Queueing Games

Moshe Haviv

Department of Statistics
The Hebrew University of Jerusalem

Based on *"To queue or not to queue: Equilibrium behavior in queueing systems"* co-authored with Refael Hassin (Kluwer, 2003).

Tel Aviv University

March, 2006

Single server M/M/1 queue

- Poisson arrivals. Effective rate of λ.
- Exponential requirements. Mean $\frac{1}{\mu}$.
- $\rho = \lambda/\mu < 1$, server utilization.
- First-come first-served (FCFS).

Single server M/M/1 queue

- steady-state, mean waiting time (service inclusive)

$$W = \frac{1}{\mu}\frac{1}{1-\rho} = \frac{1}{\mu - \lambda}$$

Single server M/M/1 queue

- steady-state, mean waiting time (service inclusive)

$$W = \frac{1}{\mu}\frac{1}{1-\rho} = \frac{1}{\mu-\lambda}$$

- for one who receives service only when the server is otherwise idle (zero-priority)

$$\frac{1}{\mu}\frac{1}{(1-\rho)^2} = \frac{1}{\mu}\frac{1}{(1-\lambda/\mu)^2}$$

To join or not to join

- Λ, potential arrival rate
- C, cost per unit of waiting time
- R, reward per service completion

Assume $R > C/\mu$ and $\Lambda < \mu$.

Unobservable queues

Edelson and Hildebrand ’75

Self-optimization (Nash equilibrium)

- Pure strategies: to join or not to join
- Randomization is allowed

Unobservable queues

Edelson and Hildebrand '75

Self-optimization (Nash equilibrium)

- Pure strategies: to join or not to join
- Randomization is allowed
- Join with probability p: A symmetric profile
- An equilibrium p_e:
  steady-state under $p_e \Rightarrow p_e$ is a best response.

Equilibrium

- $C/(\mu - \Lambda) < R \Rightarrow p_e = 1$, dominant strategy

Equilibrium

- $C/(\mu - \Lambda) < R \Rightarrow p_e = 1$, dominant strategy
- $C/(\mu - \Lambda) \geq R \Rightarrow p_e$, mixed strategy

  $p_e = (\mu - C/R)/\Lambda$ which solves
  $C/(\mu - p\Lambda) = R$.

Equilibrium

- $C/(\mu - \Lambda) < R \Rightarrow p_e = 1$, dominant strategy
- $C/(\mu - \Lambda) \geq R \Rightarrow p_e$, mixed strategy

  $p_e = (\mu - C/R)/\Lambda$ which solves
  $C/(\mu - p\Lambda) = R$.

p_e is evolutionarily **stable** strategy (ESS):
If $p \neq p_e$ is best against p_e, p_e is better than p against p.

Equilibrium

case	λ_e	p_e	$W(\lambda_e)$
$\Lambda \leq \mu - \frac{C}{R}$	Λ	1	$\frac{1}{\mu-\Lambda}$
$\Lambda \geq \mu - \frac{C}{R}$	$\mu - \frac{C}{R}$	$\frac{\mu-\frac{C}{R}}{\Lambda}$	$\frac{R}{C}$

Equilibrium

$f(p)$: the individual's difference in utility between joining and not when all use strategy p

Equilibrium

$f(p)$: the individual's difference in utility between joining and not when all use strategy p

$$f(p) = R - C/(\mu - p\Lambda) - 0.$$

Equilibrium

$f(p)$: the individual's difference in utility between joining and not when all use strategy p

$$f(p) = R - C/(\mu - p\Lambda) - 0.$$

$$f(p) \downarrow p$$

Equilibrium

$f(p)$: the individual's difference in utility between joining and not when all use strategy p

$$f(p) = R - C/(\mu - p\Lambda) - 0.$$

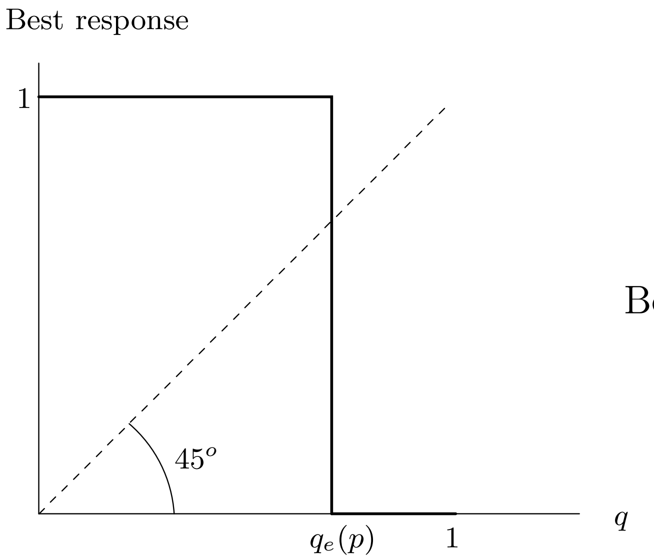
$$f(p) \downarrow p$$

The more join, the less appealing is joining.

