Call center experience optimization: A case for a virtual predictive queue

William Keith Pugh
College of Business, Department of Information Systems and Cyber Security
The University of Texas at San Antonio
San Antonio, TX 78249, USA
emailbillpugh@yahoo.com

Abstract

The evolution of the call center into contact centers and the growth of their use in providing customer-facing service by many companies has brought considerable capabilities in maintaining customer relationships, but it also has brought challenges in providing quality service when call volumes are high. Limited in their ability to provide service at all times to all customers, companies are forced to balance the costs associated with hiring more customer service representatives and the quality of service provided by a fewer number. A primary challenge when there are not enough customer service representatives (CSR) to engage the volume of callers in a timely manner is the significant wait times that can be experienced. Normally, callers are handled in accordance with a first-come, first-served (FCFS) policy with exceptions being skill-based routing to those CSRs with specialized skills. A proposed call center infrastructure framework called a Virtual Predictive Queue (VPQ) can allow some customers to benefit from a shorter call queue wait time. This proposed system can be implemented within a call center's Automatic Call Distribution (ACD) device associated with computer telephony integration. A key factor in the proposed VPQ integration is that the servicer can decide who can enter the VPQ. Another important feature of the proposed VPQ infrastructure is that it does not violate the common FCFS policy. The advanced reservation feature of the VPQ can be accounted for when providing customers in the normal service queue with an expected wait time. Deciding how many advanced reservations that should be created within the VPQ can be based on predictive analytics, past performance, or can be invoked real-time when arrivals begin to increase past expected volumes.

Keywords
Call center, call queue, queuing theory

1. Introduction

Contact center growth is steadily increasing and has become a valuable business tool for building strong customer relationships [1]. A contact center is a natural evolution of what was once termed a call center, where businesses traditionally engaged customers during inbound call operations [2]. In addition to providing service over the phone channel, the modern contact center now integrates other forms of direct customer contact such as mail, e-mail, social media, and online chat capability in one location [3], [4]. However, contact centers are not without challenges. The costs associated with operating and maintaining contact centers are significant with the majority of the budget being spent on human resources and staffing [5], [6], [7].

Additional considerations are balancing staffing levels while providing quality customer service, particularly during high-volume traffic areas, where customers may find themselves waiting for extended periods to speak to a customer service representative (CSR). Customer frustration arises when wait times exceed a person's expectation for service [8]. Excessive wait times develop because of several factors, such as understaffed call centers, unanticipated call volumes, and excessive average handle time (AHT) during customer service [9]. Regardless of the reason, callers who wait in a queue can perceive their waits differently and this phenomenon has been studied from a social justice perspective by [10]. Social justice implications can have a profound effect on customer service experience when an injustice occurs.

[11] offers a definition of social justice that equates social justice as that of *justice as fairness*. [11] also proposes a thought experiment where individuals who are deciding on what type of society to create for themselves are first placed in an *original position* where no knowledge exists about individual social status, strengths, or limitations. Under this *veil of ignorance*. [11] asserts that all individuals would choose a society where rules would

be impartial in their application and resources distributed equally. Any failure in the proposed framework would create an injustice to one or more individuals. The impact of waiting in a queue, and the perceived injustice, can arise when someone who enters a queue last, but is served before those who are waiting.

To impart a level of fairness to all customers, the typical policy for servicing customers waiting in the queue is first-come, first-served (FCFS) and call center infrastructure follows this principle by implementing a first-in, first-out (FIFO) policy [9], [12]. Call centers must consider trade-offs between efficiency and agent costs when making staffing considerations. More agents usually will increase service quality, but costs to the business can accumulate when some agents are idle or even underutilized [3], [13].

The study of waiting lines, or queues, is called *queuing theory*, and is considered a branch of operations research. The origins of queuing theory begin when A.K. Erlang proposed waiting models for the Copenhagen telephone exchange at the beginning of the twentieth century [6]. Erlang's influential work resulted in several publications that describe call arrivals as a *Poisson* process. A Poisson process is a probability distribution that describes the given number of events occurring over a fixed interval of time, provided the average arrival rate is known. In addition, a Poisson process is considered *memoryless* in that each arrival is independent of the last arrival [14], [4]. Erlang models are written in *Kendall* notation in the form A/S/c where A denotes time between arrivals in a queue, S is the service time distribution and c is the number of servers. The classical queuing model is the M/M/1, or Erlang-C model. This form denotes both arrival and service times are Markovian (M) with a single server.

Call center telephony infrastructure has advanced over the years and now incorporates many technology changes to facilitate customer contact and service resolution. For instance, Interactive Voice Response (IVR) allows customers to complete certain transactions by pressing numbers on the keypad and sending tones. Customers may also speak into the phone to accomplish certain tasks. This capability reduces the number of customers waiting in a queue to speak to a CSR. Customers who desire to speak to a CSR are routed through a device known as an Automatic Call Distributor (ACD). The ACD then routes the calls based on predetermined criteria. For instance, calls may be routed by desired service type or to CSRs with specific skill sets [6], [4], [15]. The ACD maintains a record of each CSR's skill set and routes calls to the appropriate agent when they are logged on and idle. Agent utilization is normally measured over short intervals (half-hours) during each day and is calculated as the average number of CSRs that were active during that period [6], [16], [17]. CSR utilization is one key metric in call center operations, and combined with the AHT information, call center managers can anticipate future staffing levels and gauge call center performance [6], [18]. When agents become saturated with calls, customers are placed on hold and the queue can build exponentially if the average handle time is greater than the call arrival rate. Although customer arrivals can be averaged over a specific period, they follow a Poisson probability distribution whose function is indicated by:

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ for } x = 0, 1, 2 \ldots$$

where x is the number of arrivals and λ is the *mean* number of arrivals in the time period with $e = 2.71828$. The Poisson cumulative distribution function (CDF) is used to calculate the probability of the number of arrivals being less than or equal to x and is indicated by:

$$F(x;\lambda) = \sum_{i=0}^{x} \frac{e^{-\lambda}\lambda^i}{i!}.$$

The service time represented by the M/M/1 queue is exponentially distributed and the probability density function (PDF) is indicated by:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

Integrating the above equation gives the cumulative distribution function shown here:

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

This gives the cumulative probability of having a service duration of x given arrival rate λ.

Call center queues typically have three major designs for queuing incoming calls. The first is shown in Figure 1. In this model, a customer has no idea how long the wait time will be, but chooses to remain in the queue for service and W_s is when waiting in the queue ends and waiting for service begins. The second model, shown in Figure 2, provides the customer with a recorded message on how long the expected wait will be. Realistically, the

announced expected wait time $E(W_q)$ for service will be close to the actual wait time for the customer and when service begins. In this situation, the customer chooses whether to remain in the queue or abandon the call. The third model, shown in Figure 3, is a virtual queue, where the customer is given the option to wait with an expected wait time, or to be called back when a CSR is free.

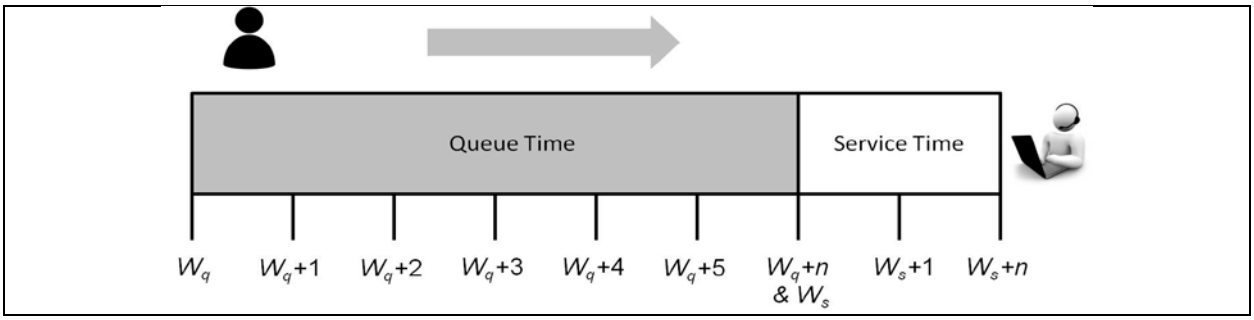


Figure 1. Normal call queue with unknown wait times.

