# "The Regulation of Queue Size by Levying Tolls" by P. Naor (1969)

## Paper presentation for BZAN605

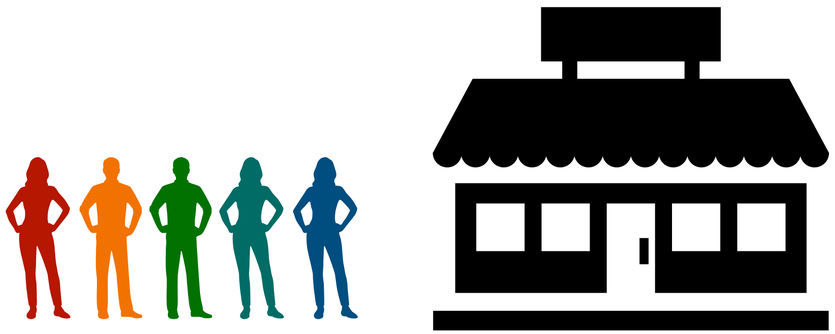
Adam Spannbauer

# A little queueing theory

# A little queueing theory

## Lines build when things arrive faster than we process

- A queue can be modeled with parameters like



# A little queueing theory

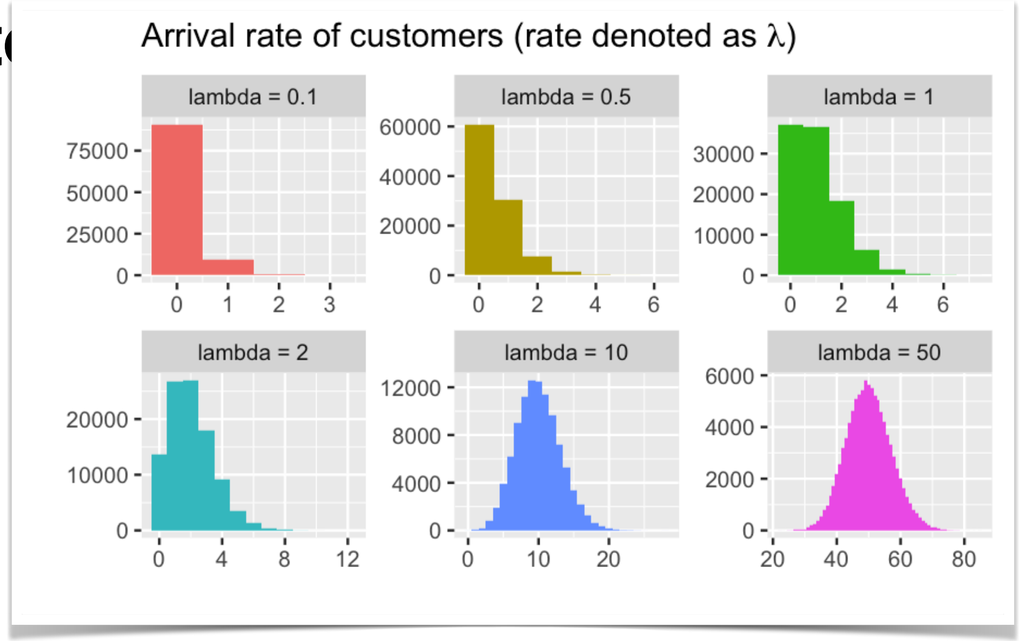
## Lines build when things arrive faster than we process

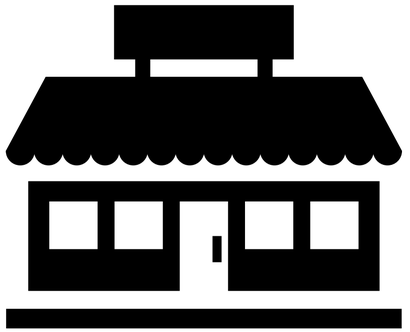
- A queue can be modeled with parameters like
  - Arrival rate

# A little queueing theory

## Lines build when things arrive fast

- A queue can be modeled with parameters like
  - Arrival rate
    - Ex: can be a Poisson process with arrival rate $\lambda$



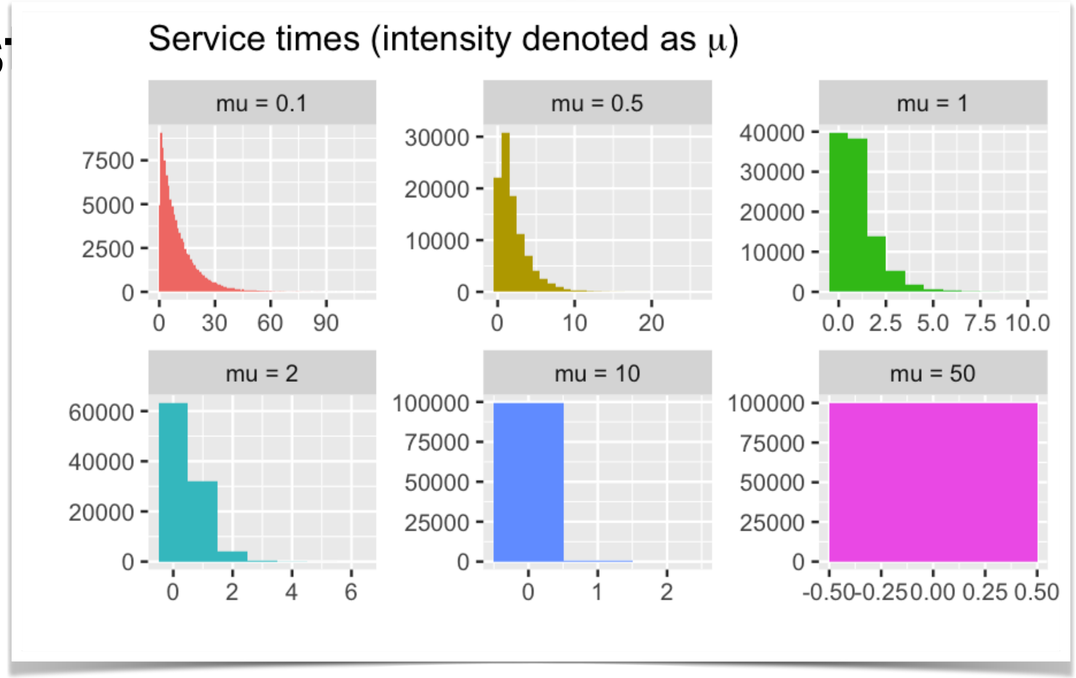
# A little queueing theory

## Lines build when things arrive faster than we process

- A queue can be modeled with parameters like
  - Arrival rate
    - Ex: can be a Poisson process with arrival rate $\lambda$
  - Service rate

# A little queueing theory

## Lines build when things arrive fas




- A queue can be modeled with parameters like
  - Arrival rate
    - Ex: can be a Poisson process with arrival rate $\lambda$
  - Service rate
    - Ex: can be exponentially distributed with intensity $\mu$

# A little queueing theory

## Metrics

- Some common ways to measure queues are
  - Expected queue length ($L_q$)
  - Average waiting time ($W$)
  - Utilization ($\rho$)
    - $\rho = \frac{\lambda}{\mu}$ (ratio of arrival rate to service rate)

Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *The Annals of Mathematical Statistics*, 338-354.

# A little queueing theory

## Metrics

- Some common ways to measure queues are
  - Expected queue length ($L_q$)
  - Average waiting time ($W$)
  - Utilization ($\rho$)
    - $\rho = \frac{\lambda}{\mu}$ (ratio of arrival rate to service rate)

Ex:

- 5 customers arrive per hour ($\lambda = 5$)
- We can service 10 customers per hour ($\mu = 10$)
- $\rho = \frac{5}{10} = 0.5$

50% of service capacity is being utilized

# A little queueing theory

## Metrics

- Some common ways to measure queues are
  - Expected queue length ($L_q$)
  - Average waiting time ($W$)
  - Utilization ($\rho$)
    - $\rho = \frac{\lambda}{\mu}$ (ratio of arrival rate to service rate)

We're in a "steady state" when these metrics have stabilized and no longer change over time.
(i.e. viewing things "in the long run")

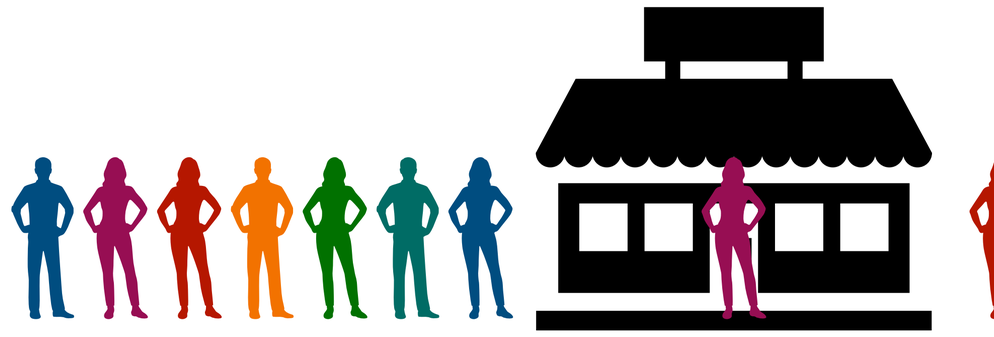
Our analyses take place with this steady state view.

# A little queueing theory

## Metrics

- Some common ways to measure queues are
  - Expected queue length ($L_q$)
  - Average waiting time ($W$)
  - Utilization ($\rho$)
    - $\rho = \frac{\lambda}{\mu}$ (ratio of arrival rate to service rate)

If $\rho > 1$, the arrival rate is outpacing service. Queues can grow indefinitely*, and the system will not reach a steady state.

