, suppose that $c = 0$. In this case, there are no congestion costs imposed by long queues, and the decision to join depends only on the information content in the queue length.

With no congestion costs, the equilibrium conditions simplify: A consumer is willing to join the queue if $\frac{1-\theta_s(n;\alpha)}{\theta_s(n;\alpha)} \leq -\frac{v_h}{v_\ell}$, or

$$\frac{\pi_\ell(n;\alpha)}{\pi_h(n;\alpha)} \quad \leq \quad -\frac{p_s}{1-p_s}\frac{v_h}{v_\ell}. \tag{6}$$

That is, the likelihood that the firm has low quality must be sufficiently low.

Now, the likelihood ratio is weakly decreasing in the queue length. Therefore, an agent with a bad signal may not join the queue if it is too short, but will join the queue if it is long enough. Depending on parameter values, there may be a case in which an agent with a bad signal is indifferent between joining and not over a non-singleton set of queue lengths, and chooses to join with certainty except at one specific queue length.

Proposition 4 *Suppose $c = 0$. In equilibrium:*
(i) $\alpha^*(g, 0) > 0$ and $\alpha^*(g, n) = 1$ for all $n \geq 1$.
(ii) there exists an $\underline{n}_b \in \{1, 2, 3\}$ such that $\alpha^*(b, n) = 0$ for $n < \underline{n}_b$ and $\alpha^*(b, \underline{n}_b) > 0$.
(iii) $\alpha^*(b, n) = 1$ for all except at most one $n \geq \underline{n}_b$.

Thus, with $c = 0$, an agent with a bad signal plays a threshold strategy of the following form: she joins the queue only if the queue length is *above* some threshold. This is akin to the "Follow the Crowd" behavior explored by Hassin and Haviv (1997). However, Hassin and Haviv considered strictly positive waiting costs and strategies in which agents join *below* a certain threshold, which is the converse of the equilibrium in our model.

We now describe an algorithm that identifies pure strategy equilibria (when such an equilibrium exists) when $c = 0$. Let a^* denote a pure strategy equilibrium. The algorithm exploits the fact that the likelihood ratio at the empty queue, $\frac{\pi_\ell(0;a^*)}{\pi_h(0;a^*)}$ lies between 1 (from Proposition 2 (i)) and $-\frac{p_g}{1-p_g}\frac{v_h}{v_\ell}$ (from equation (6)).

Let the likelihood ratio at queue length n be denoted by $\psi_n(\alpha) = \frac{\pi_\ell(n;\alpha)}{\pi_h(n;\alpha)}$. From equation (4), we obtain

$$\pi_j(n;\alpha) \quad = \quad \pi_j(n-1;\alpha)\,\frac{\lambda}{\mu}\,r_j(n-1;\alpha) \text{ for } j = h, \ell,$$

so that

$$\psi_n(\alpha) \quad = \quad \psi_{n-1}(\alpha)\frac{r_\ell(n-1;\alpha)}{r_h(n-1;\alpha)}.$$

That is, for any strategy profile α, the likelihood ratio at $n-1$ together with the actions at $n-1$ determine the likelihood ratio at n. Further, from equation (6), the likelihood ratio at n determines the optimal action at n.

Using these properties, we construct a mapping $\Phi : \left[1, -\frac{p_g}{1-p_g}\frac{v_h}{v_\ell}\right] \to \left[1, -\frac{p_g}{1-p_g}\frac{v_h}{v_\ell}\right]$. For any $\psi_0 \in \left[1, -\frac{p_g}{1-p_g}\frac{v_h}{v_\ell}\right]$, define the following pure strategy profile $a(\psi_0)$, starting with $n = 0$:

$$a(s, n \mid \psi_0) = \begin{cases} 1, & \text{if } \psi_n \leq -\frac{p_s}{1-p_s}\frac{v_h}{v_\ell} \\ 0, & \text{if } \psi_n > -\frac{p_s}{1-p_s}\frac{v_h}{v_\ell} \end{cases} \tag{7}$$

$$\psi_n = \psi_{n-1}\frac{r_\ell(n-1; a(\psi_0))}{r_h(n-1; a(\psi_0))}. \tag{8}$$

By construction, the pure strategy a ensures agents are playing best responses. Notice that, under a, agents join the queue with probability 1 when they are indifferent between joining and balking. When agents with both signals g and b join at some queue length $n-1$, $r_\ell(n-1; a(\psi_0)) = r_h(n-1; a(\psi_0)) = 1$, so that $\psi_n = \psi_{n-1}$. Hence, by construction of $a(\psi_0)$, agents with both signals g and b will join the queue at n as well, and so on for each higher queue length.

Define $\underline{n}_b = \min\{n \mid a(b, n \mid \psi_0) = 1\}$. For $n < \underline{n}_b$, only agents with signal g join the queue, so that equation (4) implies that

$$\psi_n = \psi_0\left(\frac{1-q}{q}\right)^n. \tag{9}$$

Thus, $\underline{n}_b$ exists and is finite; as n increases, for some value of n, the RHS of (9) must eventually become smaller than $-\frac{p_b}{p_b}\frac{v_h}{v_\ell}$.

Now, note that, by construction, $a(\psi_0)$ has the following structure: $a(g, n \mid \psi_0) = 1$ for all n, whereas $a(b, n \mid \psi_0) = 0$ if $n < \underline{n}_b$ and $a(b, n \mid \psi_0) = 1$ if $n \geq \underline{n}_b$. Thus, for $n < \underline{n}_b$, it follows that $r_h(n; a) = q$ and $r_\ell(n; a) = 1 - q$. Similarly, for $n \geq \underline{n}_b$, it follows that $r_h(n; a) = r_\ell(n; a) = 1$. Substituting these values into equation (5) for both firm h and firm ℓ, we obtain

$$\pi_h(0; a) = \frac{1}{1 + \sum_{n=1}^{\underline{n}_b}(\lambda q/\mu)^n + (\lambda q/\mu)^{\underline{n}_b+1}\left(\frac{1}{1-\lambda/\mu}\right)}$$

$$\pi_\ell(0; a) = \frac{1}{1 + \sum_{n=1}^{\underline{n}_b}(\lambda(1-q)/\mu)^n + (\lambda(1-q)/\mu)^{\underline{n}_b+1}\left(\frac{1}{1-\lambda/\mu}\right)}.$$

Finally, we define $\Phi(\psi_0)$ to be the resultant likelihood ratio at the empty queue. That is,

$$\Phi(\psi_0) = \frac{1 + \sum_{n=1}^{\underline{n}_b}(\lambda q/\mu)^n + (\lambda q/\mu)^{\underline{n}_b+1}\left(\frac{1}{1-\lambda/\mu}\right)}{1 + \sum_{n=1}^{\underline{n}_b}(\lambda(1-q)/\mu)^n + (\lambda(1-q)/\mu)^{\underline{n}_b+1}\left(\frac{1}{1-\lambda/\mu}\right)}. \tag{10}$$

By construction, if ψ_0^* is any fixed point of $\Phi(\cdot)$, the strategy $a(\psi_0^*)$ is a pure strategy equilibrium. Note that such a fixed point may not exist: $\Phi(\cdot)$ is a staircase function, with discontinuities at likelihood ratios ψ_0 for which, for some n, an agent with a bad signal is exactly indifferent between joining and balking. As it turns out, such discontinuities characterize equilibria in which, at some queue length, an agent with a bad signal mixes between joining and balking.

The following example clarifies the tradeoff faced by agents when $c = 0$.

Example 1 Let $v_h = 1.1$, $v_\ell = -0.9$, $c = 0$, $p = 0.5$, $q = 0.575$, $\lambda = 1$, and $\mu = 1.3$. A plot of the $\Phi(\cdot)$ mapping is shown in Figure 1.