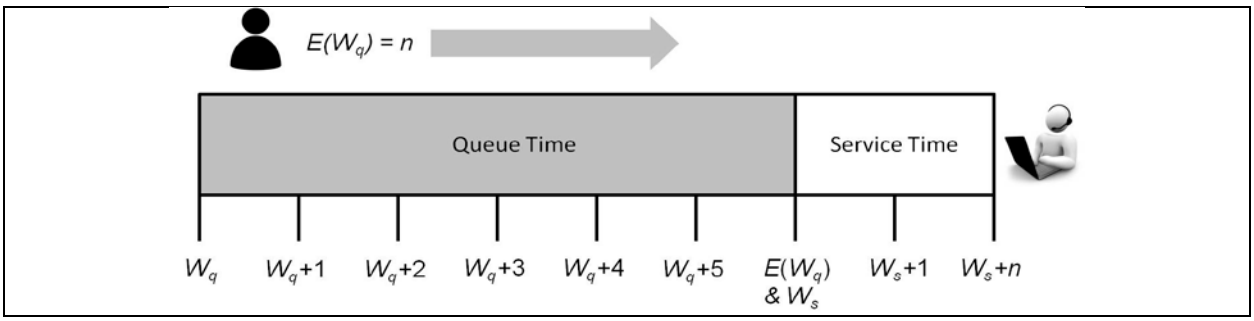


Figure 2. Normal call queue with announced wait times.

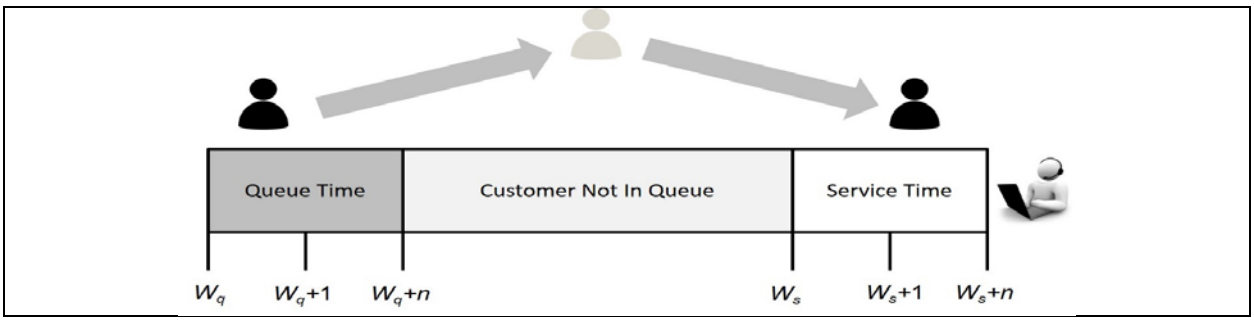


Figure 3. Call queue with callback capability.

1.1 Statement of the Problem

Customers subjected to long wait times can become frustrated and their perception of the service quality may be negatively affected by that experience [19]. When customers' expectations are not met, the psychological impact can have profound effects on their perception and feedback on the quality of service [10], [8]. In *The Psychology of Waiting Lines* [8] presents a hypothetical satisfaction formula in the form $S = P - E$, where S is satisfaction, P is perception and E is expectation. Realistically, if a person's perception were greater than their expectation, a positive satisfaction score would result. Conversely, if the expectation were greater than the perception, a negative score would result. Human impatience is a natural behavior exhibited by individuals when forced to wait in a queue for service [9]. Studies on impatience have produced interesting results on the phenomenon of waiting. [9] also explore impatience and impart a quantitative understanding through the application of an impatience function of how individuals respond to waiting for service. A common adage states: "When everyone is special, nobody is special" that intuitively suggests not everyone can be afforded priority service in all situations. Increasing the number of CSRs that are available would go far in reducing wait times. However, this would also

increase costs for a business even for a profitable endeavor. Another option is to block all callers who must wait for service, since and they would effectively get a busy signal when calling [6]. This operation would undoubtedly receive poor ratings from customers and lead to a negative experience as well. If businesses could alleviate the wait times for a select group of customers, faster service would mean higher perceived service quality and a better customer experience for those customers. With these limitations in mind, a *Virtual Predictive Queue* (VPQ) can be utilized to service a select group from a population of customers. A VPQ can be utilized in several ways to bring value to individuals, organizations, and businesses. The following applications and models illustrate potential uses of the VPQ. The Person to Business (P2B) Application, Internal to Business (INT2B) Model in depicted in Figure 4.

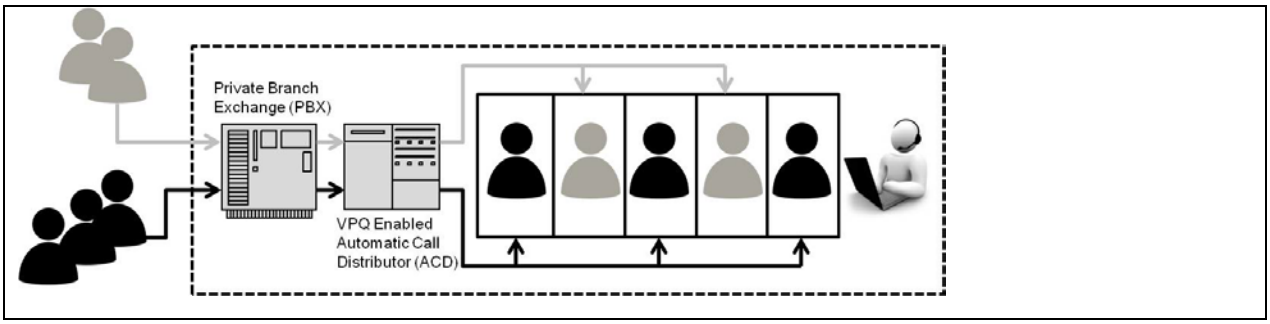


Figure 4. P2B application, INT2B model for VPQ.

In this configuration, the VPQ is enabled within the ACD inside an organization's infrastructure. This is meant to allow a deliberate attempt to provide faster service times to a select population based on server-defined requirements. The customers to utilize the VPQ can be determined within the telephone infrastructure by resolving phone number or other information. This could also be achieved through a voice over internet protocol (VOIP) gateway for those calls originating from a computer or an application on a mobile device. Making VOIP calls through the mobile device, geographic location could be resolved and customers in certain areas may be allowed to utilize the VPQ for faster service in cases such as natural disasters. This configuration is compatible with all queue models and does not violate a traditional FIFO policy in that the advanced reservation made by the VPQ is considered when announcing estimated wait times to customers. The Person to Business (P2B) Application, External to Business (EXT2B) Model is depicted in Figure 5. In this configuration, the VPQ could become enabled as a service provided to customers who are willing to pay a fee for use. The service could be used to target businesses and organizations with extremely long call wait times. In this application, the VPQ must be able to navigate the IVR system with the appropriate words or tones to allow direct entry into the queue.

