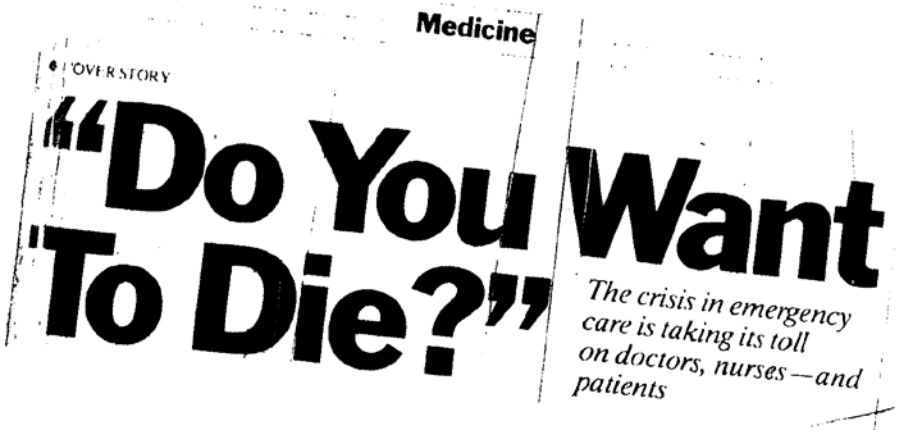
Variability and Queueing

Konstantinos (Kostas) Stouras
Assistant Professor, Operations Management
Konstantinos.Stouras@ucd.ie
www.stouras.com
@StourasK

EMERGENCY ROOM CROWDING AND AMBULANCE DIVERSION

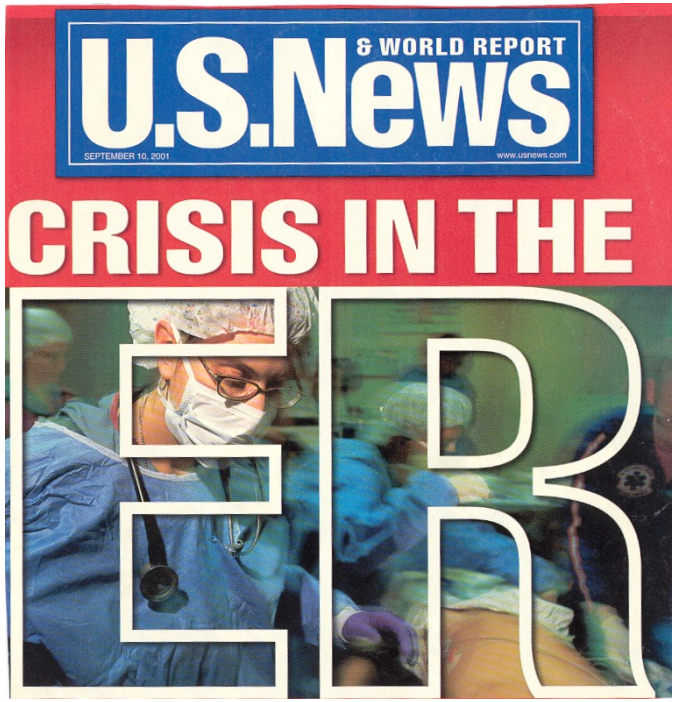

1 in 3 Hospitals Say They Divert Ambulances


Medicine

COVER STORY

"Do You Want To Die?"

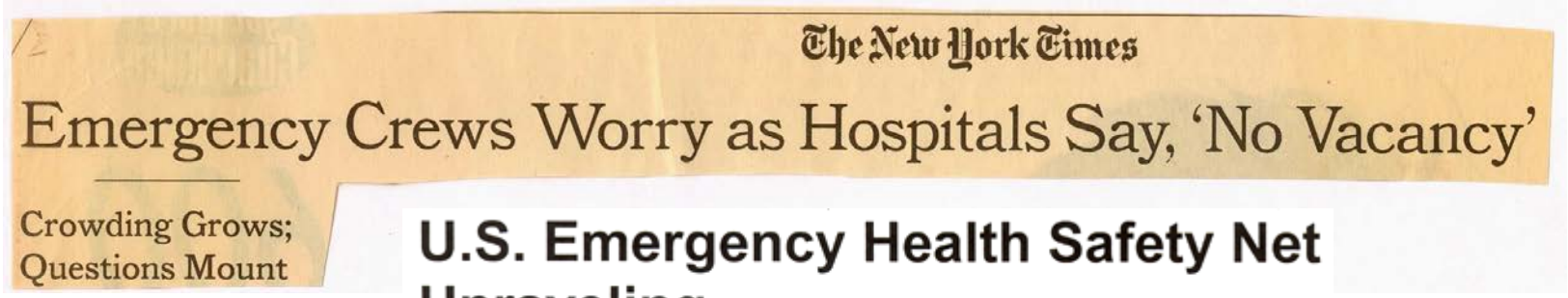
The crisis in emergency care is taking its toll on doctors, nurses—and patients


U.S.News & WORLD REPORT

SEPTEMBER 10, 2001

www.usnews.com

CRISIS IN THE ER


The New York Times

Emergency Crews Worry as Hospitals Say, 'No Vacancy'