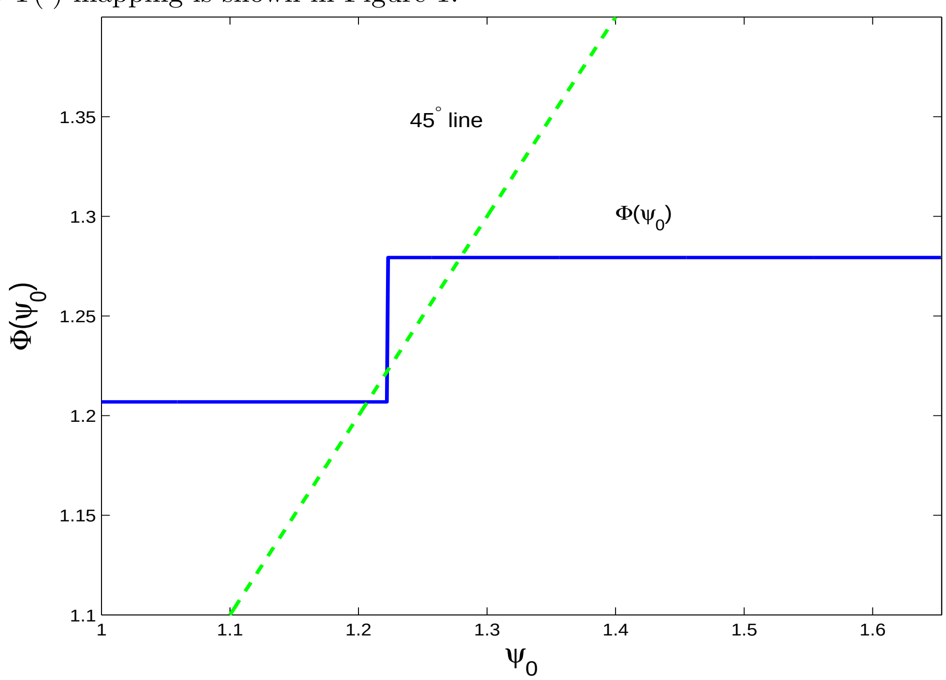

Figure 1: Pure strategy equilibria when $c = 0$, identified by the points of intersection with the 45° line.

There are thus two pure strategy equilibria, at $\psi_0^* = 1.207$ and $\psi_0^* = 1.279$. When $\psi_0^* = 1.207$, the equilibrium strategy $a(\psi_0^*)$ has all agents with good signals joining the queue, regardless of length, and agents with bad signals joining if and only if $n \geq 1$. When $\psi_0^* = 1.279$, the equilibrium strategy $a(\psi_0^*)$ has all agents with good signals joining the queue, regardless of length, and agents with bad signals joining if and only if $n \geq 2$.

In addition, there is a mixed strategy equilibrium at $\psi_0 = 1.222$, in which all agents with good signals join, and an agent with a bad signal balks at the empty queue, joins with probability 0.735 at $n = 1$, and joins with probability 1 at all $n \geq 2$.

3.2 Information and congestion externalities: $c > 0$.

Suppose $c > 0$, so agents incur a positive disutility to waiting. There is now a congestion externality in addition to the informational externality: All else equal, an agent would prefer to face a shorter queue than a longer one. The tradeoff, of course, is that queue length is still informative about the quality of the good.

As mentioned earlier, agents with good signals play threshold strategies, in which they join the queue only if it is short enough. However, with positive waiting costs, agents with bad signals no longer play threshold strategies. At any queue length at which an agent with a bad signal joins the queue with positive probability, a partial characterization of her strategy is as follows. There exists an upper threshold, $\bar{n}_b > 0$, beyond which the agent does not join the queue, since waiting costs are too high. This threshold comes from the congestion costs implied by a queue. There also exists a lower threshold, $\underline{n}_b > 0$, such that the agent will not join the queue unless there are at least $\underline{n}_b$ already waiting. This threshold comes from the informational role of the queue.

We show that, whenever $\underline{n}_b < \bar{n}_b$, an agent with a bad signal joins at some queue length, but need not join for every queue length between the two thresholds. In particular, there can be "holes," in their strategies. That is, an agent with a bad signal may not join the queue at some n, but join at $n-1$ and $n+1$. Therefore, when both the information effect and the congestion effect are present, consumer behavior cannot always be described by threshold strategies.

Proposition 5 *Suppose $c > 0$. Then,*
(i) there exists an $\bar{n}(c) \leq \frac{v_h \mu}{c} - 1$ such that $\alpha^*(g, 0) > 0$ for all $n = \{0, 1, \cdots, \bar{n}(c) - 1\}$ and $\alpha^*(g, \bar{n}(c)) = 0$.
(ii) there exist parameter values for which agents with bad signals do not play threshold strategies.

The proof of part (ii) of Proposition 5 is by example. Before presenting the example, we present an algorithm for finding pure strategy equilibria when $c > 0$. As in the algorithm for $c = 0$, start with any $\psi_0 \in \left[1, -\frac{p_g}{1-p_g}\frac{v_h}{v_\ell}\right]$, and inductively define $a(\psi_0)$, a pure strategy profile, and ψ_n, the likelihood ratio for queue length n, as in equations (7) and (8). In this case, we know that $r_h(n; a) = 1$ for all $n < \bar{n}(c)$, and $r_h(n; a) = r_\ell(n; a) = 0$ for $n \geq \bar{n}(c)$. Therefore, the resultant likelihood ratio at the empty queue may be written as

$$\Phi'(\psi_0) = \frac{1 + \sum_{n=1}^{\bar{n}(c)} (\lambda/\mu)^n \prod_{k=0}^{n-1} (q + (1-q) a(b, n \mid \psi_0))}{1 + \sum_{n=1}^{\bar{n}(c)} (\lambda/\mu)^n \prod_{k=0}^{n-1} (1 - q + q a(b, n \mid \psi_0))} \qquad (11)$$

By construction, a fixed point of $\Phi'(\cdot)$ is a pure strategy equilibrium. As in the case of

$c = 0$, such a fixed point may not exist: $\Phi'(\cdot)$ is a staircase function with discontinuities at likelihood ratios ψ_0 for which, for some $n \in [0, \bar{n}(c)]$, an agent with signal $s \in \{g, b\}$ may be indifferent between joining and balking.

Using this algorithm, we construct an example in which an agent with a bad signal does not follow a threshold strategy.

Example 2 Let $v_h = 1$, $v_\ell = -1$, $c = 0.06$, $p = 0.5$, $q = 0.65$, $\lambda = 1$, and $\mu = 1.5$. A plot of the $\Phi(\cdot)$ mapping is shown in Figure 2.

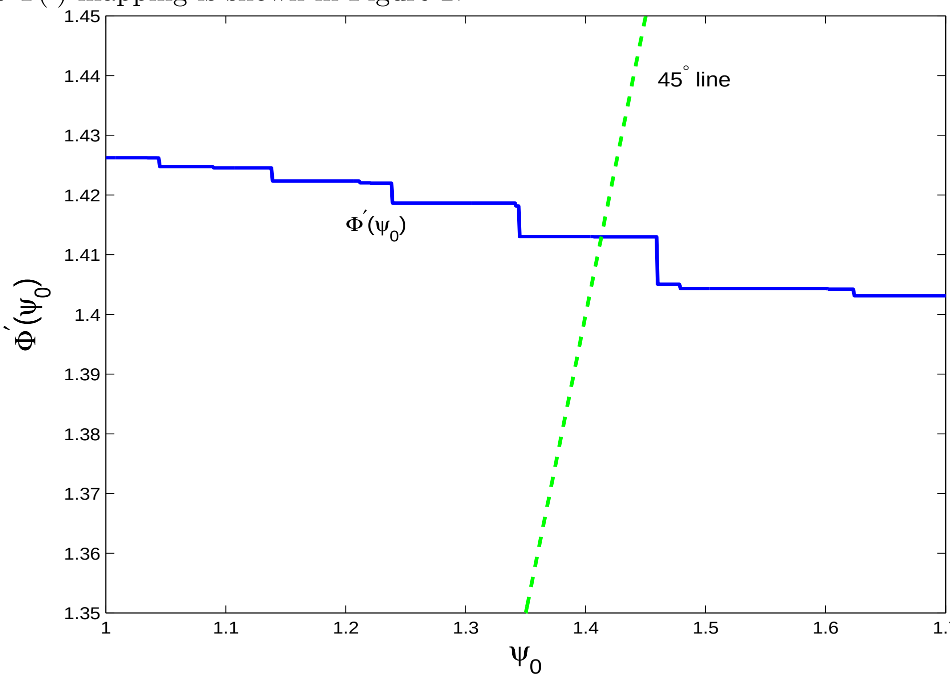

Figure 2: Pure strategy equilibrium when $c > 0$, identified by the point of intersection with the 45° line.

For these parameters, there exists one pure strategy equilibrium with $\psi_0^* = 1.413$. In the equilibrium, an agent with a good signal joins the queue if $n \leq 23$, and balks at all higher queue lengths. Therefore, $\bar{n}(c) = 24$. An agent with a bad signal plays the following non-threshold strategy: Balk if $n \in \{0, 1\}$, and join with probability 1 at all queue lengths with $n \leq 23$ except $n \in \{3, 5, 10, 15, 19, 21\}$.