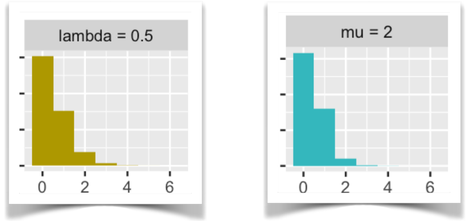


*without mechanisms like “balking”, “finite capacity”, or **tolls**

# A little queueing theory

## Metrics

- Some common ways to measure queues are
  - Expected queue length ($L_q$)
  - Average waiting time ($W$)
  - Utilization ($\rho$)
    - $\rho = \frac{\lambda}{\mu}$ (ratio of arrival rate to service rate)







"Even when the arrival rate is smaller than the service rate (so that the server can accommodate all arrivals), queues are formed due to the variability in service and inter-arrival times."

Hassin, R., & Haviv, M. (2003). *To queue or not to queue: Equilibrium behavior in queueing systems* (Vol. 59). Springer Science & Business Media.

# Background to Naor’s work

# Pay to getting in line?

## Ideas cited by Naor

- Leeman proposes tolls to manage queue size
- Saaty argues Leeman's queue toll is unjust unless queueing for a luxury
  - i.e. we shouldn't toll things like healthcare lines


Leeman, W. A. (1964). The reduction of queues through the use of price. Operations Research, 12(5), 783-785. https://pubsonline.informs.org/doi/pdf/10.1287/opre.12.5.783

Saaty, T. L., & Leeman, W. A. (1965). The Burdens of Queuing Charges-Comments on a Letter by Leeman. Operations Research, 13(4), 679–681. http://www.jstor.org/stable/167860

# Pay to getting in line?

## Ideas cited by Naor

- Leeman proposes tolls to manage queue size
- Saaty argues Leeman's queue toll is unjust unless queueing for a luxury
  - i.e. we shouldn't toll things like healthcare lines

Also suggests dynamic pricing to discourage queueing in busy times and stimulate demand in less busy times

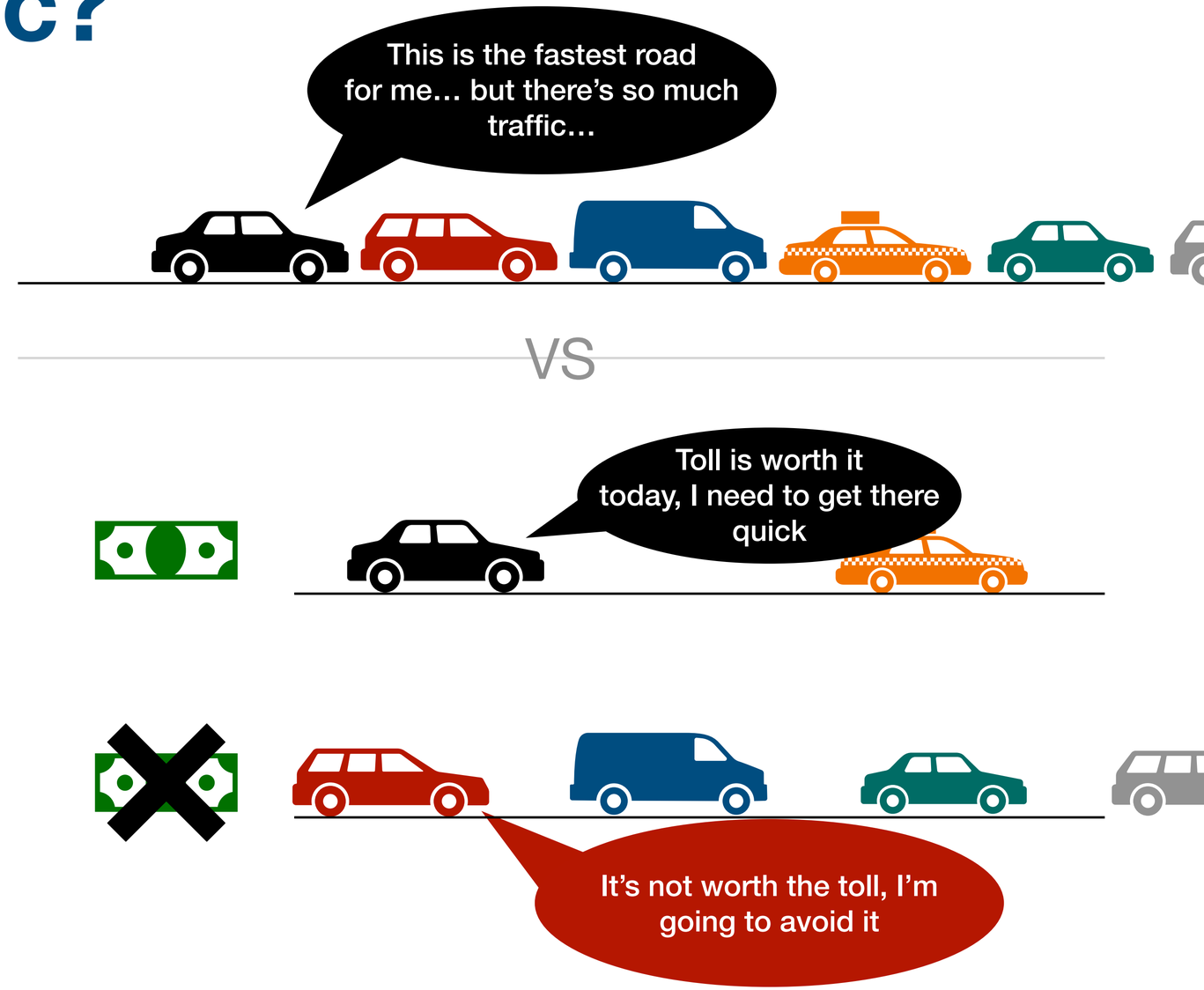

Leeman, W. A. (1964). The reduction of queues through the use of price. Operations Research, 12(5), 783-785. https://pubsonline.informs.org/doi/pdf/10.1287/opre.12.5.783

Saaty, T. L., & Leeman, W. A. (1965). The Burdens of Queuing Charges-Comments on a Letter by Leeman. Operations Research, 13(4), 679–681. http://www.jstor.org/stable/167860

# Queues are like traffic?

## Ideas cited by Naor

- Drivers focus on optimizing their own routes, this does not optimize the overall system
- Tolls on roads can help redistribute traffic?

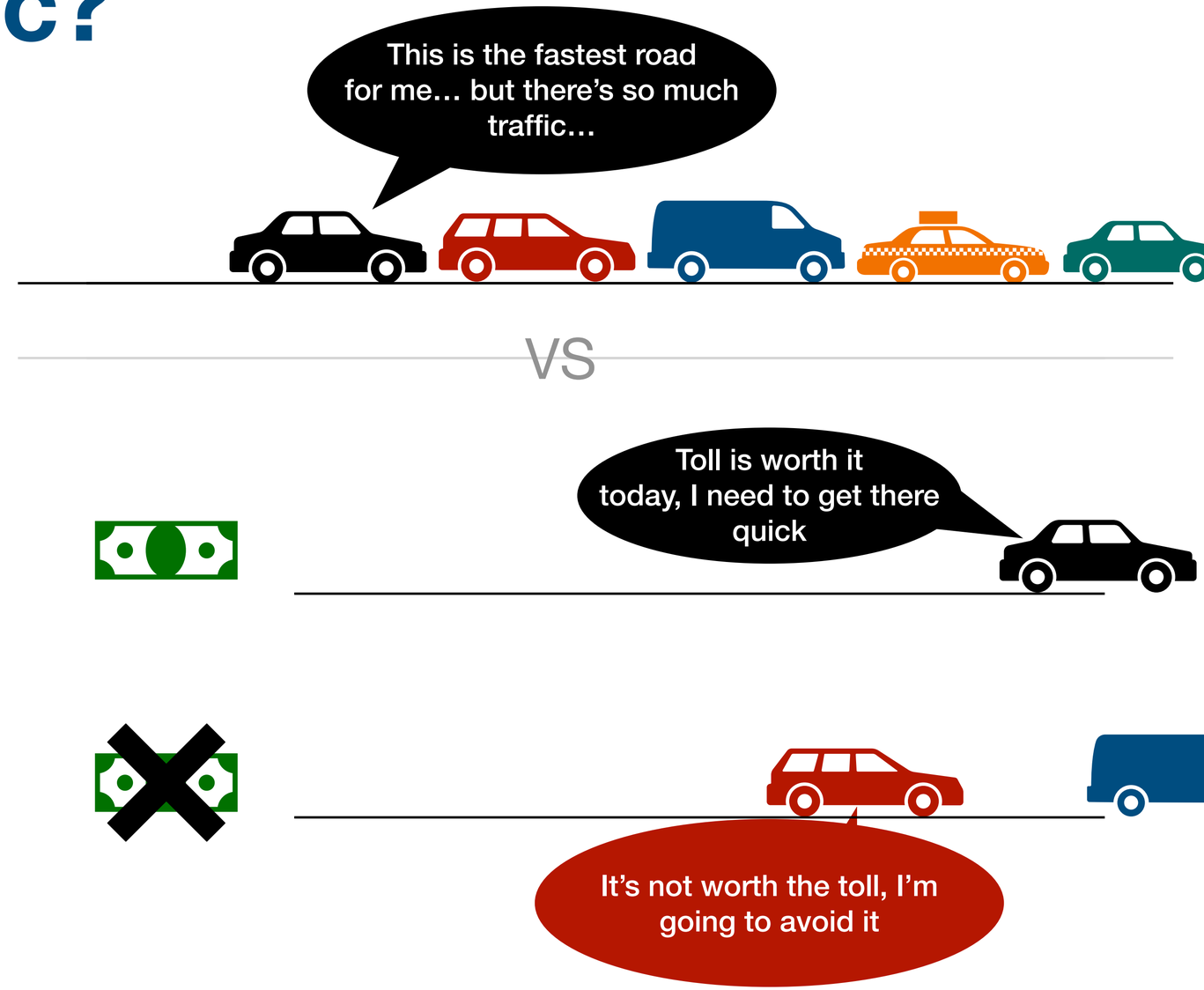



From a lecture by Professor Martin Beckmann (of Brown University and Bonn University) at Colloquium on "Decision Making in Traffic Planning" (organized in summer 1965 by Professor Arne Jensen of the Technical University, Copenhagen)