Equilibrium

Best response against p:

- $p \in [0, p_e] \Rightarrow 1$, i.e., join
- $p \in [p_e, 1] \Rightarrow 0$, i.e., do not join
- $p = p_e < 1 \Rightarrow$ any p.



Best response $\downarrow p \Rightarrow$ **Avoid the crowd (ATC)**

Choosing a transportation means

Hassin and Haviv '03

- Poisson arrivals with λ
- A bus or a cab?

Choosing a transportation means

Hassin and Haviv '03

- Poisson arrivals with λ
- A bus or a cab?
  - Bus: 5 minutes of waiting

Choosing a transportation means

Hassin and Haviv '03

- Poisson arrivals with λ
- A bus or a cab?
  - Bus: 5 minutes of waiting
  - Cabs: depart one after the other when full (7 passengers)

Choosing a transportation means

Hassin and Haviv '03

- Poisson arrivals with λ
- A bus or a cab?
  - Bus: 5 minutes of waiting
  - Cabs: depart one after the other when full (7 passengers)
- Unobservable

The bus or the cab? No regrets.

Choosing a transportation means

The cab is more appealing the more it is selected by others.

Choosing a transportation means

The cab is more appealing the more it is selected by others.

p: the probability of selecting the cab.

Choosing a transportation means

The cab is more appealing the more it is selected by others.

p: the probability of selecting the cab.

- $\lambda < 3/5 \Rightarrow p_e = 0$, dominant strategy

Choosing a transportation means

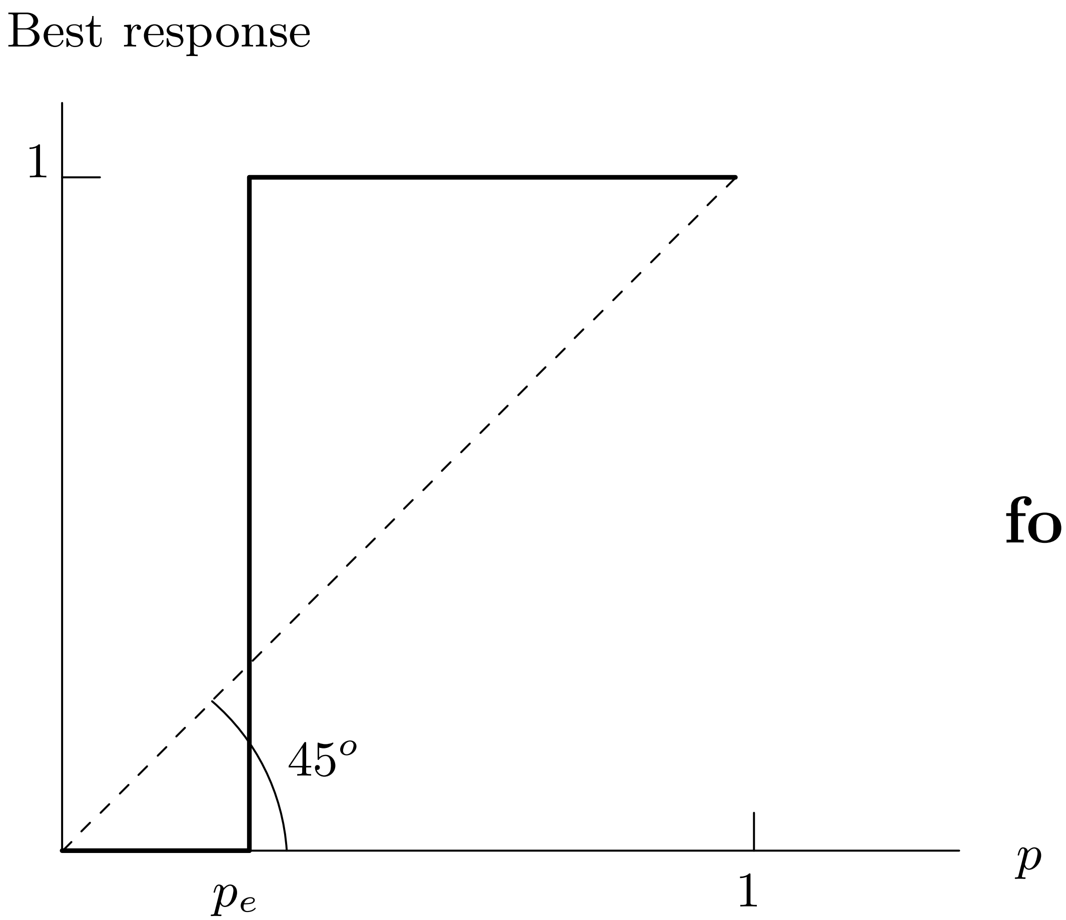
The cab is more appealing the more it is selected by others.

