

David Maister (1984): The Psychology of Waiting Lines

The 1st Law of Service: Satisfaction = Perception - Expectation

2nd " " : It's hard to play catch-up ball

[Managerial consequences:

- Managing Perception: mirrors near an elevator
- " Expectation: estimate waiting-time in a restaurant (> wait)
- Make sure the encounter starts right

~

The Principles of Waiting : Propositions about the psychology of queuers.
Each has managerial consequences.

1. Unoccupied Time Feels longer than Occupied Time
2. Pre-Process Waits Feel longer than In-Process Waits
3. Anxiety Makes Waits Seem longer
4. Uncertain Waits are longer than Known, Finite Wait
5. Unexplained Waits are longer than Explained Waits
6. Unfair Waits are longer than Equitable Waits
7. The More Valuable the service, the longer I will wait
8. Solo Waiting Feels longer than Group Waiting



A Long Line for a Shorter Wait at the Supermarket



Robert Caplin for The New York Times

Sam Baris directing customers at Whole Foods in Columbus Circle, where the long line moves quickly.

By [MICHAEL BARBARO](#)
Published: June 23, 2007

Show New Yorkers a checkout line and they'll tell you whether it's worth the wait.

Readers' Opinions

Share Your Thoughts

What are some of your recent checkout experiences?

 [Post a Comment](#)

Multimedia

[Starbucks](#) at 9 a.m.? Eight minutes, head to the next one down the street. Duane Reade at 6 p.m.? Twelve minutes, come back in the morning.

But now a relative newcomer to Manhattan is trying to teach the locals a new rule of living: the longer the line, the shorter the wait.

Come again?

For its first stores here, [Whole Foods](#), the gourmet supermarket, directs customers to form serpentine single lines that feed into a passel of cash registers.

Banks have used a similar system for decades. But supermarkets, fearing a long line will scare off shoppers, have generally favored the one-line-per-register system.

By 7 p.m. on a weeknight, the lines at each of the four Whole Foods stores in Manhattan can be 50 deep, but they zip along faster than most lines with 10 shoppers.

SIGN IN TO E-MAIL
OR SAVE THIS

PRINT

REPRINTS

SHARE

ARTICLE TOOLS
SPONSORED BY

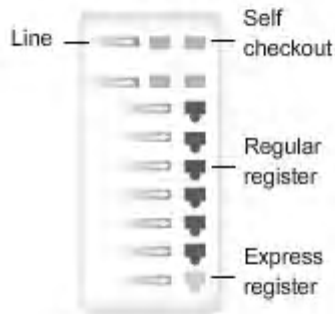
IN THEATRES JULY 6

One Afternoon, Five Lines

On a recent Sunday, a visitor checked out at five Manhattan grocery stores, which use a variety of checkout systems.



ONE LINE PER REGISTER



PEOPLE IN LINE: 10

WAIT TIME: 12 min, 54 sec

Wait times and person counts were collected in nonexpress lines from 4 to 6 p.m. on a nonholiday Sunday. Measured with a stopwatch, wait times began upon entering the line and ended when the cashier handed over a receipt.

The New York Times

Because people stand in the same line, waiting for a register to become available, there are no “slow” lines, delayed by a coupon-counting customer or languid cashier. And since Whole Foods charges premium prices for its organic fare, it can afford to staff dozens of registers, making the line move even faster.

“No way,” is how Maggie Fitzgerald recalled her first reaction to the line at the Whole Foods in Columbus Circle. For weeks, Ms. Fitzgerald, 26, would not shop there alone, assigning a friend to fill a grocery cart while she stood in line.

When she discovered the wait was about 4 minutes, rather than 20, she began shopping by herself, and found it faster than her old supermarket.

“By now,” Ms. Fitzgerald said of those competitors, “you’d think everyone else would catch onto this.”

The science of keeping lines moving, known as queue management, is a big deal to big business. Since arriving in 2001, Whole Foods stores in Manhattan have won bragging rights as the top sellers among grocery chains here, with sales of \$42 million per store last year, according to Modern Grocer, a trade publication.