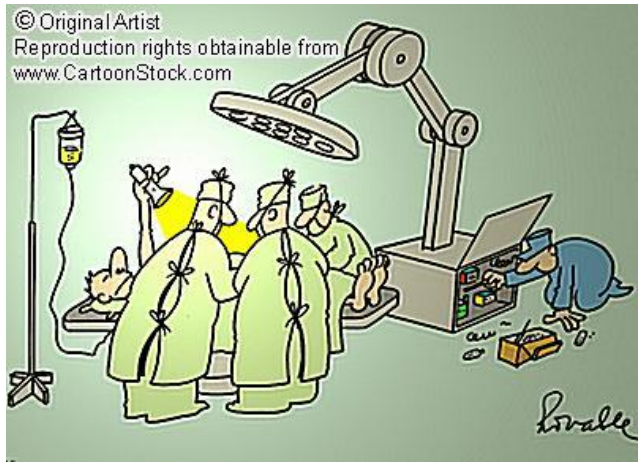
Crowding Grows; Questions Mount

U.S. Emergency Health Safety Net Unraveling

Waiting Times in the Emergency Room

Waiting Time

Service Time

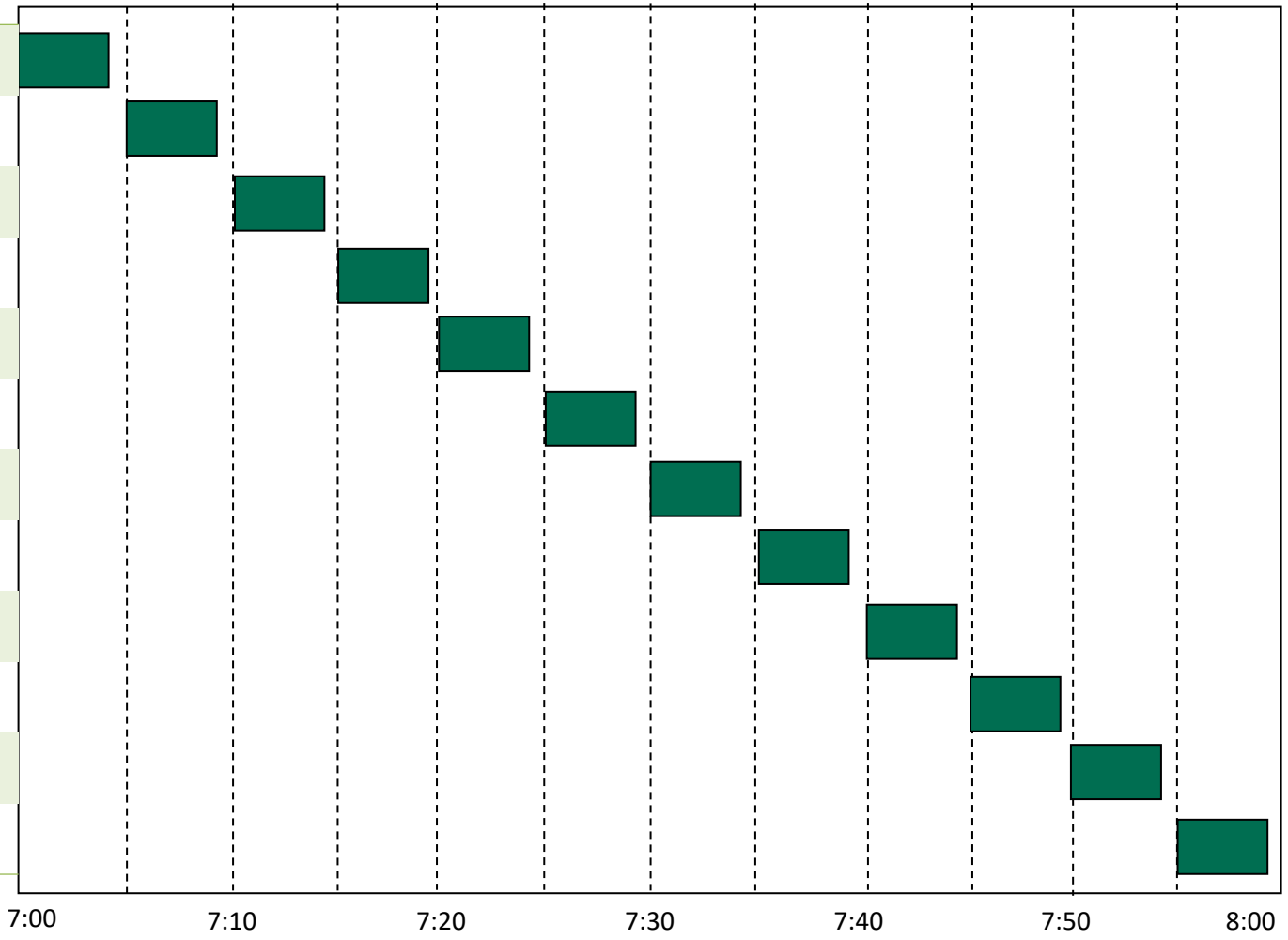


Total Time in the System

A Somewhat Odd Service Process

Patient	Arrival Time	Inter-Arrival Time	Service Time
1	7:00	0	4
2	7:05	5	4
3	7:10	5	4
4	7:15	5	4
5	7:20	5	4
6	7:25	5	4
7	7:30	5	4
8	7:35	5	4
9	7:40	5	4
10	7:45	5	4
11	7:50	5	4
12	7:55	5	4

Gantt Chart illustrating Service

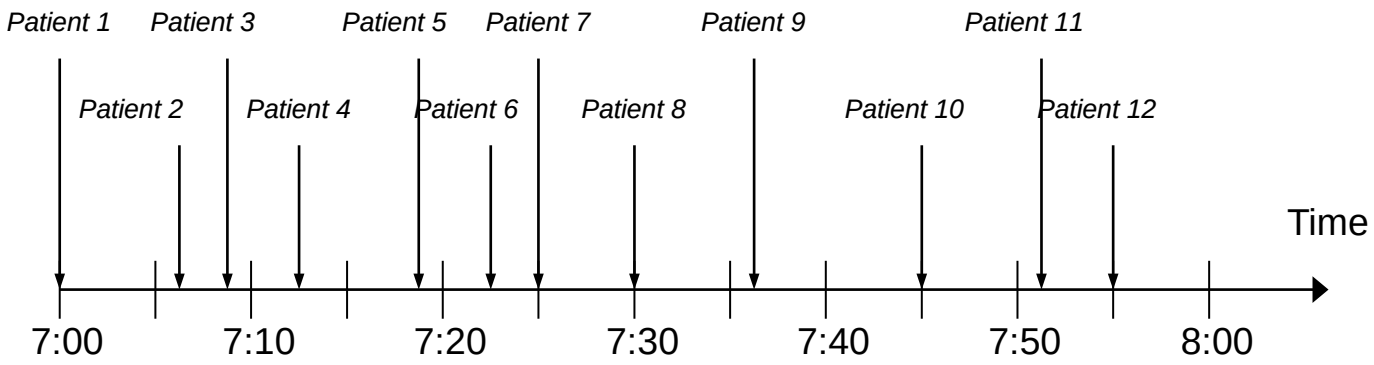


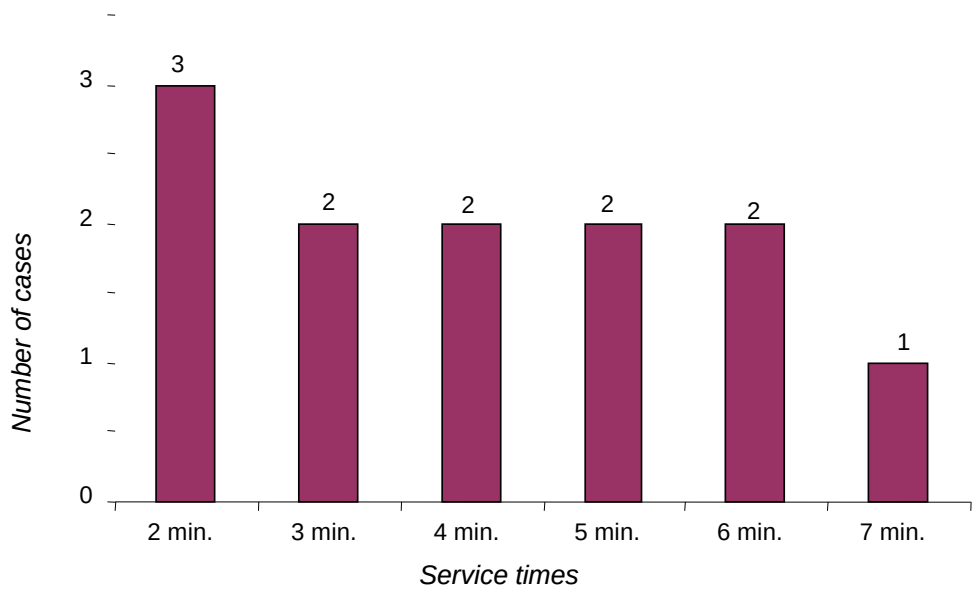
Average Inter-Arrival Time is 5 minutes, Average Service Time is 4 minutes. No Waits!

What is odd about this Service Process?