p: the probability of selecting the cab.

- $\lambda < 3/5 \Rightarrow p_e = 0$, dominant strategy
- $\lambda > 3/5 \Rightarrow$ three equilibria:
  - $p_e = 0$ (stable)
  - $p_e = 1$ (stable)
  - $p_e = 3/(5\lambda)$ (unstable)

Choosing a transportation means

The best response function.



follow the crowd (FTC)

Back to M/M/1. Social optimization

$$\lambda^* = \arg\max_{0\leq\lambda\leq\Lambda} \lambda \left(R - \frac{C}{\mu - \lambda} \right)$$

or

$$p^* = \arg\max_{0\leq p\leq 1} p\Lambda \left(R - \frac{C}{\mu - p\Lambda} \right)$$

Back to M/M/1. Social optimization

$$\lambda^* = \arg\max_{0 \le \lambda \le \Lambda} \lambda \left(R - \frac{C}{\mu - \lambda} \right)$$

or

$$p^* = \arg\max_{0 \le p \le 1} p\Lambda \left(R - \frac{C}{\mu - p\Lambda} \right)$$

- $\Lambda \le \mu - \sqrt{C\mu / R} \Rightarrow p^* = 1.$
- $\Lambda \ge \mu - \sqrt{C\mu / R} \Rightarrow p^* = (\mu - \sqrt{C\mu / R}) / \Lambda.$

Social optimization

case	λ^*	p^*	$W(\lambda^*)$	welfare
$\Lambda \geq \mu - \sqrt{\frac{C\mu}{R}}$	$\mu - \sqrt{\frac{C\mu}{R}}$	$\frac{\lambda^*}{\Lambda}$	$\sqrt{\frac{R}{C\mu}}$	$\left(\sqrt{R\mu} - \sqrt{C}\right)^2$
$\Lambda \leq \mu - \sqrt{\frac{C\mu}{R}}$	Λ	1	$\frac{1}{\mu-\Lambda}$	$\Lambda\left(R - \frac{C}{\mu-\Lambda}\right)$

Comparison

$$p_e \geq p^*$$

$$\lambda_e \geq \lambda^*$$

Equality $\Leftrightarrow p^* = 1$

Comparison

$p_e \geq p^*$

$\lambda_e \geq \lambda^*$

Equality $\Leftrightarrow p^* = 1$

While ignoring the negative externalities, customers tend to overcrowd the system.

Optimal entry fee

- T: an entry fee that makes all behave optimally

Optimal entry fee

- T: an entry fee that makes all behave optimally
- $p^* < 1 \Rightarrow T = R - C/(\mu - p^*\Lambda)$

Optimal entry fee

- T: an entry fee that makes all behave optimally
- $p^* < 1 \Rightarrow T = R - C/(\mu - p^*\Lambda)$
- $T = R - \sqrt{CR/\mu} = \lambda^* C W'(\lambda^*)$

Optimal entry fee

- T: an entry fee that makes all behave optimally
- $p^* < 1 \Rightarrow T = R - C/(\mu - p^*\Lambda)$
- $T = R - \sqrt{CR/\mu} = \lambda^* C W'(\lambda^*)$

$$\Downarrow$$

$T =$ the externalities an arrival imposes under λ^*.