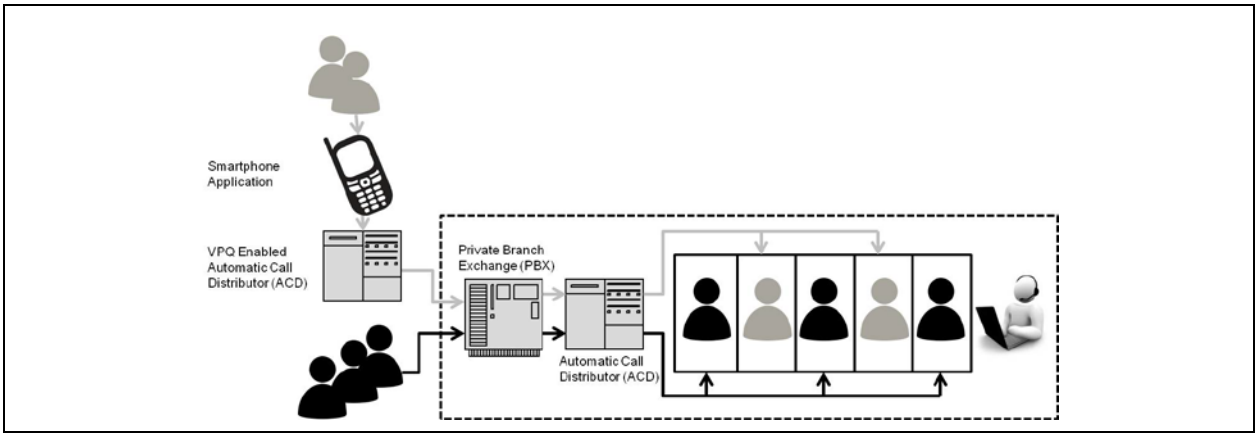


Figure 5. P2B application, EXT2B model for VPQ.

The Business-to-Business (B2B) Application, Internal to Business (INT2B) Model is shown in Figure 6. In this model, the VPQ could be used by businesses to target other organizations or businesses that have long call wait times. This could be used by back office personnel conducting regular transactional workflow where calls to external agencies are normally required. The VPQ could eliminate a significant portion of the wait time for employees making calls to other organizations and businesses while increasing productivity with financial benefits.

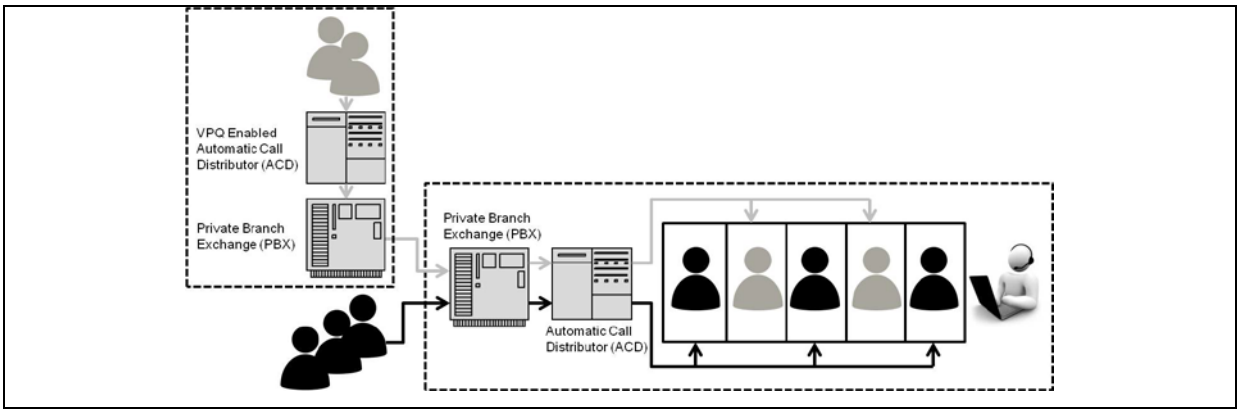


Figure 6. B2B application, INT2B model for VPQ.

1.2 Contribution to the Field of Business

The introduction of the VPQ and its effectiveness may offer resolution to business challenges in providing faster service to a select population. Businesses will have the opportunity to determine which of its valued customers may be served faster in an attempt to maintain and strengthen that relationship. The VPQ can be deployed in various configurations to benefit both customers and businesses. A VPQ service could also provide standalone value for those willing to purchase a subscription, or pay a one-time fee for service.

2. Literature Review

2.1 Definition of a Queue

A queue, or waiting line, describes a phenomenon where people are delayed in a line with others waiting for service. [19] defines waiting for service as the time a customer is ready for service until the time service is provided. Some examples of queues include waiting for a traffic light to turn green, waiting in a drive-through at a fast food restaurant, or waiting in line at a bank to deposit a check. Queues develop when the arrival rate exceeds the service rate and every arrival must be served [9]. The study of waiting lines is called *Queuing Theory* and is considered a branch of operations research and management science disciplines [20]. The quantitative portion of the study of queuing theory borrows from many mathematical areas which include discrete math, probability theory, statistics, linear programming, and matrix operations [21].

2.2 Queues as a Social Phenomenon

[22] writes a satirical dystopian reality of the Soviet citizen's life standing in never-ending service lines in *The Queue*. Written in dialogue style, the novel advances the story plot through conversation, a cacophony of random voices and whole-paragraph caesura with the intent to embed the reader in the story. In the introduction of the translated version, the Soviet propensity for queuing is illustrated by stating "the basic principle of Soviet queuing is that you join the line first and *then* ask what it's for [original emphasis]." [22] asserts that for the Soviet citizen, queuing is a natural act of a conditioned society dependent on a state economy. The willingness to queue indicates an obedient subject and residents maintain a patriotic respect for being allowed to wait in line with fellow citizens. *The Queue* offers a glimpse of a lifestyle in which many individuals might not be able to relate. However, it does represent a unique aspect of a society burdened by limited resources and is complete with conflict, expectations, happiness, and love.

Additional evidence of queuing as a social phenomenon and human behavior within queues can be found in [23] *Why Does the Other Line Always Move Faster?* [23] delivers a humorous assessment of queues and how they are ubiquitous in everyday life. [22] alludes to *The Queue*, by stating the act of queuing by the populace indicates deference to authority and a compliant citizen. The orderly arrangement of the line suggests that at some level, queue integrity is maintained through a collective effort that demands an orderly flow through a process, irrespective of the reason or desire of the individual.

2.3 Applications within Call Centers

Call centers are an increasingly important element in today's business operations and are normally considered large service entities where customer service agents provide a variety of services over the telephone [9], [17], [7]. Even with the ubiquity of the internet and electronic global communication ability, the telephone is expected to remain a primary contact channel and be the preferred method for customer service requirements [1], [9], [12], [7]. The natural successor to the call center is the contact center, where all forms of customer contact are handled [9], [24]. In addition to providing service over the phone, a contact center incorporates other multi-media customer contact channels such as e-mail, fax, internet chat and IVR units. IVRs allow customer self-service for many transactions via telephone keypad entries or voice responses [9]. Virtualization now enables seamless operations between contact centers in different geographic locations [9]. Customer initiated calls are considered *inbound* calls and are predominant during normal business operations. Call centers may also conduct outbound calls to customers for sales or marketing purposes [6], [3], [2], [25], [17].

2.4 Economic Analysis of Waiting Lines

Determining economic costs of waiting from a business perspective is easy to assess given that certain assumptions, or desired service objectives, are provided [20]. A total cost *TC* model can be constructed using the following equation

$$TC = c_w L + c_s k$$

where c_w is the waiting cost per customer per unit time and c_s is the service cost for each server. The term L is the average number of customers in the system and k is the number of servers. The assumptions for waiting cost are not direct costs to the servicer, but are related to the fact that a customer who decides not to wait, or abandon a queue, when wait times are lengthy equates to potential loss of revenue and therefore a cost to business [20].

