The Impact of Mobile Order Ahead Apps on QSR Queuing

Thomas R. Robbins
East Carolina University

Order ahead mobile apps are becoming increasingly popular at Quick Serve Restaurants. These apps allow customers skip the line when arriving at the store. We investigate the hybrid queuing system created by these applications. We define an overall structure for the hybrid queuing system. Using discrete event simulation, we examine the impact these apps can have on system performance. Our analysis indicates that overall, the mobile application increases the efficiency of the system as a whole. By completing food preparation before the customer arrives the total waiting time for all customers is reduced. The walk-in customer also sees the benefits as overall system efficiency improves although they may feel slighted as customers arrive and skip the line.

Keywords: queuing systems, simulation analysis, retail operations

INTRODUCTION

The availability and use of order ahead mobile applications has become quite common in the quick serve restaurant business. Apps are available from restaurants including McDonald’s, Chick-fil-A, Chipotle, Starbucks, Moe’s, Noodles and Co., Subway and many others. The Wall Street Journal reports that food and drink apps were downloaded more than 155 million times in 2017; a 35% increase over 2016 (Stevens, 2018). My local Starbuck’s manager reports 7.5% of her business now comes from mobile orders.

While the functions provided vary from app to app, most allow users to order and pay for their food remotely. When arriving at the restaurant the user is able to skip the line and pick up their pre-made order, often at a special Mobile Order Pick-Up location. These apps create a hybrid queuing system that alters both the real and perceived behavior of the queuing system. In this paper we explore the implications these hybrid queues have from both the customer and service provider’s perspective.

The functionality of order ahead applications varies from restaurant to restaurant. Some have built in loyalty programs that allow customers to earn discounts and free items. Most apps have integrated mapping and GPS systems to allow customers to find the nearest restaurant location. Most apps also allow users to reorder previous orders and some allow users to set up favorite items for easy reordering, and nearly all the apps support some level of menu item customization. Virtually all the apps support a pay function that manages the financial transaction as part of the ordering transaction, though some also allow a pay at the store option.

The order processing and fulfillment process also has some variations from app to app, but most follow a basic format. The user picks their restaurant and selects their menu items. The customer can then select the scheduled time for their order to be ready. Most applications have a minimum lead time, but also allow the order to be scheduled further into the future. Minimum times are generally about 10-15 minutes, and most apps allow scheduling pick up an hour or more in the future. Once the order is placed the financial transaction is processed and an order ticket is printed at the store. Some apps, such as McDonalds and Chick-fil-A, do not finalize the order until the customer is at the store and clicks the "I'm here" button. GPS tracking disables the button until the customer is at the store location. This helps prevent accidental orders to the wrong store, but significantly reduces the order ahead time and significantly alters the queuing system behavior. When the customer arrives at the store, they generally skip the line and go to a special pick up area. If their food is ready, they pick it up immediately and the service is complete. If the food is not ready, because the customer arrived early, or the order was late, they wait in a separate queue for their order to be completed.

The remainder of this paper is organized as follows. In the next section we review the relevant literature. We thee describe the Order Ahead Hybrid Queuing System and our simulation approach. The next section presents an analysis of our simulation results for an overall scanning experiment, we then present a separate experiment that isolates the impact of the order ahead percentage. We conclude with a summary of our findings.

LITERATURE REVIEW

Queuing systems are widely studied in the operations literature. Queues may form whenever jobs arrive to a service location; when all servers are busy the jobs wait in queue until a server becomes available. The behavior of the queuing system is governed by several key parameters. The arrival process characterizes the statistical distribution of new jobs presenting themselves for service. The service process characterizes the statistical distribution of the service performed on the job. Performance varies significantly based on the systems utilization, the long-term proportion of time that each server is actively working versus waiting for a new job to service, and the number of servers present in the system. Under several widely accepted assumptions, such as exponentially distributed inter arrival and service time and a first come first served protocol, closed form solutions are available and widely known (Gross, 2008). The M/M/N model describes the long run statistical characteristics of a multi-server queue with exponential interarrival and service time. Standard performance metrics for basic queuing systems include average wait time, average total service time, average queue length, and system utilization.

As systems become more complex, and assumptions are relaxed, it becomes increasingly difficult to accurately characterize queuing system behavior. The introduction of balking and/or reneging, customers who arrive but decide not to enter the queue, or who leave the queue before being serviced, leads to the M/M/N/k + M queuing model (Gans, Koole, & Mandelbaum, 2003). Customers who balk or renege are said to be impatient, and their behavior can dramatically alter the behavior of the queuing system. This phenomenon has been widely studied in the call center space. As other assumptions are relaxed, for example the service or arrival process distributions, exact closed form solutions become less viable and approximation or discrete event simulation become more popular methods of analysis (Law, 2007).

Queues can be physical, virtual, or hybrid combinations of the two. In a physical queue customer stand in line. Traditional retail outlets are served by physical queues. A key characteristic of a physical queue is the visibility of the queue's length and the speed of service. Customers enter the line with full visibility of the number of customers ahead of them and have constant visibility of the queue's length, giving them insight into the speed of service. In a multi-server system, there may be separate queues for each server, but more commonly the queues are pooled into a single queue served by multiple servers. Pooling is more efficient technically but may be perceived as a longer wait due to the longer pooled queue and this may lead to increased balking. In some cases, the choice of queues is based on a customer class, for example priority lines at airport check-in and security, or at hotel check-in.

In a virtual queue there is no physical line for customers to stand in, instead some proxy for a customer is placed in a virtual line. The quintessential example of a virtual queue is the call center, but other examples include restaurant pager queues, ticket queues, or even print queues. Hundreds of papers have been published on call centers; (Gans et al., 2003) and (Aksin, Armony, & Mehrotra, 2007) provide excellent summaries and literature reviews. An important characteristic of the virtual queue is limited visibility, and therefore limited insight into how long the potential wait might be. In a general call center, the caller has no idea how many callers are already in queue, though in some systems a system announcement will give an estimated hold time.