Optimal entry fee

- T: an entry fee that makes all behave optimally
- $p^* < 1 \Rightarrow T = R - C/(\mu - p^*\Lambda)$
- $T = R - \sqrt{CR/\mu} = \lambda^* C W'(\lambda^*)$

$$\Downarrow$$

$T =$ the externalities an arrival imposes under λ^*.

By paying T, customers internalize their externalities.

Optimal waiting costs

Change C so as the new λ_e coincides with the old λ^*.

Optimal waiting costs

Change C so as the new λ_e coincides with the old λ^*.

- $C_{new} \leftarrow C/(1 - \lambda^*/\mu)$

Optimal waiting costs

Change C so as the new λ_e coincides with the old λ^*.

- $C_{new} \leftarrow C/(1 - \lambda^*/\mu)$
- Under λ^* and C_{new}, all suffer as a zero-priority customer under λ^* and the original C

More on regulation

$$\frac{C}{\mu(1-\frac{\lambda^*}{\mu})^2} = R$$

More on regulation

$$\frac{C}{\mu(1-\frac{\lambda^*}{\mu})^2} = R$$

If all believe they are zero priority customers

$$\Downarrow$$

$\lambda^* \Rightarrow$ indifference between joining and not joining

$$\Downarrow$$

p^* is an equilibrium

More on regulation

$$\frac{C}{\mu(1-\frac{\lambda^*}{\mu})^2} = R$$

If all believe they are zero priority customers

$$\Downarrow$$

$\lambda^* \Rightarrow$ indifference between joining and not joining

$$\Downarrow$$

p^* is an equilibrium

Self-optimization = social optimization: All believe that they do not impose any externalities.

Regulation by auctioning positions

Hassin '95

- To join or not to join.
- The more one pays, the higher is one's (pre-emptive) priority.
- Ties are broken by FCFS.

Assume: $\lambda^* < \Lambda \Rightarrow$ everybody's utility is zero

Equilibrium: Join and pay with probability p^*.
Pay X, a continuous random variable on $[0, a]$ with cumulative distribution function F.

$$a = R - \frac{C}{\mu}.$$

More on auctioning

W_x mean waiting time if x is paid, $x \in [0, a]$.

$$x = R - CW_x = CW_0 - CW_x$$

$x =$ externalities imposed by an x customer.

$$\frac{x - R}{C} = W_x = \frac{1}{\mu} \frac{1}{(1 - \lambda^*(1 - F(x))/\mu)^2}$$

$$F(x) = \frac{\sqrt{\frac{C}{R-x}} - \sqrt{\frac{C}{R}}}{\sqrt{\mu} - \sqrt{\frac{C}{R}}}, \; 0 \le x \le a.$$

Profit maximization

$$\begin{array}{rl} \text{maximize} & \lambda T \\ \text{subject to} & C/(\mu - \lambda) = R - T, \quad 0 \le \lambda \le \Lambda \end{array}$$

$\lambda^* < \Lambda \Rightarrow$ Profit maximization and social optimization tolls coincide.