# Queues are like traffic?

## Ideas cited by Naor

- Drivers focus on optimizing their own routes, this does not optimize the overall system
- Tolls on roads can help redistribute traffic?
  - We'd try to set tolls in a way that aligns self-interested utility with overall utility




From a lecture by Professor Martin Beckmann (of Brown University and Bonn University) at Colloquium on "Decision Making in Traffic Planning" (organized in summer 1965 by Professor Arne Jensen of the Technical University, Copenhagen)

# Big impact of paper

# Customers act in their own self interest

**We can model this with** $v_{service} = \frac{R\mu}{C}$ (reward of service per unit of cost)

- "Naor (1969) appears to be the first to incorporate customer decisions into a queueing model."
- Naor outlines a framework for addressing $R$ewards and $C$osts for consumers


Huang, T., Allon, G., & Bassamboo, A. (2013). Bounded rationality in service systems. *Manufacturing & Service Operations Management*, *15*(2), 263-279.

# Customers act in their own self interest

**We can model this with** $v_{service} = \frac{R\mu}{C}$ (reward of service per unit of cost)

$R = \$10$

$C = \$8$ per minute

$\mu = 5$ minutes to service per person

- "Naor (1969) appears to be the first to incorporate customer decisions into a queueing model."
- Naor outlines a framework for addressing $R$ewards and $C$osts for consumers


Huang, T., Allon, G., & Bassamboo, A. (2013). Bounded rationality in service systems. *Manufacturing & Service Operations Management*, *15*(2), 263-279.

# Customers act in their own self interest

**We can model this with** $v_{service} = \frac{R\mu}{C}$ (reward of service per unit of cost)

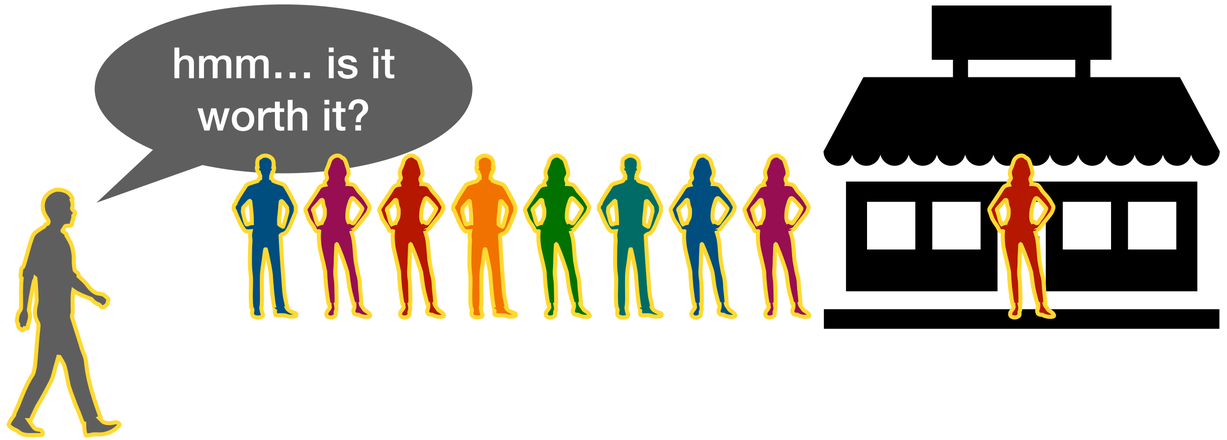
$R = \$10$

$C = \$8$ per minute

$\mu = 5$ minutes to service per person

$n = 10$ people to go including me

- "Naor (1969) appears to be the first to incorporate customer decisions into a queueing model."
- Naor outlines a framework for addressing $R$ewards and $C$osts for consumers




Huang, T., Allon, G., & Bassamboo, A. (2013). Bounded rationality in service systems. *Manufacturing & Service Operations Management*, *15*(2), 263-279.

# Customers act in their own self interest

**We can model this with** $v_{service} = \frac{R\mu}{C}$ (reward of service per unit of cost)

- "Naor (1969) appears to be the first to incorporate customer decisions into a queueing model."
- Naor outlines a framework for addressing $R$ewards and $C$osts for consumers

$R = \$10$

$C = \$8$ per minute

$\mu = 5$ minutes to service per person

$n = 10$ people to go including me

$$v_s(n) = R - \frac{Cn}{\mu}$$

(utility for customer given $n$ people wait)


Huang, T., Allon, G., & Bassamboo, A. (2013). Bounded rationality in service systems. *Manufacturing & Service Operations Management*, *15*(2), 263-279.

# Customers act in their own self interest

**We can model this with** $v_{service} = \frac{R\mu}{C}$ (reward of service per unit of cost)

- "Naor (1969) appears to be the first to incorporate customer decisions into a queueing model."
- Naor outlines a framework for addressing $R$ewards and $C$osts for consumers
  - Sometimes the juice isn't worth the squeeze and they won't queue

$R = \$10$

$C = \$8$ per minute

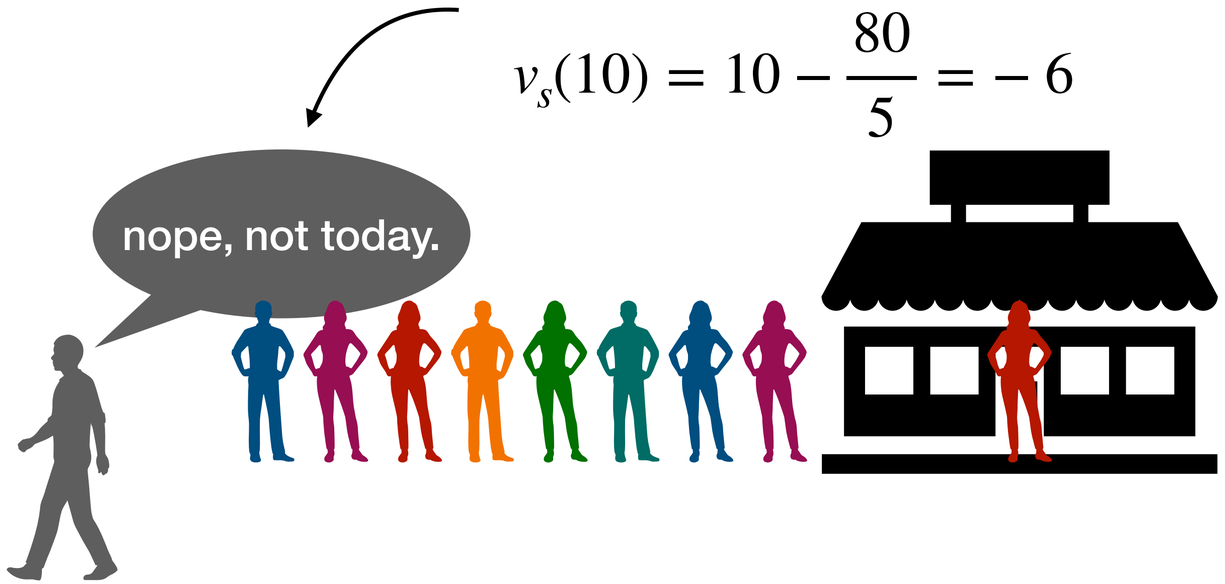
$\mu = 5$ minutes to service per person

$n = 10$ people to go including me

$$v_s(n) = R - \frac{Cn}{\mu}$$

(utility for customer given $10$ people wait)

$$v_s(10) = 10 - \frac{80}{5} = -6$$




Huang, T., Allon, G., & Bassamboo, A. (2013). Bounded rationality in service systems. *Manufacturing & Service Operations Management*, *15*(2), 263-279.

# Customers act in their own self interest

**We can model this with** $v_{service} = \frac{R\mu}{C}$ (reward of service per unit of cost)

- "Naor (1969) appears to be the first to incorporate customer decisions into a queueing model."
- Naor outlines a framework for addressing $R$ewards and $C$osts for consumers
  - Sometimes the juice isn't worth the squeeze and they won't queue