Some of its competitors acknowledge they are feeling a bit of line envy. “I should give it a closer look,” said John A. Catsimatidis, owner of the Gristede’s chain, which uses the traditional line system.

Even New York grocery chains that use a similar system but on a smaller scale admire the efficiency of Whole Foods. “It’s very impressive,” said Jon Basalone, a senior vice president at Trader Joe’s.

Lines can also hurt retailers. Starbucks spooked investors last summer when it said long lines for its cold beverages scared off customers. [Wal-Mart](#), too, has said that slow checkouts have turned off many.

And they are easily turned off. Research has shown that consumers routinely perceive the wait to be far longer than it actually is.

“We have good clocks in our heads for roughly three minutes,” said Paco Underhill, founder of Envirosell, a retail consulting firm.

“Once we get beyond that, time expands wildly,” he said. “If somebody is there for 4.5 minutes and you ask them how long they waited, they will say 15 minutes.”

In most of the United States, the wait in a grocery store checkout line is negligible — under a minute, Mr. Underhill has found.

Then there is New York City. Here, hundreds of shoppers, in grocery stores that feel as cramped as a junior one-bedroom, can wait 10 minutes or more to reach a cashier.

Whole Foods executives spent months drawing up designs for a new line system in New York that would be unlike anything in their suburban stores, where shoppers form one line in front of each register.

That traditional system, they determined, would take up too much space and could not handle the crowds they expected here.

The single-line, bank-style system was quickly chosen for its statistical efficiency. Then, Whole Foods paired the system with possibly the largest number of registers in the city, more than 30 per store, and it hired an army of cashiers to staff them throughout the day (including “floaters” to fill in for those who need a break).

The result is one of the fastest grocery store lines in the city. An admittedly unscientific survey by this reporter found that at peak shopping times — Sunday, from 4 p.m. to 6 p.m. — a line at Whole Foods checked out a person every 4.5 seconds, compared with 19.6 seconds for a line at Trader Joe’s.

Granted, it may not be an apples-to-organic-apples comparison, but when faced with a line of 50 people, it takes about 4 minutes to check out at Whole Foods, half the time it takes at competing chains with significantly shorter lines. (With a 7-person line at Zabar’s one Sunday, it took about 8 minutes to check out. With just 10 people in line, it took about 13 minutes at the Food Emporium.)

“Whole Foods has just figured it out,” said Kelli Wicker, 38, who waited less than two minutes to buy \$15 worth of groceries at the Whole Foods at Union Square, despite a line of more than a dozen people.

Perhaps the most important role players in the Whole Foods system are the “line managers,” who monitor the flow of people, direct them to a cash register and, when needed, hold up signs saying how long it will take to check out. In another innovation, color-coded digital screens are now replacing those humans.

Others have tried to copy the Whole Foods system, including Trader Joe's, a popular California grocery chain that opened its first Manhattan store last year. But with far fewer cash registers, lines often snake around the entire perimeter of the store. The wait on a typical Sunday night is about 20 minutes (which might explain why a screaming match broke out one Sunday after a customer tried to sneak into the middle of the 75-person line).

"It is something that we recognize and would like to remedy," said Mr. Basalone of Trader Joe's.

Michael Ridgway, 33, no longer shops at Trader Joe's. "The line just does not move and makes it impossible to shop in the store," he said. But every week, he and his girlfriend, Jennifer Tolan, 29, queue up, with 50 to 70 strangers, at Whole Foods in Columbus Circle. "You can't pick a slow line," Ms. Tolan said.

"Whole Foods" – Union Square, Manhattan

I wish to share with you some pictures I took during the break at "Whole Foods" supermarket in Union Square, Manhattan.

Attached are two pictures of the cash register queue. A few notes regarding:

- The registers are located at one end of the store.
- There are two groups of four queues each, one on either side of the registers area.
- Each of the four queues is assigned a color – orange, white, blue, red.
- When a register is free to accept a client, it's number appears on a screen above the group of queues, showing which "color" should approach this register. (the photo shows – "red line to register 12, orange line to register 9").
- The registers are arranged in two groups with a large aisle between them. Each group has 10 registers, 5 on each side of the group. There are 3 additional registers to the right of the right hand group that are not seen in the photo. In total – 23 registers.
- A person is called only to a register that is directly in front of him/her so that all clients walk in straight lines and no collisions should occur. (the orange and white queues will go to registers 4-9 only).
- It goes without saying that I felt I waited under 10 minutes, although the store was quite busy.