2.5 Queue Disciplines

The order in which customers are served is regarded as the queue *discipline* and other schemes, in addition to FIFO, are last in, first out (LIFO), service in random order (SIRO) and priority service based on certain customer characteristics [26]. Although FIFO is generally recognized as the predominant queue discipline, other systems have been investigated and provided interesting conclusions about their performance. [27] maintain that when organizations do not take into account customer abandonments and focus only on reducing the average wait time of those in the queue, a LIFO policy would optimize their service metric, but would also produce a negative experience for customers who joined the queue at an earlier time. [10] uses a case study from the Boston Police Department from the 1960s to illustrate the point that calls answered in random order have the same mean time spent waiting in the queue. Priority queues have been implemented in many service industries and allow a select population to receive service faster than those waiting in the regular queue, but they can also cause perceptions of social injustice in certain settings [28], [23], [29].

2.6 Consumer Psychology in Queues

The psychology of waiting in line has received considerable research in formulating consumer perceptions of service quality and experience [10], [17]. The majority of queue studies presumes FIFO is the socially just queue discipline and violations of that principle can lead to conflict [10], [8]. Individuals standing in a service line exhibit the fundamental geometry of a queue through its linear congruence where one side is for entry and the other leads to desired service.

[10], [30] explored social justice and the psychology of queuing and discovered that other factors are involved in a person's perceived value of waiting in line. [10] relates a personal experience of receiving a bicycle that was purchased at a department store pickup window. After a lengthy waiting period, [10] notices others who have arrived later received their purchases first. [30] also asserts that adherence to an FCFS policy is the universal discipline for creating social justice within a queue by stating "Queueing theorists and social scientists have long

believed that first-come, first-served (FCFS) is the socially just *queue* discipline and first-in, first-out (FIFO) the socially just *system* discipline [original emphasis]"

2.7 Customer Expectations

Research on customer service expectations has focused on understanding the nature and determinants of those expectations within a general service setting [31]. [31] proposed a model specifying three different types of service expectations, which include desired service, adequate service, and predicted service. [31] exploratory research provided a theoretical framework which included seventeen propositions and included antecedents for each of the theorized expectations. [32] investigated patience adaptation as a function of their service expectation of anticipated wait time. [32] proposed an *M/M/m* model that considers adaptive customer behavior influenced by changes in anticipated wait time. The findings suggest the model to be applicable for a steady-state equilibrium system for determining patience based on average wait time. [19] focused on understanding how delays in service affected overall customer service evaluations. Three types of waiting are defined in this study and generally, *waiting for service* is considered the time from when a customer is ready to receive service until the service begins. The three types of waits include a pre-process wait, an in-process wait, and a post-process wait. A pre-process wait would be considered waiting for a table at a restaurant. An in-process wait would be the wait between meal ordering and service and a post-process wait would be considered waiting for and paying the bill after the meal was complete.

2.8 Impatient Callers and Abandonment

Callers who initially enter a queue, but subsequently leave, are said to abandon the queue or *renege* [9]. Customers who do not enter the queue at all are said to *balk* when their impatience is greater than the perceived value of the service [6]. [33] investigated a model for rational queue abandonments with the assumption that callers have a given patience distribution for waiting in a queue. [33] findings suggested there is a rational framework for determining the queue abandonment period given that customers have no information on their position in the queue. Several queue studies have focused on abandonment as a result of customer patience [10] with consideration for time spent waiting, or what some researchers consider the "sunk cost effect" [34]. [34] researched the effect of the number of people behind in a queue on consumer psychology and the likelihood of reneging. [34] suggest that in addition to the number of people ahead in a queue, the number of individuals waiting in line behind is a key factor in the decision to renege. [34] theorize that consumers make social comparisons of their position in a queue in a downward direction. Specifically, [34] suggest individuals assess their position in a queue and seeing more people behind them leads to a more positive self-assessed affective state of well-being.

2.9 Queues and Social Justice

Queues in general are naturally assumed to follow a FIFO discipline, especially when a system maintains the generalized sociological phenomenon of an FCFS policy [28], [29], [30]. Queues are also considered in the context of the law and legal interpretations. [35] legal research report published in the *Iowa Law Review* focuses in depth on the FIFO principle with regard to case law, legal interpretation, and use as an "extra-legal norm." Within the context of the law, [35] consider FIFO an allocation method, particularly as it pertains to property law and the concept of a lien position on a collateral security. [35] main purpose was to provide a theoretical framework for evaluating FIFO rules, exceptions to policy, and practical applications. [35] also demonstrate and evaluate FIFO's role in law through investigation of its pervasiveness in case law through implementations that are typically lacking strict enforcement criteria or codified meaning. The final intent was to demonstrate that FIFO is not a compact solution for every legal application. [35] propose that FIFO applicability should be assessed subjectively for each situation and use in one context does not necessarily mean a categorical legal imperative.

A discussion about queues and fairness would not be complete without a detailed discourse of [11] seminal work titled *A Theory of Justice*. [11] interpretation of justice is from a perspective that describes social justice as being *justice as fairness* and is closely aligned with social contract. [11] asserts, as described in a thought experiment called the "original position," that a rational group of individuals who were able to decide what form of society they would create, while also being unaware of individual characteristics, social standing, or pedigree, would choose a society where rules would be impartial, and resources would be equally distributed. [11] asserts individuals in the original position are also under the influence of what he describes is the "veil of ignorance." This condition assures individuals do not know their particular social standing, pedigree, or level of education when deciding as a

community on the type of society they wish to form. Likewise, individuals in a call queue exhibit a level of blindness as to who, or how many others occupy the same queue.

2.10 Call Center Operations

Traditional call center operations are primarily focused on maintaining a balance between service quality and operational efficiency [36], [37], [24], [38]. Service quality is critical to building and maintaining strong customer satisfaction and long-term relationships. Operational efficiency attempts to minimize costs for a particular level of resource utilization [6], [32]. Most calls are typically inbound, but outbound calls are conducted for customer follow-up after service, marketing, or collections activities [2], [25], [17]. In many call centers, some agents are trained to handle calls requiring different skill sets and some call centers have specialized agents to conduct outbound calls as well [3].

2.11 Queue Simulations

Simulations have been widely used in analyzing and understanding complex systems within the realm of operations research [21]. Call queues are one area where much investigation has been focused on using simulation where some argue Erlang formulas and queuing theory fail to provide a complete understanding of queue dynamics [5], [3], [20], [18], [25]. The predominate approach to queue simulation is through a Discrete Event Simulation (DES) approach [5], [39], [40], [41]. [40] describe a unique approach to DES using spreadsheets and Visual Basic for Applications (VBA) code. [40] demonstration offers a unique method for using inherent spreadsheet functionality for data analysis and software code to manage the actual simulation. Simulations are also considered valid experimental designs in that they are specifically created to represent the natural environment where processes and activities normally occur [42].

3. Research Method

3.1 Hypotheses

Since the proposed VPQ is designed to reduce the wait time for a select population, as determined by the servicer, extended wait times and their impact on an individual's patience threshold as an antecedent to service queue abandonment was investigated. In addition, service queue abandonment, due to perceived social injustice, was also investigated [43], [29], [10], [35] as an antecedent to queue abandonment. Although a VPQ implementation would be resigned to only call queues, the first two hypotheses were researched for both apparent and phantom service queues (line-standing and call queues). Therefore, the following hypotheses were proposed to investigate customer intentions on abandoning a service queue:

H_1: Patience Threshold (PT) is negatively related to Intent to Abandon (IA) a service queue.
H_2: Perceived Justice (PJ) is negatively related to Intent to Abandon (IA) a service queue.

