

Sensitivity of Optimal Capacity to Customer Impatience in an Unobservable M/M/S Queue (Why You Shouldn't Shout at the DMV)

Mor Armony¹

Erica Plambeck²

Sridhar Seshadri³

This paper employs sample path arguments to derive the following convexity properties and comparative statics for an M/M/S queue with impatient customers. If the rate at which customers balk and renege is an increasing, concave function of the number of customers in the system (head-count), then the head-count process and the expected rate of lost sales are decreasing and convex in the capacity (service rate or number of servers). This result applies when customers cannot observe the head-count, so that the balking probability is zero and the renege rate increases linearly with the head-count. Then, the optimal capacity increases with the customer arrival rate but is not monotonic in the renege rate per customer. When capacity is expensive or the renege rate is high, the optimal capacity decreases with any further increase in the renege rate. Therefore managers must understand customers' impatience to avoid building too much capacity, but customers have an incentive to conceal their impatience, to avoid a degradation in service quality. If the system manager can prevent customers renege during service (by requiring advance payment or training employees to establish rapport with customers), the system's convexity properties are qualitatively different but its comparative statics remain the same. Most importantly, the prevention of renege during service can substantially reduce the total expected cost of lost sales and capacity. It increases the optimal capacity (service rate or number of servers) when capacity is expensive, and reduces the optimal capacity when capacity is cheap.

¹Stern School of Business, New York University, marmony@stern.nyu.edu

²Graduate School of Business, Stanford University, elp@stanford.edu

³Stern School of Business, New York University, seshadr@stern.nyu.edu