A MORE REALISTIC SERVICE PROCESS

Patient	Arrival Time	Inter-Arrival Time	Service Time
1	7:00		5
2	7:07	7	6
3	7:09	2	7
4	7:12	3	6
5	7:18	6	5
6	7:22	4	2
7	7:25	3	4
8	7:30	5	3
9	7:36	6	4
10	7:45	9	2
11	7:51	6	2
12	7:55	4	2



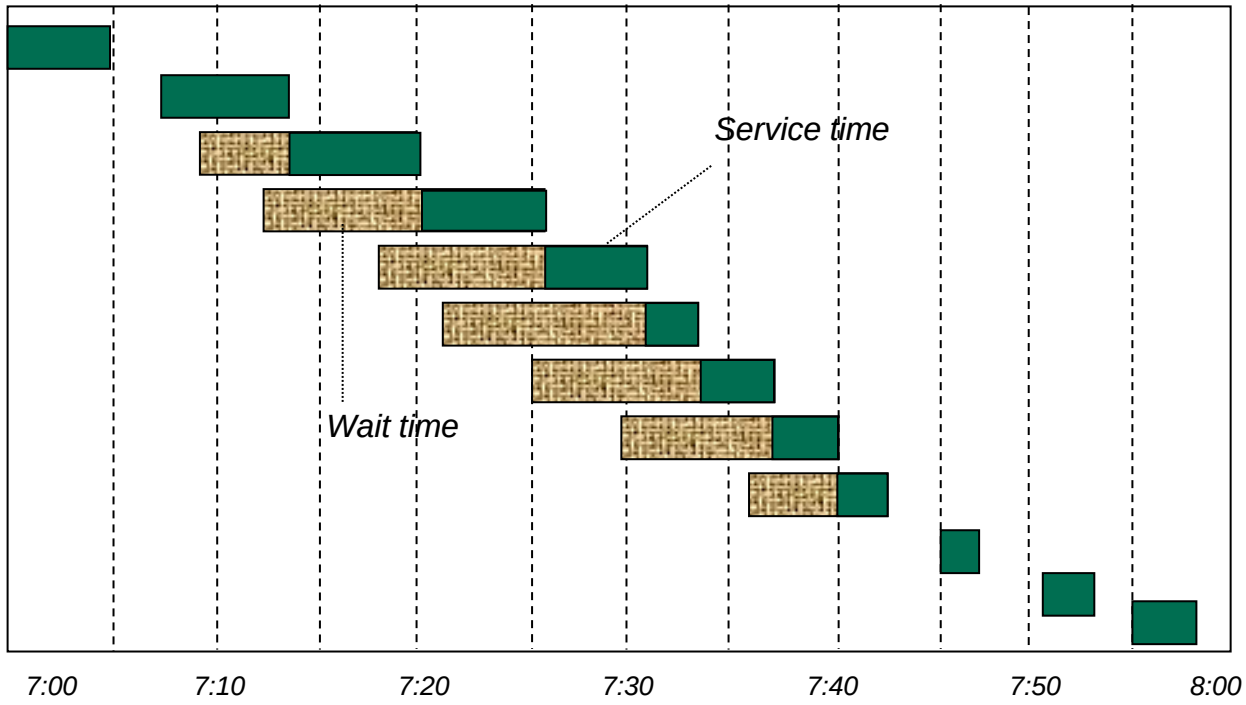


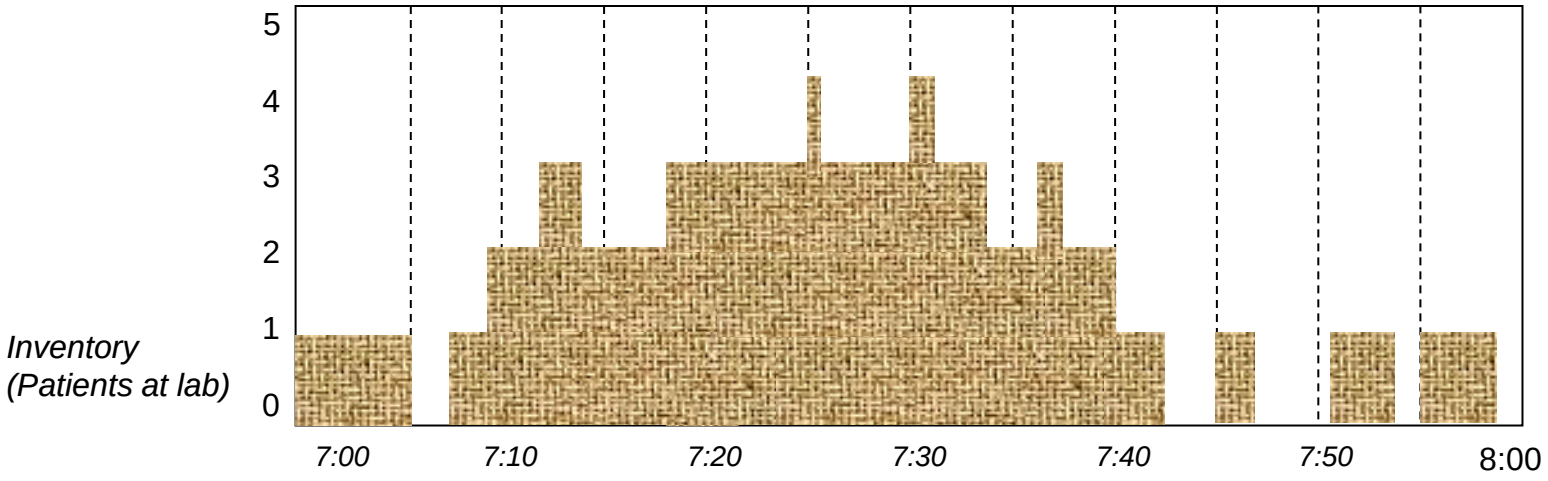
Average Inter-Arrival Time is 5 minutes, Average Service Time is 4 minutes. 12 patients arrived in this 1h.

Nothing changed it seems! Will the service performance be the same as before?

VARIABILITY LEADS TO WAITING TIME

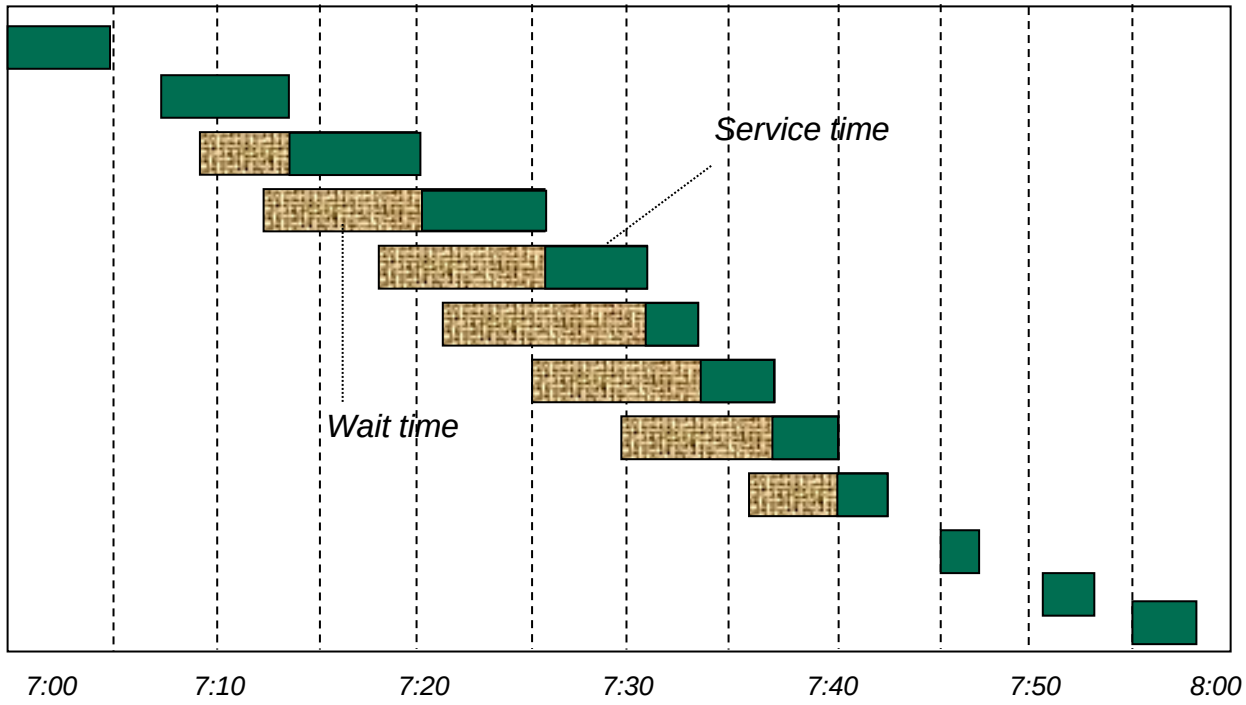
Patient	Arrival Time	Inter-Arrival Time	Service Time
1	7:00		5
2	7:07	7	6
3	7:09	2	7
4	7:12	3	6
5	7:18	6	5
6	7:22	4	2
7	7:25	3	4
8	7:30	5	3
9	7:36	6	4
10	7:45	9	2
11	7:51	6	2
12	7:55	4	2