Another interesting feature in the store has to do with it having three stories. Seeing as this is a supermarket, where people use carts, there are two types of escalators right beside each other – one for people and one for carts. The cart-escalator is built so that the wheels are locked into place and the cart is independently stable throughout the journey.

In short, an elegant and efficient service system. I was quite impressed.





בהמשך לקורס במערכת שירות שהשתתפתי בו בסמסטר שעבר, שמתי לב שהתחלתי לפתח חיבה לתורים... רציתי לשתף אותך בתמונות שצילמתי במהלך החופשה בסופרמרקט בניו-יורק, "Whole Foods", על Union Square במנהטן. זאת רשת סופרמרקטים יוקרתית מאוד, גדולה מאוד, שמוכרת הרבה מוצרי בריאות יקרים יותר ופחות.

מצורפות שתי תמונות מהתור לקופות ביציאה.

- יש קופות במקום אחד בחנות.
- יש שתי קבוצות של 4 תורים כל אחת משתי צידי מתחם הקופות.
- לכל אחד מ-4 התורים בכל צד יש צבע- כתום, לבן, כחול, אדום.
- כשקופה מתפנה, מופיע מספר הקופה אליה יש לגשת על גבי מסך מעל התור לפי הצבע של התור בו הנך עומד (רואים בתמונה, התור הכתום לקופה 9, התור האדום לקופה 12).
- הקופות מסודרות בשתי קבוצות גדולות עם מעבר ביניהם כך שבכל קבוצה 10 קופות - 5 מכל צד של הגוש. יש עוד 3 קופות שלא רואים בתמונה, מימין לתור הימני. סה"כ 23 קופות. (כפי שניתן לראות בתמונה)
- אדם נקרא רק לקופה שנמצאת מלפניו כך שעליו ללכת בקו ישר בלבד (תור כתום ולבן ילכו לקופות 4-9 למשל).
- מיותר לומר כמובן שחיכיתי בתור פחות מ- 3 דק' לדעתי, למרות שהחנות הייתה עמוסה.

עוד גימיק יפה שהיה בחנות (וחבל שלא צילמתי) נגע לכך שיש לחנות שלוש קומות. כיוון שמדובר בסופרמרקט, ובו אנשים משתמשים בעגלות, היו שני סוגים של מדרגות נעות צמודות זו לזו - האחת לאנשים והשנייה לעגלות - מבנה המדרגות הנעות היה כזה שגלגלי העגלה ננעלים והעגלה נשארת יציבה באופן עצמאי עד ליציאה בקומה מעל/מתחת.

בקיצור, מערכת שירות יעילה וגם יפה.

Pooling in a Q-Net

Pooling queues : geographic pooling (virtual service center)
servers : capacity pooling (fast vs. slow)
tasks : job design (later)

Must combine Operational + Psychological Aspects.

Rothkopf & Beth, 1987:

Common belief: combining queues is beneficial ...

e.g. banks and other counter systems.

But many operations do not combine queues

e.g. supermarkets, toll booths, rabbinate, doctors,...

In favour:

- (1) $m \times M/M/1 (\lambda, \mu)$ vs. $M/M/m (m\lambda, \mu)$
 $\forall \lambda, \forall m > 1$, the latter has smaller average wait + variance.
- (2) Share equipment (*e.g. Doctors' clinic, but see below*)
- (3) Fairness perception: no slips or skips (*Anat Rafaeli + students*)

Against

Homogeneous services

- (1) $m \times M/M/1$ not always the "right" alternative to $M/M/m$;
human (intelligent) customers jockey, join shortest queue, renege
- (2) Often alternative to $m \times M/M/1$ is $M/M/m$ with overhead,
namely $M/M/m (m\lambda, \mu - \delta)$.
- (3) Physically or psychologically prohibitive:
e.g., lines too long scare customers: cars, customers with luggage,

\Downarrow
snake-like queues

\Uparrow
airports' customs.