$R = \$10$

$C = \$8$ per minute

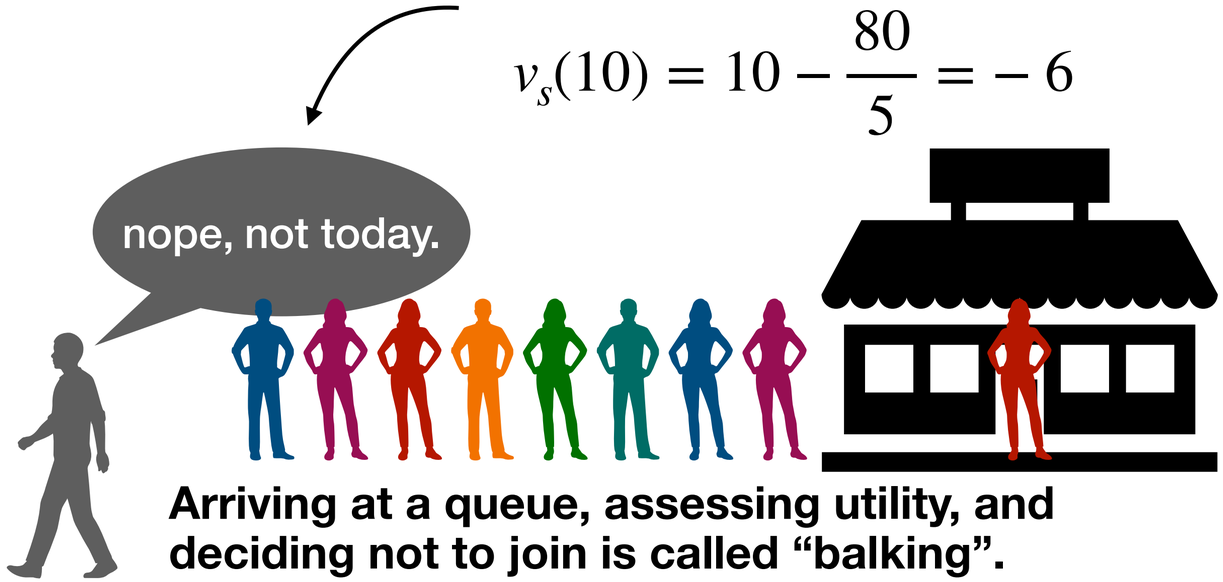
$\mu = 5$ minutes to service per person

$n = 10$ people to go including me

$$v_s(n) = R - \frac{Cn}{\mu}$$

(utility for customer given $10$ people wait)

$$v_s(10) = 10 - \frac{80}{5} = -6$$




**Arriving at a queue, assessing utility, and deciding not to join is called "balking".**

Huang, T., Allon, G., & Bassamboo, A. (2013). Bounded rationality in service systems. *Manufacturing & Service Operations Management*, *15*(2), 263-279.

# Customers act in their own self interest

**We can model this with** $v_{service} = \frac{R\mu}{C}$ (reward of service per unit of cost)

- "Naor (1969) appears to be the first to incorporate customer decisions into a queueing model."
- Naor outlines a framework for addressing $R$ewards and $C$osts for consumers
  - Sometimes the juice isn't worth the squeeze and they won't queue

$R = \$10$

$C = \$8$ per minute

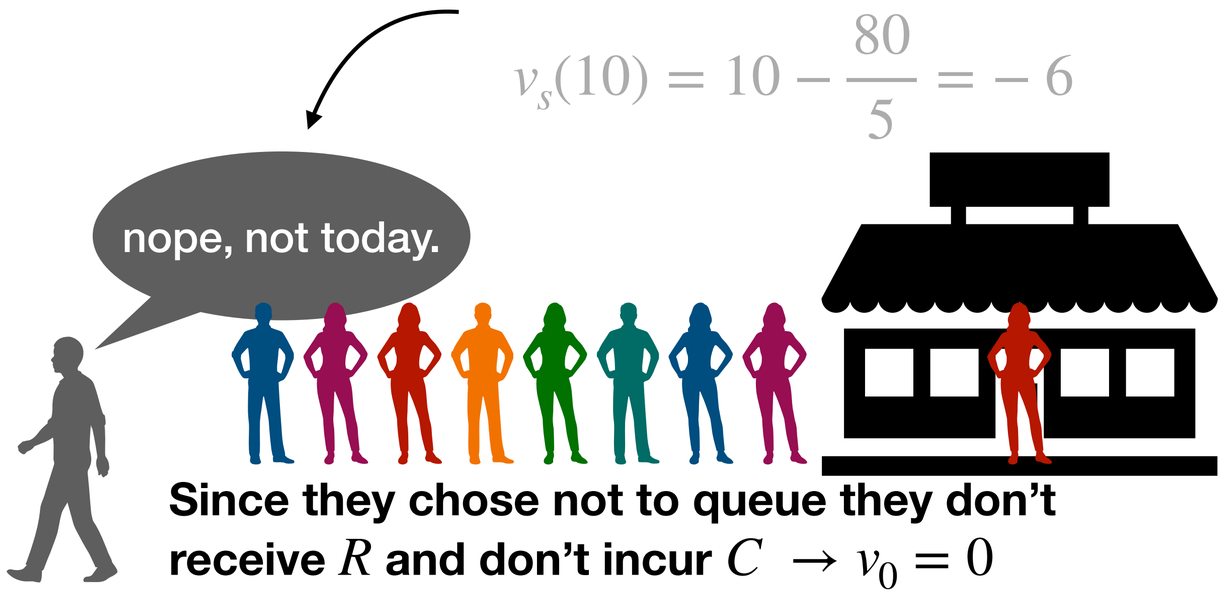
$\mu = 5$ minutes to service per person

$n = 10$ people to go including me

$$v_s(n) = R - \frac{Cn}{\mu}$$

(utility for customer given $10$ people wait)

$$v_s(10) = 10 - \frac{80}{5} = -6$$




**Since they chose not to queue they don't receive $R$ and don't incur $C$ $\rightarrow v_0 = 0$**

Huang, T., Allon, G., & Bassamboo, A. (2013). Bounded rationality in service systems. *Manufacturing & Service Operations Management*, *15*(2), 263-279.

# Naor’s model in more depth

# Problem statement

## Can queue tolls be used to optimize the system for all?

- Excessive queue lengths cause inefficiencies
  - e.g. wasted time & reduced customer satisfaction)
- Proposal: Charge customers to enter the queue
  - just like auto traffic this can reroute individuals for the overall good of the system


Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Conditions for model

$R,\ C,\ \lambda,\ \mu,\ \rho,\ v_s,\ p_i,\ g(z),\ E[i],\ \zeta,\ b$


1. There exists a public good (aka well-being) that can be maximized via some objective function.

2. “Customers are liable to be diverted from the service station”

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Conditions for model

$R,\ C,\ \lambda,\ \mu,\ \rho,\ v_s,\ p_i,\ g(z),\ E[i],\ \zeta,\ b$


1. **There exists a public good (aka well-being) that can be maximized via some objective function.**

   - "…*the expected overall profit (in unit time) accruing to arriving customers is a proper objective function representing public good*…"

   - This can have a centralized decision maker (public transportation, govt healthcare, etc) or decentralized decision makers (customers act for themselves to max their own utility)

2. "Customers are liable to be diverted from the service station"

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

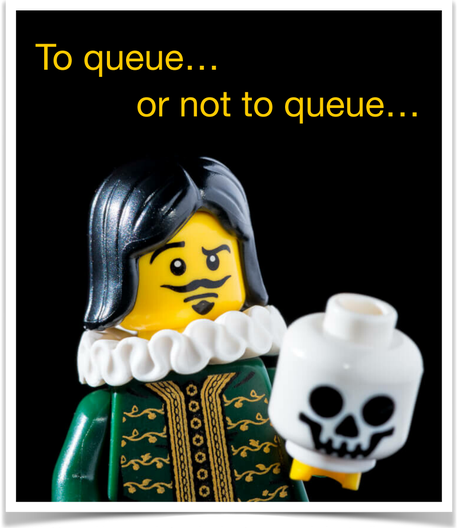
# Conditions for model

$R,\ C,\ \lambda,\ \mu,\ \rho,\ v_s,\ p_i,\ g(z),\ E[i],\ \zeta,\ b$


1. There exists a public good (aka well-being) that can be maximized via some objective function.

2. **"Customers are liable to be diverted from the service station"**

   - Big contrast to prior queueing models that assumed we serve *all customers*
   - Some will queue and some will not, those that don't have utility 0
     - A toll can influence the decision of queuing vs leaving




https://www.istockphoto.com/photo/lego-minifigures-series-8-figurine-the-thespian-gm458926969-22139332

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Defining the model

$R,\ C,\ \lambda,\ \mu,\ \rho,\ v_s,\ p_i,\ g(z),\ E[i],\ \zeta,\ b$


- Inputs
  - $R$ - reward for being serviced (e.g $)
  - $C$ - cost per unit of time for queueing (e.g. $ per minute; assumed all customer $C$s are equal)
  - $\lambda$ - arrival rate of customers (e.g. Poisson distributed)
  - $\mu$ - service rate (e.g. exponential distributed)

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Defining the model

$R,\ C,\ \lambda,\ \mu,\ \rho,\ v_s,\ p_i,\ g(z),\ E[i],\ \zeta,\ b$


- Additional variables
  - $\rho = \frac{\lambda}{\mu}$ - classic utilization factor (ratio of arrival rate to service rate)
    - Note! A different metric will be used to calculate service's "busy fraction" to account for queue length of 0
  - $i$ - queue size at a given time (randomly distributed)
  - $n$ - queue capacity (if $i > n$ a customer will not queue; a customer will queue if $i \leq n$)

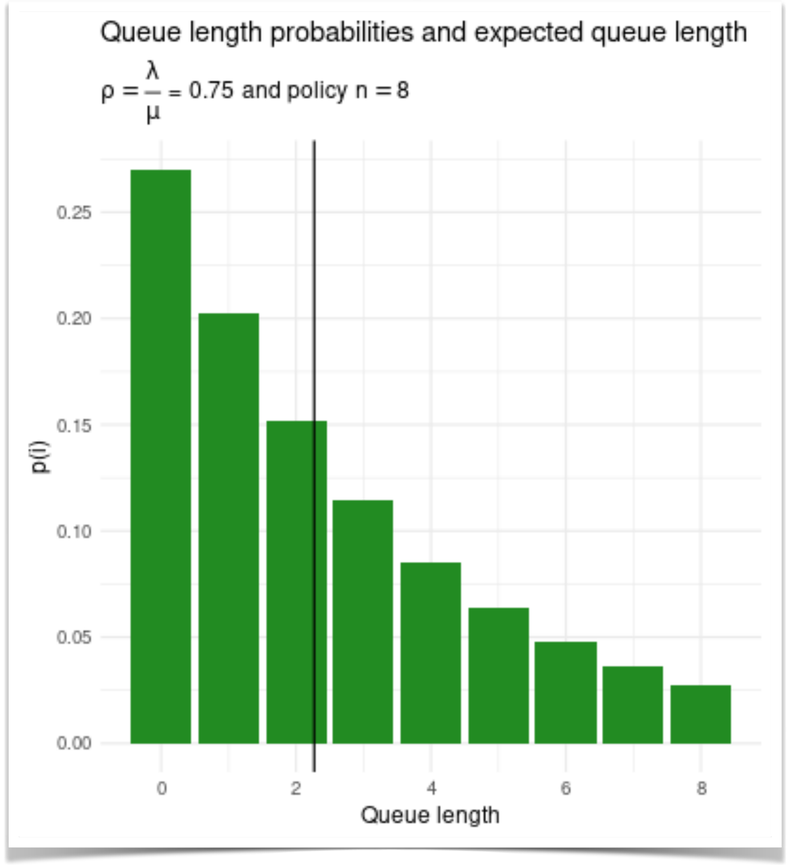
Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *The Annals of Mathematical Statistics*, 338-354.