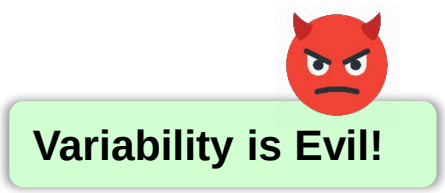




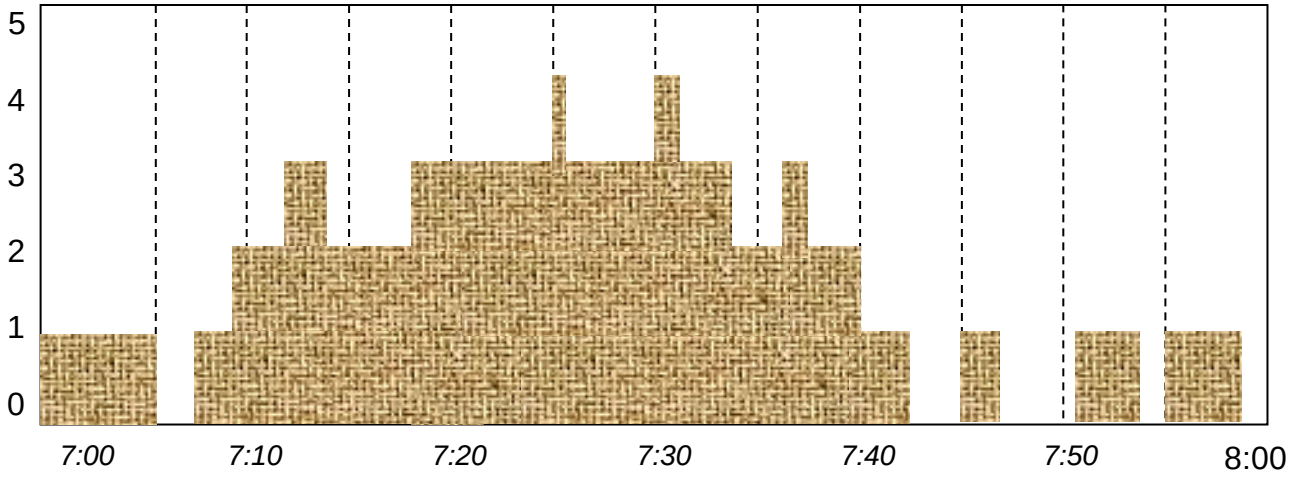
Variability Leads to Waiting Time

Patient	Arrival Time	Inter-Arrival Time	Service Time
1	7:00		5
2	7:07	7	6
3	7:09	2	7
4	7:12	3	6
5	7:18	6	5
6	7:22	4	2
7	7:25	3	4
8	7:30	5	3
9	7:36	6	4
10	7:45	9	2
11	7:51	6	2
12	7:55	4	2





Inventory (Patients at lab)



WHAT CREATES QUEUES?

Phantom Q: https://youtu.be/goVjVVaLe10?t=19

From Process to Queueing Parameters

Parameters

- *Number of servers: s*
- *Frequency of arrivals: λ*
- *Average service time: τ (Service rate: m=1/t)*
- *Utilization: ρ= λ* τ/s = λ/(μ*s)*
- *Coefficient of variation: CV = std. dev. / mean (of inter-arrival and service times)*

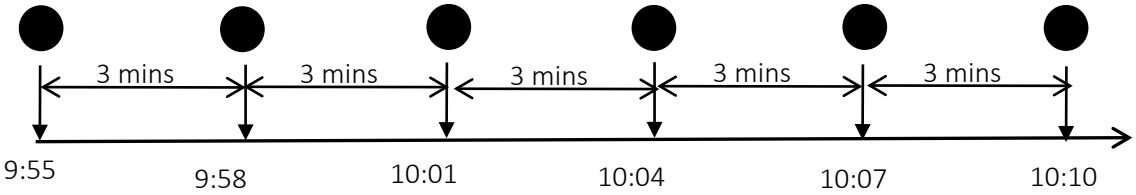
Average Performance Measures

- *Waiting time: W_q*
- *Total throughput time: $W = \tau + W_q$*
- *Number of customers in queue: L_q*
- *Number of customers in the system: L*

What Does CV Mean?

A process with CV=0

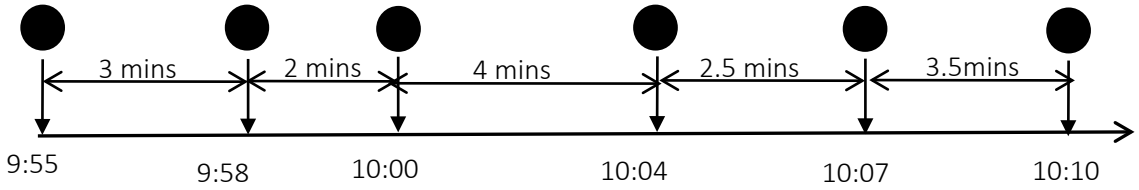
Perfectly Scheduled Arrivals, Output from an assembly line.



A process with CV=1

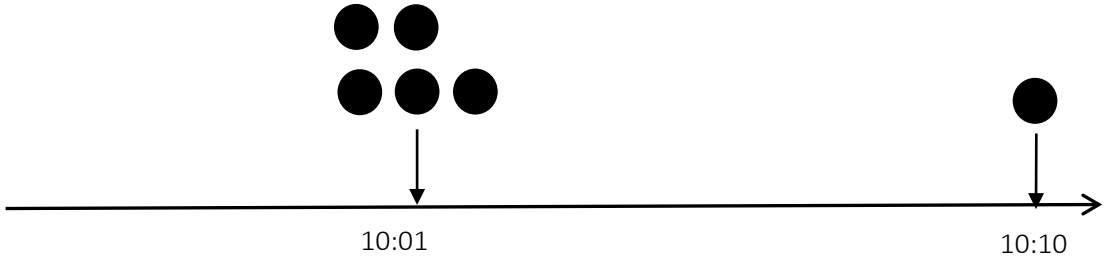
No relationship between arrivals, Independent Arrivals. Calls to a call center

Inter-arrival times are exponential, Arrivals according to a Poisson Process, Memory-less Arrivals

