When do priority queues hurt customers?

Do priority queues make customers better off? The answer is not immediately obvious. If a service provider moves from first-in, first-out (FIFO) service to some priority scheme, some customers wait more but if they pay less, the impact on the their overall welfare is ambiguous. A similar point holds for those in the high priority class; they wait less than under FIFO but pay higher prices. The net affect is again unclear.

Here we consider whether consumer surplus rises or falls when a service moves from FIFO to a priority scheme with two classes while leaving the total number of customers served unchanged. While we do not argue that maximizing consumer surplus is the most appropriate objective, we do believe our analysis provides useful insights into how consumer are affected as priority queues become more common in many marketplaces (Walker, 2012). We show, somewhat surprisingly, that in many settings consumers are worse off under priorities. That is, consumers are often better off under FIFO service than under any two-class priority even though the priority scheme more efficiently allocates delays across customers. Further, there are settings in which every customer pays a higher price under priorities than under FIFO, i.e., low priority customers pay a higher price for worse service.

We examine a service modeled as an M|M|1 queue. We suppose that service provider posts a price and expected wait (under FIFO) or a menu of prices and expected waits (under priorities) and that consumers choose whether enter and which class to buy (under priorities) without observing the current state of the system.

We consider two utility structures. Under both customers draw a type that is equivalent to their per-unit-time waiting cost. Types are drawn independently from a common continuous distribution. The distribution is commonly known, but customers are privately informed of their
realized types. The service provider must then post incentive compatible prices. The two utility structures differ in how customers value the service. Under the fixed-value model, all customers value the service at $V$ and only differ in their per-unit time waiting cost $t$. Under the increasing-value model, a customer with waiting cost $t$ has values the service at $A(t)$ for some positive function $A(t)$ such that $A'(t) > 1$. Both models have precedents in the literature. Ghanem (1975) and Gavirneni and Kulkarni (2016) use the fixed-value model while Aféche and H. Mendelson (2004) and Nazerzade and Randhawa (2017) use variations of the increasing-value model. The models result in qualitative different policies. Under a fixed-value structure, low types are more valuable to the system; they are always willing to pay more for a given level of service. The service provider must choose the highest type to admit. With increasing values, higher types are more valuable. While they find waiting more costly, their higher values mean they are willing to pay more. Now the service provider must decide what is the lowest type to admit. Finally, these models are useful base cases for analysis. The fixed-value set up is essentially the simplest way to model a market in which customers’ values for the service are independent of their waiting costs. The increasing-value model is a simple way of analyzing a market in which values are positively correlated with waiting costs.

We first show that service providers like priorities. Whether one considers revenue or social welfare, the system creates more value under any simple high-low priority scheme than under FIFO service with the same volume of customers. This is true under either consumer utility model. Results are not so sanguine for consumers. Under the fixed-value model and assuming a log-concave type distribution (a common regularity condition), consumer surplus when moving from FIFO to any two-class priority scheme falls. Under the increasing-value model, consumer surplus falls if the type distribution has a decreasing mean residual life.
function. Consumer surplus would only increase if the type distribution has an increasing mean residual life function. Note that an increasing [decreasing] failure rate implies a decreasing [increasing] mean residual life function. Thus there are non-trivial classes of distributions that satisfy both criteria under the increasing-value model.

The difference in these results is driven in part by the gap between the waiting cost of the marginal customer (which determines the incentive-compatible price) and the average waiting cost of served customers. When this changes significantly with the changes in the number of customers served (specifically, when its elasticity is greater than one), customer surplus falls when priorities are implemented. Similarly changes in the type distribution (as measured by an appropriate stochastic order) that shift this gap exacerbate the loss in moving to priorities. For the fixed-value model, greater variability implies a greater reduction in consumer surplus. In the increasing-value case, lower variability leads a greater loss.

While consumer surplus may fall, not all types are necessarily worse off under priorities. With increasing values, those with relative low or relatively high types are better off. It is the loss of those in the middle that cause overall surplus to fall. Under fixed values, it is possible that all customers are worse off. In particular, if the density of the type distribution is decreasing, the prices for both high and low priority service are higher than the price under FIFO service.

References