The following hypothesis was proposed for the simulation experiment to investigate VPQ efficacy:

The call wait times for the VPQ will be significantly less than the normal service queue.
H_3: $W_{VPQ} < W_{SQ}$
H_{3a}: $W_{VPQ} \geq W_{SQ}$

3.2 Setting and Participants

The quantitative survey portion of this research was accomplished using Amazon's MTurk platform. Respondents chose whether to complete the survey and were compensated for their effort. The computer simulation portion of the VPQ was conducted on a personal computer using a standard spreadsheet software application. Ten simulation runs were accomplished and assessed for VPQ effectiveness. Simulation results were captured individually for statistical analysis of normal service queue and VPQ wait times. Participants for the exploratory survey instrument were volunteer Human Intelligence Task (HIT) workers within the United States on the Amazon Mechanical Turk platform. Participants were compensated for their time in completing the survey and appropriate

controls were placed on the quality and reliability of MTurk respondents with only those workers with a HIT reliability rating of 95% or higher being allowed to participate.

3.3 Data Collection

Data for studying the relationship between patience thresholds, perceived justice, and intent to abandon a service queue were collected via an online survey instrument hosted on a cloud-based Google Drive platform utilizing Google Forms for the actual instrument. Survey data were collected automatically, and the end results were exported as a common spreadsheet file extension for analysis. The data were saved to a secure cloud storage location for further examination. Computer simulation data were collected after each of the 10 simulation runs and saved to a spreadsheet application on a cloud platform for further analysis of the different wait times.

3.4 Data Analysis

A total of 299 responses were gathered from the survey deployment on the MTurk platform. The data were then analyzed for outliers and incomplete responses. A total of 17 responses addressing perceptions of call queues contained missing data and 18 responses on perceptions of waiting in a line contained missing data. These responses were considered ignorable data and were determined to be missing completely at random (MACR) and therefore deleted listwise [44]. The final dataset was composed of 281 valid responses on perceptions of call queues and 280 valid responses on perceptions for waiting in a line. The data were then analyzed for descriptive statistics and reverse coded items were recoded into the appropriate values. Data analysis for the 10 computer simulation runs was conducted utilizing a statistical analysis software package to determine significant differences in hypothesized call wait times between two call queue populations.

4.0 Results

4.1 Perceptions of Service Queues – Exploratory Survey

A pilot survey was launched through Amazon's MTurk platform to assess the items developed in an initial research model on perceptions on service queues. The pilot survey addressed both line standing and call service queues. The initial theoretical model is shown in Figure 7. The pilot launch collected a total of 55 responses and after data analysis and listwise deletion of missing and outlier (straight-lining) data, 48 valid responses for perceptions of line standing queues and 50 valid responses for perceptions of call queues remained. The data were then analyzed using a common statistical analysis software package to recode the reverse coded items. A list of coded items for each construct is shown in Table 1.

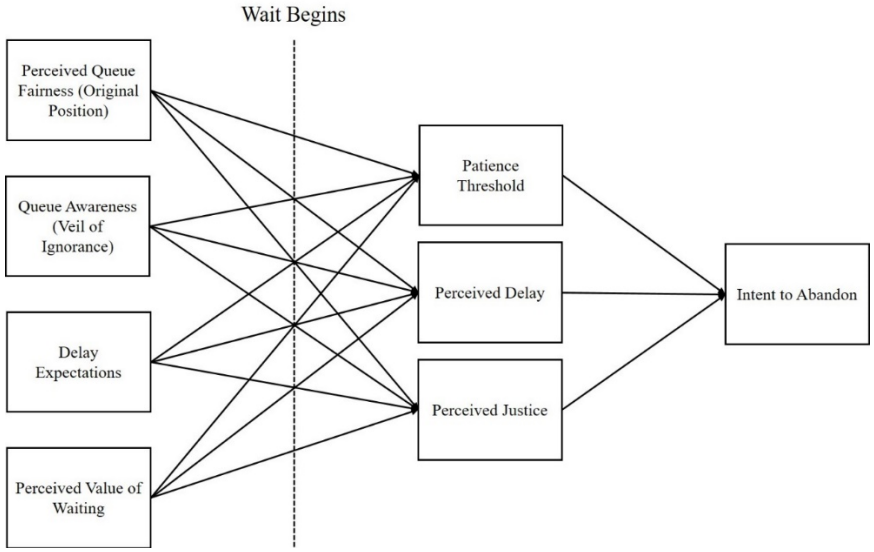


Figure 7. Initial theoretical model.

Evaluation of the constructs within SmartPLS [45] indicated several item loadings associated with QA, DE, PV, and PD were very low and were removed from the actual survey instrument. A deliberate effort was made to focus on the research questions at hand and adjust the theoretical model accordingly. After further revision, the theoretical model was refined as depicted in Figure 8.

Table 1. Coded Construct Items

Construct	Code
Perceived Queue Fairness (Original Position)	QF
Queue Awareness (Veil of Ignorance)	QA
Delay Expectations	DE
Perceived Value of Waiting	PV
Patience Threshold	PT
Perceived Delay	PD
Perceived Justice	PJ
Intent to Abandon	IA

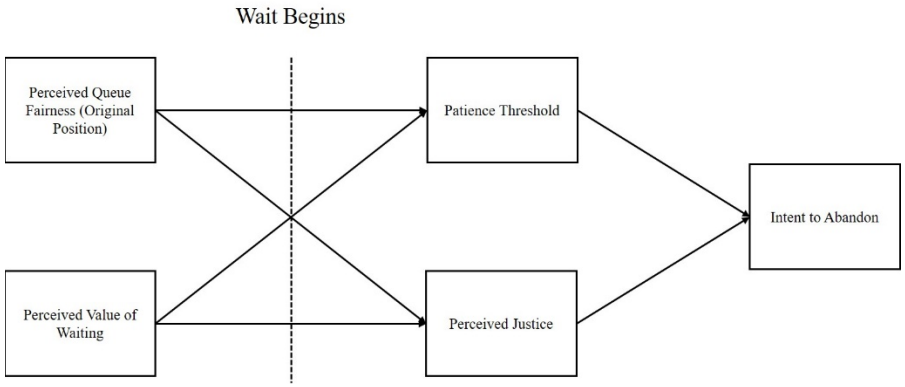


Figure 8. Final theoretical model.

Although the final survey instrument included items for QF and PV, as depicted in the measurement model, the resultant research model only contained constructs aligned with the research questions on PT, PJ, and IA, shown in Figure 9.

4.2 Perceptions of Service Queues – Line Standing

299 responses were gathered from the survey deployment on the MTurk platform. The data were then analyzed for outliers and missing values and 18 responses on perceptions of waiting in a line contained missing data. These were determined to be ignorable data and were deleted listwise [44]. The final dataset contained 280 valid responses on perceptions for standing in a service line. The data were then analyzed for descriptive statistics and reverse coded items were recoded into the appropriate values. A two-step process was used to evaluate the measurement model and the resultant structural model proposed in this exploratory research [46].