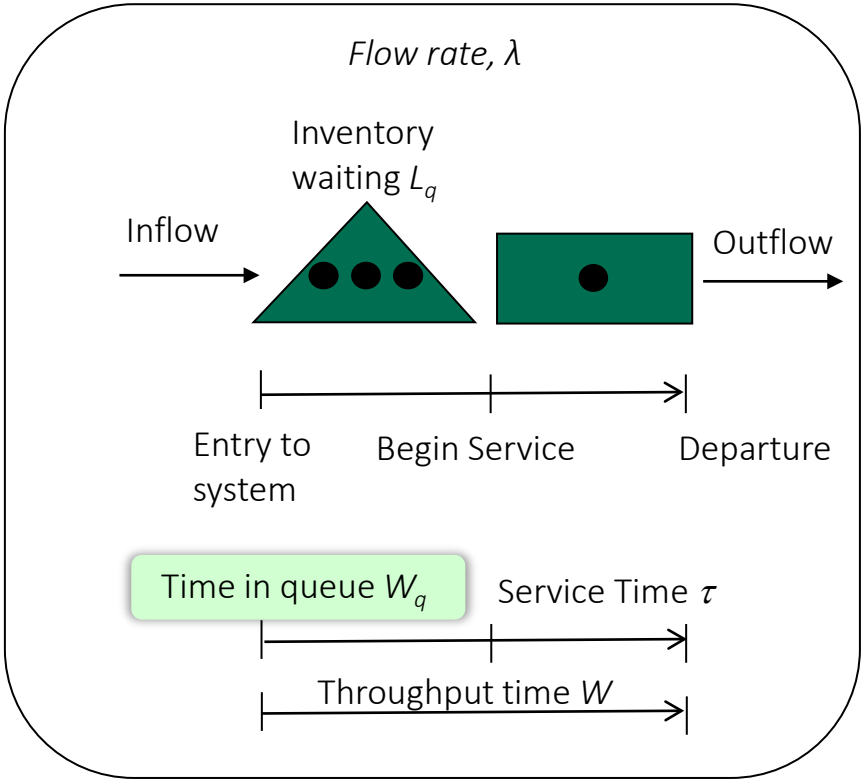


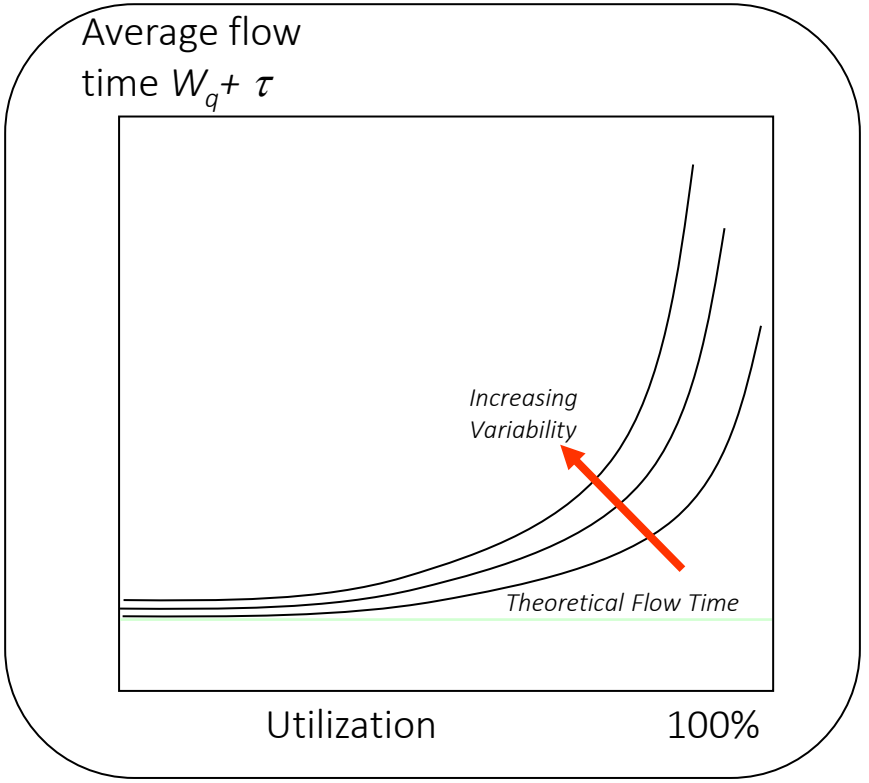
A process with CV>>1

Bunched Arrivals: Coffee orders at the bar.



The Waiting Time Formula (Approximation for Single Server)



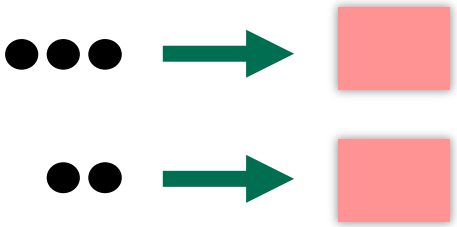


$$W_q = \tau * \left(\frac{\rho}{1-\rho}\right) * \left(\frac{CV_a^2 + CV_s^2}{2}\right)$$

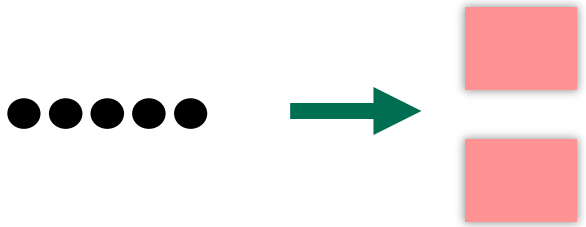
- Variability effect
- Utilization effect
- Scale effect

Good Operations for Multi-Server Systems

2 Single Server Systems



A 2 Server System



L_Q TABLE FOR MULTIPLE SERVERS

LOOKUP *s*, # OF SERVERS ON HORIZONTAL AXIS

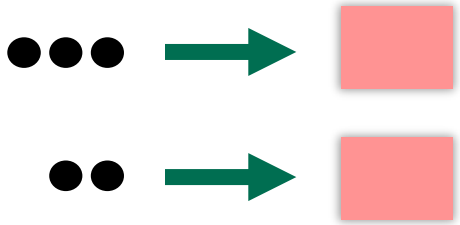
LOOKUP ρ, *COMPUTED UTILIZATION*, *TO GET* L_Q

Utilisation rate $(\rho)^*$	*Number of servers* 1	2	3
.10	.0111	.0020	.0004
.20	.0500	.0167	.0062
.30	.1286	.0593	.0300
.35	.1885	.0977	.0552
.40	.2667	.1524	.0941
⋮	⋮		
.88	6.4533	6.0414	5.7345
.90	8.1000	7.6737	7.3535
.92	10.5800	10.1392	9.8056
.94	14.7267	14.2712	13.9240

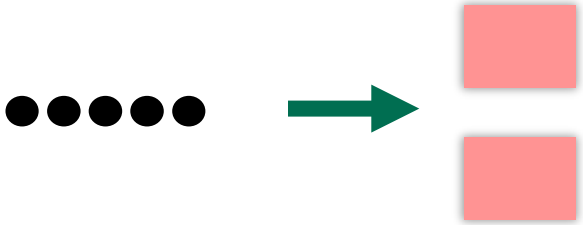
Which System Performs Better?

In theory...