A variation of the physical queue is the ticket queue. In a ticket queue the customer is given a number and they begin service when their number is called. By limiting the visibility of the queue, ticket queues have the potential to significantly alter both the customer's perception of the queue, and their behavior (Kuzu, 2015). Ticket queues have been used in environments from the DMV to the deli counter at the local grocery store. In a ticket queue the customer can compare their number to the now being served number and observe the rate at which customers are served to estimate a wait time. Pager queues are often used at full-service restaurants and an estimated wait time is often provided along with the pager, but without special effort on the part of the customer to monitor progress, no updates are available to the initial estimate. While physical queues often implement a first come first serve prioritization mechanism, virtual queues can easily implement alternative service disciplines.

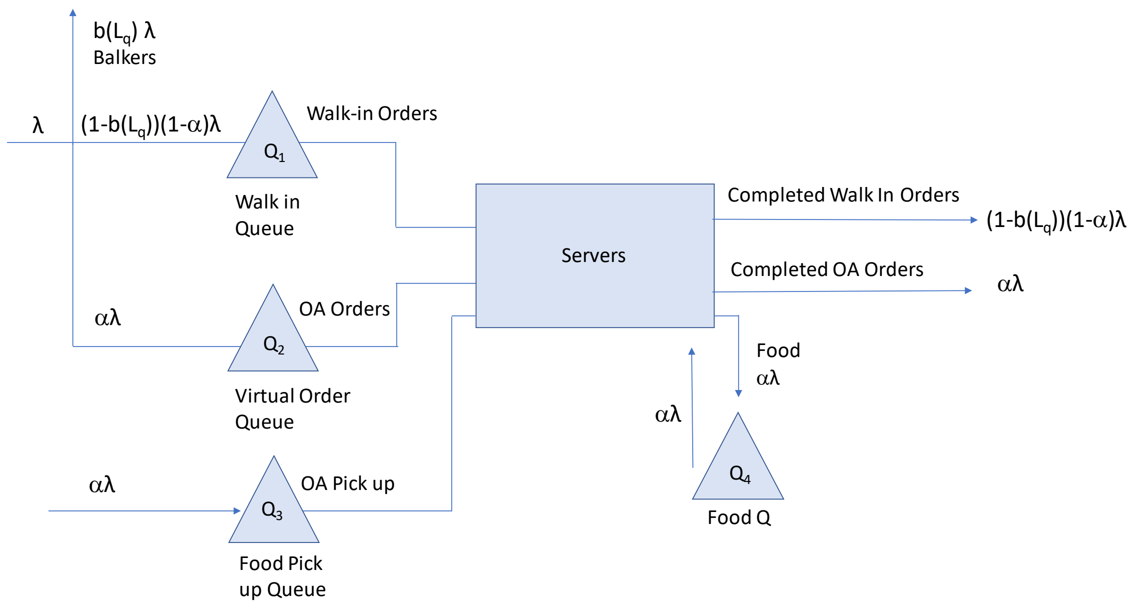
We define a hybrid queue as a linked queuing system that contains at least one physical and one virtual queue. A well-known example is the *Fastpass* system used on popular attractions at Disney theme parks. The hybrid queuing system includes a physical queue for walk up customers along with a virtual queue for Fastpass customers. The Disney Fastpass system is described in (Cope, F. Cope, & E. Davis, 2011). A hybrid queuing system for airport security checks is proposed and analyzed via simulation in (de Lange, Samoilovich, & van der Rhee, 2013). In both these systems a subset of customers are put in a virtual queue and assigned a reservation time. These customers then join the physical queue at a priority point at their reservation time. A basic version of a hybrid queueing system is used in full-service restaurants that accept reservations. Walk in customers are queued based on arrival time, while customers with reservations jump to the front of the line at their reservation time.

Order ahead apps create a somewhat different type of queuing system; an environment we refer to as the Order Ahead Hybrid Queue (OAHQ). The system contains multiple queues, physical and virtual. The system has the potential to impact the retail establishment in multiple ways. The order ahead capability will clearly impact the service experience of the order ahead customer, and potentially the walk in customer as well. (Larson, 1987) discusses the issue of social justice in queuing systems that result from what he terms slips and skips. A skip occurs when a customer arrives and is serviced before a current customer, the initial customer having been skipped over. The skipped customer feels victimized, having been treated unfairly. The slipping customer may feel a sense of satisfaction based on their reduced waiting time, but may also feel guilt for having violated a social norm. (Stevens, 2018) reports on customers who order ahead, but still wait in line to avoid the stigma associated with cutting the line. From a financial perspective the faster service time, along with the potential discounts earned, may lead the order ahead customer to increase their dining frequency. Multiple studies investigate the relationship between faster service and increased revenue (Akilimalissiga, Sukdeo, & Vermeulen, 2017; Allon, Federgruen, & Pierson, 2011; Lu, Musalem, Olivares, & Schilkrut, 2013; Perdikaki, Kesavan, & Swaminathan, 2011). Additionally, by removing some customers from the physical queue the order ahead capability may reduce the perceived waiting time for walk in customers and reduce the number of customers who balk.