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Steady state equations

$R,\ C,\ \lambda,\ \mu,\ \rho,\ v_s,\ p_i,\ g(z),\ E[i],\ \zeta,\ b$


- PMF of queue length: $p_{i+1} = p_i\rho \quad \rightarrow \quad p_i = \frac{\rho^i(1-\rho)}{1-\rho^{n+1}}$
  - The probability of line being length $i$ as a function of $\rho$ ($\frac{\lambda}{\mu}$ - utilization factor) and $n$ (queue capacity)



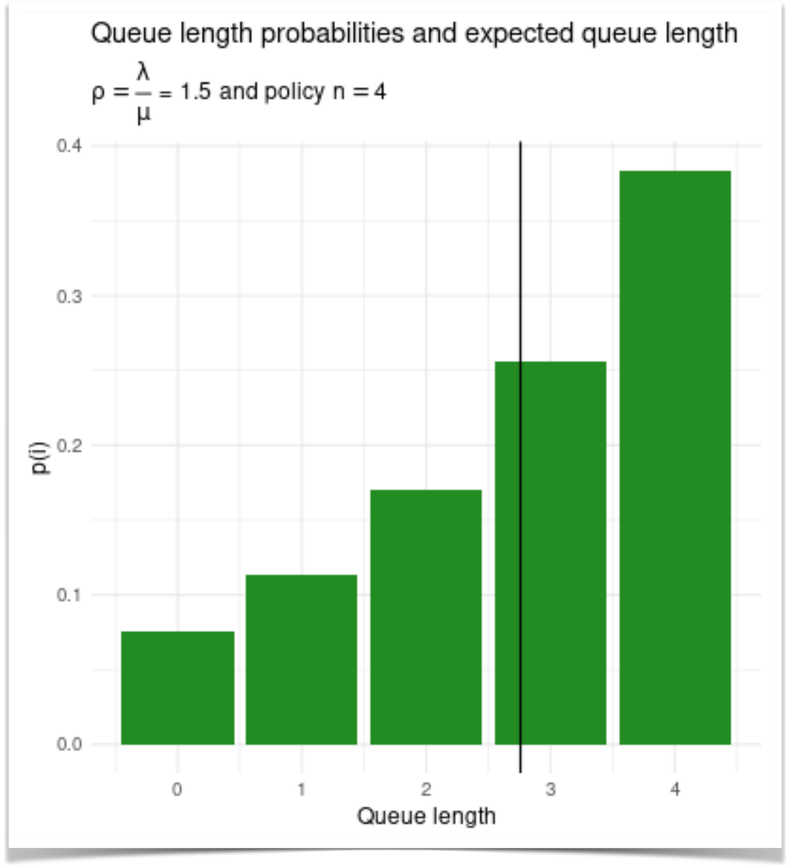

Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *The Annals of Mathematical Statistics*, 338-354.

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Steady state equations

$R,\ C,\ \lambda,\ \mu,\ \rho,\ v_s,\ p_i,\ g(z),\ E[i],\ \zeta,\ b$


- PMF of queue length: $p_{i+1} = p_i\rho \quad \rightarrow \quad p_i = \frac{\rho^i(1-\rho)}{1-\rho^{n+1}}$
  - The probability of line being length $i$ as a function of $\rho$ ($\frac{\lambda}{\mu}$ - utilization factor) and $n$ (queue capacity)




- Probability Generating Function:

$$g(z) = \sum_{i=0}^{n} p_i z^i = \frac{1-\rho}{1-\rho^{n+1}} \cdot \frac{1-(\rho z)^{n+1}}{1-\rho z}$$

- Expected length of queue:

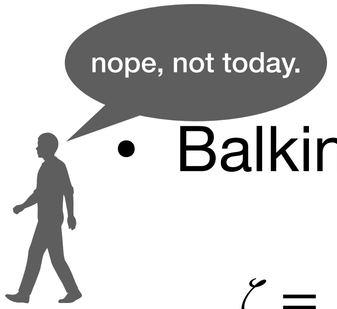
$$E[i] = \frac{\rho[1-(n+1)\rho^n + n\rho^{n+1}]}{(1-\rho)(1-\rho^{n+1})} = \frac{\rho}{1-\rho} - \frac{(n+1)\rho^{n+1}}{1-\rho^{n+1}}$$


Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *The Annals of Mathematical Statistics*, 338-354.

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Steady state equations

$R,\ C,\ \lambda,\ \mu,\ \rho,\ v_s,\ p_i,\ g(z),\ E[i],\ \zeta,\ b$





- Balking rate (arrival rate x probability of being at capacity)
  - $\zeta = \lambda p_n = \dfrac{\lambda\rho^n(1-\rho)}{1-\rho^{n+1}}$
- Effective arrival rate (arrival - rejection): $\lambda - \zeta = \lambda(1 - p_n)$
- “Busy fraction” (exclude case of no one in system)
  - $b = \sum_{i=1}^{n} p_i = 1 - p_0 = \dfrac{\rho(1-\rho^n)}{1-\rho^{n+1}}$
- Effective service rate (service * busy fraction): $\mu b = \mu(1 - p_0)$

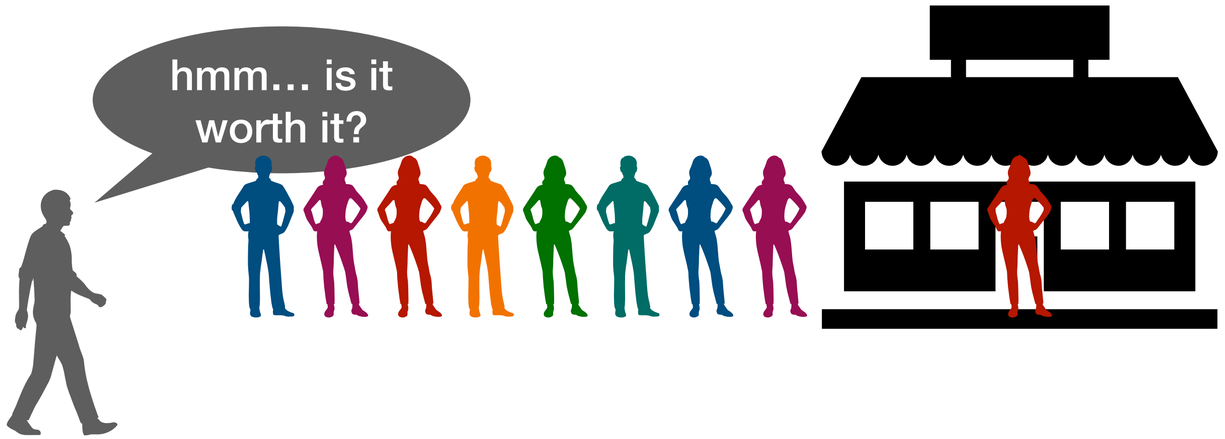

Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *The Annals of Mathematical Statistics*, 338-354.

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# The big contribution

## Modeling customers' net gain

"The newly arrived customer weighs the two alternatives - to join or not to join the queue - by the net gains associated with them."



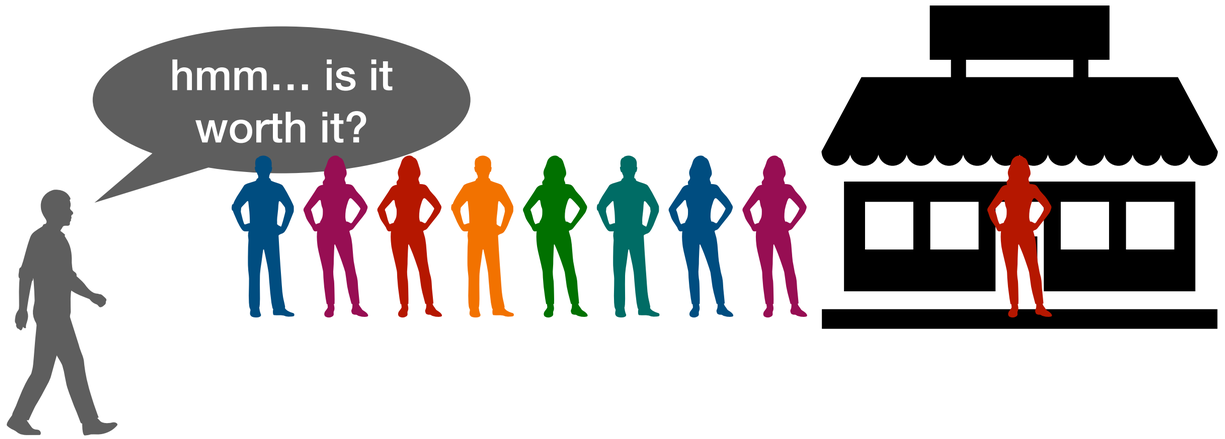

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# The big contribution

## Modeling customers' net gain

"The newly arrived customer weighs the two alternatives - to join or **not to join** the queue - by the net gains associated with them."