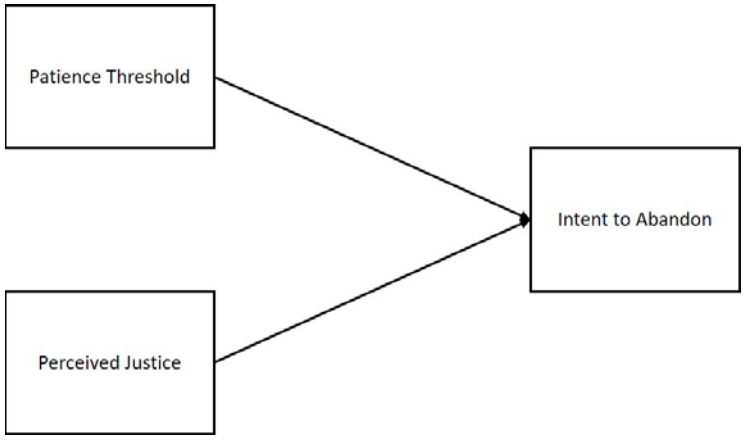


Figure 9. Research model.

The first step determines the reliability and validity of the constructs and the second step evaluates the structural model for predictive power and magnitude of relationships. Evaluation of the proposed model included both discriminant and convergent reliability. Convergent validity is an indicator that related constructs converge or have a high degree of variance in common [44]. All items in each construct for the final structural model indicated a high degree of correlation, which is an indicator of good convergent validity [47]. Discriminant validity is a measure of each construct being distinct from other constructs. All factor loadings were greater than .70, which is an indicator of good discriminant validity [48]. Additional discriminant validity is provided by the [49] criteria being satisfied, which is the square root of the Average Variance Extracted (AVE) being greater than that item with all other latent variables, as shown on the diagonal of Table 2. Internal consistency is indicated by composite reliability being higher than .70 for each factor [50]. Cronbach's Alphas were good with the exception of PT, which was .620. Although the commonly acceptable threshold is .70, values from .60 to .70 are at the "lower limit of acceptability' [44]. All AVEs were greater than .50 which is considered good for convergent validity [50], [44]. The Variance Inflation Factor (VIF) for all constructs was less than 5 which indicates multicollinearity is not a factor [50]. The R^2 value was .17 and the final structural model is shown in Figure 10. PJ and IA retain only two items each, but the strength of the relationship indicates dependable results and [51] consider a factor with two items reliable if they are highly correlated (> .70).

Table 2. Line Standing Item Loadings, AVEs, CR and Factor Correlations

Item loadings			AVEs, CR, factor correlations					
Factor	Item	Loading	AVE	CR	Cronbach's Alpha	PT	PJ	IA
Patience Threshold (PT)	PT1	.784						
	PT2	.714	.568	.798	.620	.754		
	PT5	.762						
Perceived Justice (PJ)	PJ2	.949	.765	.866	.719	.025	.875	
	PJ5	.793						
Intent to Abandon (IA)	IA2	.916	.845	.916	.817	.350	.231	.919
	IA3	.923						

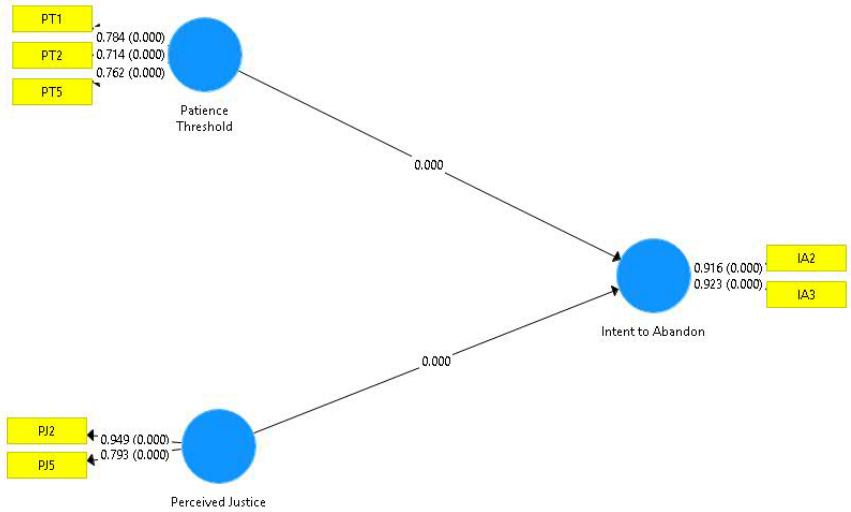


Figure 10. Line standing structural model.

4.3 Perceptions of Service Queues – Call Queues

299 responses were gathered from the survey deployment on the Mturk platform. The data were then analyzed for outliers and missing values and 17 responses on perceptions of waiting in a line contained missing data and again these were determined to be ignorable data and were deleted listwise [44]. The final dataset contained 281 valid responses on perceptions for waiting in a call queue. The data were then analyzed for descriptive statistics and reverse coded items were recoded into the appropriate values.

As with the data for perceptions on standing in a line, the factor loadings for perceptions of call queues were greater than .70. The [49] criteria was satisfied as shown in Table 3. Composite reliability was higher than 0.70 for each factor. Cronbach's Alpha for PT was again lower than .70, but meets the lower limits of reliability [44].

Table 3. Call queue item loadings, AVEs, CR and factor correlations

Item loadings			AVEs, CR, factor correlations					
Factor	Item	Loading	AVE	CR	Cronbach's Alpha	PT	PJ	IA
Patience Threshold (PT)	PT1	.730						
	PT2	.763	.602	.819	.682	.776		
	PT5	.830						
Perceived Justice (PJ)	PJ1	.766						
	PJ2	.796						
	PJ3	.830	.667	.909	.875	.163	.817	
	PJ4	.833						
	PJ5	.856						
Intent to Abandon (IA)	IA2	.925	.84	.913	.810	.396	.239	.917

	IA3	.908

All AVEs were greater than .50 and the VIF for all constructs was less than 5. The R^2 value was .19 and the final structural model is shown in Figure 11. Only IA again contained two items, however the factors were highly correlated (> .70).

4.4 Demographics

Respondent age ranges are depicted in Figure 12. The majority of the respondents were in the 22 to 34 age range consistent in part with the findings of [52] with 64% being male and 36% female. The education level of the respondents is indicated in Figure 13. A substantial proportion (37.7%) reported having a bachelor's degree. This finding is consistent with that of [53] who attribute the higher levels of education to early adopters of new technology such as Amazon's MTurk platform. The reported income levels at Figure 14 are also consistent with the findings of [53] in finding that, although the education level was higher than the general population, the income levels reported were lower than the average.