THE ORDER AHEAD HYBRID QUEUE

The Order Ahead Hybrid Queue system contains a combination of virtual and physical queues. While there are multiple variations for configuring the order ahead process, we develop a model that can represent most configurations. A diagram of the system is outline in FIGURE 1

FIGURE 1
QUEUING SYSTEM

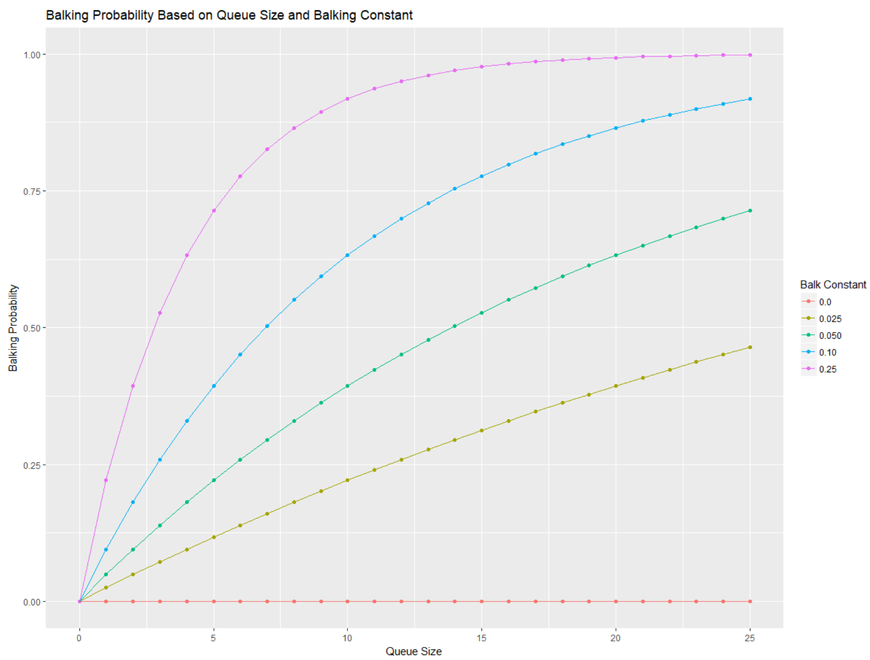


Orders arrive at a rate l. a percent of the order arrive as virtual order ahead orders, while 1-a percent of the arrivals are physical walk-in customers. Walk in customers observe the physical queue and a proportion of them determine the line is too long and they balk. The decision to balk is based on the length of the queue and the balking constant ($b(L_q)$). The balking decision is based on the number of customers currently in queue. In our model the probability the customer balks is given by

$$p_B = 1 - e^{-k_b n} \tag{1}$$

where n is the observed number of customers currently in the queue and k_B is the balking constant. If k_B is zero, then customers are infinitely patient, and none balk. As k_B increases the likelihood of balking increases. FIGURE 2 shows the probability a new customer balks based on queue size and a range of balking constants.

FIGURE 2
BALKING PROBABILITY GRAPH



The time required to process an order by an average server is a lognormally distributed random variable with mean μ^{-1} and standard deviation σ_{μ}. Walk in customers who decide not to balk enter the walk-in queue (Q1). OA Orders arrive at a rate of al. When the order arrives, it is placed in a virtual queue (Q2). Each order has an associated ready time, which we assume in our model is 20 minutes after the order is placed. Each order is assigned an early start and late start time. In our model we assign the early-start time as 4 times the average service time, and the late start time as 2 times the average service time.

The system includes 3 physical queues and 1 virtual queue.

1. Walk-In Queue: the line that walk-in customers enter to wait for service. Serviced on a first come first serve basis.
2. Order Ahead Virtual Queue: the queue of order ahead orders. Serviced based on scheduled time.
3. Pick Up Queue: a physical queue where order ahead customers wait to pick up their items.
4. Meal Queue: a physical queue where food is kept awaiting customer pickup. Serviced based on order number.

A pool of homogeneous servers service the entire queuing system. Idle servers accept work based on the following prioritization scheme.

1. OA Orders due in sooner than the late start time
2. Walk in Customers
3. OA Orders due sooner than the early start time.

In order to evaluate the performance of the queuing model, we conduct a series of designed experiments. In the first experiment we evaluate the system over a wide variety of input characteristics. Based on the assumptions for our system discussed previously, we define the following set of 6 experimental factors. High and low values are defined to cover a broad range of service scenarios.

TABLE 1
EXPERIMENTAL FACTORS

	Factor	Low	High
1	Proportion ordering ahead (a)	5%	40%
2	Offered System utilization (r)	25%	90%
3	Average service time	2 mins	8 mins
4	Balking Constant (*k*)	0	.25
5	Number of Servers (*n*)	1	5
6	Average OA Arrival Shift	-2 mins	+2 mins

The offered utilization represents the average system utilization that would occur if all arrivals were served, in other words without balking. The average OA Arrival Shift represents the mean of the distribution of the physical arrival of an order ahead customer, relative to the scheduled ready time. We assume that the physical arrival time is characterized by a normal distribution with a standard deviation of 2 minutes. So, with a zero valued shift, a customer would on average arrive at the agreed upon time, but half of the customers would arrive early. A negative shift implies a larger proportion arriving before the scheduled time.

Given the relatively large number of experimental factors, a well-designed experimental approach is required to efficiently evaluate the experimental region. A standard approach to designing computer simulation experiments is to employ either a full or fractional factorial design (Law, 2007). However, the factorial model only evaluates corner points of the experimental region and implicitly assumes that responses are linear in the design space. However, this system is likely to exhibit non-linear responses, for example queue length will have a decidedly non-linear relationship with utilization. Based on this we chose instead to implement a Space Filling Design based on Latin Hypercube Sampling as discussed in (Santner, Williams, & Notz, 2003). Given a set of d experimental factors and a desired sample of n points, the experimental region is divided into n^d cells. A sample of n cells is selected in such a way that the centers of these cells are uniformly spread when projected onto each of the d axes of the design space. We chose our design point as the center of each selected cell. This experimental design allows us to select an arbitrary number of points for any experiment.

Simulation Approach

Our model is evaluated using a straightforward discrete event simulation model at each design point. The purpose of the model is to predict the long term, steady state behavior of the queuing system. The model generates random numbers using a combined multiple recursive generator (CMRG) based on the Mrg32k3a generator described in (L'Ecuyer, 1999). Common random numbers are used across design points to reduce output variance. To reduce any start up bias we use a warmup period of 250 customers, after which all statistics are reset. The model is then run until approximately 50,000 customers have been serviced and summary statistics are collected. For each design point we repeat this process for 25 replications and report the average value across replications. Our initial analysis is based on an experiment with 1,000 design points.