What explains this seemingly strange set of holes in the strategy of an agent with a bad signal? When $c > 0$, the condition $u(s, n \mid a) \geq 0$ under which an agent with signal s is willing to join the queue becomes

$$v_\ell + \theta_s(n; a)(v_h - v_\ell) \quad \geq \quad \frac{(n+1)c}{\mu}. \tag{12}$$

A comparison of the left and right hand sides of this best response condition is shown in Figure 3. At queue lengths 0 and 1, only agents with a good signal join. Thus, the posterior belief that the firm has high quality, $\theta_s(n; a)$, increases at these queue lengths. At a queue length of 3, both agents with a good or a bad signal join. This results in $\theta_s(3; a) = \theta_s(4; a)$. That is, the belief about the quality of the good is the same at queue lengths 3 and 4. However, the costs are strictly higher at queue length 4. Therefore, for agents with a bad signal, it is rational to not join at 4.

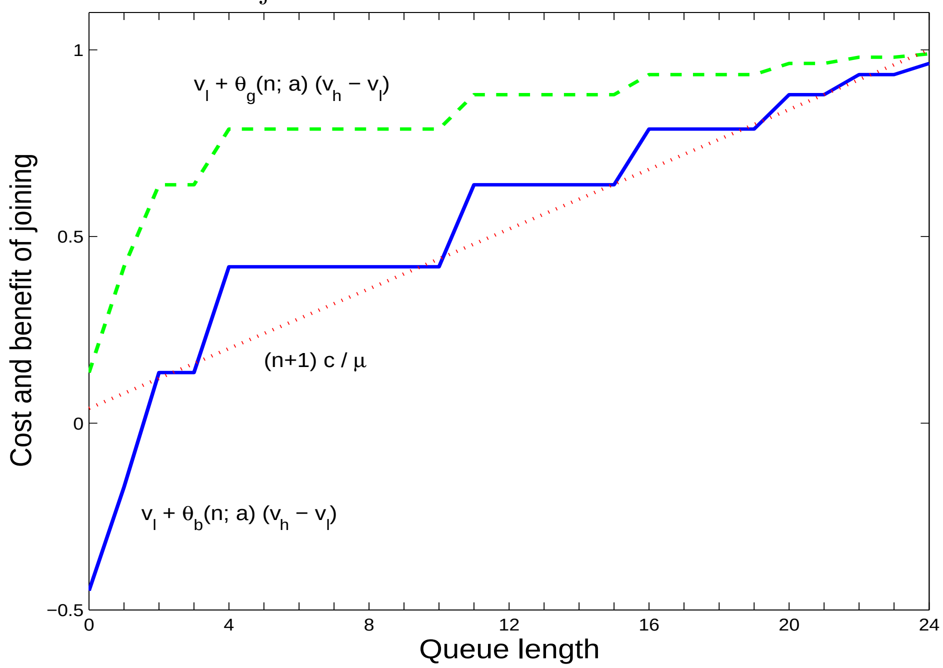

Figure 3: Equilibrium strategies when $c > 0$.

In turn, this leads to an increase in the likelihood that the firm has high quality at queue length 5 (since only agents with a good signal join at 4). Therefore, agents with both good and bad signals join at $n = 5$. As a result, the belief at queue length 6 remains unchanged. The cost of joining does increase from $n = 5$ to $n = 6$, but not by enough to deter agents with bad signals. Thus, they join at 6, and the same reasoning continues until queue length 10, when the waiting costs are higher than the valuation. Now, agents that arrive when the queue is 11 long, know that only an agent with a good signal could have joined the queue at 10. Again, this increases the likelihood that the good is of high quality. Proceeding in this way, we observe that the equilibrium joining pattern is composed of a set of "holes" in between an upper ($n = 24$) and a lower ($n = 2$) joining threshold.

Note that the equilibrium strategies constitute strict best responses. The example is therefore robust to small changes in any of the parameters.

Thus, with positive waiting costs, a bad-signal agent's strategy need not be a threshold strategy. The queue length plays a dual role. On the one hand, a longer queue implies a greater likelihood the good is of high quality. However, a longer queue also imposes higher waiting costs on an agent who joins it. In equilibrium, the decision to join the queue trades off these two effects.

4 Joining errors and herding

With positive probability agents choose an incorrect action—they either join the queue when the good has low quality (analogous to a Type II error) or balk from the queue when it has high quality (analogous to a Type I error). Since high and low quality goods generate different distributions over queue length, agents on average do learn about the quality of the good when they observe the number of agents waiting in the queue.

Given an equilibrium α^*, let $\epsilon_h(\alpha^*)$ denote the probability that a randomly arriving agent makes an incorrect decision when the quality of the good is high. This is equal to the probability that this agent refuses to join the queue, since the correct decision when quality is high is to join the queue when the length is less than $\lfloor \frac{v_h \mu}{c} \rfloor$. Similarly, let $\epsilon_\ell(\alpha^*)$ denote the probability that a randomly arriving agent makes an incorrect decision (i.e., joins the queue) when the quality of the good is low.

Recall that $\bar{n}(c)$ is the maximal queue length which can be observed in equilibrium. Exploiting the PASTA property,

$$\epsilon_h(\alpha^*) = \sum_{n=0}^{\bar{n}(c)} \pi_h(n;\alpha^*)(1 - r_h(n;\alpha^*)) \tag{13}$$

$$\epsilon_\ell(\alpha^*) = \sum_{n=0}^{\bar{n}(c)} \pi_\ell(n;\alpha^*) r_\ell(n;\alpha^*), \tag{14}$$

where $r_j(n;\alpha)$ is as defined in equation (3).

We find that there is an asymmetry in error rates whenever the equilibrium satisfies the following two properties: first, an agent with a good signal enters with probability 1 at every queue length up to $\bar{n}(c) - 1$, and second, an agent with a bad signal joins the queue at some length with positive probability. Under these conditions, agents are more likely to err by acquiring the good when the quality is low rather than by not acquiring it when the quality is high. When $c = 0$, all equilibria satisfy these conditions, as demonstrated in Proposition 4. The conditions are also satisfied in Examples 2 and 3 (below), where c is positive.

Proposition 6 *Consider an equilibrium α^* in which (i) $\alpha^*(g,n) = 1$ for all $n \in \{0, \cdots, \bar{n}(c) - 1\}$, and (ii) there exists some $\hat{n}$ for which $\alpha^*(b, \hat{n}) > 0$. Then, a randomly arriving agent is more likely to enter the queue for a low quality good than fail to enter the queue for a high quality good, or $\epsilon_\ell(\alpha^*) > \epsilon_h(\alpha^*)$.*

The asymmetry in learning with a high quality good as opposed to a low quality good is due to the one-sided nature of the servicing phenomenon. On being serviced, agents leave the queue. Therefore, a short queue may exist either because few people joined it, or because people who joined it were serviced relatively rapidly. An agent who arrives and sees a short queue cannot distinguish between these events, and is therefore more likely to err by joining a short queue.

We now define herd-like behavior when agents with a good or bad signal join the queue with the same probability (either 0 or 1) at a certain queue length. Therefore, the equilibrium strategy at such queue lengths is independent of the signal. For $c > 0$, recall that $\bar{n}(c) - 1$ is the largest value of n for which any $\alpha^*(s,n) > 0$. Therefore, herd behavior is only interesting for $n \leq \bar{n}(c)$.

Definition 2 An agent displays herd behavior at queue length $n < \bar{n}(c)$ if $\alpha^*(g,n) = \alpha^*(b,n)$.

This definition corresponds to Chamley's (2004) definition of an agent herding when his action is independent of his signal.[4] In our model, herding may occur at some queue lengths and not others (for example, Proposition 2 implies there is no herd behavior at queue length zero). Thus, herd behavior is necessarily local in our model. In our model, on the long run, the empty queue is reached infinitely often. This implies that we cannot have an informational cascade—a point of time beyond which all arriving agents herd. Starting at any queue length, random service departures will eventually reduce the queue to zero, at which point there is no herding.

Every time an agent with a bad signal chooses not to join a queue, an increase in the queue length demonstrates that the joining agent had a good signal. All participants update accordingly. In some cases, when the congestion cost is positive, signals can be sufficiently informative that agents with bad signals do not join the queue at any length. At this point, agents enter solely based on their signals.