2 single server systems 2x (s=1)



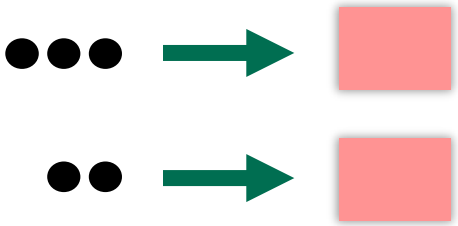
A 2 server system (s=2)



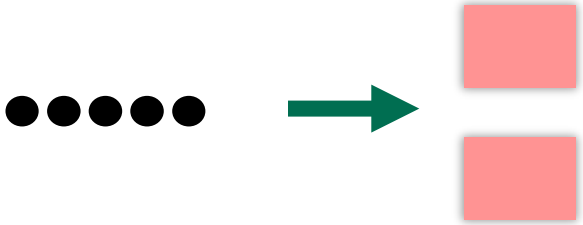
In practice...

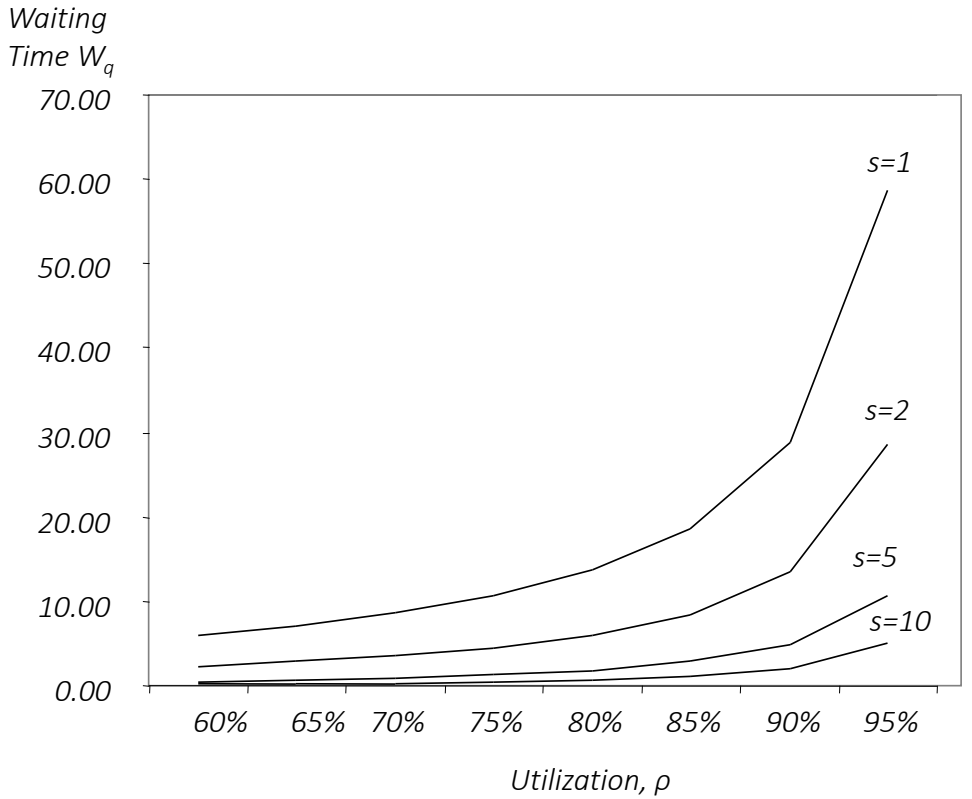
THE POWER OF POOLING

2 SINGLE SERVER SYSTEMS 2X (s=1)



A 2 SERVER SYSTEM (s=2)





Implications:

+ balanced utilization

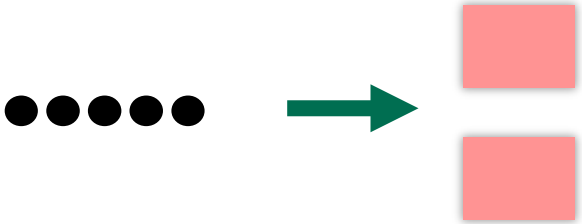
+ Shorter waiting time (pooled safety capacity)

- Changeovers and Set-ups

Pooled Systems perform better!

WAITING-IN-LINE BMIS AND INSIGHTS

- The pooled queue is better*



*All else equal!

Hauliers dismiss UK no-deal Brexit rehearsal as 'totally frivolous'

Operation Brock aimed at testing UK's ability to ease congestion near Channel ports

Mon, Jan 7, 2019, 08:23

Caveats:

1) Psychological reasons may make servers increase service rates in a dedicated system

2) Priority queues

Waiting-in-line BMIs and Insights

- The pooled queue is better
- The people behind you matter

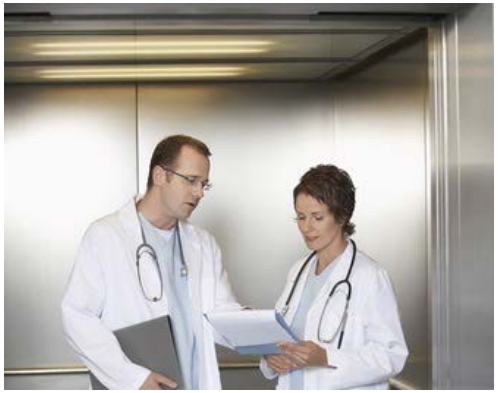
Waiting-in-line BMIs and Insights

- The pooled queue is better
- The people behind you matter
- Distractions make us queue longer

Disneyland Qs: https://youtu.be/XiVPMulew-w

Waiting-in-line BMIs and Insights

- The pooled queue is better
- The people behind you matter
- Distractions make us queue longer



Waiting-in-line BMIs and Insights

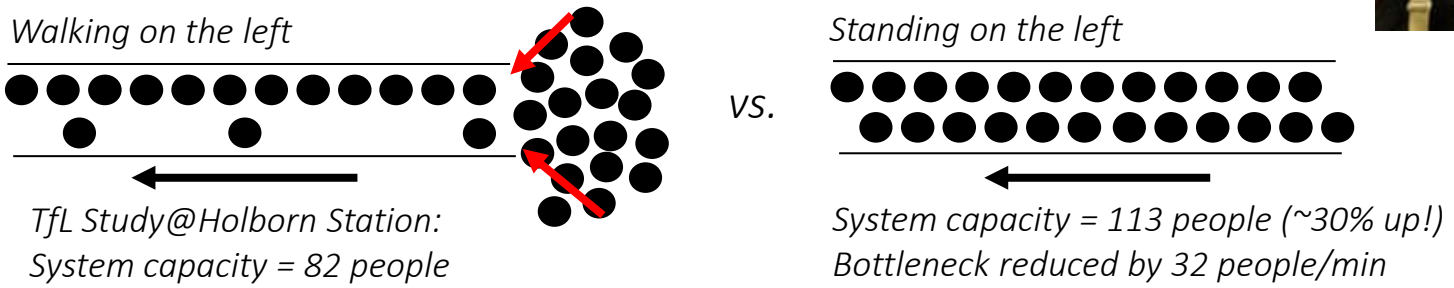
- The pooled queue is better
- The people behind you matter
- Distractions make us queue longer
- "Rule of six": People wait at most 6min,
  and don't join queues with more than 6 people

Waiting-in-line BMIs and Insights

- The pooled queue is better
- The people behind you matter
- Distractions make us queue longer
- "Rule of six": People wait at most 6min,
  and don't join queues with more than 6 people