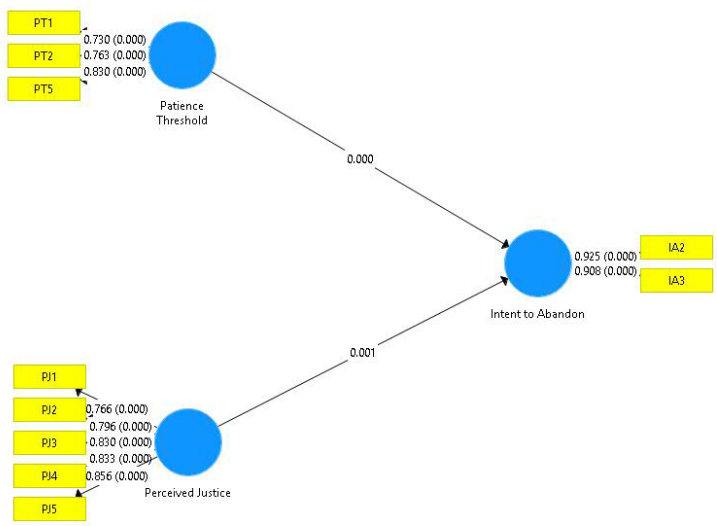
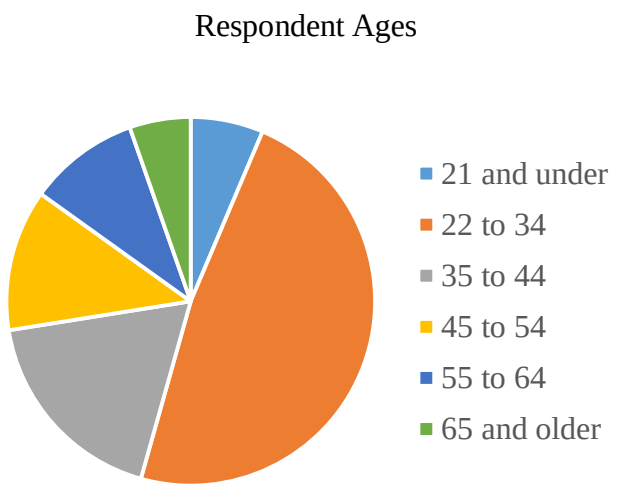
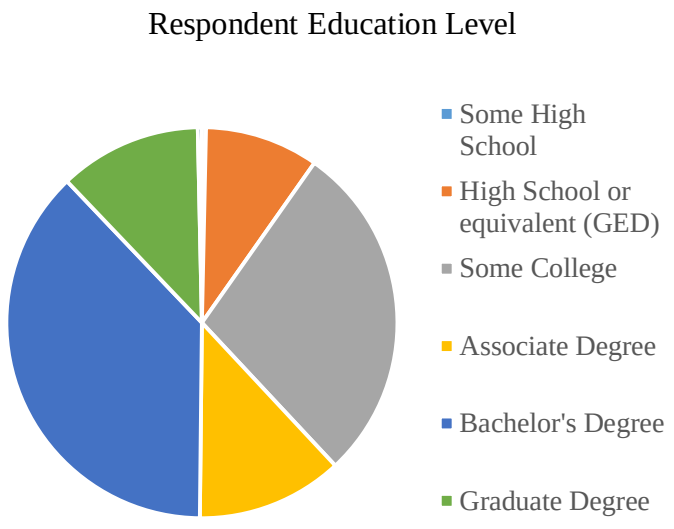


Figure 11. Call queue structural model.



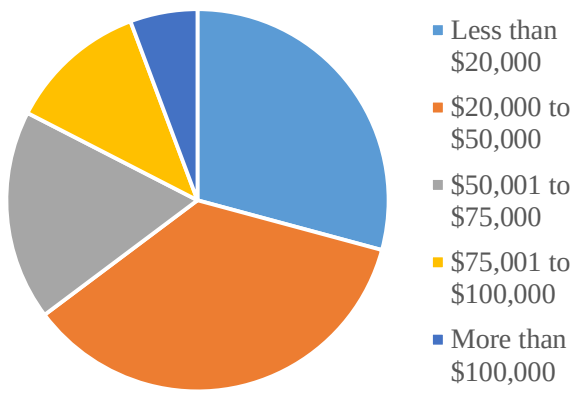
Age Range	Count	Percent
21 and under	19	6.38%
22 to 34	143	47.99%
35 to 44	54	18.12%
45 to 54	37	12.42%
55 to 64	29	9.73%
65 and older	16	5.37%

Figure 12. Respondent ages.



Education Level	Count	Percent
Some High School	1	.34%
High School or equivalent (GED)	28	9.43%
Some College	84	28.28%
Associate Degree	36	12.12%
Bachelor's Degree	112	37.71%
Graduate Degree	35	11.78%
Doctoral Degree	1	.34%

Figure 13. Respondent education level.



Income Level	Count	Percent
Less than $20,000	87	29.19%
$20,000 to $50,000	106	35.57%
$50,001 to $75,000	53	17.79%
$75,001 to $100,000	35	11.74%
More than $100,000	17	5.70%

Figure 14. Respondent income level.

4.5 Discrete Event Simulation Results Analysis

For the first DES run, the wait times for the normal service queue were not normally distributed. Normality is a fundamental assumption for many multivariate analysis tests and must be present in order to use the *t* statistic to evaluate differences in sample means [44]. A two-step method was used to transform the data in order to achieve normality [47]. The first step in the transformation required a fractional rank order of the cases be created and the second step utilized a normal inverse distribution function to achieve normality. The Kolmogorov-Smirnov test for

normality values for the initial dataset and the transformed values are shown in Figure 15. The results of all ten DES simulation runs are depicted in Table 4, showing each of the wait times for the VPQ and normal queue.

	Kolmogorov-Smirnov		
	Statistic	df	Sig.
Wait Time (Normal Queue)	.228	18	.014*
Wait Time (Normal Queue Transformed)	.125	18	.200

Figure 15. Kolmogorov-Smirnov values before and after data transformation.

Table 4. Independent samples t-test simulation results

Simulation #	VPQ Mean	VPQ SD	Normal Queue Mean	Normal Queue SD	P value
1	14.65	9.61	32.85	15.91	.001
2	16.46	6.84	39.80	16.71	.001
3	18.39	10.52	31.63	16.19	.010
4	7.16	4.53	25.75	11.10	.001
5	9.95	6.04	20.06	6.28	.001
6	22.64	13.74	35.82	22.24	.046
7	6.80	2.81	22.40	5.15	.001
8	5.01	3.01	12.71	3.97	.001
9	10.95	5.27	27.00	10.07	.001
10	15.33	5.25	46.55	15.12	.001

4.6 Summary

The results of this study indicate that perceived justice and patience threshold play a significant role in intent to abandon a service queue. A distinction was made between standing in a line and waiting in a call queue and respondents were asked a series of questions based on the respective scenarios. The exploratory instrument that was developed is promising in future research and further validation. It is noted that perceived justice (PJ) is also a factor in abandoning a call queue, even though an individual may be totally unaware of others also in the queue. Given the proposed VPQ infrastructure may reduce the wait times for a select population in a call queue, the observed relationship between patience threshold and intent to abandon a queue could justify a VPQ implementation.

The DES was successful in demonstrating significant time savings between a simulated individual in a VPQ advanced reservation and someone in the normal service queue. Ten simulation runs were conducted with varying degrees of significance, but overall significant in time savings that could be awarded to those chosen individuals. The supported hypotheses summary is shown in Table 5.

Table 5. Supported Hypothesis Summary

			t-value	P value
Apparent (Line Standing) Queue				
H_1	Customers who abandon a call queue do so because their patience time has been exceeded.	Supported	6.268	.001

H_2	Customers who abandon a call queue do so because they feel a social injustice has occurred.	Supported	4.443	.001
Phantom (Call) Queue				
H_1	Customers who abandon a call queue do so because their patience time has been exceeded.	Supported	6.470	.001
H_2	Customers who abandon a call queue do so because they feel a social injustice has occurred.	Supported	3.447	.001
	$W_{VPQ} < W_{SQ}$			
	Run 1	Supported	4.607	.001
	Run 2	Supported	5.781	.001
	Run 3	Supported	2.742	.010
	Run 4	Supported	6.664	.001
H_3	Run 5	Supported	5.588	.001
	Run 6	Supported	2.078	.046
	Run 7	Supported	10.297	.001
	Run 8	Supported	5.789	.001
	Run 9	Supported	6.153	.001
	Run 10	Supported	8.723	.001