Proposition 7 *Suppose $c > 0$ and $v_h < \frac{2c}{\mu}$. For q sufficiently close to 1, there exists an equilibrium in which an agent with a good signal joins the queue if and only if $n = 0$, and an agent with a bad signal never joins. Therefore, no herding occurs.*

It is immediate to observe that, for the equilibrium exhibited in Proposition 7, $\epsilon_h = \epsilon_\ell = 1 - q$. That is, the error probabilities are the same, regardless of the quality of the good. Suppose we choose parameters for which a no-herding equilibrium exists. Consider the effect of keeping the other parameters fixed and reducing the congestion cost c. A reduction in c has two effects: (i) the upper threshold for an agent with a good signal, $\bar{n}(c)$, increases and (ii) an agent with a bad signal joins the queue for some values of n. The former has no effect on the respective error probabilities, whereas the latter induces an asymmetry between the error probabilities for low and high quality goods. When c falls to zero, the equilibrium displays herd-like behavior at a countably infinite number of queue lengths, and the asymmetry in error probabilities is at its highest.

5 Effect of changing the service rate

Next, we study the impact of changing the service rate on the arrival rate of the high quality good. Since $\epsilon_h(\alpha^*)$ denotes the rate at which a randomly arrived consumer does not join the queue, the joining rate for the good is $1 - \epsilon_h(\alpha^*)$ per arrival, or $\lambda_h^* = \lambda(1 - \epsilon_h(\alpha^*))$ per unit of time.

First, suppose $c = 0$. Here, a queue plays a purely informational role. If the service rate μ is reduced, the queue is longer on average. As agents view long queues as positive signals, they are more likely to join at a given queue length.

However, comparative statics on μ are problematic, as there may be multiple equilibria, and a change in μ may result in a change in the nature of the equilibrium. Therefore, we focus on a particular equilibrium with $\alpha^*(g, 0) = 1$ and $\alpha^*(b, n) = 1$ for all $n \geq 1$. Such an equilibrium exists when $\frac{2p-1}{q+p-1} > \frac{\lambda}{\mu}$. We show that the joining rate for the high quality good is indeed declining in μ.

Proposition 8 *Suppose that $c = 0$, $v^h = -v^\ell$, and $\frac{\lambda}{\mu} < \frac{2p-1}{q+p-1}$. Then, the following is an equilibrium strategy: $\alpha^*(g, n) = 1$ for all n, $\alpha^*(b, 0) = 0$, and $\alpha^*(b, n) = 1$ for all $n \geq 1$. Given the equilibrium α^*, the joining rate for the high quality good, λ_h^*, is decreasing in the service rate, μ for $\mu > \lambda$.*

Next, consider overall consumer surplus. Recall that $W_j(\alpha^*; c)$ is the expected waiting time for a randomly arriving consumer when the good's quality is j, where $j \in \{h, \ell\}$. Further, if the good is of quality j, an agent who arrives and faces queue length n joins with ex ante probability $r_j(n; \alpha^*)$. Using the PASTA property and the prior of the good's quality, overall equilibrium consumer surplus per agent is defined as

$$\Omega(\alpha^*; c) \quad = \quad p\left[v_h(1 - \epsilon_h(\alpha^*)) - cW_h(\alpha^*)\right] + (1 - p)\left[v_\ell \epsilon_\ell(\alpha^*) - cW_\ell(\alpha^*)\right]. \qquad (15)$$

Consumer surplus per unit of time is then $\lambda\Omega(\alpha^*; c)$, since λ agents are expected to arrive in each time interval of length one. Now, suppose $c = 0$, and valuations are symmetric, so that $v_h = -v_\ell = v$. Then, consumer surplus per agent reduces to

$$\Omega(\alpha^*; 0) \quad = \quad v[p - \{p\epsilon_h(\alpha^*) + (1-p)\epsilon_\ell(\alpha^*)\}].$$

The term in the curly brackets is the ex ante expected error probability. We show that, under the same conditions as used in Proposition 8, if the service rate is sufficiently high, overall consumer surplus declines as the service rate increases.

Proposition 9 *Suppose all conditions of Proposition 8 are satisfied, and consider the identified equilibrium α^*. Given this equilibrium, there exists a threshold service rate $\hat{\mu}$ such that if the service rate is sufficiently high (i.e., $\mu \geq \hat{\mu}$), consumer surplus Ω is decreasing in the service rate μ.*

Do these results continue to hold when there are congestion costs, so that $c > 0$? For c sufficiently close to zero, the intuition of Propositions 8 and 9 goes through. In traditional queueing models, congestion costs are the sole factor under consideration; all else equal, all parties would prefer a higher service rate. In our model, there is a tradeoff between the informational content of a queue and the extra costs it imposes, and sometimes the former dominate.

We show via a numeric example that, even when the waiting cost c is well above zero, both the high quality good and the social planner may prefer a lower service rate to a higher one.

Example 3 The parameters are similar to those in Example 2. Consider $\lambda = 1$, $v_h = -v_l = 1$, $p = \frac{1}{2}$, $q = 0.7$, and $c = 0.14$. We consider four values for the service rate: $\mu = 1.34, 1.4, 1.72, 1.75$.

In each case, the joining rate per arrival for the high quality good and the consumer surplus per arrival are shown in Table 1.

Thus,

(i) increasing the service rate from 1.34 to 1.4 reduces the joining rate for the high quality good, but increases consumer surplus. The strategy played by agents in equilibrium remains the same. However, the increased service rate results in an increase in the steady state frequency of observing no consumers in the queue. In this state, a bad signal agent does not join, reducing the joining rate for the good.

(ii) Increasing the service rate from 1.72 to 1.75 increases the joining rate for the high quality good, but decreases consumer surplus. In equilibrium, when the service rate is 1.75,

Service rate μ	Agents' strategies $\{n \mid \alpha^*(g,n)=1\}$	$\{n \mid \alpha^*(b,n)=1\}$	Joining rate good h λ_h	Consumer surplus Ω
1.34	$\{0,1,\ldots,8\}$	$\{3,4,6\}$	0.737	0.095
1.4	$\{0,1,\ldots,8\}$	$\{3,4,6\}$	0.733	0.105
1.72	$\{0,1,\ldots,11\}$	$\{3,4,5,6,8,10\}$	0.725	0.136
1.75	$\{0,1,\ldots,11\}$	$\{2,4,5,6,8,10\}$	0.740	0.135

Table 1: Effects of changing service rate

a bad signal agent joins the queue at length 2. When the service rate is 1.72, she joins only at length 3. Since queue length 2 has higher probability, the overall effect is to reduce consumer surplus.

The example thus reinforces the notion that a queue communicates information about quality, and that this effect can dominate the effect of congestion costs. The presence of congestion costs makes the good less valuable for agents, so that they are less likely to purchase the good. Nevertheless, there are parameter values for which the joining rate for the high quality good increases with a decrease in the service rate, and similarly there are parameter values for which consumer surplus too is higher when the service rate is lower.

6 Conclusion

Queues develop for many new goods or services with unknown quality or value. These queues provide information about the quality or value of the good to subsequent agents and hence may create value. On the other hand, queues also generate waiting costs for consumers. To our knowledge, our model is the first to study this tradeoff analytically by combining theories of queueing and herding. Much of the literature on herding assumes that agents observe a truncated history (e.g., some random sample of history), rather than seeing the entire history of the game. We argue that it is natural to consider agents who observe a truncated set of actions: they observe some agents who have chosen to purchase a good, but do not see how many took the opposite decision. In this framework, a queueing model provides both truncations. Arrival and service departures are exogenous and random, so agents arriving at different times will see different queue lengths. Further, agents who have joined the queue are all purchasers, so no decisions made by non-purchasers are seen.

Our analysis of the equilibrium joining behavior reveals three insights that are not predicted by herding theory or queuing theory in isolation. First, agents with good or bad signals continue to play threshold strategies above which no agent joins, as is the case in traditional queuing models (Naor, 1969, Hassin and Haviv, 2003). Agents with bad signals may also have a threshold *below* which they do not join. That is, the expected value of agents with bad signals is high enough for them to join the queue only when the queue is sufficiently long. However, the equilibrium strategies cannot be characterized by upper and lower thresholds alone. We demonstrate that agents with bad signals may adopt strategies with "holes"—they may enter at queue lengths both below and above a particular level, but not at that level itself. At such holes, queues provide more precise information about the quality of the good.

Second, we show that there is an asymmetry between the high and low quality good when queuing is introduced: Queues for high quality goods are typically longer. Furthermore, when the waiting costs are not too high, the probability of (erroneously) joining a low quality good is higher than the probability of (erroneously) not joining a queue for a high quality good. This asymmetry is due to the one-sided nature of the servicing phenomenon. An agent that arrives when the queue is short cannot tell whether the situation arose because few agents joined the queue or because agents that joined were rapidly serviced. Since low quality goods have shorter queues, an agent is more likely to err when the good is of low quality.