The specific process for each replication is as follows. The input factors are chosen based on the experimental design. The average arrival rate is calculated based on the specified service time, number of servers, and offered utilization rate as

$$\lambda = \rho N \mu \tag{2}$$

That arrival rate is then used to generate Poisson arrivals for the replication. Each new customer generated includes an exponentially distributed interarrival time, and a lognormally distributed service time. When a customer arrives a Bernoulli random variable is used to determine if it is an OA or walk in customer. For walk-in customers the probability of balking is determined based on the current walk-in queue length and the balking constant according to equation (1). Another Bernoulli random variable used to evaluate the customer's balking decision. Our model assumes customers that join the walk-in queue remain patient and do no abandon. We also assume that since order ahead customers pay at order time they do not balk.

When the OA Customer arrives, the model determines if the food is ready. If it is ready the customer picks up the food and ends service. Our model assumes a pick-up area, so customers can pick up their prepared food without server involvement. If the customer arrives before their food is ready that are placed in queue 3 and wait until the food is finished. In this scenario services ends when the food is ready, and the total service time is equal to their wait time in Q3.

Over the course of the simulation we collect statistics on the following performance metrics:

1. **Average Queue Size**: the average size of the walk-in physical queue.
2. **Maximum Queue Size**: the maximum size of the walk-in physical queue.
3. **Realized Utilization**: the proportion of time servers are actively working on customer orders.
4. **Average Wait Time**: average time a customer spends waiting; from arrival time until their order is taken.
5. **Proportion Waiting**: the proportion of walk-in customers that must wait prior to their order being taken.
6. **Average Time in System**: the average time a customer spends in the restaurant, from arrival time until their order is filled.
7. **Proportion Balking**: the proportion of walk in line customers who balk.
8. **Number of Times Jumped**: the average number of times a walk-in customer is jumped over by an order ahead customer.
9. **Proportion Jumped:** the proportion of walk-in customers that are jumped by an order-ahead customer.
10. **Average # Jumping Over**: the average number of customers an order ahead customer jumps over.

After all design points have been evaluated, we re-run the experiments. All design factors are repeated with the exception of the OA percentage. In the second run the OA percentage is forced to zero. This allows us to determine the impact the OA customers had on the system. Common random numbers are utilized to minimize the variance and isolate the impact of the OA stream.

SIMULATION ANALYSIS

Baseline Metrics

Before examining the impact of the Order Ahead process, we review the baseline experimental region. Our initial design parameters give us a wide range of scenarios. Key performance metrics for the baseline case are shown in TABLE 2.

TABLE 2
BASELINE PERFORMANCE METRICS

	Min	Avg	Median	Max	Range	StdDev	Skewness
Average Queue Size	0.003	0.456	0.292	4.579	4.576	0.525	2.690
Max Queue Size	4.340	9.948	8.770	46.080	41.740	4.488	2.789
Realized Utilization	0.249	0.547	0.559	0.876	0.627	0.163	-0.130
Average Wait	0.302	99.04	50.57	1,265.41	1,265.11	142.19	3.522
Proportion Waiting	0.9%	31.5%	30.1%	80.3%	79.3%	18.8%	0.326
Average Time in System	125.2	399.0	376.7	1,667.0	1,541.8	195.7	1.850

Proportion Balking	0.0%	4.1%	2.4%	21.8%	21.8%	4.3%	1.267
Number of Times Jumped	0.000	0.000	0.000	0.000	0.000	0.000	NA
Proportion Jumped	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	NA
Avg Number Jumping Over	0.00	0.00	0.00	0.00	0.00	0.00	NA

The data in TABLE 2 shows that our experimental region covers a very wide range of operational scenarios. The key performance metrics cover the range of practical operating scenarios. The distribution of these metrics strongly positively skewed. They also demonstrate the significant volatility present in queuing systems. FIGURE 3 illustrates that while the median queue size is small, the distribution of avenge queue size is strongly right skewed, with average queue sizes ranging up into the range of 2. While an average queue size of 2 seems small recall that in queues the queue is volatile. FIGURE 4 shows the distribution of the maximum queue size; demonstrating that queues can balloon up to more than 20 even with a low average queue.

FIGURE 3
BASELINE AVERAGE QUEUE SIZE

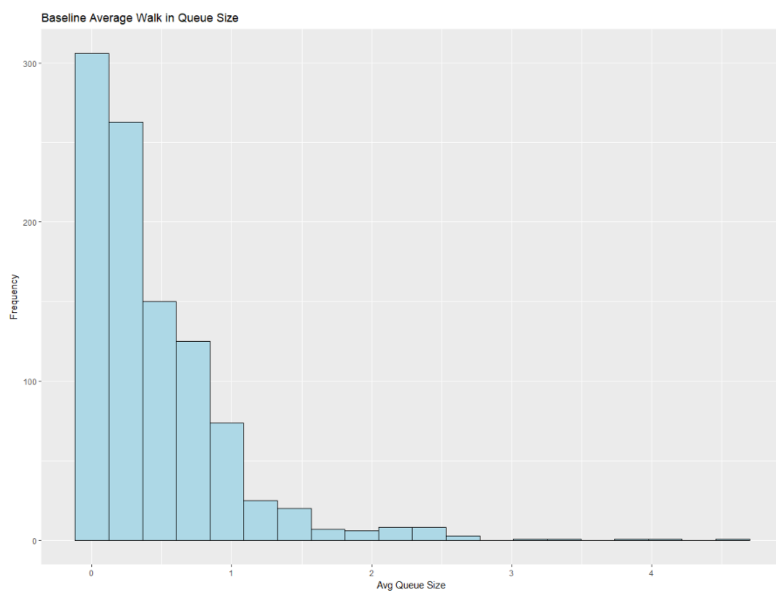
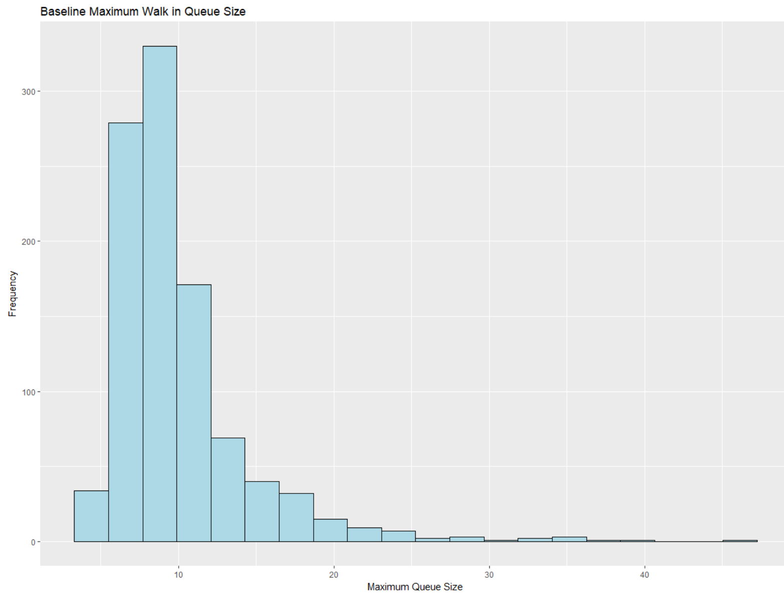