$$\Downarrow$$

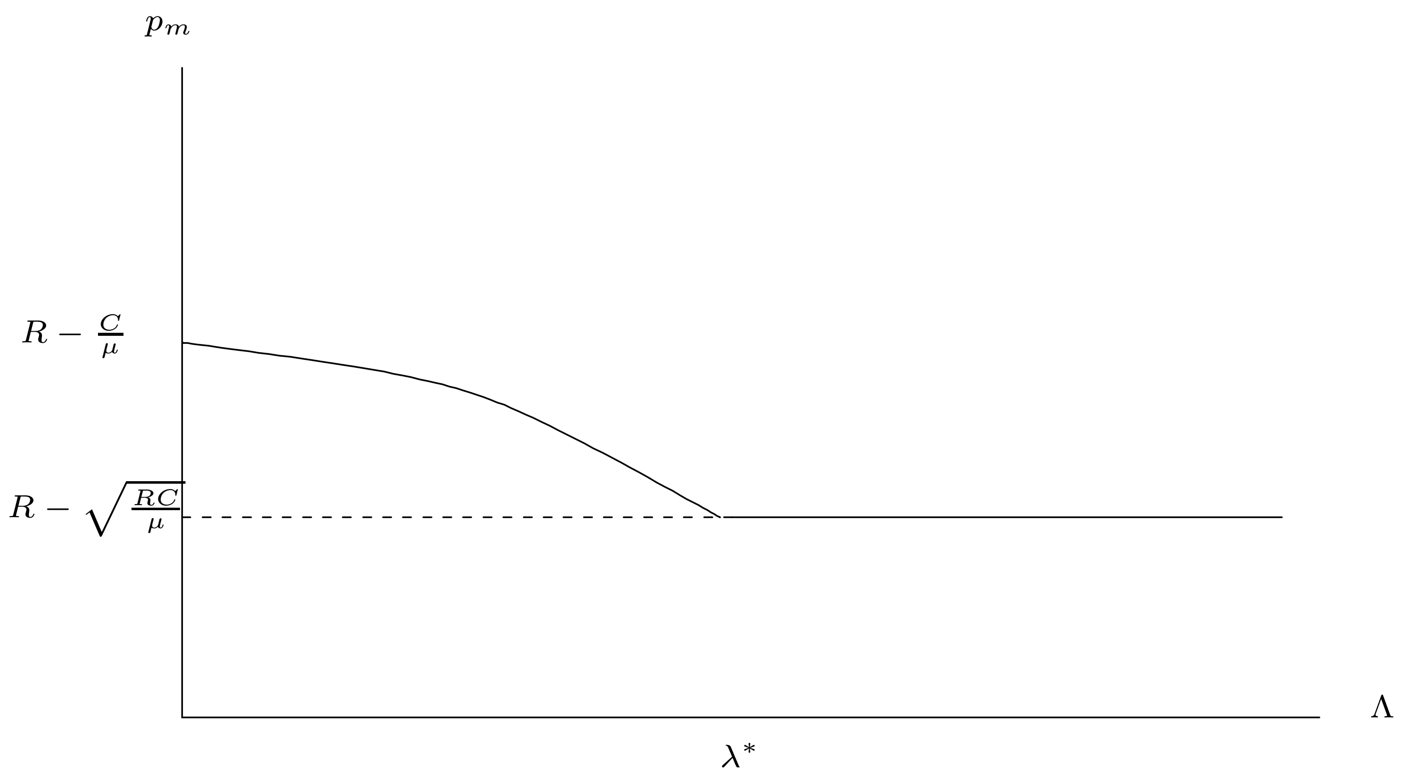
The arrival rates under both criteria coincide

Since consumers do not possess any private information, a monopoly can extract all their surplus

Profit maximization

Chen and Frank '00

Profit maximization prices are monotone non-increasing with demand.



More demand $\Rightarrow$ lower prices.

Heterogeneous service value

$V(\lambda)$ - social gain from service completions
$V(\lambda) \uparrow \lambda$ and concave

Alternatively, random reward: $R \sim F \leftrightarrow f$.

$\lambda \Leftrightarrow$ those with $R \geq r(\lambda)$ join

$\lambda = \Lambda(1 - F(r(\lambda)))$. $V(\lambda) = \Lambda \int_{r(\lambda)}^{\infty} r f(r)\, dr$

λ_e solves

$$V'(\lambda_e) = CW(\lambda_e).$$

$$\Downarrow$$

$$\{ \text{Join} \Leftrightarrow R \geq r(\lambda_e) \}$$

Social optimization

$$\lambda^* = \arg\max_{\lambda \geq 0} \{V(\lambda) - \lambda CW(\lambda)\}$$

$$\Downarrow$$

$$V'(\lambda^*) = CW(\lambda^*) + \lambda^* CW'(\lambda^*).$$

$$\{\ \text{Join} \Leftrightarrow R \geq r(\lambda^*)\ \}$$

Regulating entry fee: $T = \lambda^* CW'(\lambda^*)$

$$\Downarrow$$

new λ_e = old λ^*

Multi-class heterogeneous valuations Mendelson and Whang '90

m classes, C_i, λ_i, $V_i(\lambda_i)$ but $\mu_i = \mu$.
Priority by the $C\mu$ rule

$$\underline{\lambda^*} = \arg \max_{\lambda_1,\ldots,\lambda_m} \sum_{i=1}^{m} (V_i(\lambda_i) - \lambda_i W^(\underline{(}\lambda *)))$$

$$T_i^* = \sum_{j=1}^{m} C_j \lambda_j^* \frac{\partial W_j(\underline{\lambda})}{\partial \lambda_i}|_{\underline{\lambda}=\underline{\lambda}^*}, \ \ 1 \leq i \leq m$$

T_i = externalities

These tolls are incentive compatible.