Third, we demonstrate that *increasing* the service rate may lead to a *decrease* in the joining rate of a high quality good, and even to a *decrease* in total consumer surplus (i.e. the value net of the waiting costs) per unit of time. This feature is absent in the traditional queuing literature, and is driven by the positive information externality we have introduced. Most of the queuing literature focuses on perfect information in which increasing the service rate leads to an increase in (social) surplus. To understand our result, we develop and interpret sufficient conditions for this phenomenon in the absence of waiting costs, and then demonstrate that the phenomenon persists in the presence of small waiting costs. Obviously, when waiting costs are very high, an increase in the service rate will result in an increase in the equilibrium joining rate for the high quality good, and thereby to an increase in consumer surplus.

In this paper, we focus on understanding how queues communicate information about the quality of the good or service being provided. Our model relaxes the assumption of perfect information about the value of the product in the classical model of Naor (1969). Extensions of our model may consider more general arrival or service time distributions, multiple servers, and endogenous service rate selection and pricing. Our research suggests that obtaining analytical insights will be challenging. Nevertheless, simulation may be useful in understanding how strategic behavior by individual agents affects the overall demand for different goods and industries. Finally, our framework is amenable to laboratory experimentation to provide additional insights on how agents learn from queues.

Notes

[1]Hassin and Haviv also show knife-edge examples of non-threshold strategies that disappear with a slight perturbation on the parameters.

[2]Queueing, observing only the length of the queue, and waiting costs appear to resonate well with the canonical restaurant example of Banerjee (1992).

[3]Threshold strategies of this sort are common in queuing models, since waiting costs are strictly increasing in queue length (see, for example, Naor, 1969, Hassin and Haviv, 1997 or Hassin and Haviv, 2002).

[4]Chamley (2004), Definition 4.1.

7 Proofs

Proof of Lemma 2

$\pi_j(n;\alpha)$ is the stationary probability of observing a queue of length n when the good has quality j and agents play the strategy α. This is the long run probability of a birth-death process. The rate at which agents join the queue when the queue length is n is $\lambda r_j(n;\alpha)$. Once in the queue, agents leave at the rate μ (the death rate). Thus, given good quality j and agents' strategy α, the flow balance equations are

$$\begin{aligned} \pi_j(n-1;\alpha)\lambda r_j(n-1;\alpha) + \pi_j(n+1;\alpha)\mu &= \pi_j(n;\alpha)[\lambda r_j(n;\alpha)+\mu] \text{ for } n \geq 1 \\ \pi_j(1;\alpha)\mu &= \pi_j(0;\alpha)\lambda r_j(0;\alpha) \end{aligned}$$

Further,

$$\sum_{n=0}^{\infty} \pi_j(n;\alpha) = 1$$

Recursively solving this system of equations yields the expressions in the statement of the Lemma. ∎

Proof of Proposition 1.

We break the proof into two cases.

(i) When $c = 0$, existence follows from Theorem 5.4 of Rieder (1979).

(ii) Consider $c > 0$. Define $\bar{n}$ as the smallest integer greater than or equal to $\frac{v_h \mu}{c}$. Then, an agent will not join any queue of length $\bar{n}$ or higher, even if she knew with certainty that the good was high quality. Hence, we can restrict attention to $n \leq \bar{n}$.

Let $\mathcal{A}$ be the strategy space for an agent in the game. Recall that $\mathcal{S} = \{g, b\}$. Then, $\mathcal{A}$ is the space of mappings from $\mathcal{S} \times \{0, \cdots, \bar{n}\} \rightarrow [0,1]^{\bar{n}+1}$. Define a correspondence $\beta : \mathcal{A} \rightarrow \mathcal{A}$ as follows. For each $\alpha \in \mathcal{A}$, for each $s \in \mathcal{S}$ and $n \in \{0, \cdots, \bar{n}\}$, let $\beta(\alpha, n, s) = \arg\max_{x \in [0,1]} \; x u(s, n, \alpha)$.

Now, $\pi(v, n; \alpha)$ is continuous in $\alpha(s, n)$ for each j, s, n. Hence, the payoff function $u(\cdot)$ is also continuous in $\alpha(s, n)$ for each (s, n). Then, by Berge's Theorem of the Maximum, the correspondence β is compact, convex-valued, and upper hemi-continuous. Kakutani's fixed point theorem now implies that the map β has a fixed point.

By construction, this fixed point is a Markov perfect Bayesian equilibrium over the restricted strategy space $\mathcal{S} \times \{0, \ldots, \bar{n}\}$. For any $n > \bar{n}$, define $\theta_s(n;\alpha) = \theta_2(\bar{n};\alpha)$ and $\alpha(s, n) = 0$. ∎

Proof of Lemma 3

(i) Consider an agent who enters the system and sees a queue length n, when all other agents are playing the strategy α. The probability that she will see queue length n if the good quality is j is $\pi_j(n;\alpha)$, and is independent of her own signal (since it depends only on actions of previously arrived agents). From equation (2), we can write the posterior probability of the good being high quality as

$$\theta_s(n;\alpha) = \frac{\pi_h(n;\alpha)}{\pi_h(n;\alpha) + \frac{1-p_s}{p_s}\,\pi_\ell(n;\alpha)}.$$

Since $p_g > p_b$, it is immediate that $\theta_g(n;\alpha) > \theta_b(n;\alpha)$, and hence $u_g(n;\alpha) > u_b(n;\alpha)$.

Hence, an equilibrium strategy α^* must satisfy $\alpha^*(g,n) \geq \alpha^*(b,n)$. Now, since $q > \frac{1}{2}$, from the definitions of $r_h(n;\alpha)$ and $r_\ell(n;\alpha)$, it follows that $r_h(n;\alpha) \geq r_\ell(n;\alpha)$ for all n.

(ii) Suppose that $\alpha^*(g,n) = 0$ for some n. Since $\alpha^*(b,n) \leq \alpha^*(g,n)$ (from part (i)), it follows that $\alpha^*(b,n) = 0$. Therefore, $r_h(n;\alpha^*) = r_\ell(n;\alpha^*) = 0$. Equation (4) now implies that $\pi_h(\tilde{n};\alpha^*) = \pi_\ell(\tilde{n};\alpha^*)$ for all $\tilde{n} \geq n+1$.

■

Proof of Proposition 2

(i) From Lemma 3 (i), $r_h(n;\alpha^*) \geq r_\ell(n;\alpha^*)$ for all n. Suppose now that $r_h(n;\alpha^*) = r_\ell(n;\alpha^*)$ for all $n \in \mathbb{Z}_+$. Then, Lemma 2 implies that $\pi_h(n;\alpha^*) = \pi_\ell(n;\alpha^*)$ for all n, including $n = 0$. Therefore, the queue length is uninformative, and for each n, $\theta_s(n;\alpha^*) = p_s$. Hence, an agent with signal s will join the queue only if $v_\ell + p_s(v_h - v_\ell) \geq c(n+1)/\mu$. For $n = 0$, this condition reduces to $p_s \geq \frac{-v_\ell + c/\mu}{v_h - v_\ell}$.

From Assumption 1, part (ii), it follows that the last inequality holds strictly when $s = g$, and is violated when $s = b$. Hence, an agent with a good signal will join the queue when $n = 0$, and an agent with a bad signal will not. This contradicts the assumption that $r_h(n;\alpha^*) = r_\ell(n;\alpha^*)$ for all n. Therefore, it must be the case that there exists at least one $\hat{n}$ at which $r_h(\hat{n};\alpha^*) > r_\ell(\hat{n};\alpha^*)$.

Now, note that $\pi_j(0;\alpha^*) = \frac{1}{1+\sum_{k=1}^{\infty}\left(\frac{\lambda}{\mu}\right)^k r_j(k-1;\alpha^*)}$. Since $r_h(n;\alpha^*) \geq r_\ell(n;\alpha^*)$ with strict inequality for at least one $\hat{n}$, it follows that $\pi_\ell(0;\alpha^*) > \pi_h(0;\alpha^*)$.

(ii) Assumption 1, part (ii), implies that an agent with a bad signal who has no information from the queue will not join the empty queue. Since $\pi_\ell(0;\alpha^*) > \pi_h(0;\alpha^*)$, it follows that $\theta_b(0;\alpha) < p_b$. Hence, $u_b(0;\alpha) < 0$, so that $\alpha^*(b;0) = 0$.