FIGURE 4
BASELINE MAXIMUM WALK IN QUEUE SIZE



Impact of Order Ahead

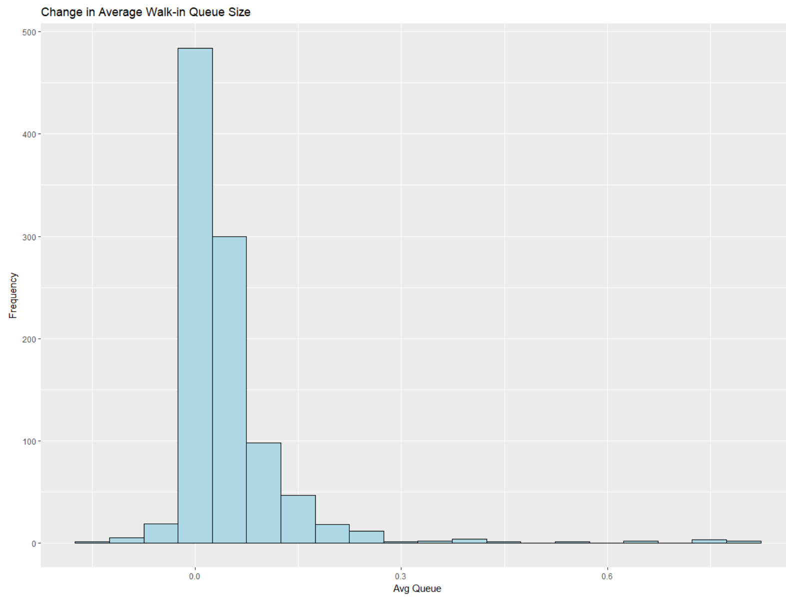
We now turn our attention to the impact Order Ahead customers have on the queuing system. We first examine the impact of OA across the experimental region, TABLE 3 summarizes the difference that occurs in each of core performance metrics as order ahead is introduced. Recall, that in our experimental design space the proportion of orders that are order ahead varies from 5% to 40%.

TABLE 3
CHANGE IN KEY METRICS WITH OA CAPABILITY

	Min	Avg	Median	Max	Range	StdDev	Skewness	pvalue
Average Queue Size	-0.130	0.048	0.024	0.817	0.947	0.085	4.761	2.59E-62
Max Queue Size	-2.320	0.239	0.200	4.320	6.640	0.618	1.173	3.14E-32
Realized Utilization	-0.039	0.000	-0.001	0.085	0.123	0.014	1.192	8.34E-01
Average Wait	0.294	113.78	55.51	1411.59	1411.30	163.01	3.266	3.03E-88
Proportion Waiting	-9.9%	-0.7%	-0.3%	8.0%	17.9%	2.0%	-0.777	5.27E-25
Average Time in System	6.9	66.1	58.6	444.8	437.9	41.1	2.127	6.30E-280
Proportion Balking	-2.5%	0.4%	0.2%	3.6%	6.1%	0.5%	1.545	1.89E-93
Number of times Jumped	-1.990	-0.174	-0.078	0.000	1.989	0.241	-2.888	0.00E+00
Proportion Jumped	-22.2%	-5.1%	-3.5%	0.0%	22.2%	4.6%	-1.143	0.00E+00
Avg Number Jumping Over	-4.65	-0.53	-0.34	0.00	4.65	0.58	-2.299	0.00E+00

FIGURE 5 shows a histogram of the change in average queue size when order ahead customers are introduced. Note that a positive number indicates a reduction in the observable (walk-in) queue length. Order Ahead reduces the queue length in 94.8% of our design points.

FIGURE 5
CHANGE IN AVERAGE QUEUE SIZE



In Figure 6 we examine the distribution of the change in the average wait time for walk in customers. This figure shows the data is highly skewed in the opposite direction. In 82.8% of the design space the average wait time increases. Taken together these two graphs show that walk-in customers face shorter but slower moving lines when Order Ahead customers are present.