- It's better to stand on both sides of the escalator

Guardian article: https://bit.ly/2OFUSBk

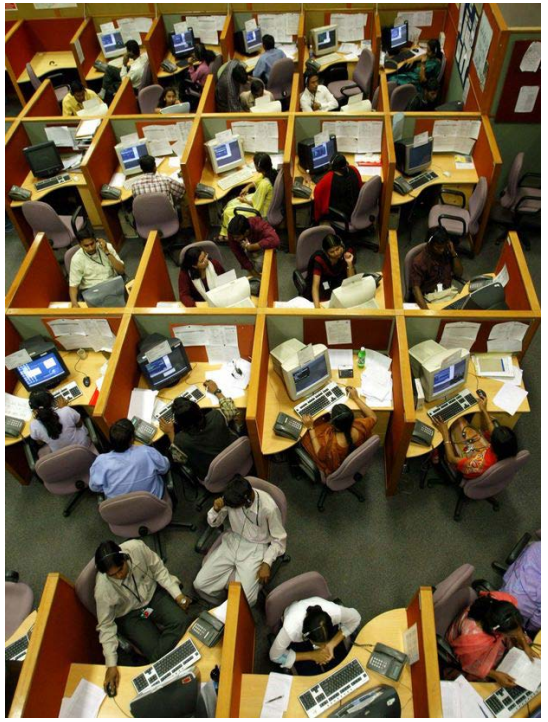


TfL Study@Holborn Station:
System capacity = 82 people

System capacity = 113 people (~30% up!)
Bottleneck reduced by 32 people/min

THE CUSTOMER CONTACT CENTER INDUSTRY

MASSIVE OUTSOURCED CALL CENTERS



Service Economics

Wide variability in demand during a shift

Shifts are staffed to manage peak demand

Typical utilization ~ 30%

Human Resources

Labor Intensive: Low Cost Countries

Relatively Low talent Labor
Sales call centers vs. service call centers

THE LIVEOPS BUSINESS MODEL*

liveops®

Calling all mompreneurs

The LiveOps Business Model

Interested Mompreneurs register and do basic on-line training. Mompreneurs when available, wait for calls (at home)

Received Calls are directed to highest Ranking Agent that is available

Mompreneur is paid for the time on the call

What is the utilization of an agent at LiveOps?

Who would work at LiveOps?

* Stouras, Girotra and Netessine: “LiveOps Inc: The Contact Center Reinvented”, INSEAD Case, 2014.

LiveOps: The Success Stories

American Red Cross

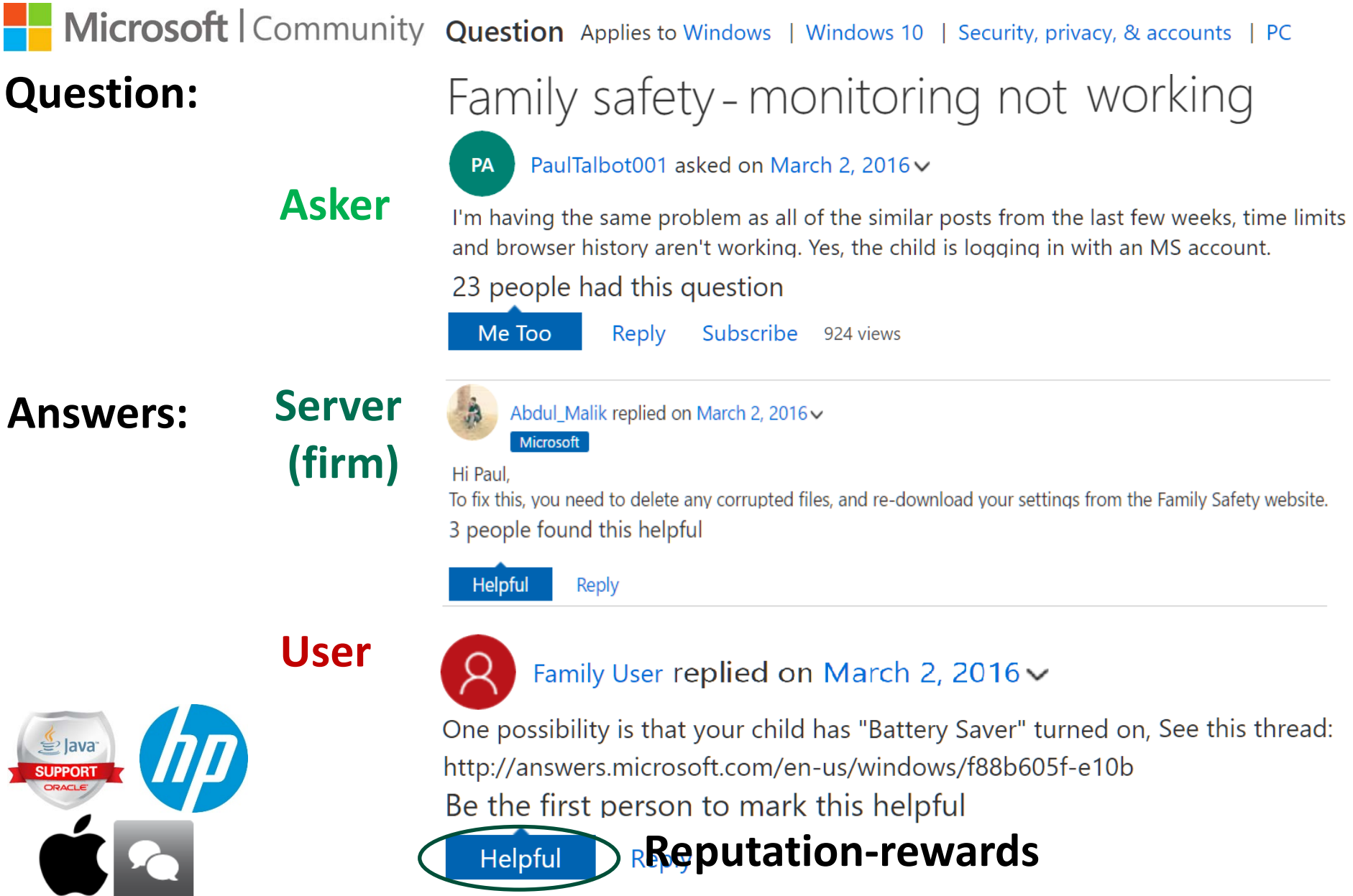
LiveOps Katrina: https://www.youtube.com/watch?v=SXH1I8erR2E