Heterogeneous customers/servers

- (4) Depersonalization (doctors, rabbinate)
- (5) Think of combining the express-lines with the rest (*M. + Reiman, 1996*)
- (6) Flexible servers expensive to hire, train, maintain.

Question: Design that "mixes" efficiency and fairness (physical queues)? *U.S. immigration.*
Business Growth (Strategic Q-Theory): Kleinrock's cycle. (1976, Classic) : *in Handout*

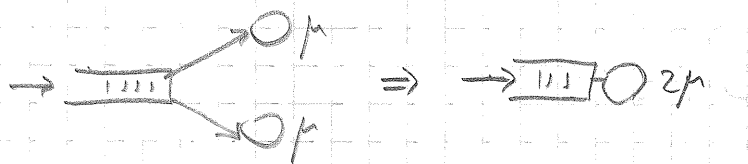
Pooling in Queuing Networks

Pooling Queues: Geographic / Telephone Pooling



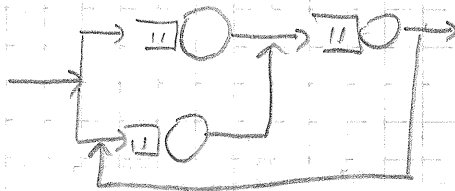
Gain: Load Balancing
(avoids a queue with
idle servers) at all times

Pooling Servers: Capacity Pooling

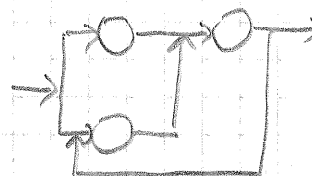


Gain: Max Capacity
at all times

Pooling Tasks: Job Design (Recall Haifa-Municipality, giving
rise to a Phase-Type service)



⇒



÷ Service Process

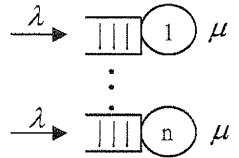
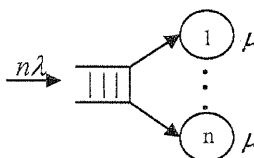
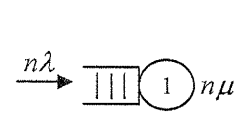
References:

Rothkopf & Beth 1982

M. & Reiman

Kleinrock's Books on Queuing (1976, classic)

Pooling Queues (one vs. many) and Servers (slow vs. fast)**Erlang-C**

	$n \times M M 1$ λ, μ	$M M n$ $n\lambda, \mu$	$M M 1$ $n\lambda, n\mu$
	Multiple queues Multiple Servers	Single queue Multiple Servers	Single queue Single server
			
Tradeoff	Separate vs. single queue		Slow vs. fast server
Enabler	Process design		Technology
Gain	Load balancing $\forall t$ (avoid a long queue & idle servers)		Maximal capacity $\forall t$ $\leq n\mu$ vs. $n\mu$
Utilization	$\frac{\lambda}{\mu} = \rho$	$\frac{(n\lambda)}{(n)\mu} = \rho$	$\frac{(n\lambda)}{(n\mu)} = \rho$
$P\{W_q > 0\}$	ρ	$E_{2,n}(\rho)$	ρ
$E[W_q W_q > 0]$	$\frac{1}{\mu} \cdot \frac{1}{1-\rho}$	$\frac{1}{n\mu} \cdot \frac{1}{1-\rho}$	$= \frac{1}{n\mu} \cdot \frac{1}{1-\rho}$
$E[W_q] =$ $E[W_q W_q > 0] \cdot P\{W_q > 0\}$	$\frac{1}{\mu} \cdot \frac{\rho}{1-\rho}$	$\frac{1}{\mu} \cdot \frac{E_{2,n}}{n(1-\rho)}$	$\frac{1}{\mu} \cdot \frac{\rho}{n(1-\rho)}$
$E[S]$	$\frac{1}{\mu}$	$\frac{1}{\mu}$	$> \frac{1}{n\mu}$
$E[W_{sys}] =$ $= E[W_q] + E[S]$	$\frac{1}{\mu} \cdot \frac{1}{1-\rho}$	$\frac{1}{\mu} \cdot \left[\frac{E_{2,n}}{n(1-\rho)} + 1 \right]$	$\frac{1}{n\mu} \cdot \frac{1}{1-\rho}$

