How to Shorten Waiting Lines

An Excerpt from *More Sex is Safer Sex*
by Steven E. Landsburg

You spend too much time waiting in lines. That's not some vague value judgment; it's a precise economic calculation. The people in front of you are wasting your time, and none of them cares. That's a recipe for a minor disaster.

Standing in front of you in line is just like dumping leaves on your lawn or ordering dessert when you're splitting the check. Because I don't feel all the costs, I'm sure to do too much of it. If I spend half a minute drinking while ten people wait behind me, I've imposed five minutes worth of costs on others. What are the odds my drink was really worth that much? Would I have stuck around for a drink if it had cost five minutes of my *own* time?

In principle, there's a market solution to this problem. If I'm in front of you, you can pay me to leave, or take up a collection among the people behind you and *then* pay me to leave. But you don't because the negotiations are a hassle, or because you're worried about "free riders" mooching off your investment, or because you don't want to look like some kind of econ geek. So you and I miss out on a mutually beneficial exchange. That's unfortunate.

Here's a different solution: Change the rules so each new arrival goes to the front of the line instead of the back. Then people near the back will give up and go home (well, actually they'd leave the line and try to re-enter as newcomers, but let's suppose for the moment that we can somehow prevent that.) On average, we'd spend less time waiting and we could all be happier.

If that sounds crazy, try an example. Imagine a water fountain in a city park with a steady gaggle of equally thirsty joggers running by. Each jogger looks at the line and decides whether it's worth joining. Because they're all equally thirsty, they all have the same cut-off for how long a line they'll join; let's say the cut-off is 12. As long as there are 12 people in line, joggers run right on by. Whenever the line length falls to 11, someone instantly joins and bumps it back up to 12.

That's disastrous. It means the line is always at the maximum length anyone will tolerate. The people in line can't be any happier than the people who look at the line and jog on—if they *were* happier, the line length would grow even longer. Since the water fountain doesn't make anyone any happier, it might as well not be there in the first place.

But what if we send newcomers to the *front* of the line? Then—because we've assumed a steady stream of new arrivals—the second guy in line never gets to drink; by the time it's his turn, someone else will cut in front of him. So as long as someone's drinking, you might as well jog right by. But if you're lucky enough to arrive just as someone else is finishing, you immediately take his place.

That's a great outcome, because nobody ever wastes time in line. You might think it has the offsetting disadvantage that a lot of people never get to drink, but that disadvantage is an illusion. Under the traditional system there are also a lot of people who never get to drink— namely the ones who never join the line because it's too long. Under *either* system the fountain is in constant use, so either system serves exactly the same number of drinkers. The only difference is the line length.

Now let's tweak the example to make it more realistic: Suppose the newcomers arrive not in a steady stream but sporadically and unpredictably.[1] Then, since newcomers go to the front of the line, it's always worth stopping for a drink. But if someone else comes along before you're finished, you'll get pushed back. If you get pushed back far enough, you'll leave.

That keeps the line short, which is good.[2] In fact, it's better than good: It's ideal. We'll always have *exactly the right line length* and here's why: Entering the line is a no-brainer. The only hard decision is whether to *leave* the line. And that decision is made by the guy at the back, who doesn't hurt anyone if he stays and doesn't help anyone if he goes.

---

[1] Note to the terminally geeky: To make this argument precise, you'll want to assume that both the arrival times and the amount of time it takes to drink are Poisson distributed.

[2] On the other hand, we wouldn't always want a line length of zero, because then the fountain might end up sitting idle.

In other words, the decision maker feels all the costs and benefits of his own actions! And that's exactly the prescription for a perfect outcome.

Now, there are a lot of assumptions here. I've assumed people have enough informationn to know when to bail out. That means they know both the current line length and the expected frequency of new arrivals. I've also assumed that everyone is equally thirsty; without that assumption we'd get bad outcomes when less-thirsty newcomers replace their thirstier counterparts.[3] And I've assumed there's a way to prevent people from leaving the end of the line and re- entering at the beginning—just as the traditional system assumes there's a way to stop people from cutting in.

Those assumptions can all be tolerably well approximated in the queues for telephone customer service. Here's how it would work: You call Microsoft for help installing Windows. An initial recording announces the average frequency of calls and explains that each new call will be placed in front of yours. Every minute or so, a new recording tells you how far back in the line you've been pushed. If you hang up and call back, Caller ID makes sure you can't get through. And for those with true emergencies (like those desperately thirsty customers at the water fountain) there can be a separate queue that you pay to join.

Sound crazy? Partly that's because you're probably not thinking about how much shorter the waiting time would be on average. It might just be crazy enough to work.

[3] So if some customers are thirstier than others, the go-to-the-front system falls short of ideal. But it's probably still better than the system we've got now.