not to join: $G_i = 0$




Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

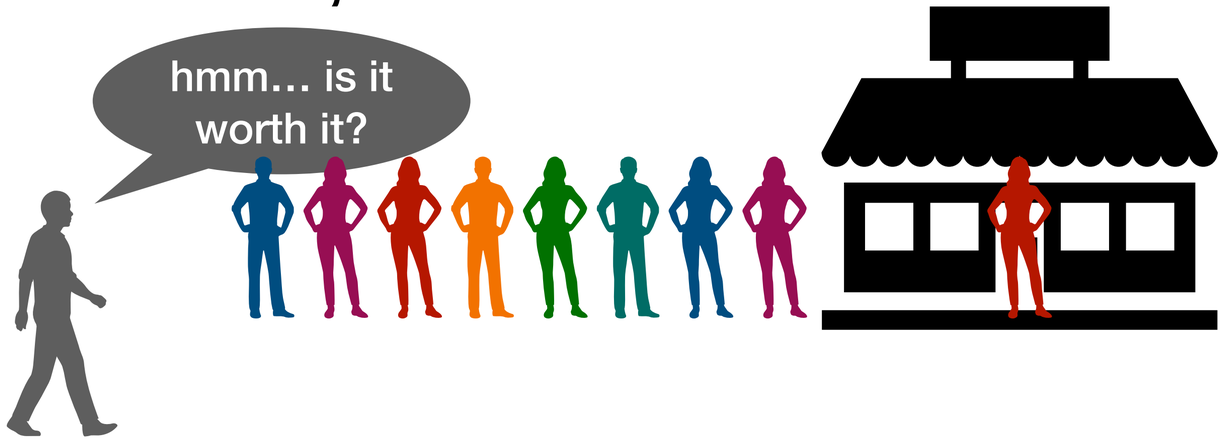
# The big contribution

## Modeling customers' net gain

"The newly arrived customer weighs the two alternatives - **to join** or not to join the queue - by the net gains associated with them."

not to join: $G_i = 0$

to join: $G_i = R - (i+1)C\frac{1}{\mu}$




Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# The big contribution

## Modeling customers' net gain

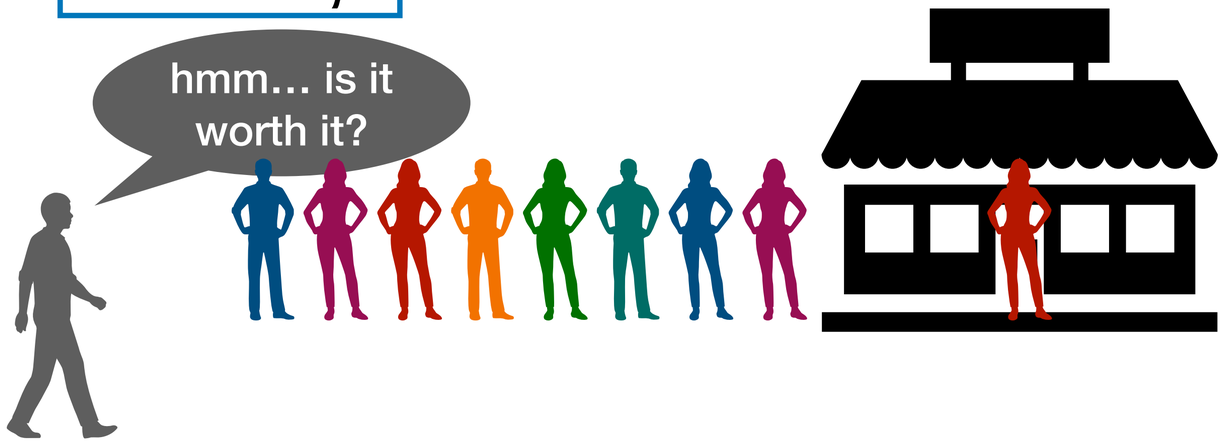
"The newly arrived customer weighs the two alternatives - to join or not to join the queue - by the net gains associated with them."

not to join: $G_i = 0$

to join: $G_i = R - (i+1)C\frac{1}{\mu}$

Reward for being serviced

Cost of waiting for everyone's service (including the one joining)




Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# “Self-optimization” leads to lines of length $n_s$

**Customers optimize for their own self-interest and the system suffers**

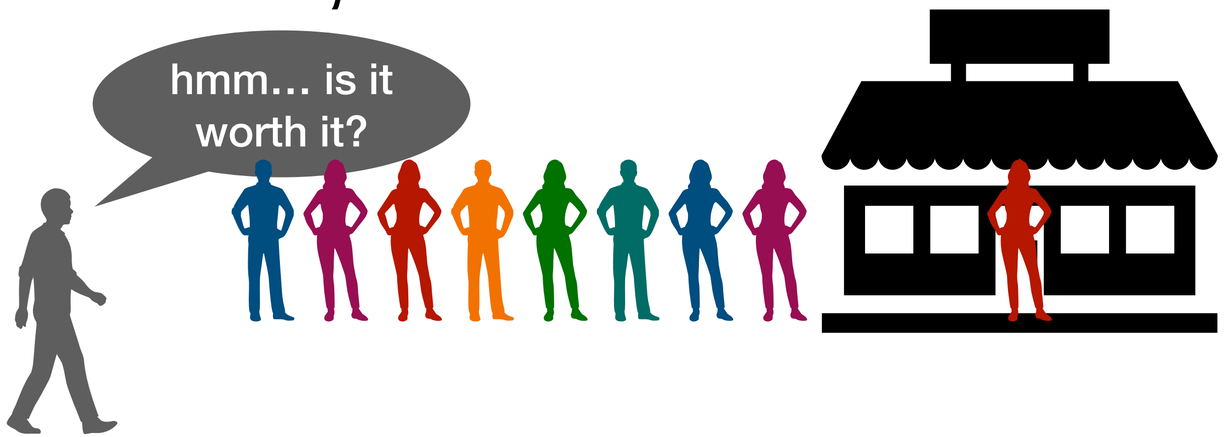
“The newly arrived customer weighs the two alternatives - to join or not to join the queue - by the net gains associated with them.”

not to join: $G_i = 0$

to join: $G_i = R - (i+1)C\frac{1}{\mu}$

**Notice this gain ($G_i$) does not consider the arrival rate ($\lambda$)!**

**If left to self-optimize, just like traffic, customers will not optimize the system overall.**




From a lecture by Professor Martin Beckmann (of Brown University and Bonn University) at Colloquium on "Decision Making in Traffic Planning" (organized in summer 1965 by Professor Arne Jensen of the Technical University, Copenhagen)

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# “Overall optimization” leads to lines of length $n_o$

## Maximize expected sum of customer net gains

“We note that the expected total net gain, $P$, under some strategy $n$ is given by”

$$P = (\lambda - \zeta)R - CE[i]$$

Reward multiplied by expected number of customers joining queue.

This assumes once customers join the queue, they will not leave (i.e. no “reneging”)

Cost multiplied by expected number of customers in queue

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# “Beneficial Toll Imposition” to lower $n_s$ to $n_o$

**Lowering self utility via toll can raise overall utility**

“a toll $\theta$ is imposed on customers joining the queue and their (individually) expected net gain is reduced in such a way that $n_o$ is the current criterion of newly arrived customers”

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

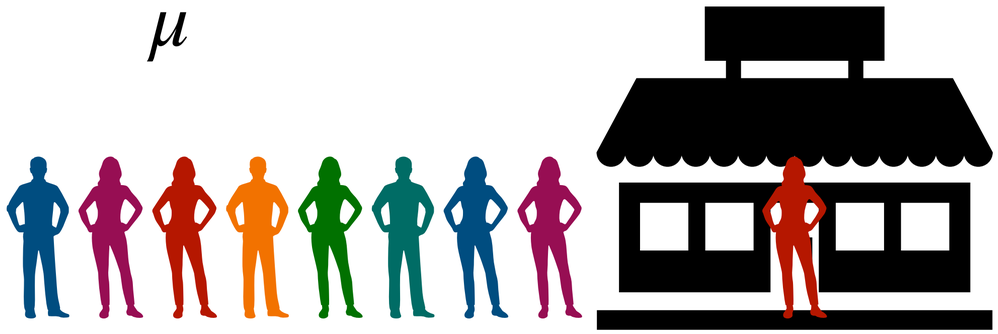
# “Beneficial Toll Imposition” to lower $n_s$ to $n_o$

**Lowering self utility via toll can raise overall utility**

“a toll $\theta$ is imposed on customers joining the queue and their (individually) expected net gain is reduced in such a way that $n_o$ is the current criterion of newly arrived customers”

$$R - \frac{C(n_o + 1)}{\mu} < \theta^* \leq R - \frac{C(n_o)}{\mu}$$

“If a toll taken from this range is levied on customers joining the queue, the combined income (in unit time) of customers and the revenue agency is maximized.”



Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# “Revenue Maximization”

## Tolls to zero out customers gain and maximize revenue

“The toll collecting agency may be completely divorced from the individual and collective economic interests of the customers. In that case the agency will seek to impose a toll, $\theta_r$, designed to maximize its own revenue rather than to optimize the whole system.”

$$M = (\lambda - \zeta)\theta$$

Expected number of customers joining queue.

