The Cost and Benefit of Skipping the Line

1 Introduction
Omnichannel retailing, an integration of brick-and-mortar stores with online channels, offers shoppers a holistic and rewarding shopping experience and has been thriving thanks to innovative IT technologies. The notion of omnichannel is also reforming the traditional service industry. For instance, Starbucks rolled out its Mobile Order & Pay application on mobile devices in September 2015. This feature allows customers to place their order in advance of their visit and pick it up later at a designated store. This application quickly becomes popular among customers for the convenience of line-skipping. It was reported that by January 2016, only 4 months after the app became available nationwide, over 10 percent of Starbucks transactions were processed via this app. The omnichannel service is now prevalent among quick service restaurants. Major chains such as McDonald’s, Dunkin’ Donuts, Subway, and Chipotle all have been establishing their omnichannel capability. Boutique coffee shops and restaurants are also able to provide the omnichannel service by using third-party online ordering platforms, e.g., ChowNow, Skip, and Ritual.

We study the omnichannel service from an operational perspective and investigate its impacts on individual customers and the system’s overall performance. We first characterize the customers’ equilibrium channel choice. We then investigate how the omnichannel service affects the system throughput, customer net utility, and the social welfare. We find that the omnichannel service increases the throughput in comparison with the conventional walk-in queueing system. The facility hence becomes more congested and customers incur longer delays before service completion. These longer delays erode both individual utility and social welfare. To address this backlash of the omnichannel service, we consider allocating total capacity to online and in-store orders dedicatedly. We show that the dedicated queues are more efficient in terms of both customers’ utility and social welfare when the system is busy.

2 Model Setup
We adopt a game-theoretical queueing framework to study the performances of an omnichannel service system. We model the facility as a single-server M/M/1. Customers arrive one at a time according to a Poisson process with a rate $\lambda$. The facility offers identical services that are exponentially distributed and are iid with a rate $\mu$ through two channels: conventional walk-in ordering and online ordering via mobile devices. When using the conventional walk-in channel, customers have to patronize the store to place their orders, but they can choose to join the line or balk after observing the queue length. In contrast, the online channel allows customers to place orders and wait without their physical presences in the queue.

---

2https://www.nbcnews.com/tech/tech-news/more-people-ordering-ahead-starbucks-n502061
Therefore, customers who choose to place orders online may incur a lower unit waiting cost when being compared with the walk-in channel.\footnote{We show that no customers would choose the online channel if the online unit waiting cost is larger than that of the conventional walk-in channel.} The online channel, however, is not able to provide customers with real-time queue length information and hence online customers have to form an expectation about the waiting time.\footnote{We know of no omnichannel service providers offering real-time delay information to online customers. For example, when a customer places orders via the Mobile Order & Pay app, Starbucks quotes an expected waiting time for the order. One obstacle to disclose real-time delay to online customers is the investment and installment of monitoring equipment which can automatically count waiting customers at local stores.}

Customers are risk-neutral and are served on a first-in-first-out basis. When service needs arise, customers strategically determine through which channel they would like to place their orders in order to maximize their individual net utility, which equals the difference between the service valuation and the total waiting cost. For simplicity, we assume that the facility has a fixed capacity and the system traffic intensity $\rho = \Lambda/\mu$.

\begin{center}
\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{channel_choice.png}
\caption{Equilibrium Channel Choice ($c_w = 2$)}
\end{figure}
\end{center}

3 Main Results and Implications

We first characterize customer channel choices in equilibrium. We show that customer online unit waiting cost, relative to the walk-in unit waiting cost, plays an important role in determining the system equilibrium status.

\begin{proposition}[Equilibrium Channel Choices] Consider an omnichannel system with a given traffic intensity $\rho$ and let $c_o$ be the online unit waiting cost. For any walk-in unit waiting cost $c_w$, there exist two thresholds $c_2 \leq c_1 \leq c_w$ such that

(i) if $c_o \in [c_1, c_w]$, all customers choose the walk-in channel in equilibrium;

(ii) if $c_o \in [c_2, c_1]$, customers use a mixed strategy in which they randomize their choices between walk-in and online channels in equilibrium;

(iii) if $c_o \in [0, c_2]$, all customers choose the online channel in equilibrium.

Moreover, the likelihood that a customer chooses the online channel decreases in $\rho$.
\end{proposition}

Proposition 1 indicates that the online channel is only a legitimate option when its unit waiting cost is small enough to offset its informational disadvantage to the queue-visibility offered by the walk-in channel. Such an informational disadvantage also causes the
attractiveness of the online channel to decline as the system turns busier. As illustrated in Figure 1, for a given $c_o$, customers choose the online channel all the time when the traffic intensity is small and use it less frequently as the traffic intensity grows. In an extreme case, customers may all choose the walk-in channel at a large traffic intensity.

We next use the conventional walk-in system, i.e., a visible queue, as a benchmark to discuss the performances of the omnichannel service. Note that when all customers choose the walk-in channel, the system has no difference from the conventional walk-in system even though the online channel is offered. Moreover, when all customers choose the online channel, the discussion resembles the comparisons of a visible queue and an invisible queue in Hassin and Haviv (2003). We, thus, focus on the case in which both channels are chosen in equilibrium.

**Proposition 2 (Impacts of OMNI-service)** *In comparison with a conventional walk-in service system, the omnix-channel improves the system throughput, but it hurts customer average individual utility and total social welfare.*

The introduction of the online channel exploits customer lower unit waiting cost in the queue. Customers who order online thus are more patient than walk-in customers and have more incentive to join the system. As a result, the omnix-service increases the system throughput (and the provider’s revenue). It, however, also results in a longer queue. In equilibrium, all customers wait longer than they do in the conventional walk-in service system. Hence, customer individual net utility declines and so does the social welfare.

Our findings are consistent with Starbucks’s observations after the brand implemented the mobile ordering app for a year. Starbucks CEO Howard Schultz said during an earnings call that there were a “growing number of stores being challenged to keep up with the increased volume demands” from mobile ordering.\(^5\) In order to address the issue of long lines and improve customer experience, we consider forming dedicated queues for different channels.

**Proposition 3 (Advantage of Dedicated Queues)** *Assume that the facility adopts dedicated queues for the walk-in and online channels respectively and has the same total capacity for allocation in comparison with a walk-in only system. Assuming the optimal capacity allocation to the dedicated queues, if system traffic intensity exceeds a threshold, this dedicated omnichannel system yields a higher customer individual utility and social welfare than the pooled one.*

Although pooling is considered to be an efficient strategy for general service systems, Proposition 3 indicates that dedicated systems may perform better in omnichannel service. By adjusting the capacity allocation between both channels, the provider is able to effectively manipulate customer channel choices and manage the congestion externality to the advantage of the overall efficiency.

\(^5\)\url{http://money.cnn.com/2017/01/27/investing/starbucks-long-lines-mobile-ordering-earnings}