Suppose $\alpha^*(g,0) = 0$. Since we have shown $\alpha^*(b,0) = 0$ as well, a queue length of zero will be absorbing, and the system will never reach a higher queue length. It follows that

$\theta_s(0,\alpha^*) = p$. That is, the revised belief the good is high quality, given queue length 0, is equal to the prior. But then $u_g(0,\alpha^*) > 0$ (by Assumption 1(ii)), which implies that $\alpha^*(g,0) = 1$, which is a contradiction. Hence, $\alpha^*(g,0) > 0$.

■

Proof of Proposition 3

(i) We first establish that for any $n, \tilde{n}$ such that $n \le \tilde{n} \le \bar{n}(c)$, $\frac{\pi_\ell(\tilde{n};\alpha^*)}{\pi_h(\tilde{n};\alpha^*)} \le \frac{\pi_\ell(n;\alpha^*)}{\pi_h(n;\alpha^*)}$.

By definition, $r_j(\hat{n};\alpha) > 0$ for all $n \le \bar{n}(c)$. Now, equation (4) implies that, for all $n \le \bar{n}(c) + 1$,

$$\frac{\pi_\ell(n;\alpha)}{\pi_h(n;\alpha)} = \prod_{k=0}^{n} \frac{r_\ell(n-1;\alpha)}{r_h(n-1;\alpha)}. \tag{16}$$

From part (i), $\frac{r_\ell(n;\alpha)}{r_h(n;\alpha)} \le 1$ for $n \le \hat{n} - 1$. Hence, for any $n, \tilde{n}$ such that $n \le \tilde{n} \le \hat{n}$, it follows that $\frac{\pi_\ell(\tilde{n};\alpha^*)}{\pi_h(\tilde{n};\alpha^*)} \le \frac{\pi_\ell(n;\alpha^*)}{\pi_h(n;\alpha^*)}$.

Now, since $\pi_\ell(0;\alpha^*) > \pi_h(0;\alpha^*)$ and $\frac{\pi_\ell(n;\alpha^*)}{\pi_h(n;\alpha^*)}$ is weakly declining in n for $n \le \bar{n}(c)$ and $\pi_\ell(0;\alpha^*) > \pi_h(0;\alpha^*)$, it follows that the queue length distribution for good h strictly first-order stochastically dominates the corresponding distribution for good ℓ. Hence, the expected queue length for good h is strictly higher than the expected queue length for good ℓ. That is,

$$\sum_{n=0}^{\bar{n}(c)} \pi_h(n;\alpha^*)n > \sum_{n=0}^{\bar{n}(c)} \pi_\ell(n;\alpha^*)n. \tag{17}$$

(ii) From equation (5), it follows that, for all n and for $j = h, \ell$,

$$\pi_j(n+1;\alpha^*) = \frac{\lambda}{\mu}\, \pi_j(n;\alpha^*) r_j(n;\alpha^*). \tag{18}$$

Hence, $\frac{\pi_j(n;\alpha^*) r_j(n;\alpha^*)}{\mu} = \frac{\pi_j(n+1;\alpha^*)}{\lambda}$, so that, for $j = h, \ell$,

$$W_j(\alpha^*;c) = \frac{1}{\lambda} \sum_{n=0}^{\bar{n}(c)} \pi_j(n+1;\alpha^*)(n+1). \tag{19}$$

Now, $\sum_{n=0}^{\bar{n}(c)} \pi_j(n+1;\alpha^*)(n+1) = \sum_{n=1}^{\bar{n}(c)} \pi_j(n;\alpha^*)n = \sum_{n=0}^{\bar{n}(c)} \pi_j(n;\alpha^*)n$. Thus, from equation (17), it follows that $W_h(\alpha^*;c) > W_\ell(\alpha^*;c)$. ■

Proof of Proposition 4

(i) When $c = 0$, the condition for joining the queue, $u_s(n;\alpha) \geq 0$, may be expressed as

$$v_\ell + \theta_s(n;\alpha)(v_h - v_\ell) \quad \geq \quad 0. \tag{20}$$

Substituting for θ_s from equation (2), we can rewrite this condition as

$$\frac{\pi_\ell(n;\alpha)}{\pi_h(n;\alpha)} \quad \leq \quad -\frac{p_s}{1-p_s}\frac{v_h}{v_\ell}. \tag{21}$$

Since $\alpha^*(g,0) > 0$, it must be that $\frac{\pi_\ell(0;\alpha^*)}{\pi_h(0;\alpha^*)} \leq -\frac{p_s}{1-p_s}\frac{v_h}{v_\ell}$. Further, since $\alpha^*(b,0) = 0$, we have $\frac{r_\ell(0;\alpha^*)}{r_h(0;\alpha^*)} = \frac{1-q}{q} < 1$. Therefore, $\frac{\pi_\ell(1;\alpha^*)}{\pi_h(1;\alpha^*)} = \frac{\pi_\ell(0;\alpha^*)}{\pi_h(0;\alpha^*)}\frac{r_\ell(0;\alpha^*)}{r_h(0;\alpha^*)} < \frac{\pi_\ell(0;\alpha^*)}{\pi_h(0;\alpha)}$. This now implies that $\frac{\pi_\ell(1;\alpha^*)}{\pi_h(1;\alpha^*)} < -\frac{p_s}{1-p_s}\frac{v_h}{v_\ell}$, so that $\alpha^*(g,1) = 1$. Finally, from the proof of Proposition 2, $\frac{\pi_\ell(n;\alpha^*)}{\pi_h(n;\alpha^*)}$ is weakly decreasing in n for each n. Hence, $\frac{\pi_\ell(n;\alpha^*)}{\pi_h(n;\alpha^*)} < -\frac{p_s}{1-p_s}\frac{v_h}{v_\ell}$ for each $n \geq 2$, so that $\alpha^*(g,n) = 1$ for each n.

We establish (ii) and (iii) simultaneously suppressing the dependence of π_ℓ, π_h on α^* for brevity. In the proof of Proposition 2 (ii), we have shown that $u_b(0;\alpha^*) < 0$, which implies that $\frac{\pi_\ell(0)}{\pi_h(0)} > -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$. As shown in part (i) above, $\frac{\pi_\ell(1)}{\pi_h(1)} = \frac{\pi_\ell(0)}{\pi_h(0)}\frac{r_\ell(0)}{r_h(0)} = \frac{\pi_\ell(0)}{\pi_h(0)}\frac{1-q}{q}$.

There are three cases to consider:

1. Suppose $\frac{\pi_\ell(1)}{\pi_h(1)} < -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$. Then $\alpha^*(b,1) = 1$, and as $\frac{r_\ell(n)}{r_h(n)} \leq 1$ for all n, $\frac{\pi_\ell(n)}{\pi_h(n)} < -\frac{v_h}{v_\ell}$ for all $n \geq 1$. Thus, $\alpha^*(b,n) = 1$ for all $n > 1$. Here, $\underline{n}_b = 1$.

2. Suppose $\frac{\pi_\ell(1)}{\pi_h(1)} = -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$. Then, $\alpha^*(b,1) \in [0,1]$. There are two sub-cases here:

   (a) If $\alpha^*(b,1) < 1$, then $\frac{r_\ell(1)}{r_h(1)} < 1$ so that $\alpha^*(b,n) = 1$ for all $n > 1$. Hence, $\underline{n}_b = 1$.

   (b) If $\alpha(b,1) = 1$, then $\frac{r_\ell(1)}{r_h(1)} = 1$, and hence $\frac{\pi_\ell(2)}{\pi_h(2)} = -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$; that is, at a queue length of 2, a bad signal agent is again indifferent between entering and not. If $\alpha^*(b,2) = 1$, then the same reasoning applies until the first queue length n at which $\alpha(b,n) < 1$, at which point for all queues greater than n, the agent will enter with probability 1. Here, $\underline{n}_b = 1$.

3. Suppose $\frac{\pi_\ell(1)}{\pi_h(1)} > -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$. Then, $\alpha^*(b,1) = 0$. Since $\alpha^*(g,1) = 1$ from part (i) above, $r_h(1) = q$ and $r_\ell(1) = 1-q$. Hence, $\frac{\pi_\ell(2)}{\pi_h(2)} = \frac{(1-q)^2}{q^2}\frac{\pi_\ell(0)}{\pi_h(0)}$. Now, observe that $\frac{p_b}{1-p_b} = \frac{p}{1-p}\frac{1-q}{q}$, and $\frac{p_g}{1-p_g} = \frac{p}{1-p}\frac{q}{1-q}$. Therefore, $\frac{p_g}{1-p_g} = \frac{q^2}{(1-q)^2}\frac{p_b}{1-p_b}$.