Basic Relations:

$$1 \leq \frac{1 - E_{2,n}}{1 - \rho} \leq n \quad \left(\begin{array}{l} \Rightarrow E_{2,n}(\rho) \leq \rho \\ \text{Intuition?} \end{array} \right)$$

Proof:

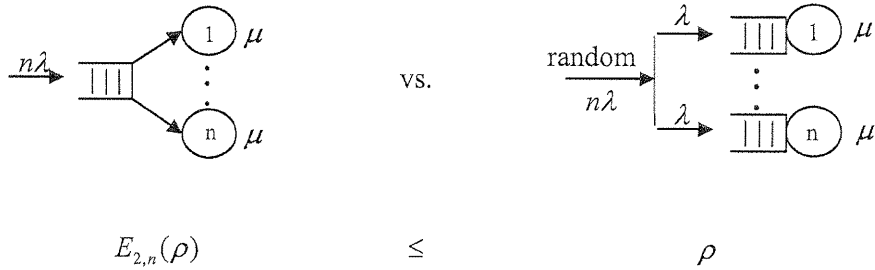
$$\begin{aligned} 1 - E_{2,n} &= P(\text{at least one server idle}) = \\ &= P\left(\bigcup_{i=1}^n \{\text{server } i \text{ idle}\}\right) \quad \left\{ \begin{array}{l} \leq \sum_{i=1}^n P\{\text{server } i \text{ idle}\} = n(1 - \rho) \\ \geq P\{\text{server } i \text{ idle}\} = 1 - \rho \end{array} \right. \\ &\qquad\qquad\qquad q.e.d. \end{aligned}$$

Corollary: $1 - \rho \leq 1 - E_{2,n} \Rightarrow E_{2,n} \leq \rho \Rightarrow$ 2 3

Corollary: $1 - E_{2,n} \leq n(1 - \rho) \Rightarrow \frac{1}{n(1 - \rho)} \leq \frac{E_{2,n}}{n(1 - \rho)} + 1 \Rightarrow$ 6

2 3	\Updownarrow	6
$EW_q(n, n\lambda, \mu) \leq EW_q(1, n\lambda, n\mu)$		$EW_{sys}(n, n\lambda, \mu) \geq EW_{sys}(1, n\lambda, n\mu)$
Multiple-slow \leq Single-fast		Multiple-slow \geq Single-fast

Intuition: $E_{2,n}(\rho) \leq \rho$



since have an earlier commitment

Erlang-C Properties

Assume an $M|M|n$ queue with arrival rate λ and service rate μ . Denote the offered load per server by $\rho = \frac{\lambda}{n\mu}$.

In steady-state,

Number in Queue:

$$L_q = \begin{cases} 0 & \text{w.p. } 1 - E_{2,n} \\ \text{Geom}_{\geq 0}(1 - \rho) & \text{w.p. } E_{2,n} \end{cases}$$

Waiting Time Distribution Given Waiting:

According to PASTA, a customer that encounter all servers busy and i customer in queue upon arrival, waits until $i+1$ customers get service ($= i+1$ i.i.d $\exp(n\mu)$ service times).

$$\Rightarrow W_q | W_q > 0 \stackrel{d}{=} \sum_{i=1}^{\text{Geom}(1-\rho)} S_i \stackrel{d}{=} \text{Exp}\left(\text{mean} = \frac{1}{n\mu} \cdot \frac{1}{1-\rho}\right)$$

Waiting Time Distribution:

$$W_q = \begin{cases} 0 & \text{w.p. } 1 - E_{2,n} \\ \text{Exp}\left(\text{mean} = \frac{1}{n\mu} \cdot \frac{1}{1-\rho}\right) & \text{w.p. } E_{2,n} \end{cases}$$

Average Waiting Time:

$$E[W_q] = E[W_q | W_q > 0] \cdot P\{W_q > 0\} = \frac{1}{n\mu} \cdot \frac{E_{2,n}}{1-\rho}$$