ADT · BEACHBODY® · Coca-Cola · ebay

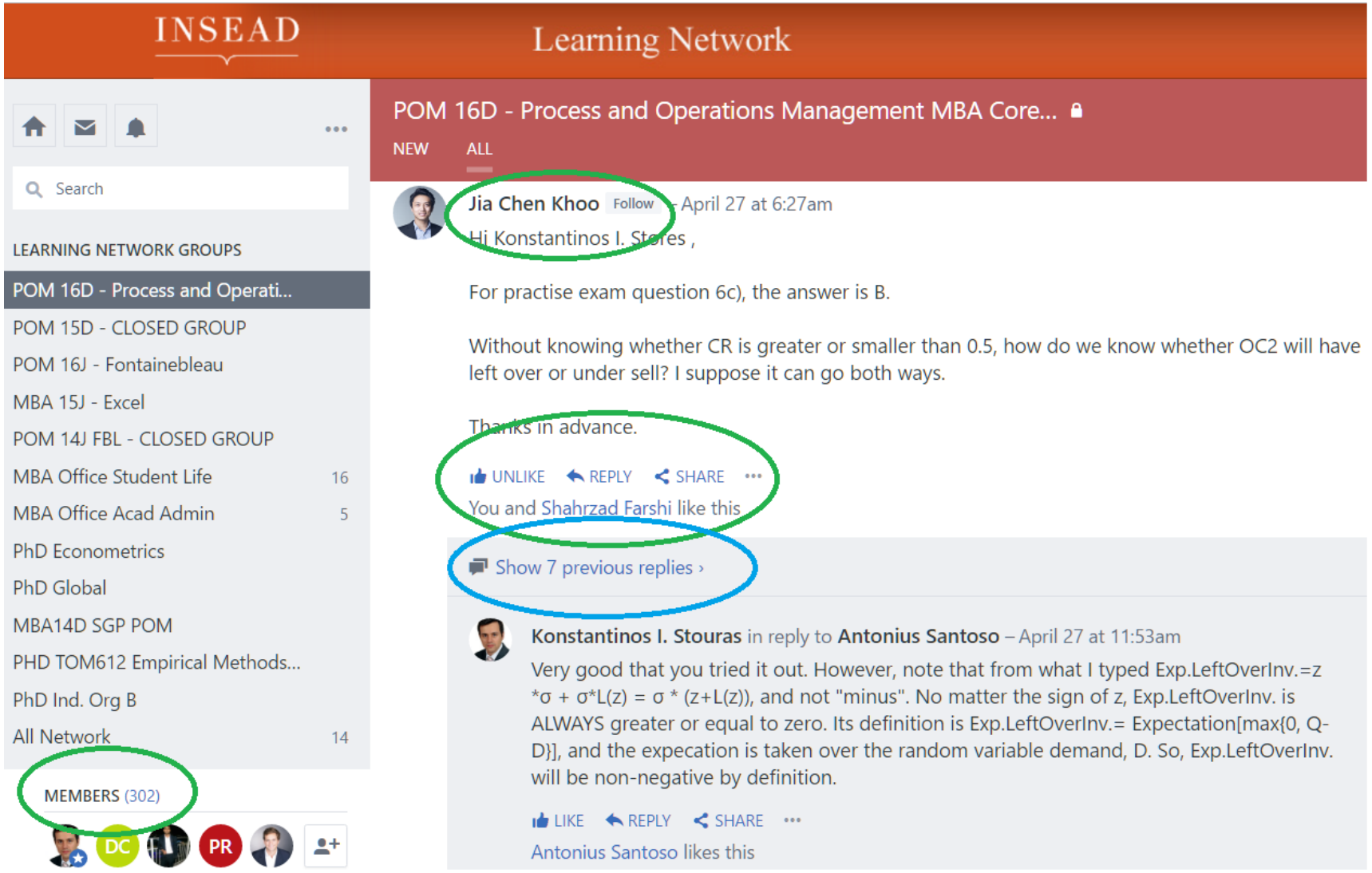
Kodak · glasses direct · Pizza Hut · West Marine

Amway · Fonterra · Murad · Snap-on

CAN WE IMPROVE LIVEOPS BUSINESS MODEL FURTHER?



THE RESEARCH IDEA: TEACHING MBAS



A. D. Frazelle and K. I. Stouras, *Managing a Question-and-Answer Forum*, Working Paper, 2021.

How Michael Dell Revolutionized the PC Industry

HP: Build To Forecast

Dell: Build To Order

Manufacture → *Demand*

Demand → *Manufacture*

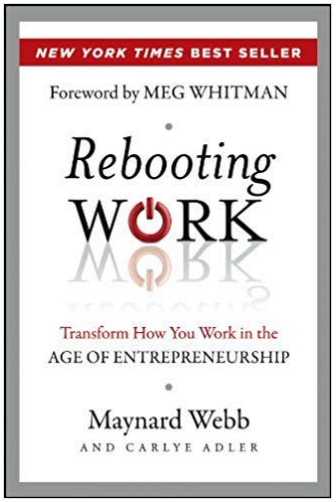
Re-sequence decisions!

THE COMMON THREAD?

The New York Times

Plugging Into the Gig Economy, From Home With a Headset

A company called Liveops has become the Uber of call centers by doting on its agents. But is the work liberating, or dehumanizing?

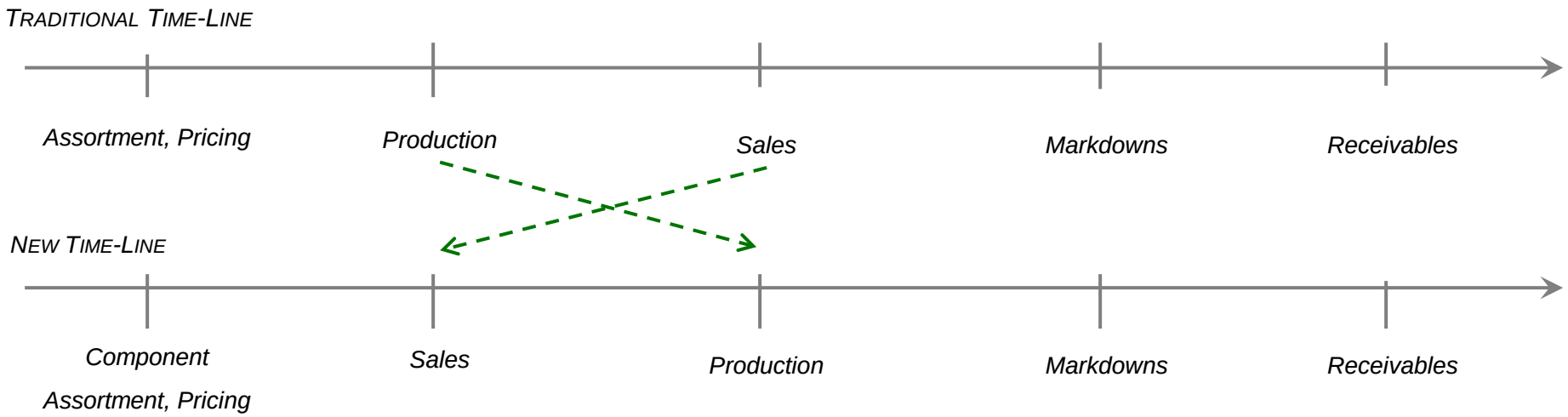


LiveOps reduces the risk of demand-supply (Calls-Agents) Mismatch

The Common Thread: Re-sequencing When Decisions are Made

Decide After Information

* Ch.4: The "*When*" strategy



The Innovation

- *Change sequence of events– Bets (Production, Hiring) and Information (Demand Information)*
- ***Changes Exposure to Sales Uncertainty Risks, Increase Costs***

Works best when

- *High mismatch costs (Demand Uncertainty)*
- *Modular product*
- *Desire for variety*
- *Limited benefits of commitment*

Realizing Business Model Innovations: Template When-Resequencing

Method I: Transfer to other Industry-Market

Pick your favorite Industry-Market

What are the consequential decisions made in this industry?
Are they made with perfect, ok, or poor information (information risk)?

When are these decisions made?
Map out important events before and after these decisions.
Has the critical decision to be made at the time it is made– Can it be resequenced?

If yes, you have a game changing idea…

Method II: Resolve pain points of an Industry

Think of industries characterized by high Information Risk.

- Frequent dramatic departures from budgeted performance metrics (sales, resource utilization, etc.);
- Wide variations in year-to-year performance;
- High exposure to prices and actions out of the control of the firm (for instance, energy prices, partners' behavior);
- Vulnerability of business performance to a few high-impact decisions subject to significant uncertainty;
- Expensive, frequently underutilized assets;

What are the critical decisions in these industries?

Can they be re-sequenced?

KEY LEARNINGS

- *Variability is Evil*
  - *Variability causes waits*
  - *Pooling helps improve system performance*
- *Business Model Innovation: On Demand Call Centers, Upside-Down R&D*
- *Template 1: Pooling and Diversification*
- *Template 2: Resequencing for BMI*