Observable M/M/1 queues

Naor ’69

- M/M/1
- same cost/reward model
- queue length inspected upon arrival

Observable M/M/1 queues

Naor '69

- M/M/1
- same cost/reward model
- queue length inspected upon arrival

To join or not to join?

Observable M/M/1 queues

Naor '69

- M/M/1
- same cost/reward model
- queue length inspected upon arrival

To join or not to join?

Equilibrium: { Join $\Leftrightarrow L < L_e$ }

$$L_e = \lfloor \frac{R\mu}{C} \rfloor$$

Equilibrium

Multi-equilibria: Prescribe anything for $L \geq L_e + 1$.

Equilibrium

Multi-equilibria: Prescribe anything for $L \geq L_e + 1$.

$\{$ Join $\Leftrightarrow L < L_e$ $\}$ is unique subgame-perfect equilibrium (SPE).

Equilibrium

Multi-equilibria: Prescribe anything for $L \geq L_e + 1$.

$\{$ Join $\Leftrightarrow L < L_e$ $\}$ is unique subgame-perfect equilibrium (SPE).

L_e is not a function of Λ.

Social optimization:

$$\{ \text{Join} \Leftrightarrow L \leq L^* \}$$

$$L^* = \max_n \{ \frac{n(1-\rho) - \rho(1-\rho)^n}{(1-\rho)^2} \leq \frac{R\mu}{C} \}$$

$\rho = \Lambda / \mu$

Social optimization:

$$\{ \text{Join} \Leftrightarrow L < L^* \}$$

$$L^* = \max_n \{ \frac{n(1-\rho) - \rho(1-\rho)^n}{(1-\rho)^2} \leq \frac{R\mu}{C} \} - 1$$

$\rho = \Lambda / \mu$

Comparison: $L^* \leq L_e$.

Regulating observable queues

1. By entry fee T:

$$L^* = \left\lfloor \frac{(R-T)\mu}{C} \right\rfloor$$

Regulating observable queues

1. By entry fee T:

$$L^* = \lfloor \frac{(R-T)\mu}{C} \rfloor$$

2. By allowing reneging (Hassin '85)

- An arrival is placed anywhere but not in the rear
- when to renege?

Regulating observable queues

1. By entry fee T:

$$L^* = \left\lfloor \frac{(R-T)\mu}{C} \right\rfloor$$

2. By allowing reneging (Hassin '85)

- An arrival is placed anywhere but not in the rear
- when to renege?

Equilibrium: $\{$Renege from the rear $\Leftrightarrow L \geq L^*\}$

Regulating observable queues

1. By entry fee T:

$$L^* = \lfloor \frac{(R-T)\mu}{C} \rfloor$$

2. By allowing reneging (Hassin '85)

- An arrival is placed anywhere but not in the rear
- when to renege?

bEquilibrium: $\{$Renege from the rear $\Leftrightarrow L \geq L^*\}$

Multi-equilibria but a unique SPE.

Monopoly profit maximization:

L_m = largest n with

$$n + \frac{(1-\rho^{n-1})(1-\rho^{n+1})}{\rho^{n-1}(1-\rho)^2} \leq \frac{R\mu}{C}$$

Optimal solution:

- server: charge (a bit less than) $R - CL_m/\mu$
- customers: $\{ L < L_m \Rightarrow$ pay and join $\}$

Result: $L_m \leq L^* \leq L_e$

A monopoly under-utilizes the server.

Purchasing priority

- two priority levels. With class, FCFS
- high priority costs θ
- no balking or reneging (no need for R)
- unobservable

Why to have priority at all?

1. to overtake ordinary customers
2. to avoid being overtaken by premium customers

Priority purchasing

Hassin and Haviv ’03

Strategy p: purchase priority with probability p

$f(p)$ = value of high priority

$$f(p) = \frac{C\lambda}{(\mu - \lambda)(\mu - \lambda p)}$$

$f(p) \uparrow p$.