   But we know that $\frac{\pi_\ell(0)}{\pi_h(0)} \leq -\frac{p_g}{1-p_g}\frac{v_h}{v_\ell}$, so it follows that $\frac{(1-q)^2}{q^2}\frac{\pi_\ell(0)}{\pi_h(0)} \leq -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$, or $\frac{\pi_\ell(2)}{\pi_h(2)} \leq -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$.

   There are now three sub-cases to consider:

(a) $\frac{\pi_\ell(2)}{\pi_h(2)} < -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$. Then, $\alpha^*(b,2) = 1$, and by the same logic as in Case 1 above, it follows that $\alpha^*(b,n) = 1$ for all $n \geq 3$ as well. Here, $\underline{n}_b = 2$.

(b) $\frac{\pi_\ell(2)}{\pi_h(2)} = -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$, and $\alpha^*(b,2) < 1$. Then, $r_h(2) > r_\ell(2)$, so $\frac{\pi_\ell(3)}{\pi_h(3)} < -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$, which implies that $\alpha^*(b,3) = 1$. Further, since $r_\ell(n) \leq r_h(n)$ for all n, $\alpha^*(b,n) = 1$ for all $n \geq 3$. Here, $\underline{n}_b = 3$.

(c) $\frac{\pi_\ell(2)}{\pi_h(2)} = -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$, and $\alpha^*(b,2) = 1$. Then, $r_h(2) = r_\ell(2)$, so $\frac{\pi_\ell(3)}{\pi_h(3)} = -\frac{p_b}{1-p_b}\frac{v_h}{v_\ell}$, which implies that $\alpha^*(b,3) \in [0,1]$, so that a bad signal agent is again indifferent between joining and not. By the same logic as in (b2) above, if $\alpha^*(b,3) < 1$, then $\alpha^*(b,n) = 1$ for all $n \geq 4$. If $\alpha^*(b,3) = 1$, the process is repeated at a queue length of 4, and so on. Here, $\underline{n}_b = 3$. ■

Proof of Proposition 5

(i) Define $\bar{n}(c)$ as the smallest queue length at which $\alpha^*(\bar{n}(c)) = 0$. If the queue length exceeds $\frac{v_h\mu}{c} - 1$, even an agent who knows the good has high quality will not join the queue. Thus, $\bar{n}(c) \leq \frac{v_h\,\mu}{c} - 1$. Further, it must be that $\alpha^*(g,n) > 0$ for all $n \in \{0, \cdots, \bar{n}(c) - 1\}$, else the queue length $\bar{n}(c)$ can never be reached.

(ii) The statement is proved via example. For the parameter values in Example 2, an agent with a bad signal does not follow a threshold strategy. ■

Proof of Proposition 6

Since $\alpha^*(g,n) = 1$ for all $n \leq \bar{n}(c) - 1$, it follows that $r_h(n) = q + (1-q)\alpha^*(b,n)$ and $r_\ell(n) = 1 - q + q\alpha^*(b,n)$ for n in this range. Further, $\pi_\ell(n;\alpha^*) = \pi_h(n;\alpha^*) = 0$ for $n > \bar{n}(c)$. From the definitions of $\epsilon_h, \epsilon_\ell$, we have

$$\begin{aligned}\epsilon_\ell(\alpha^*) - \epsilon_h(\alpha^*) &= \sum_{n=0}^{\bar{n}(c)} [\pi_\ell(n;\alpha^*)\,(1 - q + q\alpha^*(b,n)) \;-\; \pi_h(n;\alpha^*)\,(1 - q - (1-q)\alpha^*(b,n))] \\ &= (1-q)\sum_{n=0}^{\bar{n}(c)}[\pi_\ell(n;\alpha^*) - \pi_h(n,\alpha^*)] + \sum_{n=0}^{\bar{n}(c)} [\pi_\ell(n;\alpha^*)\,q \;+\; \pi_h(n;\alpha^*)\,(1-q)]\;\alpha^*(b,n))\end{aligned}$$

Since $\sum_{n=0}^{\bar{n}(c)} \pi_\ell(n;\alpha^*) = \sum_{n=0}^{\bar{n}(c)} \pi_h(n;\alpha^*) = 1$, the last equation reduces to

$$\epsilon_\ell(\alpha^*) - \epsilon_h(\alpha^*) \;=\; \sum_{n=0}^{\bar{n}(c)} [\pi_\ell(n;\alpha^*)\,q \;+\; \pi_h(n;\alpha^*)\,(1-q)]\;\alpha^*(b,n)) \;>\; 0, \qquad (22)$$

since there exists an $\hat{n}$ for which $\alpha^*(b,\hat{n}) > 0$. ■

Proof of Proposition 7

Suppose $c > 0$ and $v < \frac{2c}{\mu}$. Consider the strategy $\alpha^*(g,0) = 1$, $\alpha^*(g,n) = 0$ for all $n \geq 1$, and $\alpha^*(b,n) = 0$ for all n. We will show that if all other agents are playing α^*, it is a best response for a randomly arrived agent to also play α^*. Thus, α^* is an equilibrium strategy.

Suppose all previously arrived agents play α^*. Then, a randomly arriving agent sees a queue length of either zero or one. Further, $r_h(0) = q$ and $r_\ell(0) = 1-q$, with $r_h(n) = r_\ell(n) = 0$ for all $n \geq 1$. From Lemma 3, it follows that $\pi_\ell(0) = \frac{1}{1+\lambda(1-q)/\mu}$, and $\pi_h(0) = \frac{1}{1+\lambda q/\mu}$.

Now, an agent with signal s will join the queue at zero if $u_s(0;\alpha) > 0$, or

$$v_\ell + \theta_s(0;\alpha)(v_h - v_\ell) \quad \geq \quad \frac{c}{\mu},$$

which may be written as

$$\frac{1-p_s}{p_s}\frac{\pi_\ell(0)}{\pi_h(0)} \quad \leq \quad \frac{v_h - c/\mu}{c/\mu - v_\ell}.$$

Now, note that $\frac{1-p_g}{p_g} = \frac{(1-p)}{p}\frac{q}{1-q}$, and $\frac{1-p_b}{p_b} = \frac{(1-p)}{p}\frac{1-q}{q}$. Substituting in also the values of $\pi_\ell(0)$ and $\pi_h(0)$ exhibited earlier, an agent with a good signal will strictly prefer to join the queue at zero, and an agent with a bad signal will strictly prefer to balk, if the following two conditions are satisfied:

$$\frac{1-q}{q}\left[\frac{1-p}{p}\frac{1+\lambda q/\mu}{1+\lambda(1-q)/\mu}\right] \quad < \quad \frac{v_h - c/\mu}{c/\mu - v_\ell} \tag{23}$$

$$\frac{q}{1-q}\left[\frac{1-p}{p}\frac{1+\lambda q/\mu}{1+\lambda(1-q)/\mu}\right] \quad > \quad \frac{v_h - c/\mu}{c/\mu - v_\ell} \tag{24}$$

Given any values for other parameters, there exists a $\hat{q}$ sufficiently close to one such that, for all $q \in (\hat{q}, 1)$ both conditions (23) and (24) are satisfied. Hence, if q is in this range, and all other agents play α^*, a randomly arrived agent should also play $\alpha^*(g,0) = 1$ and $\alpha^*(b,0) = 0$.

Further, note that, since $v_h < \frac{2c}{\mu}$, it is optimal for an agent to not join a queue if the queue length is $n \geq 1$. Therefore, if all other agents are playing α^*, it is a best response for a randomly arriving agent to also play α^*. Hence, α^* constitutes an equilibrium.

It is immediate that α^* displays no herding, since $\bar{n}(c) = 0$ and $\alpha^*(g,0) \neq \alpha^*(b,0)$. ■

Proof of Proposition 8

First, we show that α^* as defined in the statement of the Proposition is indeed an equilibrium. Suppose all other players are playing α^*, and let α denote the best response of a randomly arriving player who sees n agents in the queue. Given α^*, we have $r_h(0) = q$, $r_\ell(0) = q$, and $r_h(n) = r_\ell(n) = 1$ for all $n \geq 1$. Hence,

$$\pi_h(0) \quad = \quad \frac{1}{1 + q\sum_{n=1}^{\infty}(\lambda/\mu)} = \frac{1}{1 + q\,\frac{\lambda/\mu}{1-\lambda/\mu}} = \frac{1-\lambda/\mu}{1-(1-q)\lambda/\mu}, \tag{25}$$

where the second equality uses the fact that $\frac{\lambda}{\mu} < 1$ when $c = 0$.