FIGURE 6
CHANGE IN AVERAGE WAIT TIME

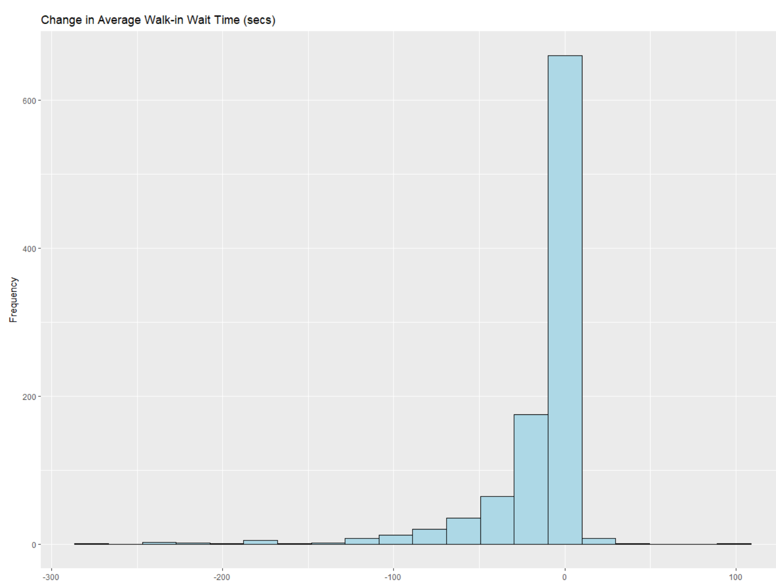
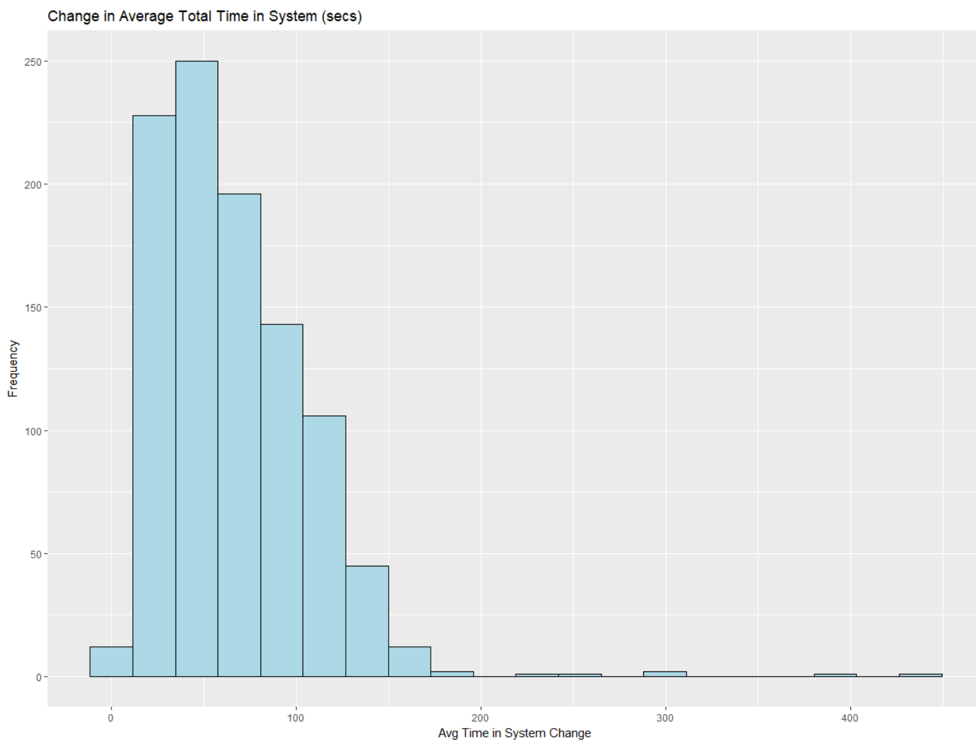


FIGURE 7 shows a histogram of the change in the Total Time in System for all customers. The overall time in system is reduced in all simulation scenarios.

FIGURE 7
CHANGE IN AVERAGE TIS



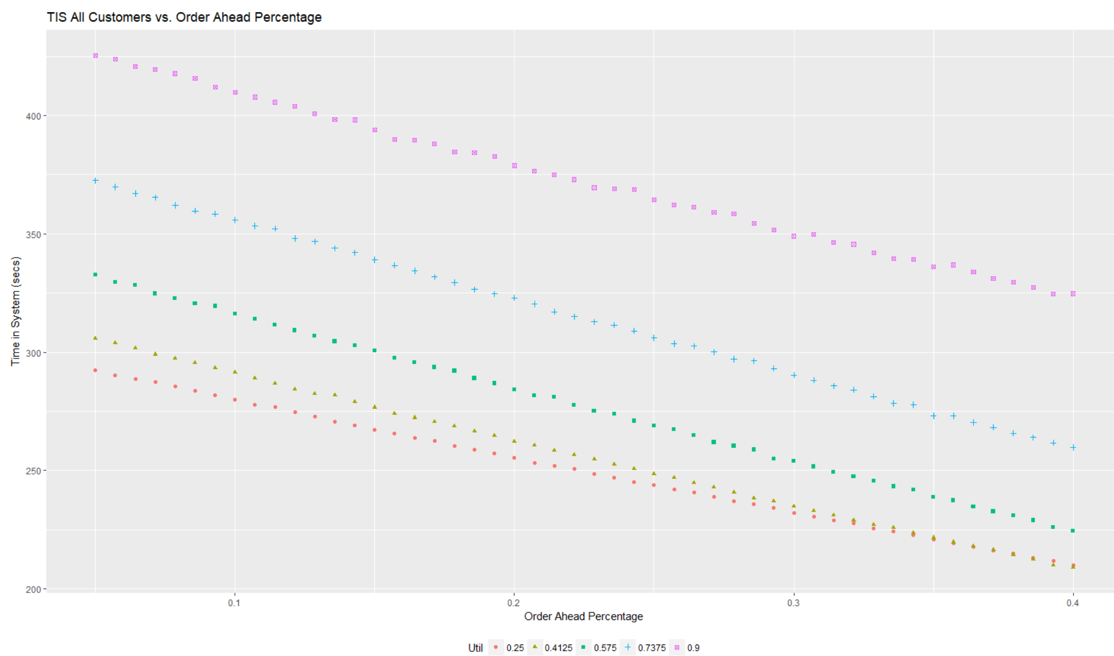
Experiment 2

We now conduct a second statistical experiment. The purpose of this experiment is to isolate the impact that the proportion of customers that order ahead has on overall operations. We use a simple experimental design with two factors, Proportion Ordering Ahead (a) and Offered System Utilization (r). We evaluate the model at 5 different utilization levels and 50 different values for the proportion ordering ahead, for a total of 250 design points. This provides us with the ability to analyze how performance metrics vary with order ahead percentage for a range of system utilizations. All the other experimental factors are set at their midpoint, with the discrete parameter Number of Servers (*n*) rounded up to 3.