References

[1] Deloitte. (2015). 2015 Global contact center survey: Executive summary.
[2] Holman, D., Batt, R., & Holtgrewe, U. (2007). The global call center report: International perspectives on management and employment.
[3] Ali, L. F. (2010). A call center simulation study: Comparing the reliability of cross-trained agents to specialized agents (master's thesis).
[4] Koole, G., & Pot, A. (2006). An overview of routing and staffing algorithms in multi-skill customer contact centers. Submitted for Publication, 1–42.
[5] Akhtar, S., & Latif, M. (2010). Exploiting simulation for call centre optimization. In WCE 2010 - World Congress on Engineering 2010 (Vol. 3, pp. 2112–2117).
[6] Aksin, Z., Armony, M., & Mehrotra, V. (2007). The modern call center: A multi-disciplinary perspective on operations management research. Production and Operations Management Society, 16(6), 665–688. https://doi.org/10.3401/poms.
[7] Weinberg, J., Brown, L. D., & Stroud, J. R. (2007). Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. Journal of the American Statistical Association, 102(480), 1185–1198.
[8] Maister, D. H. (2005). The psychology of waiting lines. Retrieved from http://www.davidmaister.com
[9] Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. Manufacturing and Service Operations Management, 5(2), 79–141.
[10] Larson, R. C. (1987). OR forum - perspectives on queues: Social justice and the psychology of queuing. Operations Research, 35(6), 895–905.
[11] Rawls, J. (1971). A Theory of Justice. Cambridge, MA: Harvard University Press.
[12] Koole, G., & Mandelbaum, A. (2002). Queueing models of call centers: An introduction. Annals of Operations Research, 113(1–4), 41–59.
[13] Fukunaga, A., Hamilton, E., Fama, J., Andre, D., Matan, O., & Nourbakhsh, I. (2002). Staff scheduling for inbound call centers and customer contact centers. Ai Magazine, 23(4), 30–40.
[14] Gans, N., Liu, N., Mandelbaum, A., Shen, H., & Han Ye. (2010). Service times in call centers: Agent heterogeneity and learning with some operational consequences. IMS Collections Borrowing Strength: Theory Powering Applications – A Festschrift for Lawrence D. Brown, 22, 99–123.

[15] Wallace, R. B., & Whitt, W. (2005). A staffing algorithm for call centers with skill-based routing. Manufacturing and Service Operations Management, 7(4), 276–294.
[16] Garnett, O., Mandelbaum, A., & Reiman, M. (2002). Designing a call center with impatient customers. Manufacturing & Service Operations Management, 4(3), 208–227.
[17] Shen, H. (2010). Statistical analysis of call-center operational data: Forecasting call arrivals and analyzing customer patience and agent service. In Wiley Encyclopedia of Operations Research and Management Science (pp. 1–10).
[18] Bouzada, M. A. C. (2009). Dimensioning a call center: Simulation or queue theory? Journal of Operations and Supply Chain Management, 2(2), 34–46.
[19] Taylor, S. (1994). Waiting for service: The relationship between delays and evaluations of service. Journal of Marketing, 58(2), 56–69.
[20] Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., Cochran, J. J., Fry, M. J., & Ohlmann, J. W. (2013). Quantitative methods for business. Mason, OH: South-Western, Cengage Learning.
[21] Winston, W. L. (1994). Operations research: Applications and algorithms (3rd ed.). Belmont, CA: Duxbury Press.
[22] Sorokin, V. (1988). The Queue. New York, NY: Readers International.
[23] Andrews, D. (2013). Why does the other line always move faster? New York, NY: Workman Publishing Co., Inc.
[24] Koole, G., & Mandelbaum, A. (2002). Queueing models of call centers: An introduction. Annals of Operations Research, 113(1–4), 41–59.
[25] Mehrotra, V., & Fama, J. (2003). Call center simulations: methods, challenges, and opportunities. In Proceedings of the 35th conference on winter simulation (pp. 135–143).
[26] Zukerman, M. (2008). Introduction to queueing theory and stochastic teletraffic models. Networks.
[27] Mandelbaum, A., & Zeltyn, S. (2007). Service engineering in action: The palm/Erlang-A queue, with applications to call centers. Advances in Services Innovations, 17–45.
[28] Alexander, M., MacLaren, A., O'Gorman, K., & White, C. (2012). Priority queues: Where social justice and equity collide. Tourism Management, 33(4), 875–884.
[29] Brady, F. N. (2002). Lining up for Star-Wars tickets: Some ruminations on ethics and economics based on an internet study of behavior in queues. Journal of Business Ethics, 38(1–2), 157–165.
[30] Larson, R. C. (1988). There's more to a line than its wait. Technology Review, 91(5), 60–67.
[31] Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1993). The nature and determinants of customer expectations of service. Journal of the Academy of Marketing Science, 21(1), 1–12.
[32] Zohar, E., Mandelbaum, A., & Shimkin, N. (2002). Adaptive behavior of impatient customers in tele-queues: Theory and empirical support. Management Science, 48(4), 566–583.
[33] Mandelbaum, A., & Shimkin, N. (2000). A model for rational abandonments from invisible queues. Queuing Systems, 36(1), 141–173.
[34] Zhou, R., & Soman, D. (2003). Looking back: Exploring the psychology of queuing and the effect of the number of people behind. Journal of Consumer Research, 29(4), 517–530.
[35] Perry, R., & Zarsky, T. Z. (2013). Queues in law. In Iowa Law Review (Vol. 99, p. 1595).
[36] Borst, S., Mandelbaum, A., & Reiman, M. I. (2004). Dimensioning large call centers. Operations Research, 52(1), 17–34.
[37] Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., & Zhao, L. (2005). Statistical analysis of a telephone call center: A queueing-science perspective. Journal of the American Statistical Association, 100(469), 36–50.
[38] Mandelbaum, A., & Zeltyn, S. (2009). Staffing many-server queues with impatient customers: constraint satisfaction in call centers. Operations Research, 57(5), 1189–1205.
[39] Bertoli, M., Casale, G., & Serazzi, G. (2009). JMT: performance engineering tools for system modeling. ACM SIGMETRICS Performance Evaluation Review, 36(4), 10–15.
[40] De Mesquita, M. A., & Hernandez, A. E. (2006). Discrete event simulation of queues with spreadsheets: A teaching case. In Proceedings of the 2006 Winter Simulation Conference (pp. 2277–2283).
[41] Mathew, B., & Nambiar, M. K. (2013). A tutorial on modelling call centres using discrete event simulation. Proceedings - 27th European Conference on Modelling and Simulation, ECMS 2013, 315–321.
[42] Sekaran, U., & Bougie, R. (2013). Research methods for business: A skill building approach (6th ed.). West Sussex, England: John Wiley & Sons.
[43] Allon, G., & Hanany, E. (2012). Cutting in line: Social norms in queues. Management Science, 58(3), 493–506.

[44] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis (7th ed.). Upper Saddle River, NJ: Pearson Education, Inc.
[45] Ringle, C. M., Wende, S., & Will, A. (2005). SmartPLS. Hamburg, Germany.
[46] Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin, 103(3), 411–423.
[47] Templeton, G. F., & Burney, L. L. (2017). Using a Two-Step Transformation to Address Non-Normality from a Business Value of Information Technology Perspective. Journal of Information Systems, 31(2), 149–164.
[48] Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In New Challenges to International Marketing (Vol. 20, pp. 277–319). Emerald Group Publishing Limited.
[49] Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50.
[50] Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. Journal of Marketing Theory and Practice, 19(2), 139–152.
[51] Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in Quantitative Methods for Psychology, 9(2), 79–94.
[52] Ipeirotis, P. G. (2010). Demographics of mechanical turk.
[53] Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on amazon mechanical turk. Judgment and Decision Making, 5(5), 411–419.

Biography

William K. (Bill) Pugh is assistant professor in the Information Systems and Cyber Security department at the University of Texas at San Antonio. Bill obtained a Bachelor of Arts in History from Louisiana Tech University, a Master of Science in Liberal Studies (emphasis in Information Systems) from the University of Central Texas (Texas A&M Central Texas), a Master of Business Administration and Doctor of Business Administration degree from the University of the Incarnate Word.