The more purchase priority, the higher its value

Purchasing priority

The best response function $\uparrow p$

$\Downarrow$

Follow the crowd (FTC)

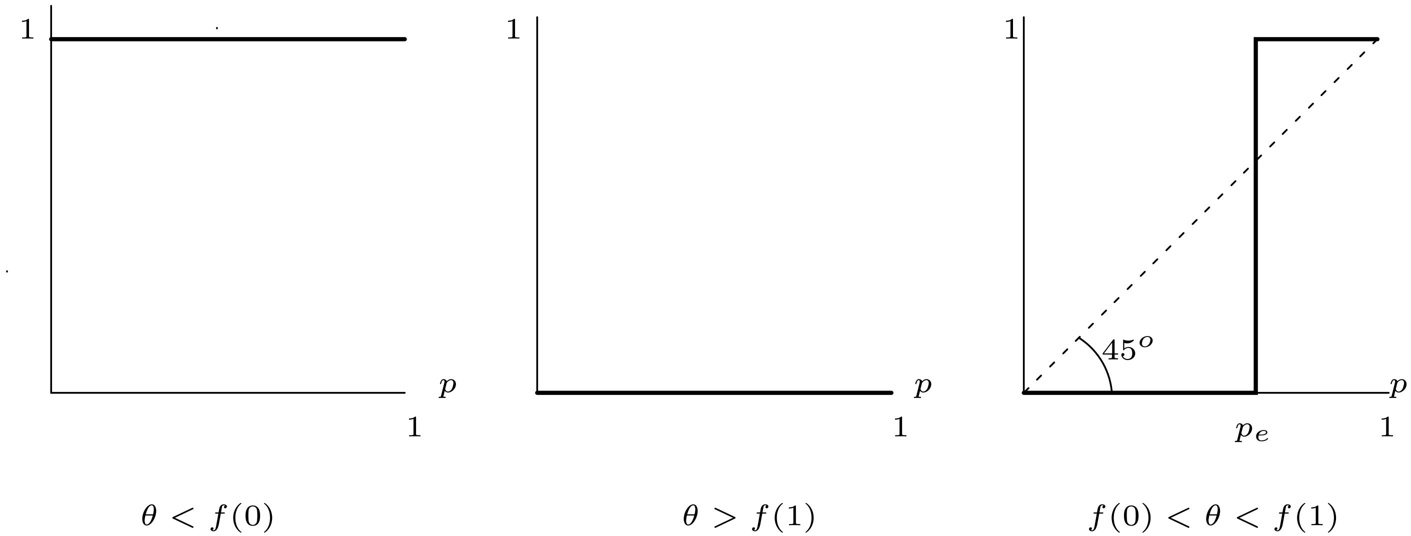
- $\theta < f(0) \Rightarrow p = 1$ is dominant
- $\theta > f(1) \Rightarrow p = 0$ is dominant
- $f(0) \leq \theta \leq f(1)$ three equilibria:
  - $p = 0$ (stable)
  - $p = 1$ (stable)
  - $p = p_e$ where $f(p_e) = \theta$ (unstable)

Purchasing priority

The best response function $\uparrow p$

$\Downarrow$

Follow the crowd (FTC)



To renege or not to renege?

Unobservable model: never renege

$$\Downarrow$$

Reneging is a possibility only if $R \downarrow$ or $C \uparrow$ with time

Bang-bang model: $R = 0$ at time T.

Equilibrium:

- join with p_e (function of λ, μ, R, C and T)
- renege at T

M/M/1 queues with retrials

- unobservable
- λ, μ and C (no balking, no reneging)
- busy signal $\Rightarrow$ try (and retry) later
- cost H per retrial

M/M/1 queues with retrials

- unobservable
- λ, μ and C (no balking, no reneging)
- busy signal $\Rightarrow$ try (and retry) later
- cost H per retrial

No information is gained while in orbit:

- on how many in orbit
- on the number of past retrials
- on the time in orbit

M/M/1 queues with retrials

- unobservable
- λ, μ and C (no balking, no reneging)
- busy signal $\Rightarrow$ try (and retry) later
- cost H per retrial

No information is gained while in orbit:

- on how many in orbit
- on the number of past retrials
- on the time in orbit

$\Rightarrow$ time between retrials $\sim$ exp(?).

Queues with retrials

Time between retrials $\sim \exp(\eta)$

- Social optimal retrial rate (Kulkarni '83)

$$\eta^* = \sqrt{\frac{C\mu}{H}}$$

Queues with retrials

Time between retrials $\sim \exp(\eta)$

- Social optimal retrial rate (Kulkarni '83)

$$\eta^* = \sqrt{\frac{C\mu}{H}}$$

- Nash equilibrium (Hassin and Haviv '96)

$$\eta_e = \frac{C\rho + \sqrt{C^2\rho^2 + 8\mu CH(1-\rho)(2-\rho)}}{4H(1-\rho)}.$$