We first examine the impact on the Time in System. FIGURE 8 shows the Time in System for all customers, while FIGURE 9 shows it for Walk In Customers and FIGURE shows it for Order Ahead customers.

TABLE 4 shows how the TIS metric changes as the order ahead percentage varies from it's low to high values. The impact of order ahead customers is relatively clear from these graphs. Overall Time in System is reduced as wait time is dramatically reduced for Order Ahead customers. The reduction is effectively linear with the proportion of order ahead customers. As demonstrated in experiment 1, the improvement for overall customers comes in part at the expense of walk in customers. Time in System for walk in customers increases as the proportion of order ahead customers increases.

FIGURE 8
TIME IN SYSTEM ALL CUSTOMERS



The increase is very slight, almost trivial for low to moderate levels of utilization. Only when utilization is high is there a meaningful increase. For lower levels of utilization, the order ahead service can be provided during idle time, even when the order ahead percentage is high. Finally, the time in system for order ahead customers, while very short, does increase with order ahead rate. As order ahead rates increase the average wait time increases roughly between 30-50%. Walk-in customer time is barely impacted under low to moderate system utilization levels. In the 2nd highest utilization level total time increases 7%, while in the highest utilization level the time in system increases about 20% as the proportion of order-ahead customers increases from 4% to 40%. The primary effect of the order-ahead capability is to reduce the overall time in system. As the order ahead percentage increase from a low (5%) level to a high (40%) level the overall time in system is reduced by about 30%. This is achieved by shifting food preparation service to the time before a customer arrives, when they are in-transit to the store.