Toll value

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# “Revenue Maximization” to lower $n_s$ to $n_o$

## Tolls to zero out customers gain and maximize revenue

Set the toll to be equal to the customers’ profit from waiting in line and being serviced

$$\theta_r = R - \frac{Cn_r}{\mu}$$

Relies on a net neutral experience being good enough to retain customers

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Examples via hands-on demo 🤞

## Queueing Theory with Naor's Model*

*as outlined in Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

Reward (R)

10

R is the reward for being served

Cost (C)

5

C is the cost per unit of waiting time

Arrival Rate (lambda)

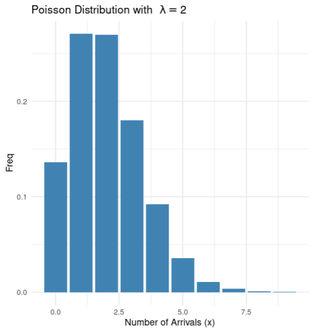
2

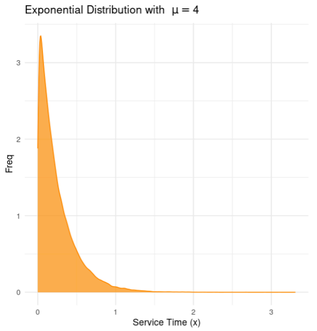
lambda is the arrival rate of customers (Poisson-distributed here)

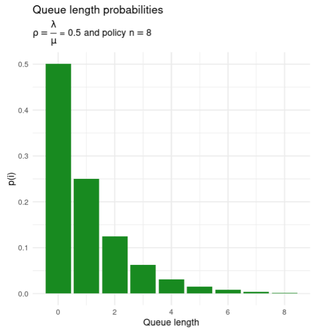
Service Rate (mu)

4

mu is the service rate (Exponential-distributed here)




### No tolls imposed:

| | Self-Optimized | Overall Optimized |
|---|---|---|
| Queue Len Policy | 8.00 | 4.00 |
| Rho | 0.50 | 0.50 |
| Service Value Ratio | 8.00 | 8.00 |
| Expected Len | 0.98 | 0.84 |
| Balk Rate | 0.00 | 0.06 |
| Busy Frac | 0.50 | 0.48 |
| Expected Leaving | 2.00 | 1.94 |
| Expected Total Gain | 15.05 | 15.16 |

### With suggested toll of 5.00:

| | Self-Optimized |
|---|---|
| Queue Len Policy | 4.00 |
| Rho | 0.50 |
| Service Value Ratio | 4.00 |
| Expected Len | 0.84 |
| Balk Rate | 0.06 |
| Busy Frac | 0.48 |
| Expected Leaving | 1.94 |
| Expected Total Gain | 5.48 |

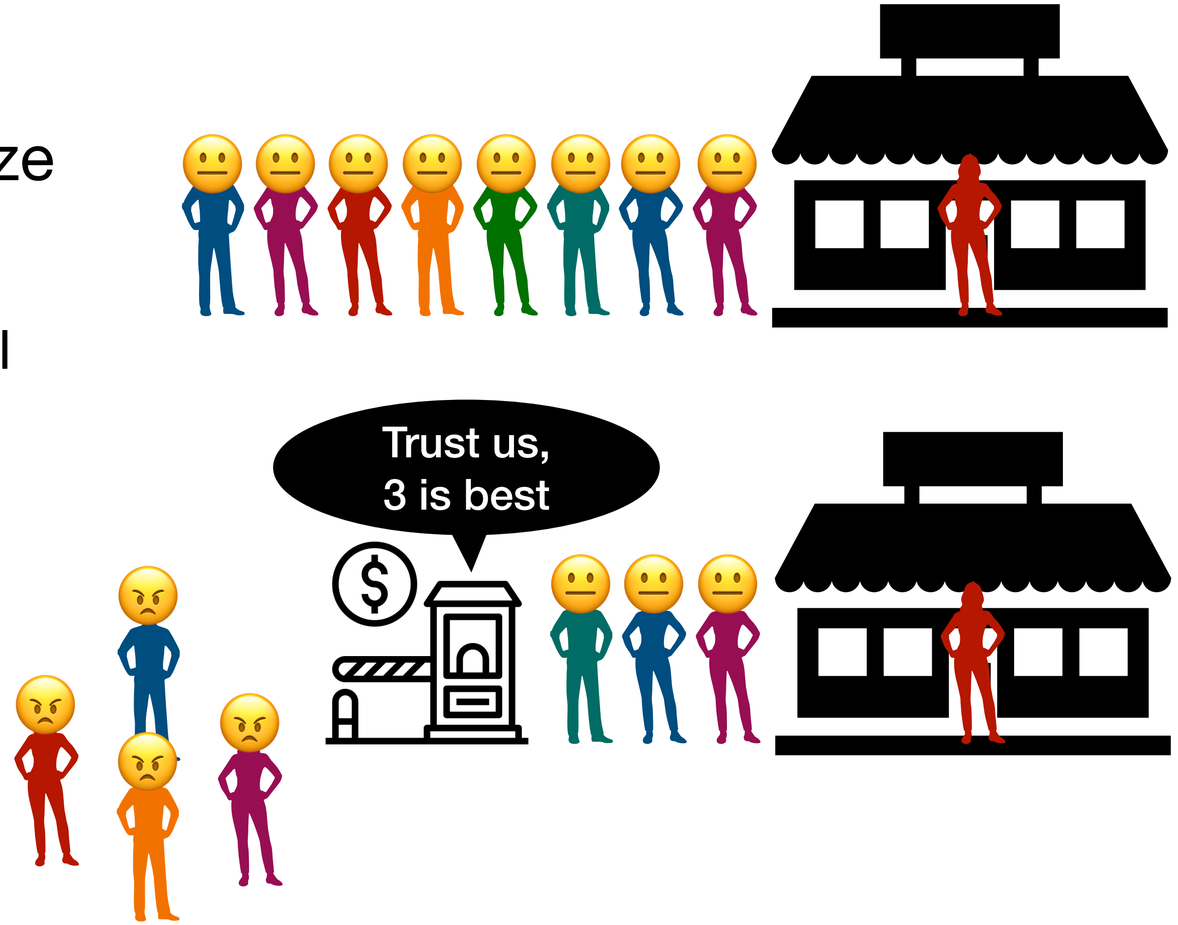
https://spannbaueradam.shinyapps.io/naor/

# Critiques

# Adam thoughts

## Before consulting the literature

- Doesn’t seem to consider how the toll affects an overall gain, just the queue size
  - We start optimizing overall gain: $P = (\lambda - \zeta)R - CE[i]$, but after setting a toll to bring down queue size we don’t readdress how this affects $P$
    - $P$ with the toll incorporated could be expressed with: $P = (\lambda - \zeta) \cdot (R - \theta) - CE[i]$

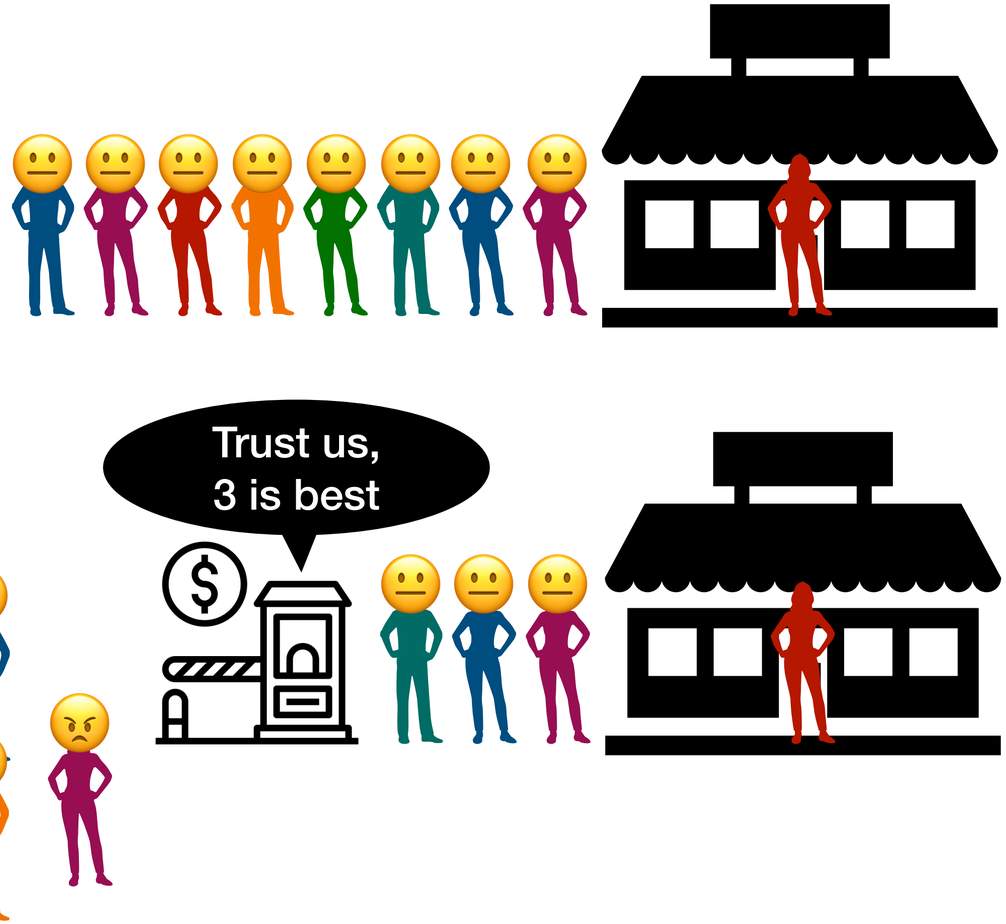



Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Adam thoughts

## Before consulting the literature

- Doesn’t seem to consider how the toll affects an overall gain, just the queue size
  - This could be considered net neutral because we’re optimizing for firm and customers?
  - “**<u>if</u>** the toll revenue may be used for redistribution of income among the population or for socially useful purposes the proposed imposition of tolls is an optimal procedure”
    - The “if” might hint at a tradeoff in firm’s favor




Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Adam thoughts

## Before consulting the literature

- When considering customer utility we don’t consider a cost for arriving early
- One of the cited works brings up the point:
  - “Some have argued that the man ahead in a queue imposes a cost on those who wait behind him and must pay for it. But I say that he has already paid for it by coming earlier.”
- But, to be fair, this seems like a useful simplification for most applications where we focus on steady state

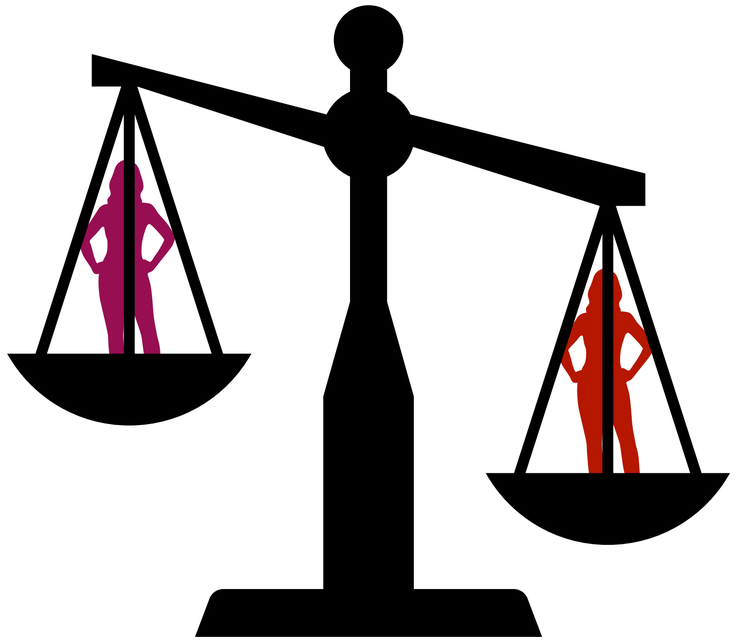
https://www.jacksonville.com/story/news/2009/11/26/stub-505/15965432007/

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Critiques

## Some other considerations

- Are customer's rewards and costs really all the same?
- Saaty's point against tolls on non-luxuries still stands - can be viewed as unjust to toll in some cases
- Relies on assumptions that customer's can gauge the cost-benefit accurately




Hassin, R., & Haviv, M. (2003). *To queue or not to queue: Equilibrium behavior in queueing systems* (Vol. 59). Springer Science & Business Media.

Huang, T., Allon, G., & Bassamboo, A. (2013). Bounded rationality in service systems. *Manufacturing & Service Operations Management*, *15*(2), 263-279.

Saaty, T. L., & Leeman, W. A. (1965). The Burdens of Queuing Charges-Comments on a Letter by Leeman. Operations Research, 13(4), 679–681. http://www.jstor.org/stable/167860

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica: journal of the Econometric Society, 15-24. https://doi.org/10.2307/1909200

# Implementations and extensions

# Real life implementations

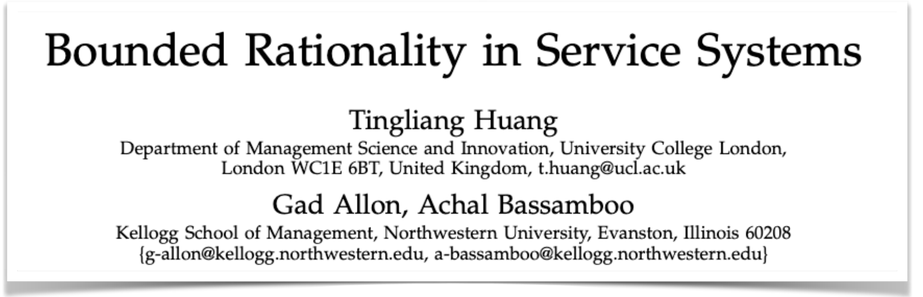
## Results, not just ideas

- Dynamic pricing in ride share
  - You observe cost before entering the queue to find a driver
  - Uber/Lyft/etc can affect queues with this dynamic pricing like Naor’s toll
- Fast track services in amusement parks
  - Pay toll to enter a special fast track queue


https://i.ytimg.com/vi/9yjZpBq1XBE/maxresdefault.jpg

# Extensions

## Provided by a familiar name


Bounded Rationality in Service Systems

Tingliang Huang
Department of Management Science and Innovation, University College London, London WC1E 6BT, United Kingdom, t.huang@ucl.ac.uk
Gad Allon, Achal Bassamboo
Kellogg School of Management, Northwestern University, Evanston, Illinois 60208
{g-allon@kellogg.northwestern.edu, a-bassamboo@kellogg.northwestern.edu}

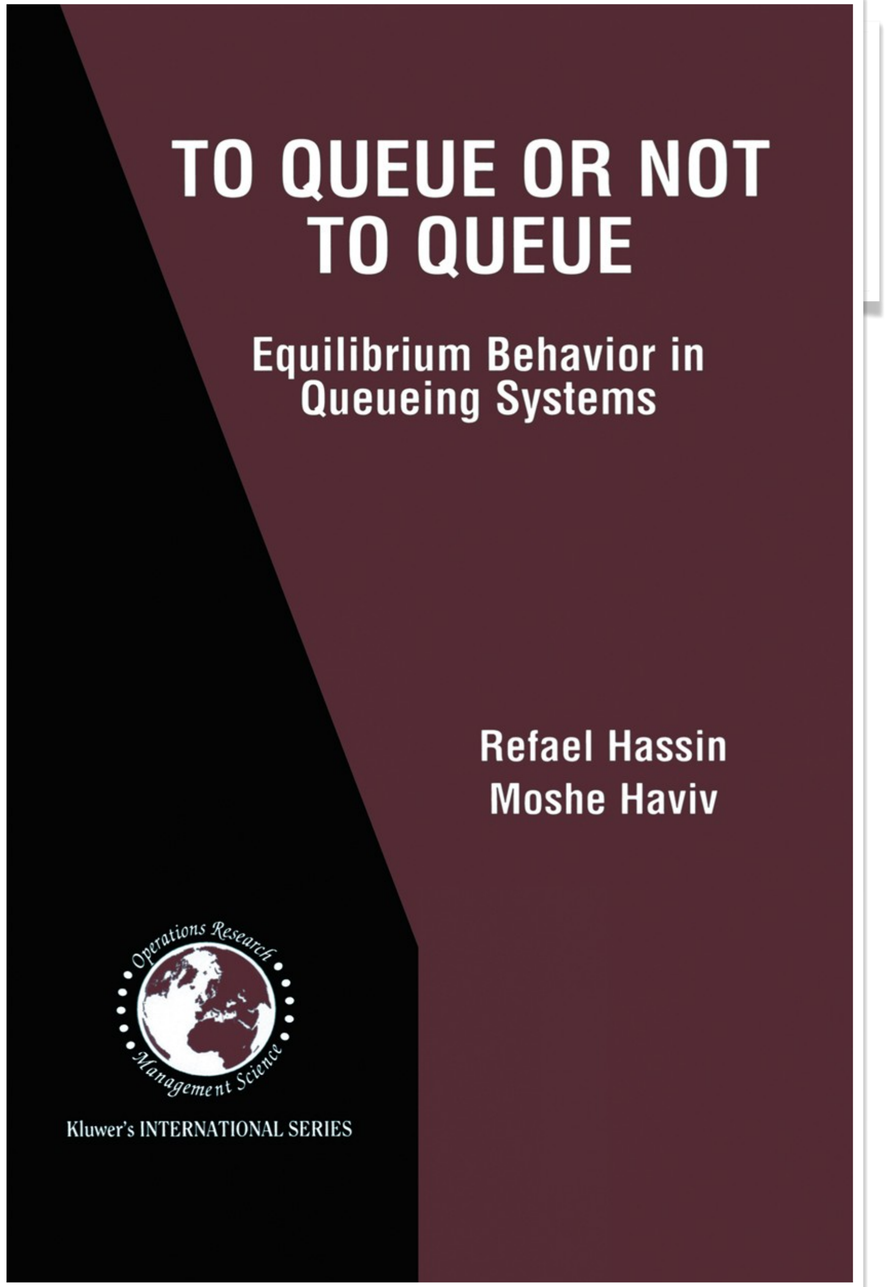
- “Yechiali (1971, 1972) extends Naor’s model to allow for GI/M/1 queues. Knudsen (1972) extends Naor’s model to allow for a multiserver queueing system in which arriving customers’ net benefits are heterogeneous. Lippman and Stidham (1977) extend the Naor model to the finite-horizon and discounted cases, showing that, in these settings, the economic notion of an external effect has a precise quantitative interpretation. Hassin (1986) considers a revenue maximizing server who has the opportunity to suppress information on actual queue length, leaving customers to decide whether to join the queue on the basis of the known distribution of waiting times. See Van Mieghem (2000), Hassin and Haviv (2003), Afèche (2004), and Hsu et al. (2009) for other extensions and a comprehensive literature review”

Huang, T., Allon, G., & Bassamboo, A. (2013). Bounded rationality in service systems. *Manufacturing & Service Operations Management*, *15*(2), 263-279.

# Extensions

## Provided by a familiar name

- “Yechiali (1971, 1972) extends Naor’s model to allo
  Knudsen (1972) extends Naor’s model to allow for a
  system in which arriving customers’ net benefits ar
  and Stidham (1977) extend the Naor model to the fi
  discounted cases, showing that, in these settings, t
  external effect has a precise quantitative interpretat
  considers a revenue maximizing server who has the
  information on actual queue length, leaving custom
  join the queue on the basis of the known distributio
  Van Mieghem (2000), **Hassin and Haviv (2003)**, A
  (2009) for other extensions and a comprehensive lit




Huang, T., Allon, G., & Bassamboo, A. (2013). Bounded rationality in service systems. *Manufacturing & Service Operations Management*, *15*(2), 263-279.

# Qs?