Similarly,

$$\pi_\ell(0) \;=\; \frac{1}{1+(1-q)\sum_{n=1}^{\infty}(\lambda/\mu)} = \frac{1-\lambda/\mu}{1-q\lambda/\mu}. \tag{26}$$

Hence, $\frac{\pi_\ell(0)}{\pi_h(0)} = \frac{1-(1-q)\lambda/\mu}{1-q\lambda/\mu}$.

Now, suppose $\frac{\lambda}{\mu} < \frac{2p-1}{q+p-1}$. Suppose further that $q+p-1<0$. Then, it must be that $p<\frac{1}{2}$, since $q > \frac{1}{2}$. Hence, $\frac{2p-1}{q+p-1} > 1$, so it cannot be (given Assumption 1 (i))be $\frac{\lambda}{\mu} < \frac{2p-1}{q+p-1}$. Hence, it must be that $q+p-1>0$.

Going back to the condition $\frac{\lambda}{\mu} < \frac{2p-1}{q+p-1}$ and multiplying both sides by $q+p-1$, we have

$$\begin{aligned} \frac{\lambda}{\mu}\,(q+p-1) &< 2p-1 \\ \frac{\lambda}{\mu}\,[pq-(1-p)(1-q)] &< p-(1-p) \\ \frac{1-p}{p}\,\frac{1-(1-q)\lambda/\mu}{1-q\lambda/\mu} &< 1. \end{aligned} \tag{27}$$

Now, note that $\frac{1-(1-q)\lambda/\mu}{1-q\lambda/\mu} = \frac{\pi_\ell(0)}{\pi_h(0)}$, $-\frac{v_h}{v_\ell} = 1$, $\frac{r_\ell(0)}{r_h(0)} = \frac{1-q}{q}$, and $\frac{1-p_b}{p_b} = \frac{(1-p)q}{p(1-q)}$. Substituting these into (27), we have

$$\frac{1-p_b}{p_b}\,\frac{\pi_\ell(0)}{\pi_h(0)}\,\frac{r_\ell(0)}{r_h(0)} \;<\; -\frac{v_h}{v_\ell}. \tag{28}$$

Hence, $\alpha(b,1)=1$ is a best response to α^*. Further, implies that $r_h(1) = r_\ell(1) = 1$, so that $\alpha(b,2)=1$, and so on for all $n \geq 1$. Hence, $\alpha(g,n) = 1$ for all $n \geq 1$.

Consider $\alpha(g,0)$. Since $q > \frac{1}{2}$, it follows from equation (27) that

$$\begin{aligned} \frac{1-q}{q}\,\frac{1-p}{p}\,\frac{1-(1-q)\lambda/\mu}{1-q\lambda/\mu} &< 1 \\ \frac{1-p_g}{p_g}\,\frac{\pi_\ell(0)}{\pi_h(0)} &< -\frac{v_h}{v_\ell}, \end{aligned} \tag{29}$$

so that $\alpha(g,0)=1$.

Hence, α^* is a best response to itself, and is therefore an equilibrium strategy.

Now, $\pi_j(n+1) = \pi_j(n)\frac{\lambda}{\mu}r_j(n)$, for each $n = 0,1,\ldots$ and $j = \ell, h$. Hence, $\pi_j(n)r_j(n) = \pi_j(n+1)\frac{\mu}{\lambda}$. Thus,

$$\begin{aligned} \lambda_h &= \frac{\mu}{\lambda}\sum_{n=0}^{\infty}\pi_h(n+1) = \frac{\mu}{\lambda}\,(1-\pi_h(0)) \\ &= \frac{q}{1-(1-q)\frac{\lambda}{\mu}}, \end{aligned}$$

which is clearly decreasing in μ. ■

Proof of Proposition 9

As shown in Proposition 8, in the identified equilibrium α^*, we have $\lambda_h = \frac{q}{1-(1-q)\frac{\lambda}{\mu}}$. Using the same technique, we can show that $\lambda_\ell(\alpha^*) = \frac{1-q}{1-q\frac{\lambda}{\mu}}$.

Set $c = 0$ and $v_h = -v_\ell = v$. Then, consumer surplus Ω reduces to $\Omega = v[p\lambda_h+(1-p)\lambda_\ell]$.

$$\frac{d\Omega}{d\mu} = v\left[p\frac{d\lambda_h}{d\mu} - (1-p)\frac{d\lambda_\ell}{d\mu}\right]. \quad (30)$$

Using the expressions for λ_h, λ_ℓ above,

$$\frac{d\Omega}{d\mu} = q(1-q)v\left(\frac{\lambda}{\mu}\right)^2\left[\frac{1-p}{(1-(1-q)\lambda/\mu)^2} - \frac{p}{(1-q\lambda/\mu)^2}\right]. \quad (31)$$

Hence, $\frac{d\Omega}{d\mu} < 0$ if and only if $\frac{1-p}{(1-(1-q)\lambda/\mu)^2} < \frac{p}{(1-q\lambda/\mu)^2}$. As $\mu \to \infty$, the latter inequality reduces in the limit to $1-p < p$, or $p > \frac{1}{2}$.

We show that the conditions of Proposition 8 imply that $p > \frac{1}{2}$. Consider the condition $\frac{\lambda}{\mu} < \frac{2p-1}{q+p-1}$. As shown in the proof of Proposition 8, if this condition is satisfied, then $q+p-1 > 0$. Since λ, μ are positive by definition, it must be that $2p-1 > 0$, or $p > \frac{1}{2}$.

Hence, as $\mu \to \infty$, in the limit $\frac{d\Omega}{d\mu} < 0$. Since Ω is continuous in μ, it follows there is a $\hat{mu}$ large enough such that for all $\mu \geq \hat{\mu}$, $\frac{d\Omega}{d\mu} < 0$ and the consumer surplus decreases in the service rate. ■

8 Acknowledgements

We thank seminar participants at Columbia University, the Canadian Economic Theory Conference (Vancouver, 2005), the Carnegie Mellon/Technische Universiteit Eindhoven Collaborative Workshop (Eindhoven, 2005), the Manufacturing and Service Operations Management meeting (Chicago, 2005) and the queuing economics seminar at the California Institute of Technology (Pasadena, 2007). The current version of this paper is maintained at http://webuser.bus.umich.edu/urajan/research/infer.pdf.

References

[1] Banerjee Abhijit (1992) "A Simple Model of Herd Behaviour," *Quarterly Journal of Economics* 107, 797-817.

[2] Banerjee, Abhijit (1993), "The Economics of Rumours," *Review of Economic Studies* 60(2), 309-327.

[3] Banerjee, Abhijit and Drew Fudenberg, (2002) " Word of Mouth Learning," Harvard working paper.

[4] Becker, Gary A. (1991) "A Note on Restaurant Pricing and Other Examples of Social Influences on Price." *Journal of Political Economy* 99(51), 1109-1116.

[5] Bikhchandani, Sushil, David Hirshleifer and Ivo Welch (1992) "A Theory of Fads, Fashion, Custom and Cultural change as Information Cascades," *Journal of Political Economy* 100, 992-1026.

[6] Celen, Bogachan and Shachar Kariv, "Observational Learning under Imperfect Information," *Games and Economic Behavior* 47(1), 72-86.

[7] Chamley, Chrisophe P. (2004), "Rational Herds", Cambridge University Press, Cambridge UK.

[8] Ellison, Glenn and Drew Fudenberg (1995) "Word-of-mouth Communication and Social Learning," *Quarterly Journal of Economics* 110(1), 93-125.

[9] Ellison, Glenn and Drew Fudenberg (1993) "Rules of Thumb for Social Learning," *Journal of Political Economy* 101(4), 612-643.

[10] Hassin, Refael and Moshe Haviv (1997) "Equilibrium Threshold Strategies: The Case of Queues with Priorities," *Operations Research*, 45(6), 966-973.

[11] Hassin, Refael and Moshe Haviv, (2003) " To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems." Kluwer Academic Publishers.

[12] Naor, Pinhas (1969), "The Regulation of Queue Size by Levying Tolls," *Econometrica* 37, 15-34.

[13] Rieder, U. (1979), "Equilibrium Plans for Non–Zero Sum Markov Games," 91-101 in *Game Theory and Related Topics* O. Moeschlin, D. Pallascke (eds.) North Holland.

[14] Smith, Lones and Peter Sorensen (2000), "Pathological Outcomes of Observational Learning," *Econometrica* 68(2), 371-398.

[15] Wolff, Ronald W. (1982) "Poisson Arrivals See Time Averages", Operations Research 30, 223-231.