FIGURE 9
TIME IN SYSTEM WALK IN CUSTOMERS

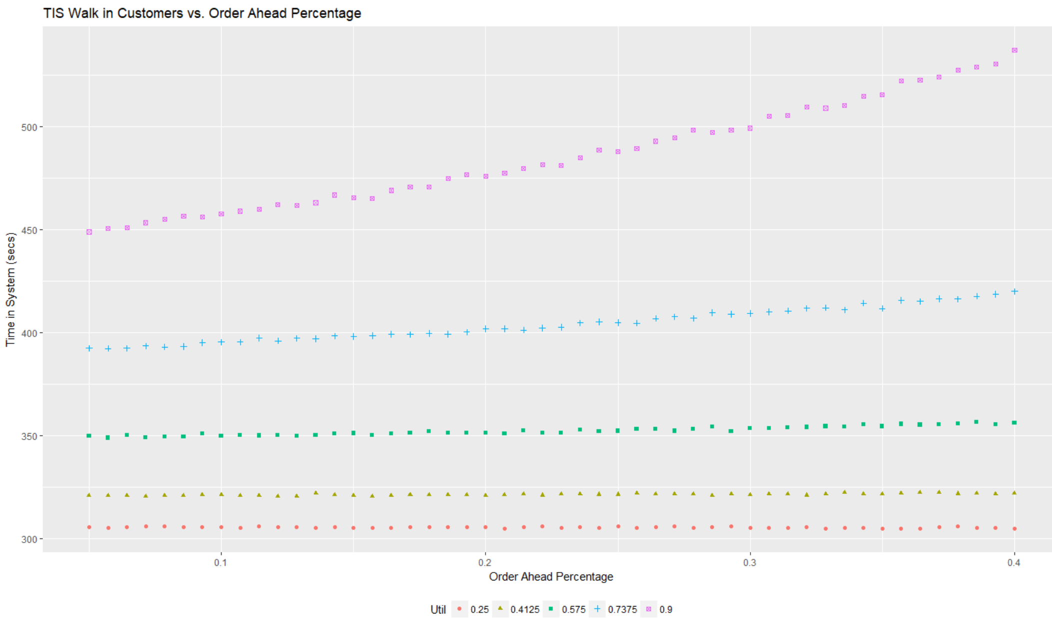


FIGURE 10
TIME IN SYSTEM ORDER AHEAD CUSTOMERS

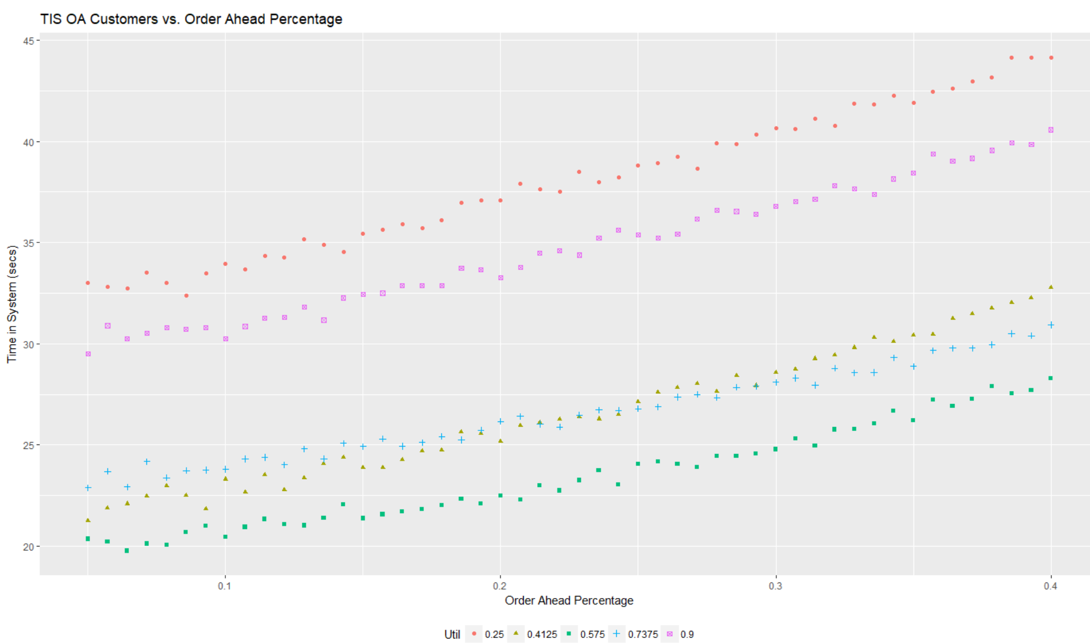


TABLE 4
TIS METRICS OVER RANGE OF ORDER AHEAD PROPORTION

	Time in System											
	All Customers				Order Ahead Customers				Walk In Customers			
Utilization	**Low**	**High**	**Chg**	**% Chg**	**Low**	**High**	**Chg**	**% Chg**	**Low**	**High**	**Chg**	**% Chg**
25.00%	292.4	210.0	-82.4	-28.2%	33.0	44.1	11.1	33.7%	305.3	304.9	-0.4	-0.1%
41.25%	305.9	208.9	-96.9	-31.7%	21.2	32.8	11.6	54.4%	320.8	322.0	1.2	0.4%
57.50%	332.7	224.3	-108.4	-32.6%	20.4	28.3	7.9	39.0%	349.6	355.9	6.3	1.8%
73.75%	372.5	259.8	-112.8	-30.3%	22.9	30.9	8.0	35.2%	392.3	419.9	27.6	7.0%
90.00%	425.4	324.8	-100.6	-23.7%	29.5	40.6	11.1	37.5%	448.8	537.0	88.3	19.7%

SUMMARY AND CONCLUSIONS

Mobile Order Ahead apps are becoming increasingly popular in the quick serve restaurant space. Customers who use these apps see significant benefits. Their orders are prepared prior to their arrival and their wait time is virtually eliminated. These customers go to the front of the line and often get their food almost immediately. Additional benefits often accrue to the customer through loyalty programs and discounts. Benefits also accrue to the restaurant as they capture customer data and potentially increase dining frequency.

Our analysis indicates that overall, the mobile application increases the efficiency of the system as a whole. By completing food preparation before the customer arrives the total waiting time for all customers is reduced. The walk-in customer also sees the benefits as overall system efficiency improves although they may feel slighted as customers arrive and skip the line.

This preliminary analysis examines the effect of order-ahead apps over a broad range of operating characteristics. Our analysis is primarily descriptive, analyzing the impact of a reasonable set of operating assumptions. From an operations perspective, further prescriptive analysis is possible to determine optimal operating policies for scheduling walk-in vs. order-ahead service. From a marketing perspective further analysis of the loyalty and purchase frequency benefits that come with order-ahead systems is warranted.

REFERENCES

Akilimalissiga, S., Sukdeo, N., & Vermeulen, A. (2017, December 10-13). *The delivery of service quality to increase customer repurchase behaviour and customer satisfaction at fast food outlets in central johannesburg, South Africa.* Paper presented at the 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM).

Aksin, Z., Armony, M., & Mehrotra, V. (2007). The Modern Call-Center: A Multi-Disciplinary Perspective on Operations Management Research. *Production and Operations Management*, *16*(6), 665-688.

Allon, G., Federgruen, A., & Pierson, M. (2011). How Much Is a Reduction of Your Customers' Wait Worth? An Empirical Study of the Fast-Food Drive-Thru Industry Based on Structural Estimation Methods. *Manufacturing & Service Operations Management*, *13*(4), 489-507. doi:10.1287/msom.1110.0343

Cope, R., F. Cope, R., & E. Davis, H. (2011). *Disney's Virtual Queues: A Strategic Opportunity To Co-Brand Services?* (Vol. 6).

de Lange, R., Samoilovich, I., & van der Rhee, B. (2013). Virtual queuing at airport security lanes. *European Journal of Operational Research*, *225*(1), 153-165. doi:https://doi.org/10.1016/j.ejor.2012.09.025

Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, *5*(2), 79-141.

Gross, D. (2008). *Fundamentals of queueing theory* (4th ed.). Hoboken, N.J.: Wiley.

Kuzu, K. (2015). Comparisons of Perceptions and Behavior in Ticket Queues and Physical Queues. *Service Science*, *7*(4), 294-314. doi:10.1287/serv.2015.0116

Larson, R.C. (1987). OR Forum—Perspectives on Queues: Social Justice and the Psychology of Queueing. *Operations Research*, *35*(6), 895-905. doi:10.1287/opre.35.6.895

Law, A.M. (2007). *Simulation modeling and analysis* (4th ed.). Boston: McGraw-Hill.

L'Ecuyer, P. (1999). Good Parameters and Implementations for Combined Multiple Recursive Random Number Generators. *Operations Research*, *47*(1), 159-164.

Lu, Y., Musalem, A., Olivares, M., & Schilkrut, A. (2013). Measuring the Effect of Queues on Customer Purchases. *Management Science*, *59*(8), 1743-1763. doi:10.1287/mnsc.1120.1686

Perdikaki, O., Kesavan, S., & Swaminathan, J. (2011). *Effect of Traffic on Sales and Conversion Rates of Retail Stores* (Vol. 14).

Santner, T.J., Williams, B.J., & Notz, W. (2003). *The design and analysis of computer experiments*. New York: Springer.
Stevens, L. (2018, May 2). Why Pre-Order Food Then Wait in Line: 'I Generally Do Not Want to Seem Like an Ass'. *Wall Street Journal*.